Effects of Regulation on Service Quality

Evidence from Six European Cases

Contract SI2.753522-573/PP/GRO/IMA/16/1131
Joint Principle Investigators: Maria Koumenta, Mario Pagliero, and Davud Rostam-Afschar
Contributers: Piotr Białowolski, Morris Kleiner, Michał Masior, Eva Pagano, Lorenzo Richiardi, Kristina Strohmaier, Mark Williams
November – 2018

EUROPEAN COMMISSION

Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs
Directorate E, Modernisation of the Single Market
Unit Directorate E.5 Professional Qualifications and Skills

E-mail: GROW-E5@ec.europa.eu

European Commission
B-1049 Brussels

EUROPEAN COMMISSION

Effects of Regulation on Service Quality

Evidence from Six European Cases

Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs

2018

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the Internet (http://www.europa.eu).

Luxembourg: Publications Office of the European Union, 2018

ISBN 978-92-79-96718-4
doi:10.2873/910094
Catalogue number: ET-04-18-894-EN-N

© European Union, 2018
Reuse is authorised provided the source is acknowledged. The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

(This page intentionally left blank)

Table of Contents

(This page intentionally left blank)

Executive Summary

Professional services are at the core of the Single Market. In January 2017 the Commission, as announced in its Single Market Strategy for goods and services of 28 October 2015[1], came forward with two important policy initiatives. The 'Guidance on Reform Recommendations for Regulation in Professional Services' identifies opportunities for reform in 7 key professions (accountants, civil engineers, architects, lawyers, real estate agents, tourist guides and patent agents) across the Member States.[2] In addition, the proposal for a Proportionality Test Directive aims to provide Member States with a clear framework when assessing the proportionality of any new or amending professional regulation before it is introduced.[3] Whereas the 'Guidance' focuses on existing overly restrictive professional regulation, the 'Proportionality Test' aims to prevent any new unnecessary barriers from arising. Against the backdrop of these policy developments and given the imperative for evidence-based policymaking, the European Commission commissioned this research to examine the relationship between professional regulation and service quality focusing on specific professions and countries. The overarching aim of this study is to provide an in-depth understanding of the potential research designs and empirical strategies that can be used to understand this important policy theme and enhance the evidence base on the economic costs and benefits of professional regulation.

In this study we:

- Review the existing economic theory and evidence pertaining to the relationship between professional regulation and quality, discussing the advantages and disadvantages of the different available empirical tools and quality indicators.
- Develop six profession and country-specific case studies to empirically assess the effect of regulation on quality and illustrate different policy evaluation approaches. Each case study exploits either a reform in the regulation status or a variation in the stringency of the regulatory regime.
- Identify and develop quality indicators for the selected professions. These include indicators of process (e.g. customer complaints and consumer ratings), indicators of outcomes (e.g. occupation-specific quality outcomes) and value-added proxies (e.g. service availability, affordability, access). Where possible, we supplement these with data on entry and exit rates, employment creation and prices.
- Provide detailed estimates of the effects of the regulatory reform (or variation in the stringency of the regulatory regime) on each of the selected quality indicators.
- Provide recommendations on the development of suitable empirical strategies to address the complex relationship between professional regulation and quality.
- Provide some preliminary suggestions on how to take policy forward (where evidence is clear) or recommendations of how to improve the evidence base (where it is not).

Up to now, little is known about the empirical relation between regulation of professions and – broadly defined – quality of services. This study makes an exceptional attempt to contribute to the debate on the quality effects of regulation. We study six markets (lawyers, architects and engineers, pharmacists, tourist guides, driving instructors and ride hailing drivers) in six EU member states (Poland, Germany, Italy, Greece, UK and Ireland) and address the issue of measuring quality in each. Moreover, we exploit specific features in each regulated market to estimate the impact of regulations on quality.

What is quality? Our work clearly shows that quality is not only difficult to measure, but that even the very definition of quality varies a lot across and within markets. The quality of some services can be verified before purchase, others during or after provision of the service, and for yet others quality can be verified only if additional costs are incurred, or not at all. Consumer satisfaction,

[1] Commission Communication - upgrading the Single Market: more opportunities for people and business, COM(2015) 550 final.

[2] Commission Communication - Communication on reform recommendations for regulation in professional services, COM(2016) 820 final

[3] Proposal for a Directive of the European Parliament and of the Council on a proportionality test before adoption of new regulation of professions, COM(2016) 822 final

complaints and peer judgments for lawyers; peer rankings for architects and engineers; product availability for pharmacists; consumer evaluations for tourist guides; service availability, pass rates and fatal accidents for driving instructors; and consumer satisfaction, hard accelerations and braking for taxi drivers are all examples of the different meanings that the idea of quality might take in different markets.

Moreover, in most markets quality is multidimensional, and even a longer list of measures might not fully capture the full experience of a customer. Still, we take on the challenge and try to measure quality using the available data. While admittedly imperfect, the long list of measures used throughout the study provides a surprising variety of sources of information on quality.

In order to estimate the impact of regulation on quality, one not only needs measures of quality, but also variability in the intensity of regulation. Our case studies focus on a number of reforms, but also exploit variability occurring across municipalities, regions and EU member states. An important lesson from our work is that there is more than one way to approach this empirical question. Taken together, our case studies make use of an extensive empirical methods toolkit: simple difference in means, difference-in-differences, synthetic control methods, panel data regressions, instrumental variable regressions and regression discontinuity designs. This variety in empirical methods reflects the variety of data used for the analysis, which came in the form of individual data from service satisfaction surveys (lawyers), peer ratings and census data (architects and engineers), confidential administrative data (pharmacists), labour force survey data and online booking website (tourist guides), administrative data (driving instructors) and even big data from a giant of the gig economy (ride-hailing drivers).

What is the impact of regulation of professions on quality? In the market for lawyers, we find little change in the overall quality of legal services following the Polish relaxation of entry requirements, yet the number of complaints per active advocate did decrease and a decline in good manners before a court as a quality component was reported. Overall quality of services provided by architects and engineers decreased with higher market concentration in response to higher insurance costs and higher service prices. In the pharmacy market, the availability of pharmacies seems to be correlated with a decrease in the number of hospital admissions related to influenza, suggesting a possible link between the availability of the services offered by pharmacists and consumer health. Tourist guides' level of education increase post reform but the guides that entered the market via the new regime are more likely to receive lower consumer ratings. An increase in the stringency of regulations pushed many driving instructors out of the market with no evidence of improved quality of instructors or of learner drivers. Finally, more stringent licensing regulations for ride hailing drivers have no effect on customer satisfaction ratings or measures of hard breaking and accelerations.

Taken together, these case studies indicate that an increase in availability of service providers and/or competition does not necessarily have negative effects on the quality of the services provided or survey measures of consumer satisfaction and well-being. To the contrary, in a number of cases we find positive effects of increased availability and competition Still, our work does not provide final conclusions. We provide a series of examples on how to approach the issue of measuring quality and how to use existing data to investigate the impact of regulation on quality. As for any empirical work, it is difficult to extrapolate from a sample and more work in this area is certainly needed.

The following table provides a summary of the six case studies, the quality indicators used in the analysis and the findings.

Table EX.1. Summary of the six cases studies, the quality indicators, and the findings.

Occupation and Country	Nature of Regulatory Reform	Quality Indicator	Summary of Findings
Advocates and Legal Advisors (Poland)	Change in restrictions relating to educational requirements (more transparent bar exam, shorter bar training, access to bar without training for selected legal professionals).	Service quality assessment from a survey on clients	No significant change in quality of legal services following regulatory changes
		Complaints and disciplinary cases against legal professionals-members of the bar	Declining number of complaints per active advocate. No change in indicators of legal professionalism
		Service quality assessment from a survey on judges	
		Measures of enterprise creation and turnover	Legal counsellors registering their firms increased from less than 100 yearly before the reform to 575 after
		Measures of employment creation	The number of advocates and legal advisors more than doubled
		Price levels	Prices of legal services providers increase more slowly than the average for all services
Architects and Civil Engineers (Germany)	Increase in price floors and ceilings Introduction of a minimum (professional indemnity) insurance coverage requirement	Score of an international peer-ranking of architectural firms	Negative effect on average score of -18%
		Exit rates	No evidence for effect on exit rates
		Continued education	No evidence for effect on continued education
		Professional-to-inhabitants-ratio	No evidence for effect on professional-to-inhabitants-ratio
		Number of firm owners	Probability of being self-employed decreased by 4%-points
		Number of employees	Number of employees increased by 6%

Pharmacists (Italy)	Relaxation of quantitative/structural restrictions	Availability of pharmacies at the municipality level	Reduction of the number of hospital admissions relating to influenza-like illness
Tourist Guides (Greece)	Relaxation of the start-up process Modifications of educational requirements (broadening access to training while increasing required entry level qualifications (university degrees in a range of subjects))	Indicators of tourist guides' labour market performance	Post reform tourist guides are more qualified and have stronger job attachment
		Quality proxies based on customers' evaluations (index) and tourist guide profile	Post reform tourist guides get lower ratings from customers
Driving Instructors (UK)	Restrictions on educational requirements	Service availability Indicators	Decrease in the number of non-fully qualified instructors, number of fully qualified instructors constant
		Indicators of student performance	Deterioration in driving test overall pass rates, pass rates at first attempt and zero faults pass rates
		Road accidents	Increase in serious accidents
		Price levels	Increase in the price of an 1-hour lesson
Ride Hailing Drivers (London and Dublin)	Variations in the stringency of the regulatory regime between Dublin (higher) and London (lower)	Customer satisfaction ratings	No evidence of statistically significant differences between London and Dublin
		Trip Safety Indicators	No evidence of statistically significant differences in hard braking between London and Dublin No evidence of smoother journeys in Dublin (as measured by hard accelerations)
		Service Availability	Higher driver to population ratio in London
		Prices	Lower prices in London

I. Introduction

According to recent estimates, 22 per cent of the EU labour force is working in a regulated occupation (Koumenta and Pagliero 2016). The introduction of such barriers to entry associated mainly (but not exclusively) with the attainment of certain skills standards are often justified by their proponents as aimed at workforce upskilling and productivity enhancements, as well as means to address market failures resulting from asymmetric information between buyers and sellers (Shapiro 1986; Law and Kim 2005). From a policy evaluation perspective, while the impact of occupational regulation on labour market outcomes is well evidenced in economic studies, the evidence base relating to the product market effects (which include quality, price, availability and access) is less developed (especially at EU level) and more inconclusive. This report is structured as follows: we begin by providing an overview of the available theory and evidence on the relationship between regulation and quality, we then synthesize the results of this review and explain our empirical strategies, with specific reference to the six case studies that are the focus of our study. Each case study is then presented in a separate chapter. In the conclusion we summarize our findings and discuss their implications for the study of professional regulation, quality and policy more generally.

I.A. Theory and Evidence

I.A.1. Theoretical Background

The key public policy justification for professional regulation in general, and licensing in particular, is its presumed ability to protect consumers and the wider public from incompetent and unscrupulous practitioners (Humphris, Kleiner and Koumenta 2011). Consumers cannot easily obtain information or lack the knowledge to assess the quality of the product or service prior to its purchase, particularly where the provision of a technical service requiring specialist knowledge and skills is involved. Through setting minimum qualifications requirements for entry to occupations and making various postulations regarding work experience and continuous professional development, occupational licensing is expected to raise average skills/competence levels in the occupation, since low-quality providers will presumably be unable to meet the new qualification requirements and are driven out of the occupation (Pagliero 2013). Reducing minimum qualification requirements, have been documented to foster entry of providers who worked to a large extent as employees in their previous occupation thus increasing competition in the occupation (Rostam-Afschar 2014; 2015). Provisions on entry requirements into occupations are further supplemented with provisions relating to conduct upon entry. Typical conduct regulations include the regulation of process and fees (fixed prices, minimum or maximum prices), regulation of advertising and marketing, regulation of location, restrictions on inter-professional co-operation and restrictions on the forms of business that can be adopted (Paterson et al. 2003).

Taken together, entry and conduct regulations are expected to ensure that consumers receive a more homogeneous and high-quality product while the resulting higher investments in training have the potential to enhance the skills base in the economy (Shapiro 1986). Quality is supposed to be further ensured through the regular monitoring of performance standards, deviations from which can lead to ‘punishments’ such as financial penalties or exclusion from practising the occupation. Finally, professional associations’ activities relating to encouraging members to discuss and promote positive aspects of work experiences, disseminate information about how to do the job better, engage in job-specific training, promote ethical standards or devise methods of adjudicating disputes between consumers and producers all have the potential to positively affect service quality. With regards to other forms of regulation, minimum skills standards are a key feature of certification and accreditation (i.e. protected titles without reserves of activities) schemes. Since such schemes make stipulations regarding competence which could in turn be positively related to human capital characteristics (or propensity to invest in their acquisition) one would expect some impact on quality, although the extent of this would depend on the demand for certified/accredited practitioners in the market (Koumenta et al. 2014).

In practice, the effect of regulation on quality might not be as straightforward. Skill- and/or qualification specific regulations pertaining to entry to occupations, if properly set and monitored by competent authorities and adhered to by practitioners, should raise the human capital stock of the service provider. Hence, the assumption is that human capital acquisition (i.e. increases in skill and/or qualifications) will be positively correlated with the service that consumers receive, or else that input productivity results in output productivity. However, as Kleiner (2006) points out, licensing tests measure competence, not job performance, and given the little to no research on the relationship between performance on the licensing exam and an individual’s ability to perform on the job, we have no way of knowing whether regulation correlates with subsequent performance.

The effect of regulation on service quality can also be negative. Quality is not only linked to skill but also quantity supplied. To explore such an effect, it is useful to consider the imposition of barriers to enter occupations which are cumulatively imposed over time on occupations. Examples of such barriers include compulsory membership of professional associations, artificial limits on the number of professionals that are allowed to operate in the market, restrictions on corporate forms, shareholding requirements, restrictions on joint exercise of professions, incompatibilities of activities, etc. If an increase in quality through better-trained practitioners results in a subsequent fall in their supply (due to aspiring practitioners not meeting the entry or exercise requirements), the service actually received by the consumer suffers for the following reasons (Pagliero 2011). First, if a decline in the number of available practitioners leads to an increase in price of the product or service, then some consumers might opt for lower-quality services. In a context of licensing, such substitutes are confined to 'do-it-yourself' services (Friedman 1962; Kleiner 2006). Price increases can also be driven by consumers themselves. Regulation can reduce uncertainty or the likelihood of poor quality practitioners in the market. As a consequence, consumers perceive the service to be of higher quality and demand more of the service, thus pushing up the price (Akerlof 1970).

A more extreme unintended consequence of licensing could involve the decision not to consume the service at all, which could be a health and safety risk in itself. Such an effect is likely to be more pronounced among low-income consumers, meaning that any improvement in quality is only felt by those at the middle and upper quartiles of the income distribution (Shapiro 1986). Overall, the effects of regulation should be analysed not only in relation to improvements in skill levels but also price and availability of services. For example, while one might receive a better quality service from a licensed pharmacist, such effects cannot be realized if such individuals are in short supply and therefore access to pharmaceutical services is restricted. Finally, licensing takes the form of a minimal human capital requirement to practise the occupation and often provides no incentives for human capital development after entry. It is therefore possible that the 'minimum' skill standard imposed by licensing becomes the 'maximum' across the occupation. Coupled with the fact that it restricts competition among practitioners, licensing can reduce the pressure to compete on quality, thus leading to a fall in the overall service quality received by consumers (Carroll and Gaston 1981).

Turning into the impact of licensing on prices, if raising the entry requirements via occupational licensing (a) limits the supply of labour to a profession and (b) increases the entry costs for practitioner (e.g. financial investment in education and training), then the effect on the price of the product or service will depend on the price elasticity of demand. The more price inelastic the good is, the more scope there is for licensed producers to increase its price. Price elasticity will depend on the price and availability of substitutes and whether these are also subject to licensing. If there is a strong substitution effect with unlicensed products, then producers will be less inclined to increase price. Furthermore, producers will have more scope to increase prices for goods and services that consumers perceive as necessities rather than luxuries. As such, if the good is highly income inelastic, demand is likely to be relatively unresponsive to price changes. The proportion of income that is devoted to the purchase of the good or service is also an important consideration. The lower the proportion of consumer income spent on the service, the more scope for licensed producers to increase prices without experiencing a proportionate fall in demand. Of course, licensing can be accompanied by other regulations of professional conduct. In particular, the theoretical literature has shown that a combination of price floors and entry regulation can increase welfare if the consumer cannot verify the quality of a service (Atkeson et al. 2015). According to this theoretical model, in the absence of these regulations buyers would not be in a position to learn whether a firm made an investment in high quality, and an extreme "lemons problem" develops, where no firm invests, and the market shuts down.[4] Often, in addition to a price floor, price ceilings are imposed for specific professions in order to prevent overcharging. In such a case, it is an empirical question to measure how variations in these prices affect the quality provided.

To summarize, the effect of occupational regulation on quality has been put forward as the main justification for its existence. However, as this analysis has demonstrated, because of the corresponding effects of regulation on labour supply, it is difficult on theoretical grounds to determine a priori its impact on the quality of the service provided. Similarly, any net effect on price will depend on the characteristics of the product and service in question. It is therefore a question that can only be addressed empirically. But this is not always a straightforward task. For example, estimates of service quality can either be difficult to measure (e.g. the quality of a visit to

[4] However, it is worth noting that the authors do not consider heterogeneous consumer preferences, in the sense that regulation would not harm those with preference for low quality.

the chiropodist) or data might not always be available (e.g. customer satisfaction surveys). The following section explores these challenges in more detail, discusses how researchers have attempted to overcome them and presents their findings.

I.A.2. Empirical Background

Defining quality poses several complex problems of definition and measurement (Parasuraman et al. 1985; Cronin et al. 1992). Beginning with defining quality, it is generally accepted that quality is related to an absence of defects or a degree of excellence. However, in practice, conceptualizing quality is a very challenging task. While consumer satisfaction is a key dimension of quality and – as we will show below – an important way in which quality is assessed, issues relating to subjectivity in consumer expectations as well as the asymmetric information about what constitutes a good quality service (which is one of the key justifications for regulation) mean that solely relying on consumer satisfaction can provide inaccurate information on quality. It is for this reason that studies on regulation have commonly adopted more expanded definitions of quality. These have taken into account broader notions such as consistency, reliability, continuity, non-excludability and affordability in the delivery of a service. As such, typically, the link between regulation and quality has been investigated by looking at the following: (a) measures of process or the procedures involved in providing services, (b) measures of the outcome of the services provided and (c) other value-added proxies.

Measures of process

Process measures refer to the specific steps in the process of service delivery that are used as indicators of service quality. In the regulation literature, these have included elements such as number of customer complaints, consumer ratings of practitioner behaviour, malpractice insurance premiums, number of malpractice lawsuits and number of disciplinary actions. Maurizi (1980) compares consumer complaints received by the California Contractors State License Board to restrictiveness of entry measured by the annual proportionate increase in the number of licences. His evidence shows that as the stock of licences increased, so did the number of complaints per licensee. Holen (1978) also finds low pass rates on entry examination (his measure of restrictiveness) to be related to low malpractice insurance rates. The author finds similar results when he compares customer complaints for 32 licensing boards in the US (Maurizi 1974). Using pharmacist malpractice suits as a measure of quality, Martin (1987) on the other hand finds no links between restrictiveness of entry and the incidence of malpractice. More recently, Kleiner (2013) compares premiums in regulated versus unregulated states and uses that as a good proxy for service quality (since a reduction in malpractice lawsuits should lead to lower premiums). Empirical work by Cordes (2005) and Kleiner and Kudrle (2000) shows that licensing has no effect on reducing the risk of a high payout for occupational therapists, nurses, clinical psychologists and counsellors.

Such measures of quality have the advantage of being easy to access. A number of regulatory bodies and professional associations keep records of such incidents and in theory should be reliable. However, many have pointed out the disadvantages of such data as proxies for quality. Maurizi (1980) for example accepts that voiced complaints are an imperfect measure of customer dissatisfaction and shows that factors such as proportion of income devoted to the good or service, frequency of purchase and the damage resulting from poor quality are likely to be stronger predictors of customer complaints. For example, a typical consumer is most likely not going to voice a formal complaint for receiving a bad haircut, but might do so if he receives bad service from an architect or a lawyer. Further, filing formal complaints might not only be a function of receiving sub-standard service but also ease and opportunity to voice such complaints. If consumers face difficulties in doing so, or the regulator has developed a reputation of failing to adequately address grievances, then individuals are likely to be put off from doing so. Indeed Maurizi's own work shows an increase in complaints following the regulator opening more branch and district offices. Others have noted that there is a positive link between the complaints procedure being advertised or mass-media publicity about cases of malpractice and the number of complaints being received, and warn for caution when employing such measures (Schuck 1980). Similarly, malpractice insurance rates can be affected by lawyers' willingness to represent clients, the size of the damage awards and statutory restrictions on size of the awards. According to Hirschman (1970), when consumers are faced with deteriorating quality of a good, the exercise of voice (in this case complaints) would also depend on the availability of alternatives (perfect or near perfect substitutes in the market). The more monopolistic the market, the less able consumers will be to use the exit option and thus the more likely they will be to voice their concerns (an argument clearly in favour of licensing when the latter is having such an effect in product markets). In such studies, it is therefore advisable to combine such measures with additional sources of quality data, an approach we follow in this study.

Measures of Outcomes

Outcome measures typically look at the quality of the end product or service to consumers. The assumption is that since occupational regulation improves the stock of skills, then on-the-job performance and quality should also improve. Humphris and Koumenta (2014) for example examine the quality effects of licensing for security guards and nursery school workers in the UK using local crime data and government inspection reports of nurseries respectively. In the US, Angrist and Guryan (2003) investigate the effect of the requirement imposed by some states for teachers to pass a standardized test before practising in a public school on the subsequent quality of their teaching (the latter measured by student exam grades). Kleiner and Petree (1988) use data on the educational attainment of teachers in the licensing exam and student achievement scores, and Kleiner (2013) looks at child development and early childhood academic achievement scores to assess the impact of licensing childcare services in the US. Outcome measures, if properly chosen, can be direct and objective indicators of quality. However, in some cases they come with their caveats. For example, in the case of teachers, it is also likely that the student's effort and ability affect achievement scores (over and beyond licensing), so ultimately the robustness of findings would rest on the ability to control for such factors through sophisticated econometric techniques.

'Value-Added' Proxies and Structural Measures

In an attempt to overcome some of the weaknesses of the above measures but also to capture some of the unintended outcomes of regulation discussed earlier, researchers have adopted more expanded definitions of quality and devised proxies to measure them. Central in these endeavours is the belief that the following issues should be considered:

- The degree of availability of the product or service in question to consumers
- The extent to which consumers can easily access it
- The extent to which it is provided with continuity and at an affordable cost
- The effect on the prices of products and services

Studies have subsequently devised the following proxies to assess such outcomes.

(a) Substitutes for licensed services such as 'do-it-yourself' options: according to Carroll and Gaston (1983), monopoly pricing resulting from restrictions to entry could force consumers to substitute expert for own services. Examples of occupations in which such measures can be employed include retail sales of hair cutting equipment and retail sales of home electrical repair equipment to replace barbers and electricians respectively. For example, in their study, the authors find that the stock of qualified plumbers and electricians is negatively correlated with the retail sales of plumbing equipment and accidental deaths by electrocution.

(b) Availability of practitioners in regulated versus non-regulated regions/states: practitioner availability is therefore linked to service or product availability in that, for example, other things equal, one would expect a fall in the number of dentists to lead to a fall in the dental health of the population (see for example Kleiner 2013). For example, a US study of veterinarians by Carroll and Gaston (1978) shows that the stricter the state barriers in obtaining a veterinary licence, the fewer the practitioners, which has led to an increase in cases of rabies and brucellosis not being detected. A recent study of the removal of licensing for various occupations in Greece by Athanassiou et al. (2014) finds strong effects on employment growth for these occupations compared to other occupations in the labour market that were not subject to regulatory changes.

(c) Access to the service or product by different income groups: Kleiner and Kudrle's (2000) study on dentists, for example, finds that higher income groups are the main beneficiaries of the effects of licensing on dental health, while Kleiner and Todd (2007) in their study of mortgage brokers find that occupational licensing results in negative outcomes for consumers such as a greater percentage of high interest rate mortgages possibly affecting low-income groups disproportionately.

(d) Existing estimates of the effect of licensing on price commonly use average prices charged for the service. Shepard (1978) and Kleiner and Kudrle (2000), for example, calculate the average fees charged by dentists for various dental procedures from a national data set produced by the American Dental Association. In terms of findings, most studies of such estimates show that prices rise as a result of licensing (Kleiner 2006).

Estimating Product Market Effects in the EU

As our discussion of the literature has shown, examining the impact of regulation on quality is challenging both in terms of specifying the suitable indicators and in terms of accessing such data. However, the challenges are also methodological. In particular, there may be non-random and unobservable factors which are confounding attempts to identify a causal effect of occupational regulation through cross-sectional analysis. Studies from the US exploit between-state variation in licensing regimes which enables researchers to undertake comparisons of the same occupation while controlling for human capital characteristics, year and state fixed effects. In the case of the EU, such variations are hard to come by, as professions are regulated in very different ways across EU member states, so researchers have to seek ways to address common issues related to identification and endogeneity.

A series of six independent studies contracted by the European Commission cover liberal professions (e.g. architects, engineering consultants, accountants/tax advisers and lawyers) in Austria and 22 professions (e.g. lawyers, legal advisers, and notaries, taxi drivers, real estate agents, and tourist guides) in Poland, 94 crafts professions in Germany, 11 professions in Greece (lawyers, notaries, auditors, accountants, tax consultants, dentists, physiotherapists, taxi drivers, shipping agents, tourist guides, chartered valuers and real estate agents), pharmacists and legal professions in Italy, as well as nursery school workers and security guards in the UK.[5] Examples of recent studies include Noailly and Nahuis' (2010) and Tavares and Rodrigues' (2013) work on the liberalization of entry in the profession of notaries in the Netherlands and Portugal respectively; Janczak's (2015) on deregulation of tourist guides in Poland; as well as an older literature on regulatory reforms for architects (e.g. Button and Fleming 1992) and legal professions in the UK therein. The most credible results are from studies that have sought to overcome such concerns by looking at switchers, namely occupations that have incurred a change in their regulation status (e.g. from unregulated to licensed, or from licensed to accredited), as a form of a 'natural experiment' of before and after. While methodologically such approaches are more robust, evidence has consistently shown that the effects of regulation can be lagged (i.e. it takes a certain amount of time before they emerge),[6] which delays policy evaluation initiatives. The second challenge for EU studies is that of obtaining occupation-level suitable data, a well-documented issue in the study of occupational regulation.

In the chapters that follow, for each case study we outline ways in which we meet the objectives of this study while also addressing the methodological and data challenges outlined above. Due to the unique nature of each occupation in the labour market and the high heterogeneity in how quality can be defined and measured, our approaches should be viewed as illustrative examples of how the relationship between regulation and quality can be examined. We do not offer a one-size-fits-all solution to assessing the impact of regulation on quality. Instead, our aim is to draw attention to the challenges and limitations of undertaking research of this kind, while at the same time through the six case studies empirically illustrating ways in which one can address this topic. In addition to meeting the European Commission's requirements as regards the occupational categories to be analysed, the geographical coverage and timeframe of reforms, we believe a key strength of our approach is the inclusion of diverse forms of regulatory reforms, the study of a broad spectrum of quality indicators and data sources and the use of a variety of methodological approaches. Overall, the series of occupation-specific case studies that we have selected provide a comprehensive picture of the forms that regulation (and de-regulation) takes, while our analytical strategies and choice of quality measures are representative of the current knowledge in the field of occupational regulation and econometric analysis. Our chosen approaches came with limitations and caveats, and these are discussed in detail in the relevant sections.

[5] Details of these studies available here: http://ec.europa.eu/growth/content/effects-reforms-regulatory-requirements-access-professions-country-based-case-studies-0_en

[6] See, for example, Forth et al. (2013); Humphris and Koumenta (2014); and Kleiner (2013).

(This page intentionally left blank)

1. Lawyers in Poland

1.1 Introduction

It has been 40 years since Carlson (1977) claimed that '"there is a record of concern about the quality of legal services. And there is a record of the attempt to ensure quality through restricting the admission of persons to the practice of law. But beyond this, there is no 'system' for, nor for that matter any theory about, assuring the quality of legal services." Although time has passed, legal professions are still largely dependent on state-endorsed self-regulation and are overwhelmingly driven by the assumption that greater restriction to entry to the professions is focal to the security of consumers and the quality of services. This assumption, although to some extent justified in the literature, has never been thoroughly tested. Recently, the European Commission has put forward a proposal which provides a framework for Member States to assess the proportionality of any new or amending professional regulation they intend to introduce. [7] A key aim of this initiative is to ensure that any new professional regulation will be based on legitimate public interest grounds of justification and will be proportionate, in line with the general principles of EU law. In this light, an evidence base for various regulatory changes and their link to the quality of services is especially important. Recent regulatory reforms in Poland relating to the relaxation of access rules to the legal professions serves as an interesting case study to examine such effects.

Significant changes oriented in introducing a more transparent and inclusive regulatory regime for legal professions have been implemented in Poland over the past decade. Restrictions on access to roles in advocacy and legal advice have been relaxed as the consequence of less strict regulation applied to work experience and bar exams. Required work experience was shortened and relaxation of access-rules opened paths to the professions of advocate and legal advisor for other legal practitioners. More importantly, new admission processes became more transparent. Bar exams were harmonized across the country and the number of points required to pass exams was reduced. The legal documents that serve as a basis for questions in the exams are now communicated in advance. These changes certainly improved public access to legal expertise. However, their impact on quality has never been tested.

In this study we attempt to identify arguments for relaxation of access to legal services, contrasting them with arguments in favour of regulation. We approach the issue of regulation from the perspective of legal service quality. Although in legal professions quality is clearly ensured by the high skills of practitioners, in the case of rapidly changing markets, such as in Poland over the past decade, quality also depends on availability and accessibility to professional services. Lack of access could push clients to resort to inferior solutions, such as dispensing with a trainee lawyer for their representation, which might consequently lead to suboptimal outcomes.

The literature on the quality of legal services serves also as a benchmark for the analytical approach proposed. Our study on the shifting regulatory environment for legal professions in Poland utilizes measures of consumer satisfaction and dissatisfaction with quality of service as well as attempts at objective measures of quality. We survey consumer satisfaction with legal services using a method based on the SERVQUAL approach, i.e. a multidimensional tool oriented on the assessment of quality of services with a number of easily assessable indicators for each dimension (Parasuraman, Zeithaml, & Berry, 1988); but we also accommodate quality of legal services by investigating dissatisfaction with legal service quality, proxied by complaints and disciplinary proceedings. As an objective proxy for legal service quality, we study the opinions of judges. This group consists of professionals able to provide valid opinions on legal service quality who can also be considered impartial in their assessment. This group seems to constitute the best arbiters of change in the quality of legal services in Poland.

7 http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2016:822:FIN Proposal for a directive of the European Parliament and of the Council on a proportionality test before adoption of new regulation of professions, COM/2016/0822 final – 2016/0404 (COD), accessed 23.09.2017.

1.2 Literature Review

Arguments for Regulation and the Side-Effects of Regulating Market Legal Services

Chaserant & Harnay (2013) point out three potential problems in the quality assessment of legal services. First, consumers cannot be certain whether they need the service. Second, they cannot assess the quality before using the service. Third, even after the event, they do not know whether their lawyer exerted sufficient effort to win the case and most often they cannot be certain if the lawyer was sufficiently knowledgeable to address the issue at hand. These characteristics of legal services validate the argument for their regulation. Pearce & Nasseri (2013) cite additional arguments to support regulation of entry to the legal profession. Restricted entry should limit provision of legal services to highly skilled lawyers, judicial efficiency should be improved, consumer dissatisfaction measures (such as complaints of lawyer misconduct) should be reduced, and due to regulation of entry lawyers' livelihoods should also be protected.

However, as a result of limited access to the profession, side-effects can easily be found in the area of legal services. Carlson (1977) indicates that one of the primary effects not intended by regulation is homogenization of legal services and award of a monopoly power to the suppliers. Economic theory predicts that monopoly always acts towards limiting the supply to capture consumers willing to pay the highest fees. Consequently, the needs of the less wealthy are likely not to be met.

Another perilous outcome, not intended initially by regulation, is related to the fact that legal professionals (as well as other professionals) are reluctant to reject tasks for which they are overqualified, simply because it undermines their business. Even if an aspect of business is routine and can easily be outsourced to less skilled individuals, giving it up would imply that a regulated profession engages in a costly search for other opportunities instead of benefitting from the current state of affairs.[8]

According to Love & Stephen (1997), self-regulation can be responsible for at least three types of side-effects: restrictions to entry, restrictions on advertising and restrictions on commercial competition. When one of the outcomes of regulation is lower supply, the natural response of self-regulating bodies is to set a maximum price for legal services. It is likely however, that maximum price becomes a reference point and leads to clustering of actual charges to clients around the maximum level (Chaserant & Harnay, 2013).

A positive side-effect generated by restrictions to access was noted by Pagliero (2013) and relates to stability. Although a larger potential supply of candidates willing to enter the legal profession can push the licensing requirements higher and lower the pass rates, it also implies that the market is sheltered from supply shocks. This relates to the argument for *protection of the interests of lawyer households*.

The Reforms of Legal Services and Quality

Reflecting on the considerable issues ensuing from regulation of legal professions, one may ask why access to the legal professions should not be deregulated completely. It seems that justification for regulation swings between the extremes of public interest and private interest theory. Pagliero (2011) identifies the arguments for both these explanations, while demonstrating stronger empirical support for private interest theory. Licensing translates to higher salaries and lower availability of lawyers, leading on to a reduction in consumer welfare (Pagliero, 2011). Studies by Kleiner (2000) and Kleiner, Gay & Greene (1982) confirm the effect of higher salaries for professions with regulated access.[9] Pearce and Nasseri (2013) argue that lawyers fulfil vital functions in a liberal democracy and thus barriers to enter the profession should be set as low as reasonably possible. Broader access to legal services implies that individual rights can be better protected, and the rule of law can be better enforced.

---

[8] Love & Stephen (1997) present the example of the loss of the conveyancing monopoly (i.e. legal transfer of property rights in the UK) in 1985, which despite being simple task from a technical perspective, met a high level of resistance from the solicitors, who were previously solely authorized to take part in conveyancing work.

[9] It is, however, not always the case that higher licensing requirements lead to immediate increase in the level of salaries. Law & Marks (2013) show that mandatory licensing had little to no effect on wages and participation rates of nurses.

Although lower regulation might prove beneficial for competition in the legal services sector, problems of adverse selection (poor lawyers replacing good ones) and moral hazard (providing poor quality service not identifiable by the consumer) might become more apparent (Chaserant & Harnay, 2013). There is still very limited research on which standards should be maintained in order to avoid instigation of those problems (Pearce & Nasseri, 2013).

Lower regulation provides an interesting alternative, especially for less complicated legal tasks. Non-lawyers tend to outperform their licensed counterparts in providing legal assistance for civil matters, like welfare benefits, debt and employment (Winston, Crandall, & Maheshri, 2011), but it also seems that clients value hands-on experience over formal training in the law (Kritzer, 1998).

Regulation is primarily imposed by the stringent exams to qualify to practise law. The justification for such reforms (oriented on lowering the threshold of a bar or equivalent exam) is that it broadens access and, via inclusion of the population excluded from access to expensive "top notch" legal services, increases the average quality of legal services available to clients. Lowering the exam threshold seems to lead to supply-side effects, which in some circumstances may compensate for lower requirements. Changes to the Japanese bar exam demonstrated that lowering the threshold from a level at which only 2 per cent of applicants passed to a situation in which around 40 per cent qualified, significantly increased the attractiveness of legal education and training (Ramseyer & Rasmusen, 2015). With very strict requirements the bar exam was a gamble with a very low probability of success. Consequently, very few educated individuals with legal skills decided on a path associated with such a high risk of failure and pursued other professional careers. With much higher pass rates, pursuit of a legal career became much less elusive and consequently attracted many more people with the requisite high potential and skills.

In a similar vein, Canton et al. (2014) showed that less strict regulation improved allocative efficiency. Niches with highly restricted supply tended to open, attracting new people and reducing the cost for end clients. Relaxation of regulation affected also the investment component. In a less regulated environment, larger incentives to invest are present.

Measuring quality of services in legal professions

Carlson (1977) divided approaches to measuring quality into three fundamental groups: input measures, process measures and outcome measures.

The most common approach to measuring quality relies on measuring inputs. It is the easiest way to formalize and objectivize the approach to measurement of quality. It relies on exams during education, plus the bar exam for entering the legal profession. The main drawback to this form of measurement is that it is extremely prone to moral hazard, as once a person is admitted to the profession, their skills are not always subject to verification. Additionally, even if a lawyer's knowledge does not deteriorate, this form of assessment does not require continuous development and legal competences acquired long ago might not correspond to the current state-of-the-art. Input assessment in the form of the bar exam measures only one important feature of a good lawyer – a high degree of knowledge – omitting other competences from the assessment.

A second approach, according to Carlson (1977), is focused on processes. This approach relies on peer-review, verified membership of professional associations and a requirement for continuing education. All these aspects can be measured and are often measured under self-regulation of the profession. This approach relies on the assumption that quality can be assured if certain steps are followed during the course of a career. It assumes that continuing education and membership of professional bodies will help lawyers to keep in touch with the most up-to-date knowledge and will also be a guide to the most suitable procedures for handling specific cases.[10] Process measures, although free from some of the input measure errors (e.g. inability to assess if a lawyer is relying on outdated knowledge), are still subject to a variety of moral hazard issues (e.g. providing a legal service without due diligence).

Process measures can be also readily obtained. Professionals treat their clients with dignity, lack of bias and leave them with the perception that the service they received was thorough (Barendrecht, Mulder, & Giesen, 2006). These measurable features of a legal service can serve as markers of quality. Additionally, according to the Legal Services Consumer Panel (2010), the following are considered to be aspects of good quality legal service: technical competence, good client care and utility of service. Although a client may not be able to assess the technical competence of a lawyer,

[10] As Carlson (1977) points out, as in a medical procedure for which the standard is to follow predefined steps in order to comply with best possible practice.

he or she will surely have an impression of whether a lawyer *appeared to him or her* as knowledgeable, experienced, professional, trustworthy, empathetic, moral, honest, good-mannered, courageous, determined, available when needed, communicative, efficient, effective and even innovative.

Outcomes of legal services are even more difficult to measure. Nevertheless, misconduct measures, like complaints or disciplinary actions, are available as indicators of quality. Although client satisfaction can easily be biased simply by an unfavourable outcome of a case, despite the lawyer's best efforts and skill, when client's dissatisfaction manifests itself in the form of a complaint it is subsequently investigated by other lawyer which increases the chances that the final outcome is impartial.

Optimal System of Measuring Quality in Legal Services

In light of the arguments presented earlier, designing a system for quality measurement of legal services is an intrinsically difficult task. As outlined in the introduction to this report, the traditional approach to look at quality as a variable related to regulation requires tools for measuring the process of service rendering or, if feasible, tools for measuring outcomes of service as they are even more direct and objective measures of quality. To address difficulties with process and outcome measures, the literature proposes also investigating proxies related to the value-added of service, such as its availability, ease of access, continuity of provision, low cost and price.

This approach correlates with the consumer markets scoreboard devised by the European Commission as a tool to measure market performance (European Commission, 2015). The components of the consumer markets scoreboard as far as the process measures are concerned are related to overall expectations (consumer satisfaction or the extent to which suppliers live up to clients' expectations), detriment (proportion of consumers who experienced problems and the degree to which they caused detriment) and complaints (propensity to complain to a lawyer or other party). Subsequently the following should be treated as the value-added proxies of quality: comparability of services on offer, choice (consumers' satisfaction with the number of legal aid suppliers to choose from), trust that service providers would comply with consumer protection rules and ease of switching service providers.

Parasuraman, Zeithaml, and Berry (1988) proposed an additional approach to quality measurement. They identified core dimensions of service quality, i.e. tangibles, reliability, responsiveness, assurance and empathy. These dimensions contain many characteristics of the service and service provider, which generally are process measures (consumer satisfaction). In this study we refine them, by adapting the approach of the Anglo-Welsh regulator of the market for legal services (Legal Services Consumer Panel, 2010), for the purpose of quality assessment in legal services.

Based on these concepts, feasibility of the research techniques and availability of the data, we employ the following approaches to capture the quality of legal services on the Polish market as thoroughly as possible:

(a) Process measures: frequency and the structure of complaints by type, filed against practising lawyers to the bars; number of disciplinary actions taken against them relative to the number of practising lawyers to offset the effect of their growing population.

Our approach allows for the identification of the most egregious cases. For a large number of cases and the law of large numbers in place, a safe assumption can be made that a variation in the share of cases ending up with a complaint or disciplinary action is a proxy for a variation in the overall quality. It also corresponds with the measures of quality as indicated in the consumer markets scoreboard.

(b) Process measures: consumer satisfaction survey

Our survey builds on the developments from consumer markets scoreboard (see Parasuraman *et al.,* 1998; and Legal Services Consumer Panel, 2010). It focuses on the dimensions of service quality as defined in the SERVQUAL approach and expands it by addressing overall expectations, detriment (the proportion of consumers who encountered problems in the process) and complaints (share of consumers who filed a complaint and the awareness of how to do it).

(c) Process/outcome measures: survey of judges

Our proposed approach is based on the assessment of the professionalism of legal representatives from the perspective of judges. Process and outcome measures were obtained as the judges expressed their opinions on the prowess of lawyers in the court representation (process measure from the perspective of a client), judged validity of their arguments in court and assessed whether they acted in line with the clients' interests (closer to outcome measures). Their opinions were assumed to be neutral as judges have no interest in providing a favourable assessment (which thus precludes the private interest theory-argument). Their perspective allowed for a detailed record of the most difficult part of legal aid: representation before the court.

All approaches are complementary. More details on the specific concepts are provided in the discussion about the questionnaires and the results.

1.3 The Regulatory Reforms

Poland regulates currently 350 professions.[11] In the years 2013-2015 the restrictions on entry to 246 professions, including lawyers, were relaxed. However, the first reforms in access to the professions of advocate and legal advisor were introduced already in 2005. In addition, since 2004 the set of exclusive reserved activities for those professions has entered gradually into a pool of shared reserved activities with other legal professionals.

Legal aid is a legal term describing services provided by advocates and legal advisors. It consists of drawing up legal opinions, rendering legal counsel, drafting legal acts and appearing before a court or an administrative body. Before 2004 these activities remained in the pool of exclusive reserved activities of advocates. The reforms transferred most of them into the pool of the reserved activities shared with other legal professionals. Legal advisors before 2004 were authorized to render legal aid with the exception of family, custody and criminal (incl. delinquency and fiscal crimes) cases. Advocates traditionally specialized in individual, that is civil and criminal, cases, whereas legal advisors specialized in administrative, labour and economic issues.

In the course of the reforms three major types of changes to the regulatory regime were introduced:

- Substantial reduction in the scope of exclusive reserved activities of advocates and legal advisors; since 2004 they were limited to court representation in administrative, criminal and civil cases (see also a change in 2005 in Table 1.1);
- in 2005 exclusive reserved activities of advocates related to family, custody and delinquency cases entered into the pool of shared reserved activities with legal advisors; in 2015 criminal and fiscal criminal cases followed suit;
- entry to the professions of advocate and legal advisor was relaxed for the holders of a master's degree in law in 2005, and then in 2009; for other legal professionals the access was relaxed in 2004, 2005, 2009 and 2013.

Candidates for advocates and legal advisors have to obtain a master's degree in law, which in Poland is awarded after 5 years at a university. Before 2005, there used to be a single path of entry to the profession, i.e. passing an exam to bar training organised by one of the 24 local bar councils of advocates or 19 district chambers of legal advisors. The examinations were organised independently by the self-governed professional organisations and the decisions of the examination boards were non-contestable. Recruitment was highly selective, subject to diverse and unclear rules. Most applicants failed and had to choose a different career or retake the examination the following year. The bar training lasted 3.5 years and was concluded with an exam (so-called advocate and legal advisor exam). Similarly to the entrance exam, it was organised locally by the bar councils of advocates and district chambers of legal advisors. Only after successful passing of these two examinations, one could enlist as an advocate or legal advisor in a chosen bar council or

11 Regulated professions database: http://ec.europa.eu/growth/tools-databases/regprof/index.cfm?action=regprofs&id_country=23&quid=1&mode=asc&maxRows=*#top%20 (accessed 25.03.2018).

12 The last recruitment to bar training took place in the mid-2003. The trainees accepted at these exams would continue their training until the end of 2006 and then would pass their professional exams in the beginning of 2007, being able to commence their practice in the mid-2007.

district chamber. The reform of 2005 standardised nationally the entrance exams to bar training and granted supervision over their organisation to the Ministry of Justice. It also eliminated limits on entry established earlier by the bar councils and chambers (see a detailed description of the reforms below). It shortened bar training to 3 years, which slightly reduced the length of training before admission to the profession. It also introduced standardised payment for bar training linked to the minimum wage in the economy. The reform of 2009 standardised the professional examinations to take place at the end of the bar training and put them under the supervision of the Ministry of Justice. Motivation for the reform was not to lower the requirements for the candidates to the profession but to have their qualifications checked fairly. It was achieved by reducing the authority of the advocates' and legal advisors' self-regulatory bodies. The reform of 2005 introduced also new paths to the profession for other legal professionals. They took the form of exemptions from bar training and exemptions from the professional exams. Subsequent amendments of 2009 and 2013 slightly extended the scope of exemptions. Legal professionals, often prosecutors, notaries and academics, were granted possibility to circumvent traditional bar training and professional exams.

Table 1: Important amendments to the rules of entry to the professions of advocate and legal advisor in Poland between 2004 and 2015

Year	Description of change
2004	On 18.02.2004 Polish Constitutional Tribunal issued a judgment (P 21/02) stating that the law associations were not allowed to set limits on the entry to the profession of advocate and legal advisor as it was the case. By doing so, they breached the freedom of enterprise and freedom to choose a profession (failing proportionality test), as set in the Polish Constitution of 1997. The Tribunal called for the legislator to clarify the rules of entry to bar training and to obtain a professional title (prerequisites and criteria of selection). As the legislator (the Sejm and the Senate – Polish Parliament) worked on it for more than a year, the law associations stopped admitting new lawyers to the bar training because of uncertainty of the rules during the transition period.[12]
2004	Since 1.05.2004 foreign lawyers (from the EU or outside) were allowed to render regular legal services in Poland on the condition they are enlisted (upon completion of a special procedure) as advocates or legal advisors. The lawyers from a different EU Member State can represent a party in front of a court.[13]
2004	On 27.08.2004 the act on the freedom of economic activity entered into force,[14] abrogating at the same date the prohibition to render legal aid on other grounds than the act on advocates and the act on legal advisors.[15] It also explicitly classified professional activity (including legal aid) as an economic activity thus guaranteeing freedom of its exercise, even if it is rendered in an organised, continued manner and for profit, like law offices do. Consequently, the act allowed practically everybody to provide legal aid, including law graduates without the professional title of advocate or legal advisor. The only exception refers to the right of representation in front of a court, to which only advocates, and legal advisors remained entitled.
2005	Starting from 5.02.2005 everybody (incl. law graduates) can be authorised to

[12] The last recruitment to bar training took place in the mid-2003. The trainees accepted at these exams would continue their training until the end of 2006 and then would pass their professional exams in the beginning of 2007, being able to commence their practice in the mid-2007.
[13] Ustawa z 5.07.2002 o świadczeniu przez prawników zagranicznych pomocy prawnej w Rzeczypospolitej Polskiej [act of 5.07.2002 on rendering legal aid in the Republic of Poland by foreign lawyers (Journal of Laws 2002 vol. 126, item 1069)].
[14] Ustawa z 2.07.2004 o swobodzie działalności gospodarczej [Act of 2.07.2004 on the freedom of economic activity (Journal of Laws 2004, vol. 173, item 1807)].
15 Such prohibition existed as the art. 87 in the previous act of 19.11.1999 on economic activity: Prawo działalności gospodarczej (Journal of Laws 1999, vol. 101, item 1178), largely abrogated with the act on the freedom of economic activity.

	represent a party before a lower court in a civil case as his or her mandatary without a professional proxy.[16] Advocates and legal advisors kept the monopoly to act as legal representatives before higher civil courts and in criminal and administrative courts (it is shared to a small extent with tax advisors and patent agents).
2005	On 30.06.2005 the amendments to the act on advocates and to the act on legal advisors were voted.[17] The reform comprised inter alia: abolition of limits to the number of candidates accepted to a bar training; transferring competence to organise entrance examination to bar training and professional exam from the law associations to the Ministry of Justice; standardising entry and professional exams – instead of heterogeneous forms as applied by different local bars, the exams took the form of a written anonymised test, the same nationally; authorisation for legal advisors to take family, custody and delinquency cases; easing mobility for lawyers across country – local bars were no longer entitled to indicate main office, since then advocate or legal advisor wishing to change their main office has only been required to notify the bars and has not been subject to their approval; allowing access to the profession (i.e. professional title) to more legal professionals thanks to an exemption from bar training and professional exam for professors, doctors in law with habilitation (i.e. post-doctoral degree) and persons who passed the exam for a judge, prosecutor, notary and, correspondingly, legal advisor or advocate; allowing access to the profession for more legal professionals thanks to a newly established exemption from bar training (candidates were still required to pass a professional exam) for: (1) holders of PhD in law and the holders of a master's degree in law who within 8 years prior to the application for at least 5 years: (2) worked on a labour law contract in applying or creating law, (3) worked continuously on civil law contracts rendering services in the application or creation of law, (4) operated their own registered business consisting in providing legal aid,[18] (5) worked as court referendaries or assistants to a judge; fee for the bar training was introduced to offset its costs to law associations.
2006	The Constitutional Tribunal found some provisions of the act of 30.06.2005 amending act on advocates and some other acts *unconstitutional insofar as they allowed the entry into the profession of the unexperienced persons*[19]*.* The Tribunal obliged the legislator to modify the rules of entry to guarantee that law associations could influence the content of the exams and sit in the examination commissions.

[16] Ustawa z dnia 2 lipca 2004 r. o zmianie ustawy – Kodeks postępowania cywilnego oraz niektórych innych ustaw [Article 1 points 12-13 of the Act of 2.07.2004 amending Code of civil procedure and some other acts (J.L. vol. 172, item 1804)].
[17] Ustawa z dnia 30 czerwca 2005 r. o zmianie ustawy - Prawo o adwokaturze i niektórych innych ustaw [Act of 30.06.2005 amending act on advocates and some other acts (Journal of Laws 2005, vol. 163, item 1361 with amendments)]. Some of the amended provisions entered into force on 10.09.2005, some on 1.01.2006.
[18] Points 2-4 were found unconstitutional by the Constitutional Tribunal in its rulings of: 19.04.2006 (K 6/06) – referring to advocates and of 8.11.2006 (K 30/06) – referring to legal advisors. These provisions lost their binding force correspondingly on 4.05.2006 and 31.12.2006.
[19] Judgment of the Constitutional Tribunal of 19.04.2006 K 6/06 – referring to the act on advocates and the judgment of the Constitutional Tribunal of 8.11.2006 K 30/06 (J.L. 2006 No. 206, item 1522) – referring to the act on legal advisors.

2009	The act of 20.02.2009 amending act on advocates, act on legal advisors and act on notaries[20] further modified the rules of entry to the profession of advocate and legal advisor: rules governing entrance examinations to bar training were clarified: the passing threshold was lowered from 75% to 66%, obligation to publish legal acts required for the exam 90 days prior to its date was introduced, tests for advocate and legal advisor training were standardized to set the same requirements for both training; rules governing professional exams were revisited: their form was modified by abolishing oral exam to guarantee transparency, anonymity, equal examination conditions and rules of grading, obligation to publish legal acts required for the exam 90 days prior to its date was introduced, central examination committees in charge of formulating questions were introduced, examination commission of the 2nd grade was created to consider the appeals against the results of the examination; bar training was shortened from 3.5 to 3 years (which applied to the new trainees since 2010), but the overall minimum period from the moment of graduation from law studies to the moment of taking an oath as an advocate or legal advisor remained at 4 years; more persons were exempted from the bar training and the professional exam with at least 3-year practice in these positions: advisors and senior advisors of the State's Treasury Solicitor's Office, judges, prosecutors, notaries, advocates or legal advisors, correspondingly, and persons who either as holders of PhD in law or after passing the exam for a judge or a prosecutor worked at least 3 years on selected positions (e.g. as judge assessors, assistants to a judge) or rendered legal aid; more persons were exempted from the bar training in addition to holders of PhD in law and court referendaries and assistants to a judge, i.e.: (1) persons who passed the exam for a judge, prosecutor or notary, (2) advisors and senior advisors of the State's Treasury Solicitor's Office, and holders of a Master's Degree in law who within 10 years prior to the application for at least 5 years: (3) worked as court referendaries, senior court referendaries, judge trainees, prosecutor trainees, judge-prosecutor trainees, assistants to a judge, assistants to a prosecutor, (4) exercised duties demanding legal knowledge directly linked to legal aid rendered by advocate or legal advisor, (5) worked in public administration exercising duties demanding legal knowledge directly linked with legal aid rendered for the administration.
2013	The last major amendment to the rules of entry to the profession of advocate and legal advisor was introduced on 13.06.2013.[21] It increased access to the profession for other legal professionals extending the exemptions from the bar training and professional examination: more persons were exempted from the bar training and professional exam: (1) persons who worked for at least 3 years as bailiffs; (2) persons who passed the exam for a judge or a prosecutor after 1.01.1991 or a notary exam after 22.04.1991 and worked for at least 3 years (within 5 years prior to the application): (a) in the Supreme Court, the Constitutional Tribunal or in an international court, in particular in the Court of Justice of the EU or the European Court of Human Rights performing tasks of assistants to a judge, (b) rendering legal aid in a partnership limited by shares, (c) based on labour law contract in public administration creating projects of law; the exemption covered also (3)

[20] Ustawa z dnia 20 lutego 2009 r. o zmianie ustawy – Prawo o adwokaturze, ustawy o radcach prawnych oraz ustawy – Prawo o notariacie [Act of 20.02.2009 amending act on advocates, act on legal advisors and act on notaries] (Journal of Laws 2009, vol. 37, item 286). It entered into force on 25.03.2009.

[21] Ustawa z dnia 13 czerwca 2013 r. o zmianie ustaw regulujących wykonywanie niektórych zawodów [Act of 13.06.2013 amending acts regulating certain professions] (Journal of Laws 2013, item 829). It entered into force on 23.08.2013.

	holders of a PhD in Law who worked for at least 3 years (within 5 years prior to the application) in public administration creating projects of law; more persons were exempted from the bar training: (1) persons who worked for at least 4 years (instead of 5) within 6 years (instead of 8) prior to the application to take the legal advisor examination as court referendaries, senior court referendaries, assistants to a judge or assistants to a prosecutor or in the Supreme Court, the Constitutional Tribunal or in an international court, in particular the Court of Justice of the EU or the European Court of Human Rights performing tasks of assistants to a judge, (2) holders of a Master's Degree in Law who after their studies in law rendered legal aid based on labour law or civil law contracts in a law firm for at least 4 years (instead of 5) within 6 years (instead of 10) prior to the application to take the exam, (3) holders of a Master's Degree in Law who after their studies in law gained at least 4 years' (instead of 5) experience working on a labour law contract in public administration providing legal aid within 6 years (instead of 10) prior to the application to take the advocate exam, (4) persons who completed the legislative training and worked for at least 4 years within 6 years prior to the application to take the legal advisor exam in public administration providing legal aid, (5) persons who passed the exam for a bailiff; the amendment also raised educational requirements for candidates by modification of the form of the professional exam: a test with closed-ended questions was replaced with pleadings or legal opinions to write (corresponding with real-life tasks of a lawyer) and an exercise in professional ethics was added.[22]
2015	Starting from 1.07.2015[23] legal advisors are authorised to defend a party in criminal and fiscal crime cases, which means their scope of rights has become equal to advocates.

1.4 Quality Indicators

An important question that arises in the aftermath of the supply-side oriented reforms introduced into the Polish legal system is whether increase in the quantity of legal services (availability of legal aid) came at the expense of their quality. On one hand, competitive pressure from new entrants and increased industry rivalry might have driven up the standards of customer service. On the other hand, the entry to the profession of legal professionals without practical experience in legal aid and insufficient bar training as well as the entry of legal counsellors and other service providers possibly even without prior legal studies or bar training might have put the customers at disadvantage. Also, new examinations might be worse or better suited to check for the candidates' knowledge and skills than the exams before organised by the peers.

Process Quality with an Investigation of Trends in Complaints and Disciplinary Cases

In order to identify evolution of quality in legal services, we investigate the number of reported complaints and disciplinary cases. With this set of data we address the impact of entry of new less-experienced legal professionals on average level of professionalism as perceived by their peers. The statistics for the two professions, advocates and legal advisors, were analysed separately following differences in the data sets available for analysis for those two professions.

As mentioned in section 1.3 the process measures referring to the most egregious cases, when a legal professional has not met the quality standard, can be considered one of the indicators of quality of legal service. Self-governing bodies such as the law associations provide unique data for this purpose: statistics of disciplinary cases. Authorized and obliged by law, they control professional conduct of their members (advocates, legal advisors or trainees), examining

[22] Information about the deregulation in Polish can be found in Napiórkowska-Piłat (2016).
[23] Ustawa o zmianie ustawy – Kodeks postępowania karnego i niektórych innych ustaw [The act of 27.09.2013 amending the code of criminal procedure and some other acts (Journal of Laws 2013, item 1247)].

complaints filed against their members and, if necessary, holding them liable for violation of the rules of conduct. Codes of ethics regulate the legal practice, relations with clients, courts, colleagues and legal authorities. The obligations include, among others, honesty, factuality, confidentiality, avoidance of conflicts of interests, courageous and honourable defence of clients' interests and informing clients about the case on a regular basis. Thus, as already mentioned in section 0, the quality of legal services can be proxied by the frequency and structure of wrongdoings committed by the professionals.

Measures of Consumer Satisfaction

In a complex, multi-faceted service, like legal aid, quality assessment requires multiple indicators. Several methods for testing quality of service are available for this task. Among them are SERVQUAL, SERVPERF, scales for measuring customer satisfaction and loyalty, tech cases, observational studies, group discussions and in-depth interviews. Although there are divergent views as to which procedures are the most valid, there is a general agreement that comparison of customer expectations with their perception of service (Gilmore, 2006, p. 44) is of particular importance.

SERVQUAL is a non-direct approach to measuring quality. It can be instrumental for quality assessment of complex services, like the legal ones. Legal services are difficult to assess based solely on client's experience. SERVQUAL concentrates on the measurement of various aspects of service and relates them to the expectations of clients. The approach based on SERVQUAL addresses a number of sub-components that encompass the quality of the service actually provided to customers. In the synthetic version, it assesses five dimensions of the quality of service: material characteristics of the service, reliability of service, readiness to help the customer and to provide the service quickly, ability to inspire a sense of confidence that the service will be well performed and empathy of the service provider (Parasuraman et al., 1988).

To better adapt the SERVQUAL tool to legal services we draw on the study ordered by the Anglo-Welsh market regulator of legal services, the Legal Services Board (Legal Services Consumer Panel, 2010). Six characteristics of high-quality legal service emerged from this study and closely correspond to the ones identified in original SERVQUAL methodology:

- empathy – treating clients with dignity and respect; understanding their situation and relating to it;
- efficient services process – ensuring smooth and timely progress of the case;
- achievement of the desired results – winning a case, processing a transaction (e.g., sale of real estate) or drafting a legal document (e.g., a will);
- clarity – explaining the scope of legal service in a way understandable to the client, signposting of what should happen and when, and guidance on charges;
- proactive use of legal knowledge – explicitly suggesting alternative solutions, tailoring advice to individual customer needs;
- professional appearance (physical characteristics of service) – ensuring high quality personal appearance e.g., in terms of dress, but also high quality physical attributes of the office environment.

This framework provided a base for developing a questionnaire for clients of lawyers. We asked respondents to assess various indicators of the service they received in the near or more distant past (see Appendix 1.1 for details). In order to measure the differences in the perception of legal services we use propensity score matching (Heckman & Ichimura, 1998; Rubin & Thomas, 1996) and then estimate average treatment effects with treatment defined as being exposed to legal services in the period 2011–2017 instead of 2000–2005 or 2006–2010.

Objective Measures of Legal Service Quality

We collect additional information on the quality of legal services from a questionnaire distributed amongst judges (see Appendix 1.2 for details). As professional lawyers who are in a unique position to observe and analyse lawyers' work in court (speeches, pleadings, motions), they seem to be well-suited to assess the quality objectively and knowledgeably. The judges were asked to retrospectively evaluate different aspects of the behaviour of various legal representatives in court: advocates, legal advisors, their trainees, legal counsellors and clients without any legal representation.

Microeconomic Context: Dynamics of Employment, Number of Businesses and Prices

Along with the impact on quality, changes in the regulatory framework related to entry to legal professions were instrumental in stimulating higher interest in legal professions and lead to higher turnover within the profession. Increased competitive pressure was hypothesized to lower the prices or at least curb their upward trend.

Control variables: Number and Length of Court Proceedings

Additional check for any correspondence between the reforms in the legal profession and the efficiency of the legal system was performed with information on the length of court proceedings and the number of cases taken to court.

1.5 Data Sources

Administrative Data on Complaints and Disciplinary Cases

The statistics on the number and the character of complaints and disciplinary proceedings in the advocates' profession were obtained from the reports published every three years by the Polish Bar Council. They comprise information from the Disciplinary Prosecutor of the Polish Bar Council, the Higher Disciplinary Court and the Bureau of Complaints. The data regarding legal advisors do not distinguish complaints – they are included in the statistics of disciplinary cases. The data are provided in the reports of the analogous bodies: disciplinary prosecutors of the local chamber of legal advisors, the Main Disciplinary Prosecutor, district disciplinary courts, and the Higher Disciplinary Court. The statistics on disciplinary proceedings describe both the local and national level, which corresponds to the two-instance process of holding a licensed lawyer liable for his or her misconduct.

Consumer Satisfaction Survey

To address the question of legal service quality two studies have been commissioned in two different public pollster companies: TNS Polska S.A. and Pactor S.A. Nationwide samples comprised 320 and 400 respondents, respectively. We targeted respondents who had seen a lawyer after the year 2000. Most of them (679) used legal services provided by advocates and legal advisors, i.e. professional lawyers associated with bars. The detailed survey questionnaire is provided in Appendix 1.1. The first study conducted by TNS Polska took the form of CAWI, namely computer-assisted interview using web pages, while the second study – conducted by Pactor – used the methodology of CATI, i.e. computer-assisted telephone interview. It is important to point out that there was no 100 per cent overlap between the two surveys. Some of the indicators were explored in either first or the second study only, while some indicators were collected on both occasions.

Although the regulations related to advocates and legal advisors have been altered significantly since 2003, leading to a change in the market composition, not a single study has ever been undertaken to assess the impact of the discussed amendments. For this reason, the data covering the period before the reforms were not available. Consequently, the analysis performed here is of a retrospective nature.

Survey of Legal Services Quality – Opinions of Judges

The association of Polish judges, Iustitia, substantially facilitated the study by directly mailing its members with a request for participation in the study aiming to assess the quality of legal services in Poland and their evolution after the reforms. Judges, having day-to-day contact with legal representatives but also being very knowledgeable in the legal field, served as ideal arbiters for the quality of legal services provided by advocates and legal advisors. Their opinions, developed over time, on the quality of legal services are based on numerous interactions with representatives of the profession. Hence, the opinions of judges can be considered equal to those provided by qualified experts and assumed to be prone to much lower measurement error than opinions of unknowledgeable clients of legal services.

The study of judges was conducted in early November 2017. It was a web survey with the link provided only to invited members of Iustitia – the largest in Poland association of judges. In the

survey, a total of 60 responses were collected. The total number of professional judges in the Polish judiciary amounts to 10,096 (Council of Europe, 2016), out of which around one-third, that is more than 3,500 are members of Iustitia. Our survey was then filled by ca. 1.7% of the total number of Iustitia-members and 0,5% of all judges. However, judges should be perceived as professionals with high ability to assess the quality of service. The number of judges participating in the study does not preclude statistical reasoning.

Microeconomic Context: Dynamics of Employment, Number of Businesses and Practitioners, and Prices

The data on businesses operating in the market for legal services were employed to investigate the microeconomic context of the reforms.

Control Variables: Cases in Courts

The statistics on the functioning of Polish courts are provided by the Ministry of Justice and the Institute of Justice.[24]

1.6 Results

Number of Practitioners, Price Dynamics and Availability of Service

New entries and increased market rivalry

The reforms limited the exclusive rights of advocates and legal advisors to the representation before the courts. In addition to that, the exclusive rights of advocates were gradually shared with legal advisors and the entry of foreign lawyers was facilitated. The reforms concentrated on the bar training. The abolishment of quotas and introduction of nationally standardized examinations, the same for both professions, guaranteed that fulfilling objective criteria was sufficient to enter the profession. These changes alone caused a dramatic increase in the number of trainees and subsequently advocates and legal advisors. More candidates passed a predefined, standardized threshold on the exam and were accepted to the respective bars, either as advocate or legal advisor trainees (Tables 1.2–1.4).

As demonstrated in the table below the average pass rates increased when the entrance exams were taken over by the state. It attracted more candidates, which further increased the number of law graduates entering the bar training.

Table 2: Pass rates and the number of successful candidates before and after the introduction of the new entrance examinations to the bar training in 2005

	2000-2004	2005-2017
Average pass rate - entrance exam for advocate training	31.6%	49.1%
Average pass rate - entrance exam for legal advisor training	26.9%	41.3%
Average number of candidates who passed the examinations for advocate trainee	293	1,583
Average number of candidates who passed the examinations for legal advisor trainee	760	2,307

Source: Ministry of Justice

First candidates for advocate trainee and legal advisors trainee passed their entrance exams under the new regulation in December 2005 (1,822 out of 4,526, i.e. 40 per cent on average). They started 3.5-year training in 2006 and subsequently sat the new professional examinations in mid-

[24] https://isws.ms.gov.pl accessed 15.02.2018.

2010. 1,423 out of 2,086, or 68 per cent, passed it and could be enlisted as advocates or legal advisors by the end of 2010. Other cohorts followed. Thus, since 2011 can the full effects of the reform of 2005 be observed.

Average pass rates both for the advocate and legal advisor exams stood at 71 per cent (data for 2009–2017). Multiplied by the pass rates for the new entrance examinations to the bar training they can serve to calculate the chances of a law graduate candidate in the entrance examinations becoming (four years later) an advocate or legal advisor, *ceteris paribus*. These are: 35 per cent for the candidates for advocates and 29 per cent for the candidates for legal advisors, which is for both professions three percentage points more than before the reform of 2005.

Furthermore, as a result of the reform, some candidates were exempted from the bar training and an additional group was also exempted from the professional examinations. The exemption from the bar training by itself opened a path to both professions for 1,279 non-trainees who passed the professional examinations (from the first professional examinations in 2009 to 2015). They account for 6 per cent of the total number of lawyers who passed the professional examinations.

Table 3: Number of lawyers who passed the professional examinations with or without prior bar training in 2009–2015

Number of candidates taking:	With a prior bar training	Without prior bar training	Total
Advocate examination	6,953	354	7,307
Legal advisor examination	12,383	925	13,308
Total	19,336	1,279	20,615

Source: Ministry of Justice ; Note: In December 2009 only lawyers without prior bar training could seat professional examinations.

Even more lawyers (4,559) took advantage of the second exemption, namely being exempt from both the bar training and the professional examinations. They account for 15 per cent of the total registered advocates and legal advisors between 2009 and 2015 (Table 1.4).

Table 4: Number of entries into the registries of advocates and legal advisors with and without a prior professional examination in 2007-2015

Number of entries into the registry of:	With a prior professional exam	Without a prior professional exam	Total
Advocates	8,623	2,208	10,831
Legal advisors	16,538	2,351	18,889
Total	25,161	4,559	29,720

Source: Ministry of Justice

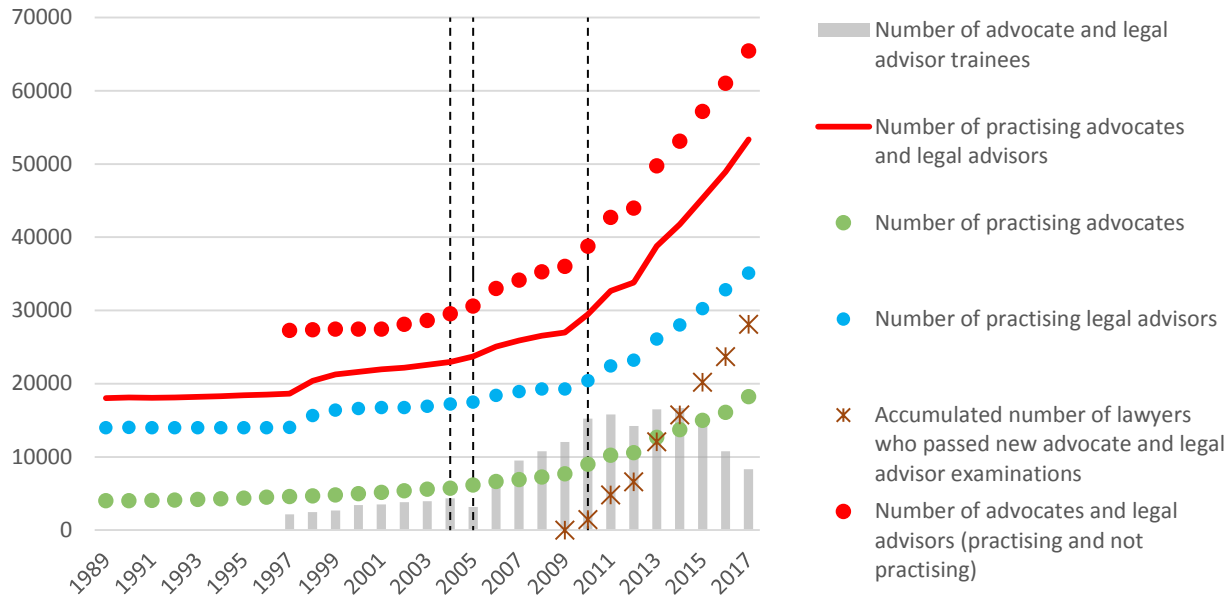
In 2017 lawyers who obtained their professional title according to the amended, more transparent rules, constituted around half of the total number of advocates and legal advisors.

Number of Practitioners and Availability of Legal Service

The number of practising legal professionals (advocates and legal advisors) almost doubled between 2005 and 2015. The number of advocates increased from 6,191 to 15,028 and that of legal advisors from 17,501 to 30,274. Overall, by the end of 2017 the increase is estimated at 125 per cent. The effects of the reforms are clearly visible in the number of candidates for the entrance examination to advocate and legal advisor training. The number fluctuated between 3,000 and 4,000 before the reforms came into force. Opening the access to legal professions impacted the situation dramatically. The numbers have been steadily rising since 2005, reaching over 12,000 in the peak year 2009, although later when they started to decline but remained close to 10,000. These numbers clearly show an increased interest of law graduates in pursuing a career in the profession of advocate or legal advisor. We interpret this as evidence that there was an accumulated potential for growth amongst law graduates who were discouraged from pursuing this career before the reforms, an explanation which closely matches the outcome observed by Ramseyer & Rasmusen (2015) of the reforms of the lawyer profession in Japan (see section 0). This large number of new candidates seeking entry to advocate and legal advisor training

translated into a growing number of trainees. In 2014 over 16,500 law graduates were in the training process.

Figure 1: Growing population of professional lawyers in Poland, 1989-2017



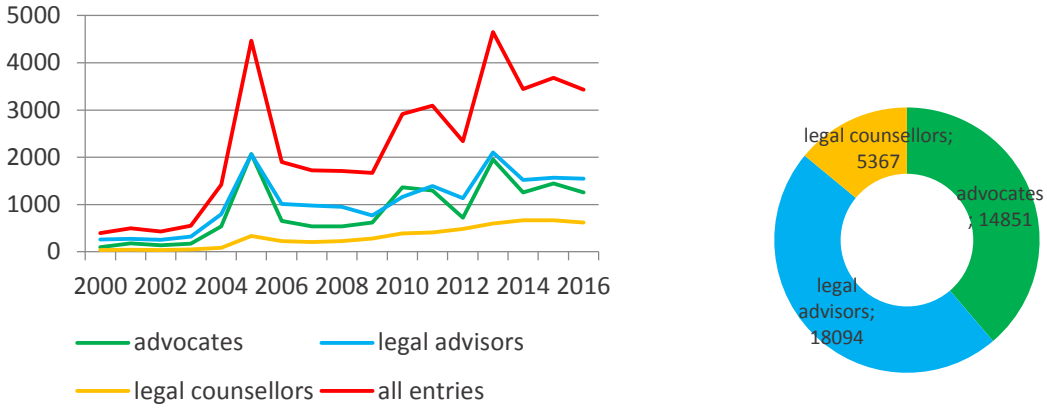
Source: Polish Bar Council, National Chamber of Legal Advisors, Ministry of Justice. *Note*: Dashed lines mark: reforms allowing, inter alia, law graduates to compete with professional lawyers (2004), first standardized entrance examinations for bar training (end of 2005), first standardized professional exams for the new trainees (mid-2010). The number of legal advisors for 1989, 1991-1996, 2016-2017 as well as the number of advocates for 2017 are approximate.

Higher supply of legal services was generally matched by the demand. In 2015 – compared to 2003, i.e. before the reforms were started – the share of non-practising advocates and legal advisors slightly decreased from 21.2 per cent to 20.8 per cent. Had there been any problems with sufficient demand, new entrants would have been forced to search an alternative career path.

At the end of 2003 the number of practising advocates in Poland amounted to 5,623 and the number of practising legal advisors to 16,954. This positioned Poland at the 22nd place in the EU-28 (Council of Europe, 2004) with 59 professional lawyers per 100,000 inhabitants. After the reforms described below their total number rose more than two-fold, to 45,302 (by the end of 2015; and reached 53,350 by the end of 2017). In the CEPEJ report on the European judicial systems, edition of 2016 (2014 data), the availability of professional lawyers rose to 137 per 100,000 inhabitants and Poland ranked 13th in the EU-28 (Council of Europe, 2016). Thus, it almost reached the EU average (147). However, it seems there is still significant potential for growth of the legal profession. There are 391 practising lawyers per 100,000 inhabitants in the United States, 315 in England and Wales and 202 in Germany. Upsurge in the entries to the industry took place exactly after the first reforms of 2004 and 2005 came into force (over 2,000 entries for each profession) and subsequently when the first wave of new trainees obtained their title in 2010–2011 (2013 record number of entries mirrors high enrolment to bar training in 2009).

Similar conclusions can be drawn from the analysis of the number of firms rendering legal activities identified by their statistical code NACE in the registry of entrepreneurs, CEIDG.

Figure 2: New entries of lawyers to the Central Registration and Information on Business (registry of natural persons, CEIDG), 2000-2016 (left panel). Right panel presents the total number of entries in 2000-2016 by the type of legal professional.



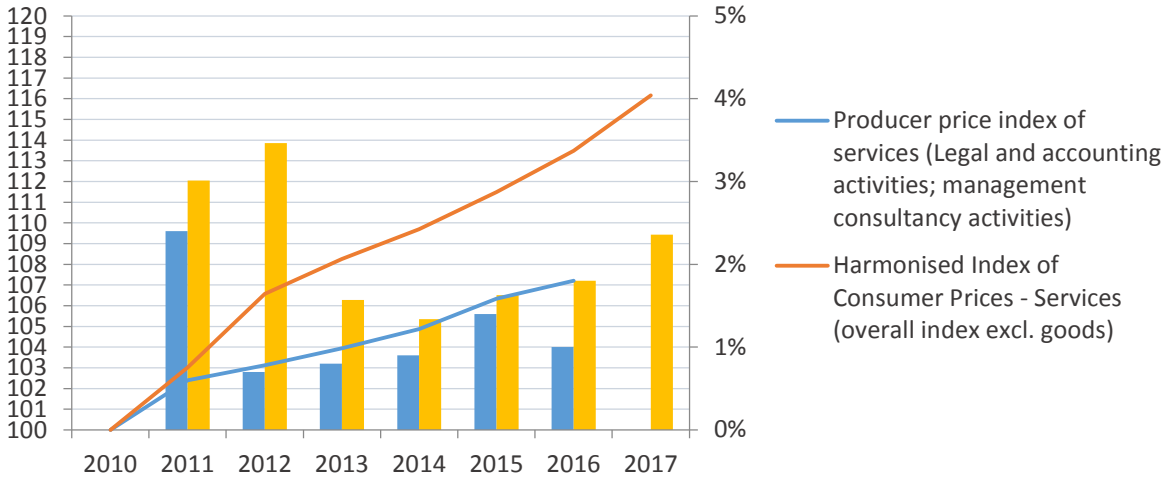
Source: Ministry of Economic Development, CEIDG

The study of consumers – clients of legal professionals – revealed a significant increase in the perceived availability of legal services in Poland with respect to the time before the reforms took place. The share of respondents judging availability of legal services as 'very large choice of lawyers' or 'large choice of lawyers' was 55.8 per cent for those who used legal services before the reforms and 66.6 per cent for those who utilized legal services from the year 2011 onwards. The difference between the availability before and after the reforms appears to be significant at p=0.025.

Price levels

Price levels with respect to the reference year 2010 (=100) reveal a slower growth of prices of legal services (including also services provided by accountants and managers) than the average for all services. The period 2010–2016 corresponds with the inflow of new lawyers following the reforms.

Figure 3: The index of producer prices in services (Legal and accounting activities, management consultancy activities) and the index of consumer services prices (HICP without goods) for Poland, reference period 2010 = 100 (left axis, lines); annual growth rates (right axis, bars).



Source: Central Statistical Office of Poland

Based on the results of the consumer study, the average price of legal representation declined after the reforms. It was estimated at the level of 2486 PLN before 2010 compared to 1826 PLN after 2010. The difference was significant at $p<0.01$.

We conducted an additional study amongst lawyers (how they perceive the reform), gathering a total of almost 300 responses. By answering a specific question, practising lawyers reported a change in real and nominal fees for their services. Of the respondents, 32.16 per cent reported a decline in fees in nominal and real terms, 40.99 per cent declared a decline in fees in real terms, 19.08 per cent reported stability of fees in real terms and 7.77 per cent reported an increase in fees in real terms.

The results presented above indicate that the removal of disproportionate barriers coincided with a substantial increase in the availability of legal services in Poland. The larger availability of services was reinforced by lower price dynamics observed among legal service providers than the ones observed for other services during the period.

Disciplinary Proceedings and Complaints

The dynamics and the structure of complaints and disciplinary cases informed us about the professionalism of lawyers before and after the reforms, at least as measured from the perspective of the most egregious cases.

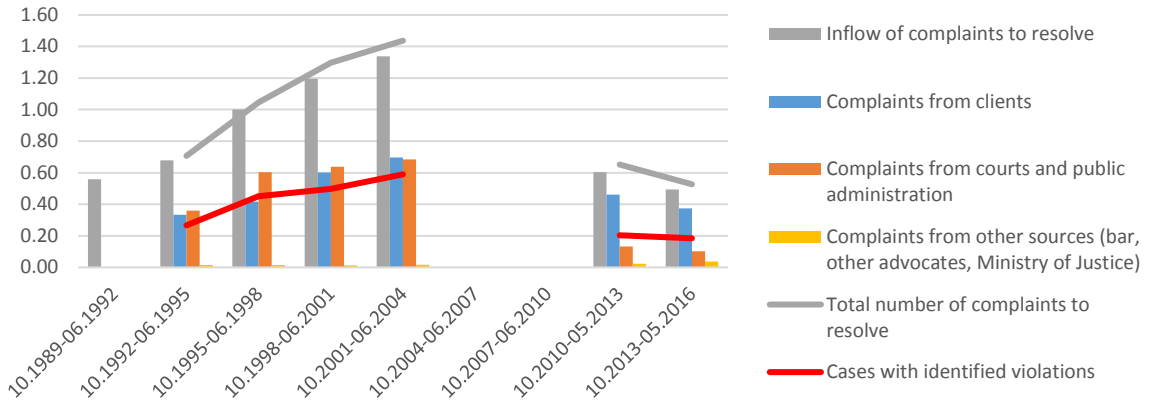
Advocates

Clients dissatisfied with the service received could complain to the bar of their legal representative. In the advocates' bar, the complaints are addressed in the so-called bureaus of complaints, which decides whether there is any sign of a violation of the professional conduct and if so, they pass the cases for investigation to a disciplinary prosecutor, who is also a member of the bar. In the legal advisors' association, a complaint is handled directly by a disciplinary prosecutor who decides whether to initiate a disciplinary investigation or not.

Complaints or notifications of improper professional conduct are filed not only by the lawyers' clients but also by the courts, public administration, authorities of the bar, the Ministry of Justice or other members of the bar (often the opposing party in a legal dispute).

The number of complaints filed against practising advocates (calculated relative to their number) and the number of cases with identified breaches of the rules of professional conduct show that the discontent about the advocates' performance gradually declined between 2004 and 2016. Especially judges and clerks in the public administration were less likely to complain against advocates but this trend was also observed among clients. In the last reporting period 2013-2016 (Figure 1.4), 7,100 complaints were filed to 24 district bar councils. It translates to 0.5 complaints per practising advocate over the three-year period (or 0,16 per advocate per year).

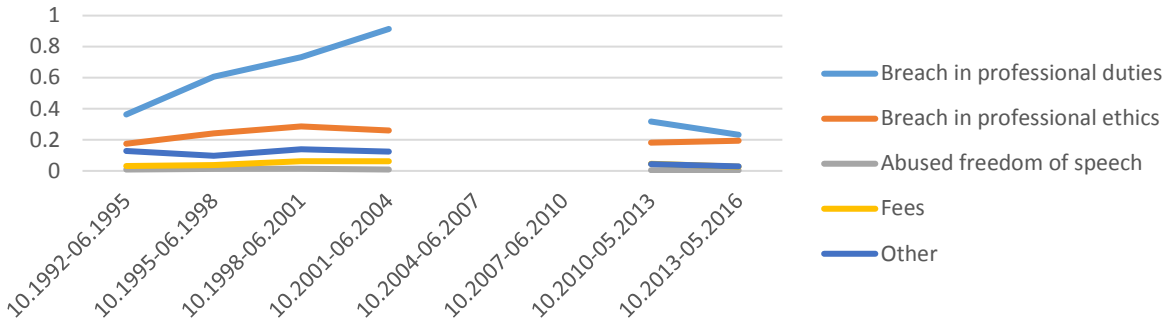
Figure 4: Complaints against advocates filed in all district bar councils of advocates (per practising advocate[25])



Source: Polish Bar Council

The analysis of complaints by the type of reported misconduct indicates that the frequency of the most common allegation, a breach of professional duties, fell by two thirds following the reforms and opening of the profession to newcomers (Figure 1.5). The frequencies of other types of reported breaches, including unethical behaviour, remained stable or slightly declined.

Figure 5: Complaints on advocates by type of allegation (per practising advocate)

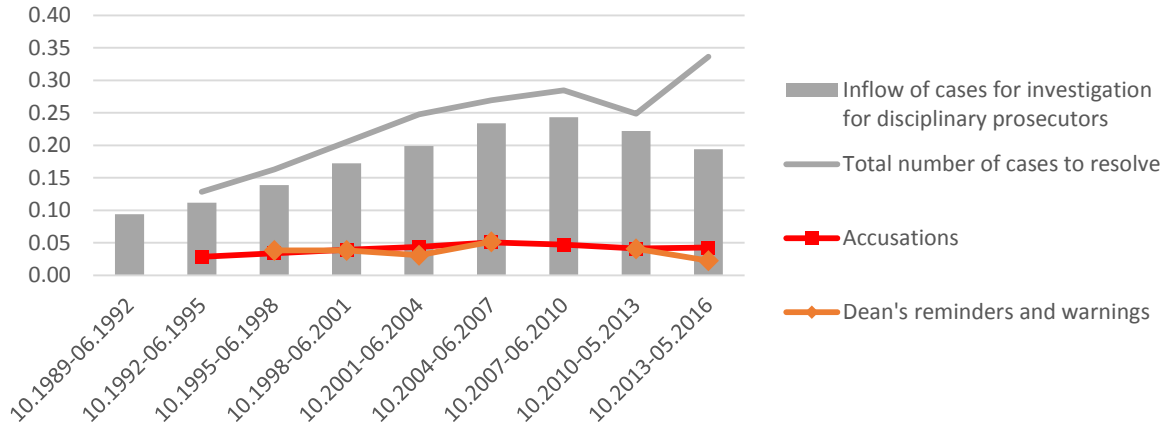


Source: Polish Bar Council

A misconduct was identified by the bar in 2,700 (37%) cases, whereas for the remaining complaints no justification was found. The most troubling cases were usually resolved directly with the complainant, although an increasing number of them was referred to disciplinary prosecutors for further investigation (their share grew from 22% in 1992-1995 to 54% in 2013-2016). The number of motions to initiate a disciplinary investigation was depicted in Figure 1.6. Its peak was noted in the period 2007–2010 (0.24 motions per active advocate). Upon completion of their investigations, disciplinary prosecutors rarely found enough evidence to bring an accusation against an advocate. A decrease in the number of cases for investigation (to 0.19 in 2013–2016) but also slightly lower number of accusations compared to the peak of 2004–2007 (drop from 0.051 to 0.043 accusations per advocate) suggest that despite broader access to the profession, the reform might have disciplined advocates.

[25] The denominator is a three-year average of the number of practicing advocates. The Polish Bar Council reports national data only once in 3 years for the full intervals.

Figure 6: Disciplinary investigations against practising advocates in all bar councils (per practising advocate)

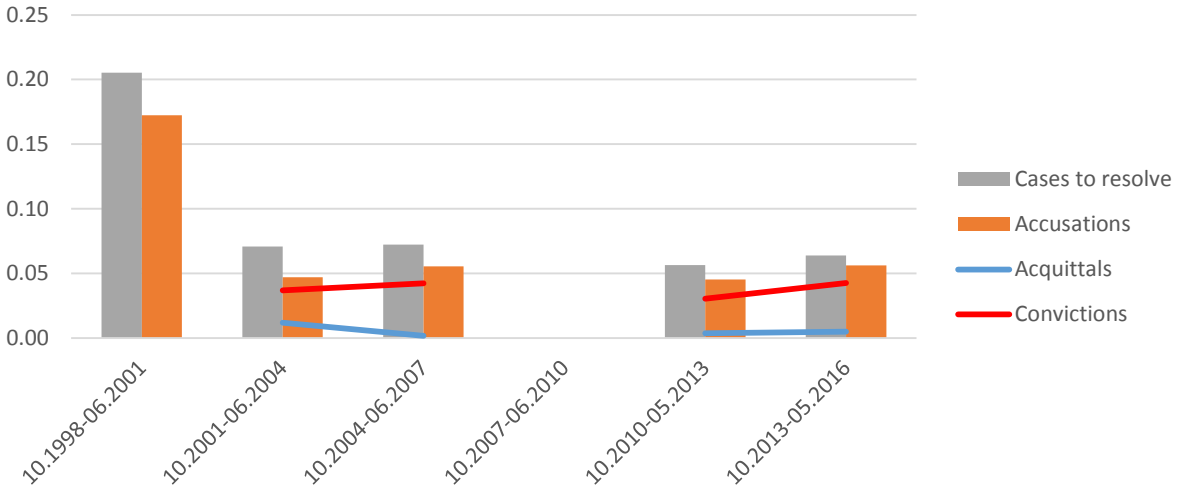


Source: Polish Bar Council

A case of minor importance might have been passed on to a dean of the bar. The dean could issue a warning or a reprimand to a lawyer. Although the lawyer fault was confirmed by this action, he or she avoided the risk of a disciplinary penalty (incl. a fine, suspension or expulsion). The dean could also refer the case to a disciplinary prosecutor. The frequency of dean's warnings and reminders remained quite stable over time.

More serious issues were taken to the district disciplinary courts (Figure 1.7). There was not any significant change in the number of disciplinary cases per advocate over the 2001–2016 period. The same is true for the structure of convictions. The ratio of inflowing cases (around 0.05 per advocate), convictions (0.03–0.04 per advocate), and acquittals (up to 0.01) also remained quite stable.

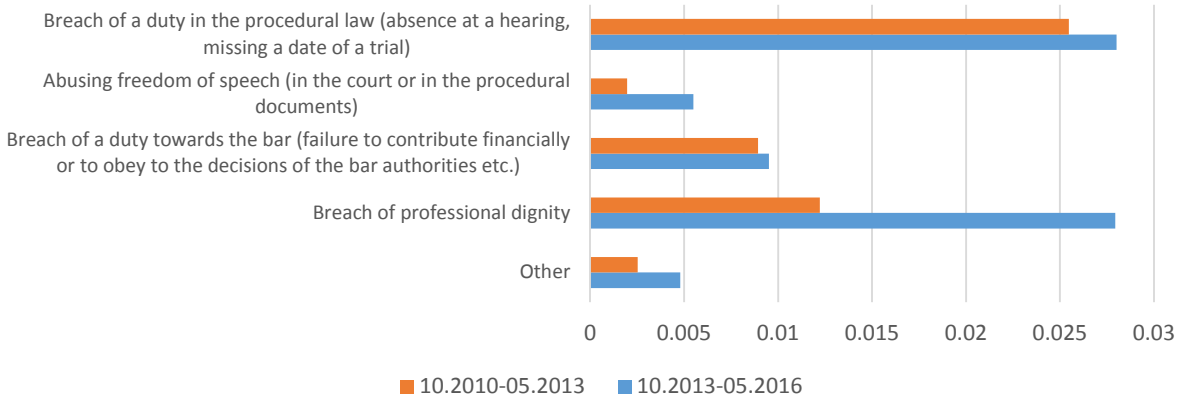
Figure 7: Cases against advocates in the district disciplinary courts including accusations brought by disciplinary prosecutors and the rulings issued by courts (per practising advocate)



Source: Polish Bar Council

The advocates were most often found guilty in the recent years of breaches in the procedural law. Failure to comply with the law put cases of their clients at risk. This type of misconduct includes absence during an obligatory hearing, errors in calculating fees or in the statements lawyers make, either orally or in writing. However, what may be more upsetting is an increase in the number of breaches in professional dignity after 2013 (in absolute terms from 130 in 2010-2013 to 402 in 2013-2016). They usually refer to failures to meet ethical model rules of conduct in various situations. More convictions in this realm may mean that the advocates in general face a fiercer competition due to the entry of younger lawyers to the profession or may be the result of courts' stricter penal policy (the rules of conduct did not change). Yet, it should be underlined that the relative importance of such breaches is marginal. Even taking into account an increase in the incidence of the breaches of duty in procedural law and breaches of professional integrity, still less than 1% of active advocates per year were subject to disciplinary actions related to those infringements.

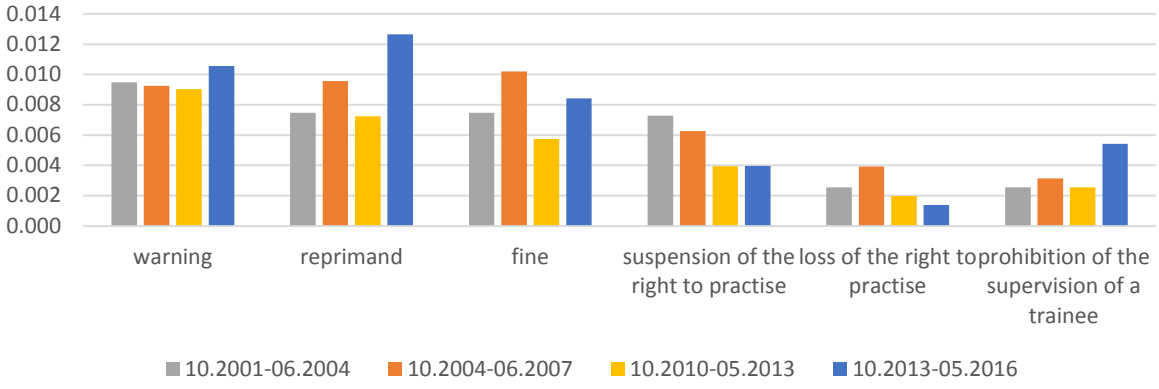
Figure 8: Breaches of the disciplinary rules and principles committed by advocates as identified by the disciplinary courts (per practising advocate)



Source: Polish Bar Council

Relative frequency of penalties issued by the district disciplinary courts (Figure 1.9) indicates that the most common disciplinary measures adopted were those from the lighter end of the spectrum: warning, reprimand and fine. Serious penalties – suspension of the right to practise as advocate or expulsion from a bar – were much less frequent in 2010–2016 than in 2001–2007. The courts were only more likely to issue a prohibition of the supervision of a trainee, which may be interpreted as a precautious move to ensure proper education of advocate trainees and/or a ban on low-cost labour for professional advocates. It might be also consequential to the larger number of trainees. The frequency of penalties has changed little in time.

Figure 9: Penalties issued by the disciplinary courts against advocates in the first instance (per practising advocate)



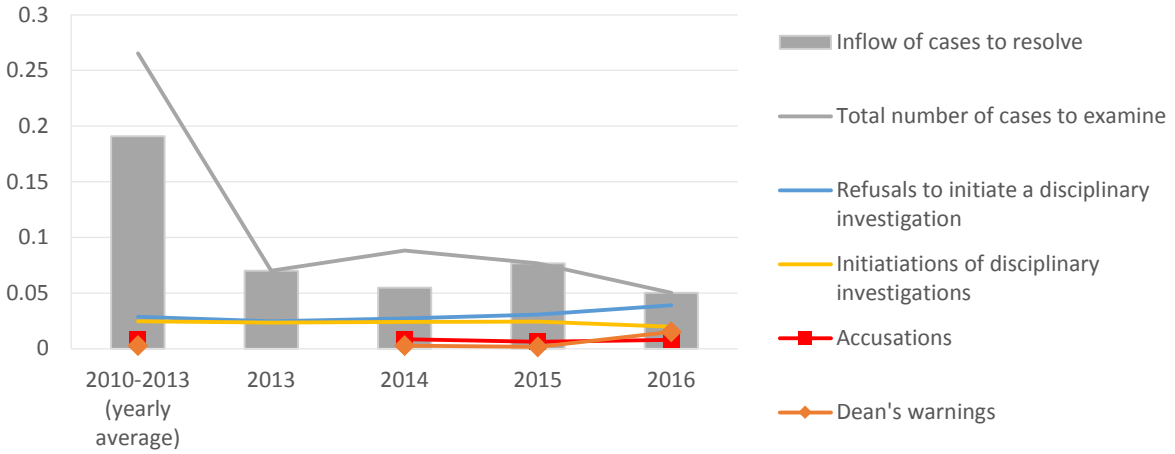
Source: Polish Bar Council

Legal Advisors

The complaints against legal advisors are resolved by disciplinary prosecutors in the district chambers of legal advisors. Their validity is examined, and the prosecutor assigned to the case decides whether to: initiate an investigation, refuse to do so, pass the case to a dean of the chamber (in less severe instances) or to a different chamber, try to reach an amicable settlement or drop the case. Once he or she decides to enquire into a case, he or she can also suspend it or quash an enquiry. Should he or she find enough evidence against a legal advisor, he or she files an accusation to a disciplinary court. Similarly, should the dean conclude that a legal advisor committed wrongdoing, he or she can issue a warning to him or her.

A comparison of the numbers of incoming cases between advocates (0.52 per practising advocate in the period - mid-2013–mid-2016) and legal advisors (yearly inflow multiplied by three, i.e., 0.15 for the three-year period 2014-2016) shows that the frequency of cases to resolve against legal advisors is much lower. One of the reasons may be their lower engagement in litigation. Despite short span of the times series, it is noticeable that there has been a decline in the number of cases filed to the chambers between 2010-2013 and 2013-2016 per practising legal advisor, which may stem from a better conduct of legal advisors. Other indicators of the lack of professionalism, i.e. accusations and warnings did not exhibit any significant variation (Figure 1.10).

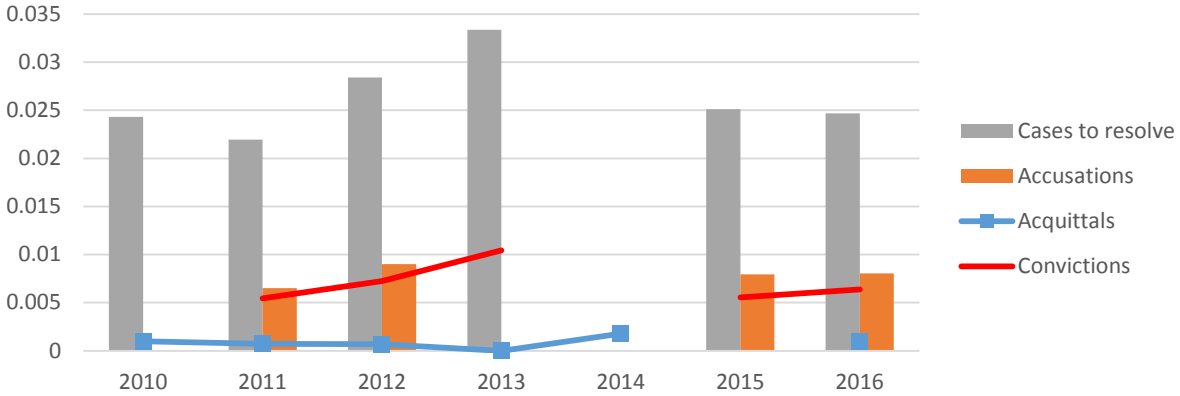
Figure 10 Number of cases to investigate by the district disciplinary prosecutors (per practising legal advisor)



Source: National Chamber of Legal Advisors

District disciplinary courts of legal advisors received 500–900 cases yearly during the period 2010–2016 (Figure 1.11). The inflow remained steady at the level of 2.2 per cent to 3.3 per cent relative to the population of active legal advisors. Most of the inflow should be attributed to the appeals from their refusals to accuse (1.2–1.6 per cent).

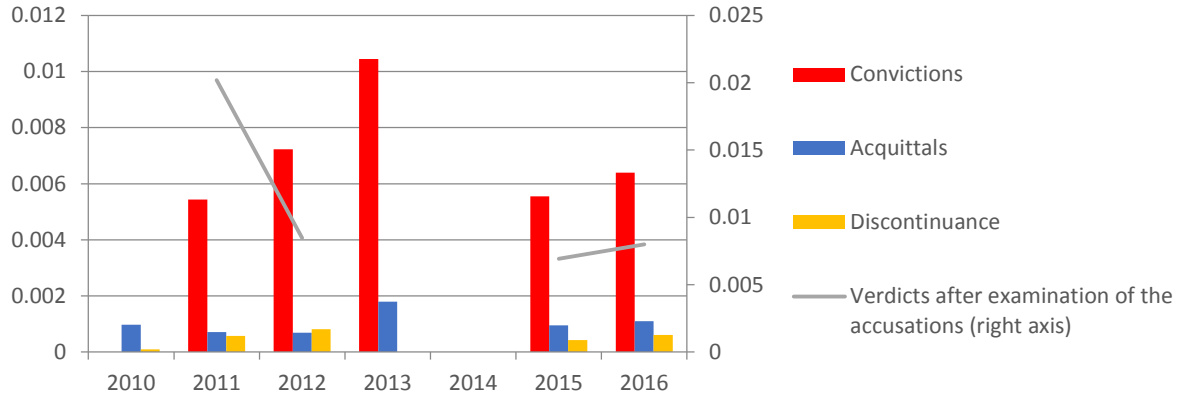
Figure 11: Inflow of cases to district disciplinary courts of legal advisors (per practising legal advisor)



Source: National Chamber of Legal Advisors

The structure of rulings issued by the disciplinary courts (Figure 1.12) indicates that they mostly corroborated findings of the disciplinary prosecutors: the number of acquittals was low, amounting to 27 on average per year (0.001 per practising legal advisor), similar to the number of rulings of discontinuation of the proceedings – 13 on average per year. Consequently, the number of convictions per year stood on average in 2010–2016 at around 0.007 per active legal advisor. These statistics remain relatively stable in time suggesting no deterioration in quality. Additionally, the share of legal advisors affected by those rulings is extremely low suggesting that the problem is limited and its extent for sure cannot point towards any deterioration following the reforms.

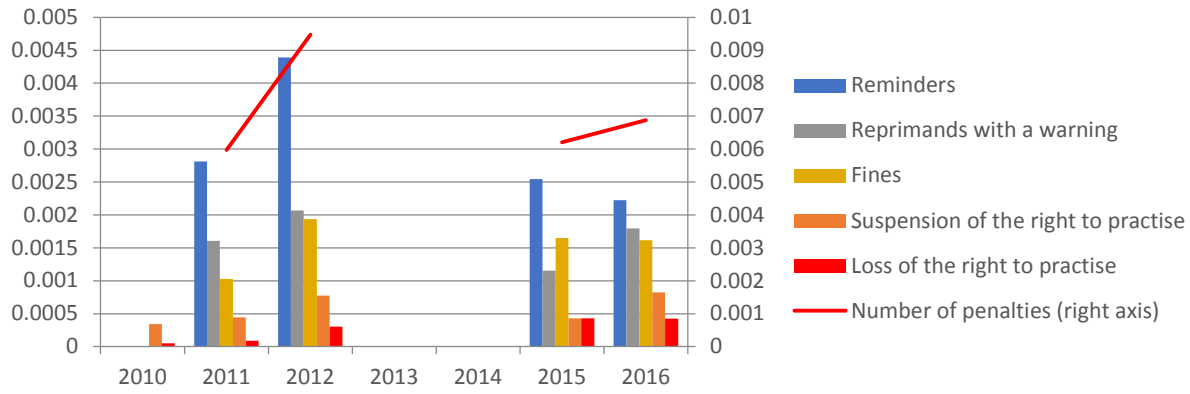
Figure 12: Rulings of district disciplinary courts (per practising legal advisor)



Source: National Chamber of Legal Advisors

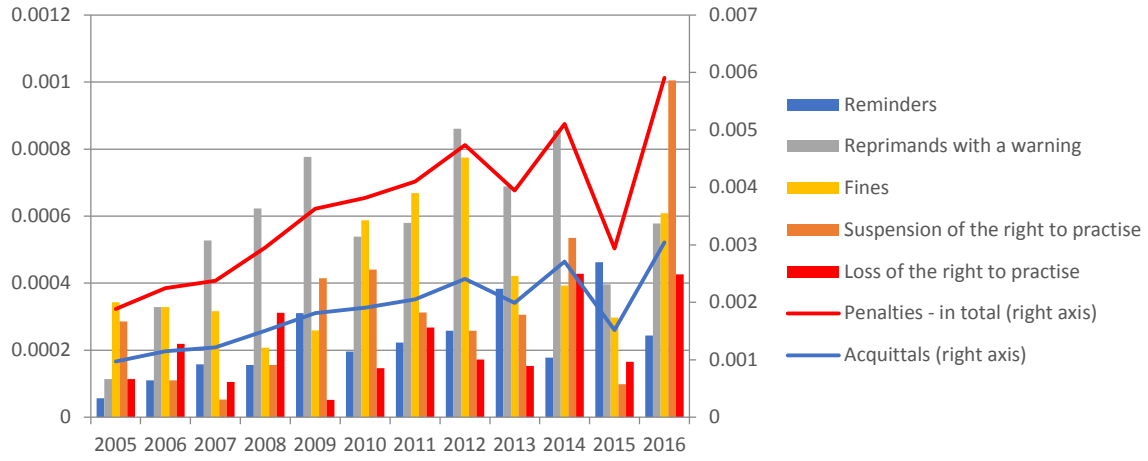
In the case of conviction (Figure 1.13), courts tended to impose reminders (the most frequent penalty over the 2010–2016 period: 60–100 per year), then reprimands with warnings (ca. 40–60 per year) and fines (ca. 20–50 per year), i.e. light penalties. However, the number of penalties remained steady at around 0.007 per active legal advisor in a given year (or 130–230 in absolute terms).

Figure 13 Penalties imposed by district disciplinary courts on legal advisors ruling in the first instance (per practising legal advisor)



Source: National Chamber of Legal Advisors

Figure 14: Rulings of the Higher Disciplinary Court (per practising legal advisor)



Source: National Chamber of Legal Advisors

In the second instance, the Higher Disciplinary Court of legal advisors verified the rulings issued by the district disciplinary courts (Figure 1.14). The analysis of the structure of its rulings from 2005 to 2016 indicates an increase in the ratio of issued penalties between 2005 and 2012 (from 33 to 110 in absolute terms or from 0.002 to 0.005 per active legal advisor). Reprimands with a warning, fines and – recently – suspension of the right to practice are the most commonly applied penalties. Rising incidence of stricter penalties may also be the result of the policy of the court. However, also in the case of the second instance rulings, one might notice that the percentage of legal advisors affected by those rulings is extremely low. Although the number of penalties per legal advisor went up from 0.002 to 0.005, one needs to bear in mind that it is still merely one in every two hundred legal advisors that has been affected by a penalty during the last year.

To sum up, we conclude that:

- Complaints are rarely filed against advocates. Only a small minority of complaints is found justified by the bar councils of advocates.
- A very small ratio of complaints on advocates and legal advisors lead to an accusation and later to a conviction by a disciplinary court and a penalty or a dean's warning.
- The number of complaints per practising advocate declined across all groups of complainants following the reforms. A decrease is particularly significant for the complaints from courts and public administration, which may point to a higher assessment of professionalism of advocates.
- Despite the fact that relatively lower share of complaints on advocates referred to the breaches of professional duties in 2013-2016 than in 2010-2013, the advocates were found guilty more often of this type misconduct. More recently an increase in the number of reported breaches of professional ethics was observed. It may point to a minor deterioration of the quality of advocates' work in this respect.
- Despite an increased inflow of cases for investigation to the disciplinary prosecutors in the bar councils of advocates between 2004 and 2013, the number of accusations remained quite stable. It may point to little substance in those accusations.
- The relative number of penalties imposed on advocates did not change over the 2001–2016 period. Their structure remained quite similar although the least severe penalties (warnings and reprimands) were imposed slightly more frequently. The harshest penalty, loss of the right to practise, has been imposed only around 100 times since 2001.
- For legal advisors: neither the dynamics of cases for investigation by the district disciplinary prosecutors in 2013–2016 nor the dynamics of cases registered in the district disciplinary courts in 2010–2016 exhibit any upward or downward trend. The share of acquittals and convictions in the rulings of the district disciplinary courts between 2010 and 2016 remained quite stable, pointing to no deterioration in service quality.

- The time series of rulings of the Higher Disciplinary Court of legal advisors in 2005–2016 allows for an identification of a stable upward trend in the number of penalties issued during the period 2005–2010. However, since effects of the reforms became apparent (i.e. 2011) the number of penalties has been fluctuating, not exhibiting any particular trend.

First limitation: the lack of data on disciplinary proceedings in the district disciplinary courts of legal advisors for the period before 2013 implied that the analysis for this profession (accounting for two thirds of the lawyers in Poland) was necessarily limited to the investigations by disciplinary prosecutors and the proceedings at a national level before the Higher Disciplinary Court. Comparability with the period before the reforms was not possible.

Second limitation: there are no similar disciplinary data for non-professional lawyers, i.e. legal counsellors, who given their lack of prior bar training and the findings of the study among judges, seem to perform in a less professional manner.

Results from the Consumer Survey

The study of consumer satisfaction based on the SERVQUAL methodology covered 720 individuals. They used legal services during one of the following periods: 2000–2005, 2006–2010, 2011–2016 or 2017. The survey sample included 180, 169, 267 and 63 individuals reporting the use of legal service of legal advisors or advocates in the respective periods. 41 individuals reported experience with notary services – this group was excluded from further analysis.

The set of indicators from the SERVQUAL study of legal services quality, following the methodology described in detail in section 1.5.2, was divided into indicators referring to specific domains. The assignment of indicators to specific domains is described in Table 1.5. Following the construction of each individual scale, an overall indicator of quality, based on six dimensions of quality as presented in the SERVQUAL methodology for legal service quality, was constructed.

Each domain of legal service quality was covered by a specific set of indicators. The number of indicators ranged from three for *clarity* and *efficient service provision* to as many as 12 indicators for *professional appearance*. For each set of indicators, dimensionality assessment was performed and the reliability of the indicator for each specific dimension was assessed (Table 1.6). For this purpose, first, Cronbach's alpha, and subsequently, principal component analysis were applied. Cronbach's alpha is a measure of reliability of scale. It is also known as an internal consistency estimate of reliability of test scores. Because intercorrelations between items (question responses) are maximized for items depicting the same construct, Cronbach's alpha should indirectly indicate whether there is a single unidimensional latent construct behind the data. In order to calculate indicators for each dimension of quality, principal component analysis was used. First principal component served as a proxy for individual level satisfaction from the quality of legal service.[26]

[26] As not all indicators were measured in both parts of the study (CATI and CAWI), multiple imputation techniques were used to account for the missing of data. Thanks to the use of multiple imputation techniques, it was possible to obtain consistent indicators in each of the domains of legal service.

Table 5: Allocation of indicators to different dimensions of legal service quality

Dimension of quality	Indicators
Empathy	I trusted my lawyer
	The lawyer was honest
	The lawyer was good-mannered
	The lawyer was kind
	The lawyer was empathic
Efficient services process	It was easy to meet the lawyer
	It was easy to reach the lawyer by phone
	It was easy to contact the lawyer by email
Achievement of the desired results	I am satisfied with co-operation with the lawyer
	I am happy with outcome of the case
	I could recommend the lawyer
	Service quality exceeded my expectations
Clarity	It was easy to understand the lawyer
	The costs were defined upfront
	The lawyer provided me with information during the case
Active use of legal knowledge	The lawyer was engaged in the case
	The lawyer was efficient
	The lawyer informed me thoroughly about the case
	I could co-decide on the development of the case
Professional appearance (physical characteristics of service)	The lawyer worked professionally
	The lawyer was knowledgeable in his field
	The lawyer had a profound experience
	The law firm made a good impression
	The lawyer had professional appearance
	The lawyer operated innovatively
	The lawyer was respecting the deadlines
	The lawyer was present during the trials
	The lawyer kept his promises
	The lawyer behaved ethically
	The lawyer was expressing in a balanced way
	I had no problems with the settlement

Table 6: Reliability assessment in different dimensions of quality of legal services

Dimension of quality	Cronbach's alpha	Percent of variance explained by the first principal component
Empathy	.877	67.32
Efficient services process	.749	67.28
Achievement of the desired results	.881	74.34
Clarity	.729	65.84
Active use of legal knowledge	.816	64.59
Professional appearance (physical characteristics of service)	.913	52.15
Overall quality indicator	.952	80.75

All scales produced sufficiently high levels of Cronbach's alpha. The Cronbach's alpha for each scale was above 0.7 and for the scale of *professional appearance* it was as high as 0.913. We also observed a very significant contribution of the first principal component to the total variance. It further points towards unidimensionality of the measured concept. The lowest share of explained variance by the first principal component was observed in the dimension of *professional appearance*. However, taking into account that this dimension consists of 13 indicators, 52.15 per cent of the variance explained by the first principal component was still large. In the case of *achievement of the desired results* the first principal component was responsible for explaining almost ¾ of the total variance. Even higher overall quality of the indicator was observed for the overall score. The indicator constructed based on the six dimensions of legal service quality from SERVQUAL was characterized by the Cronbach's alpha equal to 0.952 and the percentage of variance explained by the first component was over 80 per cent.

In order to measure differences in perception of the six dimensions of service quality over time, propensity score matching was used. Propensity score matching (PSM) (Heckman and Ichimura 1998) attempts to eliminate selection bias by conditioning outcome on confounding variables. A set of control variables, potentially influencing perception of the legal service, was introduced to the analysis. Specifically, we control for sociodemographic variables gender, education and age. These have previously been shown to correlate with psychological well-being (Grossi, Blessi, Sacco, & Buscema, 2012), which might also be relevant to the subjective assessment of the quality of legal service. Additionally, we controlled in the analysis for the characteristics of the legal service. The assessment of legal service quality was performed by respondents who were matched with respect to the form in which they used legal services (business or individual), character of the case for which they took a lawyer (civil, family, penal), his/her age, and the cost of service.

Based on their propensity scores, statistical twins were found using the five nearest neighbours matching algorithm with replacements. This strategy allowed balancing of the treatment and control groups with respect to observed characteristics (details in Appendix 1.3) and the possibility of calculating average treatment effects on the treated (ATT): a measure of difference between a quality outcome observed for the treatment group (legal service obtained in 2011–2017) and the control group (legal service obtained in either 2000–2005 or 2006–2010). The results of average treatment effects (ATT) are presented in Table 1.7.

Table 1.7 depicts average treatment effects, where treatment is defined as using legal services in the period 2011–2017 and the control group is either those who used legal services between 2000 and 2005 or those who used them between 2006 and 2010. The major finding is that from the consumer perspective there were almost no discernible differences in the quality of legal services before and after the reforms. Not only did we not observe a change in the overall quality but also within given dimensions there were no significant differences between the periods before and after the reforms took place. Also, for specific questions there were no visible differences between before and after the reforms (details presented in Appendix 1.4).

The results, although showing lack of any change in the perception of quality of legal services, are informative for the study. As we matched respondents using variety of their characteristics but also variety of characteristics of the service and the lawyer who provided the service, we have a quite

valid position to establish lack of change in the quality of legal service as perceived by clients of those services.

Table 7: Change in customer satisfaction with the legal services over time

Outcome of legal service	Average treatment effects		95% confidence intervals	
	2011-2017 (treatment) vs. 2000-2005 (control)	2011-2017 (treatment) vs. 2006-2010 (control)	2011-2017 vs. 2000-2005	2011-2017 vs. 2006-2010
Empathy	-0.026	0.035	(-0.312;0.259)	(-0.187;0.257)
Efficient services process	-0.113	-0.042	(-0.386;0.161)	(-0.274;0.189)
Achievement of the desired results	-0.071	-0.113	(-0.376;0.234)	(-0.342;0.115)
Clarity	-0.055	0.025	(-0.33;0.221)	(-0.202;0.252)
Active use of legal knowledge	0.026	-0.013	(-0.265;0.317)	(-0.224;0.198)
Professional appearance (physical characteristics of service)	-0.023	0.029	(-0.31;0.264)	(-0.183;0.242)
Overall quality	-0.046	-0.013	(-0.338;0.245)	(-0.222;0.195)

Note: For each construct a standardized first principal component was used. Consequently, the units of measurement are standard deviations of respective concepts in the population. Due to the scale orientation, positive values indicate improvement between the reference period (either 2000-2005 or 2006-2010) and the period when the reforms already entered into force (2011-2017).

Results from the Survey of Judges

The survey of judges[27] concentrated on three broad areas in the assessment of change in the quality of legal services between the period prior to 2006 (prior to the reforms) and present. Our respondents – judges – first evaluated change in the level of professionalism of various groups of legal professionals between today and the period prior to 2006. Then they assessed various qualities of legal service as perceived from their perspective. Finally, they focused on accessibility of legal services as measured by the propensity to appear in front of a judge without qualified legal representation. In addition to the investigation in these areas, we also provided judges with an open question, where they could express their opinions about changes in the quality of legal services. The results of responses related to the level of professionalism are presented in Table 1.8.

[27] The results were based on opinions of 60 judges. The sample size is not significantly different than sample sizes used in surveys oriented on gathering expert opinions. In economics sample size in surveys of professionals usually oscillates around 50 (see, e.g., Chew & Price, 2008). In medicine surveys based on as little as 24 opinions have been published in prestigious journals (see, e.g., Slick, Tan, Strauss, & Hultsch, 2004). It has a profound justification. Experts are much more reliable in their opinions and have a very good understanding of procedures and, as in the case of judges, legal framework.

Table 8: Assessment conducted by judges of professionalism of various legal service providers in Poland

Group of legal professionals	Period of interest	Professionalism (% of response in the given category)					P-value Stuart–Maxwell test for marginal homogeneity
		High	Rather high	Average	Rather low	Low	
Advocates	Nowadays	6.67	38.33	50	5	0	.3381
	Before 2006	5	55	35	3.33	1.67	
Legal advisors	Nowadays	5.17	29.31	48.28	13.79	3.45	.7831
	Before 2006	3.57	21.43	55.36	16.07	3.57	
Legal counsellors	Nowadays	0	0	20.83	43.75	35.42	.375
	Before 2006	0	0	21.74	34.78	43.48	
Advocate trainees	Nowadays	1.67	20	60	13.33	5	.1393
	Before 2006	0	33.33	53.33	11.67	1.67	
Legal advisor trainees	Nowadays	0	19.3	52.63	19.3	8.77	.2517
	Before 2006	0	19.64	55.36	17.86	7.14	
Individuals without representation	Nowadays	0	0	29.31	43.1	27.59	.0037
	Before 2006	0	0	17.54	42.11	40.35	

Note: The p-values show the difference in distribution between the period before 2006 and today.

We used the Stuart–Maxwell test (Maxwell, 1970; Stuart, 1955) for the comparison of shares derived from paired assessment of legal service quality. The results clearly indicate that there was no change in professionalism following the reforms. Tests for all groups of professionals yield insignificant results. There is only one outlier in the data. It seems that, although assessed at a rather low level both today and in the past, the professionalism of individuals without legal representation improved significantly. It is probably due to improved legal awareness but also much better access to legal knowledge by outsiders.

We gathered opinions of judges about specific indicators of quality. The comparison between the perception of current performance of legal professionals and their performance before 2006 (i.e. before the reforms) is depicted in Table 1.9.

Table 9: Assessment conducted by judges of various legal service quality indicators by legal representatives

Indicator of legal service quality	Period of interest	Frequency of action (%)				P-value Stuart–Maxwell test for marginal homogeneity
		Always	Usually	Rarely	Never	
Actions in the best of client's interest	Nowadays	18.33	71.67	10	0	.5215
	Before 2006	25	70	5	0	
Good command of legal procedures	Nowadays	1.67	76.67	21.67	0	.034
	Before 2006	13.33	71.67	13.33	1.67	
Good command of substantive law	Nowadays	3.33	86.67	10	0	.1156
	Before 2006	15	73.33	11.67	0	
Selection of effective tactic	Nowadays	1.67	66.67	28.33	3.33	0.768
	Before 2006	1.67	81.67	16.67	0	
High care in drafting pleadings	Nowadays	8.62	63.79	27.59	0	1.00
	Before 2006	8.62	63.79	25.86	1.72	
High skills in formulating claims, demands and allegations	Nowadays	1.67	63.33	35	0	.3083
	Before 2006	8.33	61.67	30	0	
High skills in questioning witnesses and expert witnesses	Nowadays	3.33	35	61.67	0	.0044
	Before 2006	5	56.67	38.33	0	
High care in running cases assigned by court	Nowadays	6.78	47.46	40.68	5.08	.1126
	Before 2006	3.39	40.68	52.54	3.39	
High care in running cases of choice	Nowadays	18.33	76.67	5	0	1.00
	Before 2006	16.67	78.33	5	0	
Impeccable manners before a court	Nowadays	5	70	25	0	.000
	Before 2006	43.33	55	1.67	0	
Impeccable manners in the relationship with clients	Nowadays	12.28	80.7	7.02	0	.0781
	Before 2006	24.56	75.44	0	0	
Proper contact with other participants of proceedings	Nowadays	1.75	64.91	33.33	0	.0144
	Before 2006	8.77	73.68	17.54	0	
Respecting	Nowadays	20	71.67	8.33	0	.718

deadlines	Before 2006	21.67	65	13.33	0	
Good instructions for the client	Nowadays	1.72	53.45	44.83	0	.3018
	Before 2006	3.45	60.34	34.48	1.72	
Good instructions for trainees how to represent a client	Nowadays	1.72	51.72	44.83	1.72	.5298
	Before 2006	1.72	62.07	34.48	1.72	
Courage in reacting to abnormalities	Nowadays	15.25	50.85	30.51	3.39	.2586
	Before 2006	11.86	44.07	42.37	1.69	

Note: The p-values show the difference in distribution between the period before 2006 and today.

Out of the 16 indicators presented to judges for assessment, 12 of them showed no significant differences in the perception of a given skill or given action. Most notably, there were no differences noted in the good knowledge of material law, the ability to formulate efficient tactics and draft pleadings. However, for four indicators significant differences were noted. It appears that the lawyers providing services today have slightly less knowledge of the procedure and are on average slightly less skilled in questioning witnesses. Additionally, opinions of judges demonstrate that the legal representatives today more often act less courteously before a court than prior to 2006[28]. They also indicate that the legal representatives today do not keep to the same standard of contacts with other participants of the proceedings as they did few years ago. With the exception of a decline in good manners before the court, those differences, although significant, are not substantial. Lower quality of questioning requires scrutiny and its negative evolution should raise concerns. Nevertheless, it should also be noted that following the reforms, a significant inflow of new professionals resulted in the decline of the average age of a legal professional. This in turn implies that the legal representatives have on average less experience than their counterparts from the period before 2006.[29]

Finally, using the data set collected from judges, we also compared the average perception of accessibility of legal services today and before 2006. The results of responses to the qualitative question are presented in Table 1.10.

Table 10: Assessment conducted by judges of professionalism of various legal service providers in Poland

Accessibility	Agreement					P-value of the t-test for no change
	Strongly agree	Rather agree	Neither agree nor disagree	Rather disagree	Strongly disagree	
The individuals without legal qualifications appear in court more often today than before 2006.	1.69	25.42	30.51	25.42	16.95	.0352

Note: The p-values show whether the difference between average score of 3 (neither agree nor disagree) and the actual score is significantly different

[28] In particular, the judges were asked to assess how frequently they encounter the legal representatives characterised by impeccable manners (translating directly from Polish: « high personal culture ») in their attitude towards the court. This question was intended to capture the extent to which the lawyers display professionalism and courtesy when addressing the court.

[29] Due to the lack of detailed data we could not have compared legal professionals with the same experience as those in 2005 or earlier.

The results indicate that the access to legal representation has improved. Judges disagree with the claim that 'today more people appear in court without a legal representative'. The average response to this question is significantly below the 'neither agree nor disagree' threshold (p-value=.0352). It confirms broader access to legal services in Poland.

To sum up, the results obtained from the survey on judges corroborate earlier findings of the lack of change in the professionalism between before and after the reforms. Among the indicators of quality of legal services there is a strong consensus pointing to no change in the average quality. However, judges' opinions point to a higher frequency of less courteous behaviour of lawyers towards the judges. Out of the professional quality indicators judges point to lower skills in questioning witnesses and expert witnesses but also to the worse knowledge of the procedures. These indicators might be connected, however, to the decline in the average number of years of professional experience of lawyers today. Due to a substantial influx to the profession, the average number of years of professional experience decreased. Consequently, this decrease in the perception of quality in experience-related skills should not be associated with a lower general quality of legal professionals but rather with a changing age structure within the representatives of the profession.

In response to the question 'If there is any additional area where you have observed changes and would like to share your comments, please do', 21 out of 60 judges provided their thoughts on the consequences of the reforms, mentioning various behaviours. Some of them drew attention to currently observed lower speaking and writing competencies of lawyers and trainees. In particular, they were thought to exhibit lack of clarity and precision in their (lengthy) court statements as well as inadequate and excessive references to previous court rulings and quotations from legal literature, although the later might be driven by the growing body of law in force.

Many of those who provided general remarks complained about the disrespectful attitude of lawyers towards the court and other participants of the trial, treating cases of their clients as though they were their own, which leads to a lack of professional distance in a case, putting forward delayed motions to adduce further evidence, delaying court proceedings by unnecessary motions or defending a party in a criminal trial at any price. These complaints, (although might be considered troubling from the perspective of judges as they result in a larger amount of workload) should not necessarily be considered as indicators of lower quality in general but rather as representing higher engagement of lawyers with their case.

According to judges, lawyers are now more likely to mislead a court (being aware that the facts are different), act disrespectfully towards a victim or a witness, ignore obligations assumed on them by a court in order to delay a ruling, be un-loyal to other parties (ignore obligations towards them) and abuse procedural intricacies in order to overthrow a ruling without a prior notice to the court etc.

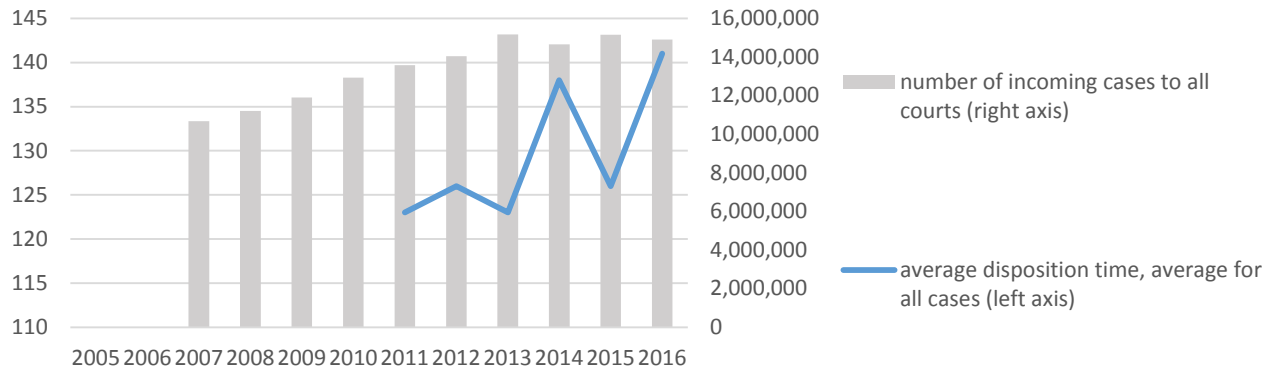
Furthermore, two judges pointed to the insufficient abilities of legal advisors to act as a defender in criminal cases (they obtained this right only in mid-2015), which harms their clients. One judge noticed that the trainees working in big law firms and acting as substitutes for the professional lawyers are often unprepared for the trial as opposed to their counterparts from smaller law offices. Small law firms acted before 2006 and still act very professionally in his or her view (no matter what the age of an advocate or legal advisor who works there), whereas big law firms are more likely to accept the cases that cannot be won. These clients, solicited by the big firms, are neither informed about the small chances of winning nor about the consequences of failure, and the rules of a civil trial or the expectations the court would have towards them.

There were also substantial arguments supporting the reforms. Two judges were convinced that the regulatory changes contributed to a more competent legal aid. One of our respondents was convinced that thanks to the internet clients he or she became more aware of the law and thus more knowledgeable in distinguishing competent from incompetent lawyers.

Results from the Analysis of the Court Data

After a surge from 2007 to 2013, the number of incoming cases to Polish courts stabilized at around 15 million yearly. The length of proceedings for all cases in Polish courts (excluding administrative courts and the Supreme Court) increased slightly between 2011 and 2016. In 2016, it took on average 4.6 months to conclude a case as opposed to 4.1 months in 2011. However, there is no discernible trend over the past years (Figure 1.14).

Figure 15: Number of incoming cases to Polish courts and the disposition time (length of proceedings), 2005-2016.



Source: Ministry of Justice

The data on lengthiness of court proceedings cannot serve as an indicator of legal service quality because there are numerous factors beyond the control of legal professionals that affect this outcome. However, a decline in the share of convictions in criminal cases may imply both that legal aid became more available after 2010 and/or that legal representatives started to be better prepared to serve their accused clients. It is pointing to better availability and/or better quality of legal service at the onset of the reforms.

1.7 Conclusions

The literature review on the regulation of legal services presents two contrasting views. A major strand depicts regulation as a gatekeeper to safeguard consumers against inadequate legal expertise. Yet, the second describes regulation as a tool for helping the private interests of legal professionals. Although most arguments seem to point towards prevalence of private interests, this aspect has not been subject to sufficient empirical scrutiny. The studies discussed in the literature review are critical of the most common approach to ensure quality in legal professions, namely measuring inputs. Despite these concerns in the literature, input quality is still the prevalent approach adopted by bar associations for member selection. Measuring quality of legal services via evaluation of processes and outputs provides a complementary approach to understanding the relationship between professional regulation and quality.

We used a number of tools to study the quality of legal services in Poland. The primary methodology for our study involved the use of surveys to obtain indicators proxying quality. We indirectly measured consumer satisfaction with legal services based on the SERVQUAL approach. The resulting analysis revealed little change in the quality of legal services following the reforms. More specifically, while the demand for legal services increased rapidly following the improvement in the financial position of Polish households, to maintain a relatively stable cost of the service, the supply had to keep up. Matching the rapid increase in demand with a supply increase but without any deterioration in the quality of service is probably a major success of the reforms on the access to legal professions in Poland. The findings from the SERVQUAL study of consumers were also supported by the analysis of the most severe instances of lawyers' misconduct, reflected in the complaints and disciplinary proceedings. The declining number of complaints per active advocate, especially those originating from courts, also marks a significant professional improvement that accompanied the reforms. Performance of legal advisors matched that of advocates. The survey of judges served as an additional tool for measuring the quality of legal services in Poland. It detected no change in legal professionalism and most of the quality indicators remained unchanged from the period prior to the reforms. However, in the opinions of judges, some quality indicators deteriorated. This was largely reflected in the measures of good manners of the legal representatives in court but also touched on certain skills that are closely related to experience. However, slightly lower performance in these areas might be expected among lawyers recently admitted to the professions.

The results remain tentative at this stage but some provisional conclusions might be appropriate. Firstly, the Polish reform case indicates no evidence of a decline in quality following a two-fold increase in the number of legal professionals. However, one should be cautious in extrapolating. In Poland, the rapid increase in the number of legal professionals occurred from a very low base. With an initial low number of legal professionals, a large talent pool was available. In these

circumstances, it could be argued that opening up of legal professions could be conditional on the availability of a large pool of good candidates, especially if the profession deals with highly specialised matters.

Secondly, the conditions for liberalization were especially favourable in Poland because the rules for admission to the legal professions before the reforms were mostly arbitrary and the access could have been restricted without any justification other than the private interest of lawyers. The proportionality principle requires that a profession intending to keep regulation in place should provide clear justifications for restrictions to entry, which was not the case in Poland. Moving partially away from pure self-regulation to a form of co-regulation proved successful. It urged law associations to perform better to live up to the expectations of new trainees, new members of the profession and, especially, their clients. When contemplating different regulatory models it has to be borne in mind that lawyers play an important role in the proper administration of justice and the rule of law in a state, and their independence should be ensured.

Finally, the impact of the external factors facilitating entry of new lawyers to the market and resulting in growing demand for their service should also be taken into account. The developing economy, adoption of the EU law and the rapid growth in the volume of the national legislation were among the most important factors pushing up the demand for legal services. All these circumstances helped the reforms to be apposite for the epoch.

References

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, *46*(3), 399–424.

Barendrecht, M., Mulder, J., & Giesen, I. (2006). How to Measure the Price and Quality of Access to Justice? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.949209

Canton, E., Ciriaci, D., Solera, I., European Commission, & Directorate-General for Economic and Financial Affairs. (2014). *The economic impact of professional services liberalisation*. [Luxembourg]: [Publications Office].

Carlson, R. J. (1977). Measuring the quality of legal services: an idea whose time has not come. *Law and Society Review*, 287–317.

Chaserant, C., & Harnay, S. (2013). The regulation of quality in the market for legal services: Taking the heterogeneity of legal services seriously. *The European Journal of Comparative Economics*, *10*(2), 267–291.

Chew, J., & Price, C. (2008). Introducing : An Industry Classification for the Survey of Professional Forecasters.

Gebel, M., & Voßemer, J. (2014). The impact of employment transitions on health in Germany. Adifference-in-differences propensity score matching approach. *Social Science and Medicine*, *108*, 128–136.

Gilmore, A. (2006). *Usługi. Marketing i zarządzanie*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

Grossi, E., Blessi, G. T., Sacco, P. L., & Buscema, M. (2012). The Interaction Between Culture, Health and Psychological Well-Being: Data Mining from the Italian Culture and Well-Being Project. *Journal of Happiness Studies*, *13*, 129–148.

Heckman, J. J., & Ichimura, H. (1998). Matching As An Econometric Evaluation Estimator. *Review of Economic Studies*, *65*, 261–294.

Kleiner, M. M. (2000). Occupational Licensing. *Journal of Economic Perspectives*, *14*(4), 189–202.

Kleiner, M. M., Gay, R. S., & Greene, K. (1982). Barriers to Labor Migration: The Case of Occupational Licensing. *Industrial Relations*, *21*(3).

Kritzer, H. M. (1998). *Legal Advocacy: Lawyers and Nonlawyers at Work*. University of Michigan Press.

Law, M. T., & Marks, M. S. (2013). From Certification to Licensure: Evidence from Registered and Practical Nurses in the United States, 1950-1970. *European Journal of Comparative Economics*, *10*(2), 177–198.

Legal Services Consumer Panel. (2010). *Quality in Legal Services*.

Love, J. H., & Stephen, F. H. (1997). Deregulation and Professional Boundaries: Evidence from the English Legal Profession. *Business and Economic History*, *26*(2), 792–804.

Maxwell, A. E. (1970). Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry*, *116*, 651–655.

Napiórkowska-Piłat, K. (2016). Otwarcie dostępu do zawodów prawniczych. *Na Wokandzie*, *28*. Retrieved from https://nawokandzie.ms.gov.pl/numer-28/aplikacja-i-kariera-14-numer-28/otwarcie-dostepu-do-zawodow-prawniczych.html

Pagliero, M. (2011). What is the objective of professional licensing? Evidence from the US market for lawyers. *International Journal of Industrial Organization*, *29*(4), 473–483.

Pagliero, M. (2013). The impact of potential labor supply on licensing exam difficulty. *Labour Economics*, *25*, 141–152.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL Multiple-item Measuring Perceptions of Service Quality. *Journal of Retailing*, *64*(1), 12.

Pearce, R. G., & Nasseri, S. (2013). The virtue of low barriers to becoming a lawyer: promoting liberal and democratic values. *International Journal of the Legal Profession*, *19*(2–3), 357–378.

Ramseyer, J. M., & Rasmusen, E. B. (2015). Lowering the Bar to Raise the Bar: Licensing Difficulty and Attorney Quality in Japan. *The Journal of Japanese Studies*, *41*(1), 113–142.

Rubin, D. B., & Thomas, N. (1996). Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*, *52*(1), 249–264.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology*, *19*(4), 465–473.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, *42*, 412–416.

Winston, C., Crandall, R. W., & Maheshri, V. (2011). *First Thing We Do, Let's Deregulate All the* Lawyers. Washington D.C.: Brookings Institution Press.

Appendices

Appendix A – Questionnaire SERVQUAL in the Survey Among Clients

Table A.1: Survey Among Clients

1. Have you ever seen a lawyer (that is an advocate or a legal advisor)?	a. Yes, before 2000 b. Yes, between 2000 and 2005 c. Yes, between 2006 and 2010 d. Yes, between 2011 and 2016 e. In 2017 f. Never	2. Did you see only one lawyer in that time?	a. Yes b. No c. Don't know, hard to say
3. In what type of a case did you consult with the lawyer? If you worked on more issues, please indicate only the most important one.	a. Civil – real estate b. Civil – consumer case c. Civil – damages d. Civil – inheritance e. Civil – other f. Family – divorce g. Family - alimony h. Family – other i. Criminal or delinquency j. Administrative k. Business – registry l. Business – other m. Tax n. Employment o. Social Security p. Hard to say	4. Has the matter been completed?	a. No, the case is not over yet b. Yes, the case is over
5. Did you consult the lawyer in a private or a business case?	a. Private b. Business	6. What was the professional title of the lawyer you saw?	a. Advocate b. Legal advisor c. Advocate trainee d. Legal advisor trainee e. Law graduate (no title) f. Don't know, hard to say
7. What was the age of the lawyer?	a. Up to 30 years old b. 31-40 years old c. 41-50 years old d. 51-60 years old e. Over 60 years old f. Hard to say	8. In what type of organisational entity did the lawyer practise?	a. Partnership b. Individual law firm c. Employee in a non-legal company d. Own business e. Other f. Hard to say
9. What convinced you to take this lawyer? (multiple choice possible)	a. Price a. Convenience of where the office was located b. Declared delivery date c. Reputation d. Specialisation in my area e. Supervision of a legal authority f. Recommended by someone g. Good quality of	10. How important were the following factors when your lawyer: (very important – fairly important – neither – not very important – not at all unimportant)?	b. Price c. Convenience of where the office was located d. Declared delivery date e. Reputation f. Specialisation in my area g. Supervision of legal authority h. Recommended by someone i. Good quality of service

	service h. Earlier co-operation with the lawyer i. Trusted brand name j. Offers an online service k. The lawyer was my friend l. Other		j. Earlier co-operation with the lawyer k. Trusted brand name l. Offers an online service m. The lawyer was my friend n. Other
11. Did you compare offers of different lawyers before purchasing the legal service?	a. Yes b. No c. I wanted to compare but didn't know how d. Don't know, don't remember	12. Did you try to solve the legal issue on your own?	a. Yes b. No
13. Where did you look for aid to solve the legal issue?	a. On the Internet b. Among family members c. Among friends d. Followed experts' advices e. In the books f. In press g. Other	14. Do you consider you had a big choice of lawyers?	a. Very big choice b. Big choice c. Small choice d. No choice e. Don't know
15. To what extent would you agree with the following statements (strongly agree – rather agree, neither – rather disagree – strongly disagree)?	a. I could recommend the lawyer b. Service quality was beyond my expectations c. Lawyer worked professionally d. I trusted my lawyer e. The lawyer was honest f. The lawyer was a good man g. The lawyer was knowledgeable in his area h. The lawyer had a vast experience i. The lawyer was kind j. The lawyer was empathic k. The lawyer was active in the case l. The lawyer put client's interest in the first place m. The lawyer was effective n. The lawyer worked fast o. The lawyer respected deadlines p. The lawyer was good-mannered	16. The price for service was:	a. Fixed b. Hourly rate c. Percent of an agreed sum d. Success fee e. I didn't pay f. Other arrangement
17. What was the total cost of the legal service?	a. Up to 100 PLN b. 100-200 PLN	18. Have you co-operated with the same lawyer	a. Yes b. No

	c. 200-500 PLN d. 500-1000 PLN e. 1000-2000 PLN f. 2000-5000 PLN g. 5000-10000 PLN h. Over 10000 PLN i. Don't know	again?	
19. To what extent would you agree with the following statements (strongly agree – rather agree, neither – rather disagree – strongly disagree)?	a. It was easy to meet the lawyer b. It was easy to reach the lawyer by phone c. It was easy to contact the lawyer by email d. It was easy to understand the lawyer e. The costs were defined upfront f. There was a good price to quality ratio for the service g. Legal service was priced favourably h. I was charged for the service after i. The lawyer informed me thoroughly about the co-operation j. The lawyer provided me with information during the case k. I could co-decide on development of the case l. I left the freedom of choice to the lawyer m. Law firm made a good impression n. The lawyer operated innovatively o. The lawyer advertised himself	20. Size of the city in which you used services:	a. Over 500 thousand inhabitants b. Between 100,000 and 500,000 inhabitants c. City between 50,000 and 100,000 inhabitants d. Town below 50,000 inhabitants e. Rural area
21. If you were not satisfied with the service, would you know where to file a complaint?	a. Yes b. No c. I am not sure	22. Have you ever filed a complaint on the lawyer you saw?	a. Yes, I have b. No, I haven't c. Not applicable
23. Considering the quality of the legal services rendered by lawyers – how important are the following characteristics of a lawyer (very important – fairly important – neither – not very important – not at all unimportant)?	a. Good manners b. Good, noble character c. Reliability d. Honesty e. Specialisation in a particular area of law f. Experience g. Innovativeness h. Empathy, sensitivity to the needs of a client i. Determination and being active in a case j. Putting client's	24. Sex of the respondent	a. Female b. Male

	interests first k. Effectiveness l. Quickness m. Availability for a client n. Clear language employed o. Clear rules of cooperation with a lawyer p. Low price q. Favourable financial arrangement		
25. Education of the respondent	a. Elementary b. Lower secondary c. Upper secondary d. Higher		

Appendix B – Questionnaire in the Survey among Judges

Table A.2: Survey among Judges

1. Please evaluate the professionalism of advocates and legal advisors standing before a court: a. Now and b. Before 2006 (very high – rather high – average – rather low – very low).	a. Professionalism of advocates b. Professionalism of legal advisors c. Professionalism of legal representatives who are neither advocates nor legal advisors (so-called legal counsellors) d. Professionalism of advocate trainees e. Professionalism of legal advisor trainees f. Professionalism of the persons without a legal representative	2. Please evaluate the professionalism of legal representatives (advocates and legal advisors) appearing in court: a. Now and b. Before 2006 (always – usually – rarely – never).	a. Actions in the best of client's interest b. Good command of legal procedures c. Good command of substantive law d. Selection of effective tactic e. High care in drafting pleadings f. High skills in formulating claims, demands and allegations g. High skills in questioning witnesses and expert witnesses h. Hgh care in running cases assigned by court i. High court in running cases of choice j. Impeccable manners before a court k. Impeccable manners in the relationship with clients l. Proper contact with other participants of proceedings m. Respecting deadlines n. Good instructions for the client o. Good instructions for trainees how to represent a client p. Courage in reacting to abnormalities
3. To what extent do you agree or disagree with the following statements? (1 – strongly agree, 2 - rather agree, 3 – neither agree nor disagree, 4 – rather disagree, 5 – strongly disagree)?	a. There are more legal representatives without a professional title of advocate or legal advisor who act before court now than before 2006. b. Advocates and legal advisors are now more professional than before 2006. c. Advocate trainees and legal advisor trainees are now more professional than before 2006.	4. If there is any additional area where you have observed changes and would like to share your comments, please comment below.	/open question/

5. What is your age (in years)?	a. 34 or less b. 35-39 c. 40-44 d. 45-49 e. 50-54 f. 55-59 g. 60-64 h. 65-69 i. 70 or more	6. What is the length of your professional experience as a judge?	a. Less than 5 years b. 5-9 years c. 10-19 years d. 20 years or more
7. Please choose your current job placement from the list.	a. District court b. Regional court c. Appellate court d. The Supreme Court e. Regional administrative court f. The Supreme Administrative Court g. Military court	8. Please indicate in which court's department you worked the longest.	a. Criminal b. Civil c. Family and minors d. Economic e. Land and mortgage registry f. Labour and social security g. Administrative court h. Military court i. Other
9. The court in which you work is located in a city counting:	a. Over 500 thousand residents b. Between 200 and 500 thousand residents c. Between 100 and 200 thousand residents d. Between 50 and 100 thousand residents e. Less than 50 thousand residents		

Appendix C – Balancing properties of the sample in the survey of the quality of legal services among consumers

Matching procedure is based on probit regression with a binary dependent variable describing the state of being in the treatment group.[30] Covariate balance between the treatment and control groups was also observed using both standardized differences and percentage bias reduction (Austin, 2011; Gebel & Voßemer, 2014). Percentage bias and the bias reduction were reported in Table A1.1.

Table A.3: Propensity score matching for reduction of sample bias

Control variable	Model with treatment representing clients from 2011-2017 and control being clients from 2006-2010		Model with treatment representing clients from 2011-2017 and control being clients from 2000-2005	
	%bias before (after)	bias reduction due to matching	%bias before (after)	%bias reduction due to matching
Female	-5.8 (-12.4)	-112.4	-6.9 (5.2)	23.7
Secondary education (ref. primary or	-23.2 (-2.3)	90.1	-25.9 (3.0)	88.3

[30] Although in our case the treatment group always consisted of those individuals who used legal services between 2011 and 2017, the control group consisted in one analysis of individuals using legal services between 2000 and 2005 and in the latter analysis of individuals benefitting from those services between 2006 and 2010.

vocational)				
Higher education (ref. primary or vocational)	31.2 (-1.7)	94.4	48.9 (-3.5)	92.8
Respondent age 36-50 (ref. up to 35)	21.4 (5.9)	72.7	19.1 (-2.0)	89.3
Respondent age 51 or more (ref. up to 35)	-36.3 (-2.5)	93.0	-64.3 (0.6)	99.1
Used service of one lawyer only	-33.8 (2.2)	93.5	-22.2 (-7.5)	66.2
Case type – family (ref. civil)	-25.4 (4.7)	81.3	-21.3 (-9.2)	57.0
Case type – penal (ref. civil)	5.9 (-1.3)	78.6	-5.4 (-2.0)	62.8
Case type – other (ref. civil)	6.5 (-9.0)	-37.4	2.0 (2.4)	-19.5

The biases for the 'before' period show substantial differences between the samples used for evaluation of the legal service quality before and after deregulation. In most cases – like the age or education levels of respondents, type of the case (especially for civil) and lawyer age – the biases for the samples before matching were substantially above the 10 per cent threshold. The treatment and control groups did not differ considerably after matching and almost all differences were well below 10 per cent for matched individuals. This implies that all predictors, which prior to matching were above the 10 per cent threshold, showed significant improvement in bias as a result of matching. The only predictors that demonstrated increased bias were those which initially showed very low bias but even after the increase remained below the limit. Good matching is confirmed also by the outcome of the likelihood ratio tests. For the unmatched sample of legal service quality assessment in the sample of legal services in the period 2006–2010 (control) and 2011–2017 (treatment), the LR statistic shows 48.72 yielding $p=0.000$, while for the matched sample the outcome is $LR=7.89$ with $p=0.895$. The same comparison for the sample of legal services in the period 2000–2005 (control) and 2011–2017 (treatment) of unmatched and matched samples yielded $LR=118.9$ ($p=0.000$) before matching and $LR=5.09$ ($p=0.984$) after matching. Both these results imply that the hypothesis of similarity between unmatched samples should be rejected, while, due to PSM, almost all differences between the treatment and the control group were accounted for.

Appendix D: Change in Customer Satisfaction with Legal Services over Time – all SERVQUAL Indicators

Table A.4: Average treatment effects

Outcome of legal service	Average treatment effects	
	2011-2017 (treatment) vs. 2000-2005 (control)	2011-2017 (treatment) vs. 2006-2010 (control)
I trusted my lawyer	0.029	0.03
The lawyer was honest	-0.025	-0.079
The lawyer was good-mannered	0.066	0.004
The lawyer was kind	0.008	-0.096
The lawyer was empathic	0.015	0.027
It was easy to meet the lawyer	0.037	-0.095
It was easy to reach the lawyer by phone	0.09	0.018
It was easy to contact the lawyer by email	0.157	0.217*
I am satisfied with co-operation with the lawyer	-0.019	0.051
I am happy with outcome of the case	0.041	0.063
I could recommend the lawyer	0.015	0.133
Service quality was beyond my expectations	0.271	0.194
It was easy to understand the lawyer	-0.021	-0.051
The costs were defined upfront	0.23*	0.029
The lawyer provided me with information during the case	-0.033	-0.022
The lawyer was engaged in the case	-0.09	-0.085
The lawyer was efficient	-0.075	-0.021
The lawyer informed me thoroughly about the co-operation	0.088	0.008
I could co-decide on development of the case	0.008	0.146
Lawyer worked professionally	-0.032	-0.102
The lawyer was knowledgeable in his field	0.042	-0.091
The lawyer had a vast experience	0.1	-0.086
Law firm made a good impression	-0.009	0.037
The lawyer had professional	-0.012	-0.072

appearance		
The lawyer operated innovatively	0.151	0.084
The lawyer was always respecting the deadlines	-0.007	-0.071
The lawyer was present during all trials	0.18	0.157
The lawyer kept all his promises	-0.007	-0.036
The lawyer behaved ethically	-0.106	-0.006
The lawyer was expressing in a balanced way	-0.098	-0.062
I had no problems with settlement	0.107	0.151

Note: For each item a response score according to the following scale (1=strongly agree, 5=strongly disagree) was used. Consequently, the units of measurement are related to question answers. Due to the scale orientation, positive values indicate deterioration between the reference period (either 2000-2005 or 2006-2010) and the period when deregulation was already in place (2011–2017). * indicates significance $p<0.1$

(This page intentionally left blank)

2. Architects and Engineers in Germany

2.1. Introduction

A recent communication from the European Commission (European Commission, COM (2016) 820, January 2017) showed that architects and civil engineers[31] are two of the most heavily regulated professions. In terms of a restrictiveness indicator that measures the intensity of regulation on a scale from 0 (very unrestrictive) to 6 (very restrictive), architects face higher regulation intensity compared to civil engineers across the European Union, though in some states like Denmark regulation of both architects and civil engineers is low.[32] Figures 1 and 2 show that within the profession of architects and civil engineers, those operating in Germany are amongst the most heavily regulated in general. The index comprises information on the regulation of activities, protection of professional titles, entry restrictions like qualification requirements, and exercise restrictions. Similarly, in the World Bank Doing Business data for the years 2015 to 2017, Germany received a score of three out of four in the professional certifications index, while Denmark had a score of one in each year. In 2016, for example, it took 1.5 times as long for a business in the construction industry to complete all procedures to build a warehouse in Germany compared to Denmark, while the quality control of buildings was higher in Denmark. According to Figures 1 and 2, the presence of exercise requirements explains a large part of the high overall regulation level of Germany, reflecting that architects and civil engineers operating in Germany are subject to the most restrictive laws imposing these specific kinds of requirements in the European Union. In this chapter, we focus on two kinds of exercise restrictions affecting all architects and most engineers in the construction sector: (i) requirements on professional indemnity insurance and (ii) price restrictions.

Motivated by the importance of regulation for the services of architects and engineers as suggested by the regulation index levels for Germany, we focus on two major reforms as a means to explain the circumstances that led to the current level of regulation and to shed light on the effects of regulatory changes on service quality of architects and civil engineers. This is of primary importance, since quality improvements seem to be a main driver of job creation, investment behaviour[33] and economic growth in the construction sector, although their contribution is unobserved in cost-based official statistics. This is true for many other sectors in developed economies as well. Therefore, in this chapter we contribute a small step towards a better measurement of economic growth in general and present several strategies to evaluate the effects of regulation.

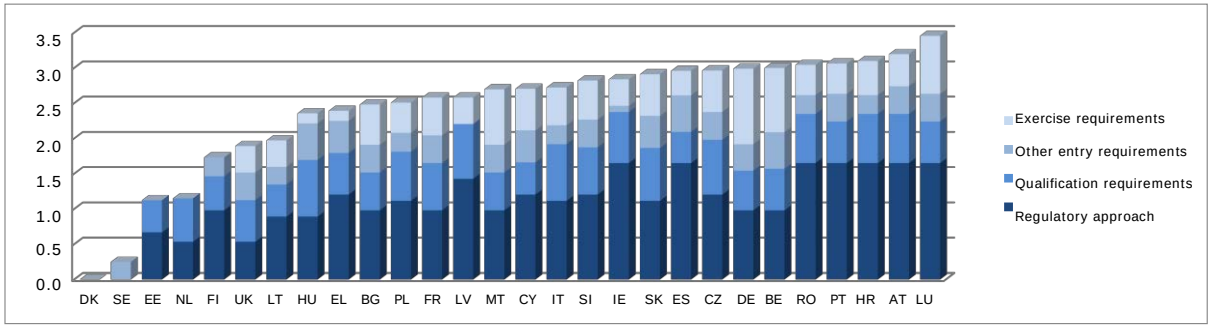
Setting a minimum level of quality is the main justification for regulation in many sectors, particularly for architects and civil engineers as well as for craftsmen working in construction sites (see Rostam-Afschar, 2014, 2015). Interestingly, there is a major difference in terms of the perceived quality of architectural and engineering services from the consumers' perspective depending on whether the contracting authority is private or public. According to a survey conducted in the years 2010 and 2011 and presented in a study commissioned by the Federal Institute for Occupational Safety and Health (2012), 96 per cent of public contracting authorities are satisfied with the quality, compared to only 56 per cent of private consumers. Moreover, a large share (38 per cent) of private consumers is not satisfied at all. The difference could be due to the fact that public consumers usually do not use the contracted building themselves. However, the data show that private consumers are unhappy because of delays and prices, while for public consumers only delays seem to be problematic, albeit to a lesser extent. Even though damages do not seem to be the prime source of dissatisfaction with quality, many studies find that damages arise most frequently from planning mistakes and flaws in construction site management, both of which are price-regulated project phases in Germany. The corresponding damages are covered to some extent by the professional indemnity insurance requirements. In this chapter, we take a closer look at these two exercise restrictions, since recent policy changes may help to understand how quality is affected.

[31] We use the term civil engineers to describe all engineers that engage in non-military activities.

[32] This index is a refinement of the OECD Product Market Regulation Indicator and an alternative to the indices provided by the World Bank Doing Business database.

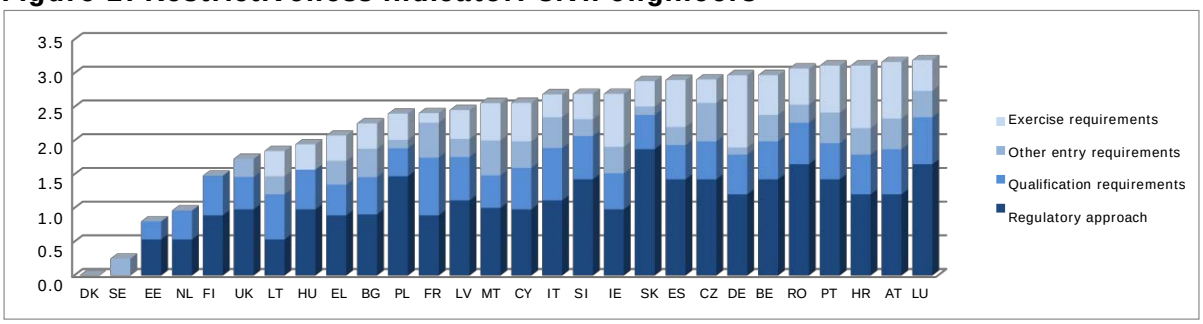
[33] Investment behaviour may in turn influenced by regulation not only in the form of licensing but also in the form of taxes; see Fossen et al. (2017).

Figure 1: Restrictiveness indicator: Architects



Source: European Commission, COM (2016) 820, January 2017.

Figure 2: Restrictiveness indicator: Civil engineers



Source: European Commission, COM (2016) 820, January 2017.

To give a comprehensive measure of quality, we examine aesthetic quality (using an international ranking of architectural firms), business formation and market concentration, as well as demand for labour and continued education as main outcomes that could have been affected by the reforms. Moreover, we show how service availability has changed. The particular two reforms we analyse are legislative amendments of the Insurance Contract Act (VVG, 'Versicherungsvertragsgesetz') and the German Fee Structure for Architects and Engine (HOAI, 'Honorarordnung für Architekten und Ingenieure'). The VVG sets, among other things, the legal framework for professional indemnity insurance for architects and civil engineers and thus defines exercise requirements. The HOAI is a fee scale for architects and engineers that restricts exercise by imposing price floors and price ceilings on services provided by architects and civil engineers who engage in specific activities detailed in the law.

The VVG-Reform 2008 increased professional indemnity insurance requirements to 1 million euros per year which affects mainly small and medium-sized firms that previously contributed 0.5 million euros per year. The HOAI-Reform 2009 increased prices (to adjust for inflation) by 10 per cent for architects and by 8 per cent for civil engineers.

These exercise requirements are part of a broader set of regulatory measures for architectural and some civil engineering services that also include entry regulations. In fact, exercise requirements and entry regulation are in some cases intertwined, e.g. parts of the fee scale regulation apply to specific activities that are subject to entry requirements. The most important regulatory measures other than mandatory professional indemnity insurance and fee scales are the requirement to obtain a university degree (four to five years of full-time study) plus two years of work experience, protection of professional titles such as architect ('Architekt'), engineer ('Ingenieur') and consulting engineer ('Beratender Ingenieur'), enrolling in the occupational register, mandatory membership in the chamber of architects/engineers, and requirement to prove continued education.

There exist three main minimum requirements for a policy evaluation to allow the measurement of causal effects of a reform. First, exogenous variation in the legislation (e.g., generated by an unanticipated reform) that changed the environment in which architects and civil engineers operate must be present. Second, measures of service quality for the particular case of architects and civil engineers must be defined. Third, individual observations must be assigned to a group affected by the reform (treatment) and—more challenging—another group not affected by the policy change

(control)[34]. In the first part of our application, we measure effects on service quality using German architectural firms as the treatment group and non-German firms as the control group. In the second part, we study the regulatory effects on other outcomes: entry and exit to self-employment, self-employment rates, firm size and continued education. In this part, the treatment group consists of architects and engineers in the construction sector, while all other civil engineers form the control group.

Whether these requirements can be met depends on the availability of high-quality information. In the case of architects and civil engineers in Germany, this is a challenging but not an impossible task. The main reasons for the lack of high-quality data are unwillingness or inability to contribute,[35] costs of elicitation, data protection restrictions and ethical reasons that preclude randomized assignment. Moreover, even in a laboratory setting, it is not clear how to measure quality of services as these may have a strong subjective component. We suggest simple measures that try to take this into account.

This chapter presents two complementary strategies to make the best possible use of available data in order to evaluate the effects of regulation on the quality of the service using different data sets. First, we apply the synthetic control method to an authoritative firm ranking database for architects based on peer ratings that we link to data from the Architects' Council of Europe. Second, we use data from the German microcensus on architects, civil engineers in the construction sector and other civil engineers in a quasi-experimental analysis.

We find that architectural quality measured by peer ranking score decreased in Germany due to the HOAI reform in 2009, which exogenously increased prices for architectural services. This is supported by further evidence that shows that the market concentration of architects and engineers subject to the HOAI increased in response to the VVG and HOAI reforms, complementing recent evidence that also income inequality increased (see Rostam-Afschar and Strohmaier, 2018). While the higher prices of the HOAI encouraged new start-ups, the effect of increased insurance costs of the VVG more than offset this effect. At the same time, a stronger VVG effect than the effect of the HOAI on exits could have reinforced this and further reduced the number of businesses. While the effects on exits are not statistically different from zero, those on the probability of being self-employed are significant and support the view that the reforms made businesses formation harder. Our results also show that the former self-employed partly became employees: firm size increased due to the HOAI reform by 6 per cent. A high market concentration may lead to economies of scale but may also have negative effects on quality of services. The extraordinarily high share of self-employment among architects and construction engineers suggests that too much market power is unlikely to pose a concern. Finally, we look at training on the job as an indicator of service quality and find that the probability of attending continued education increased after 2008, while the average hours spent on continued education in the last year decreased by 4 hours. Since the control group experienced a particularly steep increase in hours spent on continued education, the measured effects seem not to be caused by the amendment to the HOAI alone.

Even though most of our results are consistent with theoretical predictions, the reliability of our results is limited by the quality of the data. Therefore, our results must be re-examined when informative data becomes available in the future.

2.2. Literature Review

While recent policy reports try to contribute to the discussion mostly by confining themselves to claims that an indicator is biased (without proof or suggestion of better measures), three strands of the scientific literature provide results that are relevant to the question on how regulation affects quality in the case of architects and civil engineers: the literature on markets with information asymmetry; on misallocation of talent, corruption and rent-seeking; and on the empirical measurement of quality.

[34] This terminology originates from medical trials: the treatment group (randomly selected group of people) gets the new medicine (or is subject to the reform) and the control group (randomly selected group of people, which should be similar to the distribution of people in the treatment group) does not (and gets a placebo drug in medical trial).

[35] In particular, some individual insurers, the Federal Chamber of German Architects, and the German Sustainable Building Council could not provide us with information. We are very grateful to the German Insurance Association and the Chamber of Architects, Baden-Württemberg, for making some statistics available to us.

The early literature classified three types of qualities of services (and goods): search, experience and credence qualities (Nelson 1970, Darby and Karni 1973). Search qualities can be ascertained in the search process prior to purchase. Experience qualities can be discovered only after purchase. Credence qualities cannot be evaluated in normal use but require additional costly information. While some services of architects and engineers have search quality (e.g., access to amenities, transport, design) most have experience (e.g., firmness, heat protection, fire safety, sound insulation, energy and water consumption) or credence qualities (e.g., cost estimate, environmental impact, building life cycle costs). In practice, there are methods to turn experience qualities into search qualities. For instance, by the provision of model houses, the consumer can verify qualities that would have been experience qualities in absence of model houses even before purchase. Due to a lack of reliable data on quality of services, the early literature focused on theoretical models of market behaviour and compellingly showed that it can be optimal to regulate entry or prices to guarantee a minimum quality standard. Recently, Atkeson et al. (2015) showed that a combination of entry taxes and a price floor, like in the case of architects and construction engineers in Germany, may lead to higher overall production and higher average quality if investments of architectural firms in high-quality inputs are not, or only with a delay, observable by outsiders, while a sufficiently small but positive entry fee alone may already lead to higher average quality. The theoretical literature also finds that in other related situations fixed prices may cure market failure. For instance, Pesendorfer and Wolinsky (2003) show that a price floor may lead to higher welfare if a consumer gathering multiple opinions cannot verify if an expert invested great effort to find the correct service for the consumer's need. However, in an environment where quality is imperfectly observed, it may be hard for the regulator to find the optimal fees in practice. Moreover, such regulation usually limits the variety of quality levels which might be undesirable for some other consumers. In the case of architectural services, one should take account of the high demand for partly redeveloped older buildings that were constructed before the introduction of price regulation like Art Nouveau or Bauhaus houses. If there is such demand for both high and low quality, according to Meran and Schwarze (2010), average service quality would never decrease and could even increase further after a deregulation of fixed prices.

These arguments refer mainly to price floors, which were not the primary objective of the introduction of the HOAI, since according to the German Monopolies Commission (2006) it was not in response to actual problems with quality. Instead, price ceilings were considered more important, in order to reduce overcharging. In fact, overcharging and under treatment seem to be the most important arguments for price regulation. These price caps, however, may limit investments in higher quality, business growth, and employment.

A very important but often unmentioned problem of the price regulation for architects and engineers in Germany is that it leads to a misallocation of talent that could be very costly. Misallocation of talent occurs when talented persons decide whether to engage in a productive or an unproductive activity. As Murphy et al. (1991) put it, if the most talented people start firms, they innovate and foster growth, but when they become rent seekers, they only redistribute wealth and reduce growth. In their paper, they show that countries with a higher proportion of engineering college majors grow faster as opposed to countries with a higher proportion of law students who proxy rent seekers. These are, of course, only very rough proxies, but in practice not just lawyers but also some engineers seem to be rent seekers. The reason is that regulations like the HOAI create opportunities for rents that did not exist before 1977, when the regulation only stipulated price ceilings or was non-binding (see next section).[36] In fact, some architects and engineers make a living from explaining the regulation to their peers. For instance, the biggest German railway company, Deutsche Bahn AG, recently posted a job advertisement for the position explicitly titled 'HOAI expert'. Required qualifications are, among others, profound knowledge of the HOAI and skilful application of its rules. This means, instead of applying their talent to construct buildings or other useful infrastructure, HOAI experts explain legislative texts. This is an example of misallocation of talent, since it seems highly unlikely that a talented architect or engineer could contribute more to economic growth by translating the HOAI legislation compared to working, for example, in other member states, where there is demand for productive architectural and engineering services.

In addition to that, the existence of regulation requires enforcement which is why many architects and engineers work as bureaucrats, e.g., to issue building approvals and ensure quality. These

[36] Since 1871 professional associations have issued non-binding fee scales for architects and since 1878 for engineers. Since 1935 (since 1950 because of the Second World War) price ceilings but not price floors were issued for architects by a government institution, whereas the fee scales for engineers only served as recommendations.

bureaucrats perform useful activities including provision of public goods and correction of market failures. However, they might abuse their power and ask for excessive fees or even accept bribes (cf., e.g., Acemoglu and Verdier 1998). If these bribes or efficiency wages to prevent corruption are large enough, there will be misallocation of talent because it is not necessarily those with a comparative advantage for the public sector who are working there (Acemoglu and Verdier 1998). Our empirical analysis shows some suggestive evidence along these lines, since the relative prices of self-employment compared to working as a civil servant have changed.

In addition to price regulation, we investigate the effects of the mandatory professional indemnity insurance. This insurance that costs at least 250,000 euros per quarter constitutes from the perspective of an architectural or engineering firm a fixed cost. According to Branstetter et al. (2013), an increase in fixed costs leads in the short run to decreased business formation, decreased employment, higher quality and productive entrants with larger offices.

To summarize, the theoretical literature finds that the effect of entry and price regulation on service quality depends on the heterogeneity of consumer preferences, on how much opportunity for corruption and misallocation of talent the regulation brings with it, and on the verifiability of the quality of investments of the firm.

Of course, a reliable relationship between regulation and service quality requires a good measure of quality of architectural and engineering services in the first place. While for other professions, service quality seems relevant enough to be monitored by involvement of the respective professional association, e.g. the German Midwifery Association has measured and published since 1999 various statistics on nonclinical births for quality management, such data is to our knowledge not available for architects and engineers. Some attempts to empirically pin down quality from other countries include a quality of architecture index based on experts' judgement or value of character of architectural design based on property price differences between designated conservation areas and neighbouring outside areas. These studies often focus on how quality of architectural design affects property prices or office rents (e.g., Gat, 1998, Fuerst, 2012). For example, Gat (1998) finds that one additional building quality point on an expert's scale increases rent more than 5 per cent compared to office rent in Tel Aviv. An example of a study that presents evidence of the influence of regulation on the work of architects is the survey of UK architects in Imrie and Street (2009), which was conducted in 2006 and 2007. In this survey, 32 per cent strongly agree and 47 per cent agree that they spend more of their time on red tape and bureaucratic procedure than five years ago. Almost half, 45 per cent, strongly agree and 40 per cent agree that regulation is becoming more difficult to deal with. Of course the time spent on red tape may reflect a form of misallocation of talent. Ahlfeldt and Holman (2016) show with data from the UK that regulation costs of taxes, subsidies, height restrictions, zoning, etc. have a negative impact on property values of about 10 per cent but lead to an increase in distinctiveness of architectural design that exceeds the direct regulatory effect.

2.3. Reforms

VVG-Reform 2008

The VVG Act was introduced in 1908 and remained the same until 2008. The key aim of its reform in 2008 was to improve consumer protection. It was also designed to be in accordance with the directives of the European Union (Directive 2002/92/EG). Overall, it is expected that the reform could have led to a higher quality of products and higher prices due to higher liability and cover requirements. Moreover, the number of architects could have declined in the short run, which may have led to a higher market concentration. More specifically, it is possible that the reforms have raised insurance premiums for architects and engineers. Schwarze (2007) writes of a doubling or tripling of premiums. Similarly, Krause-Allenstein (2017) predicts an increase in insurance premiums for small and medium-sized firms. In particular, the new requirement of a minimum insurance coverage of 1 million euros per year seems a drastic change for small and medium-sized firms. According to the German Insurance Association, the average insurance sum increased from about 620,000 euros in 2005 to 720,000 euros in 2010.

Table 1. The Elements of the VVG-Reform 2008.

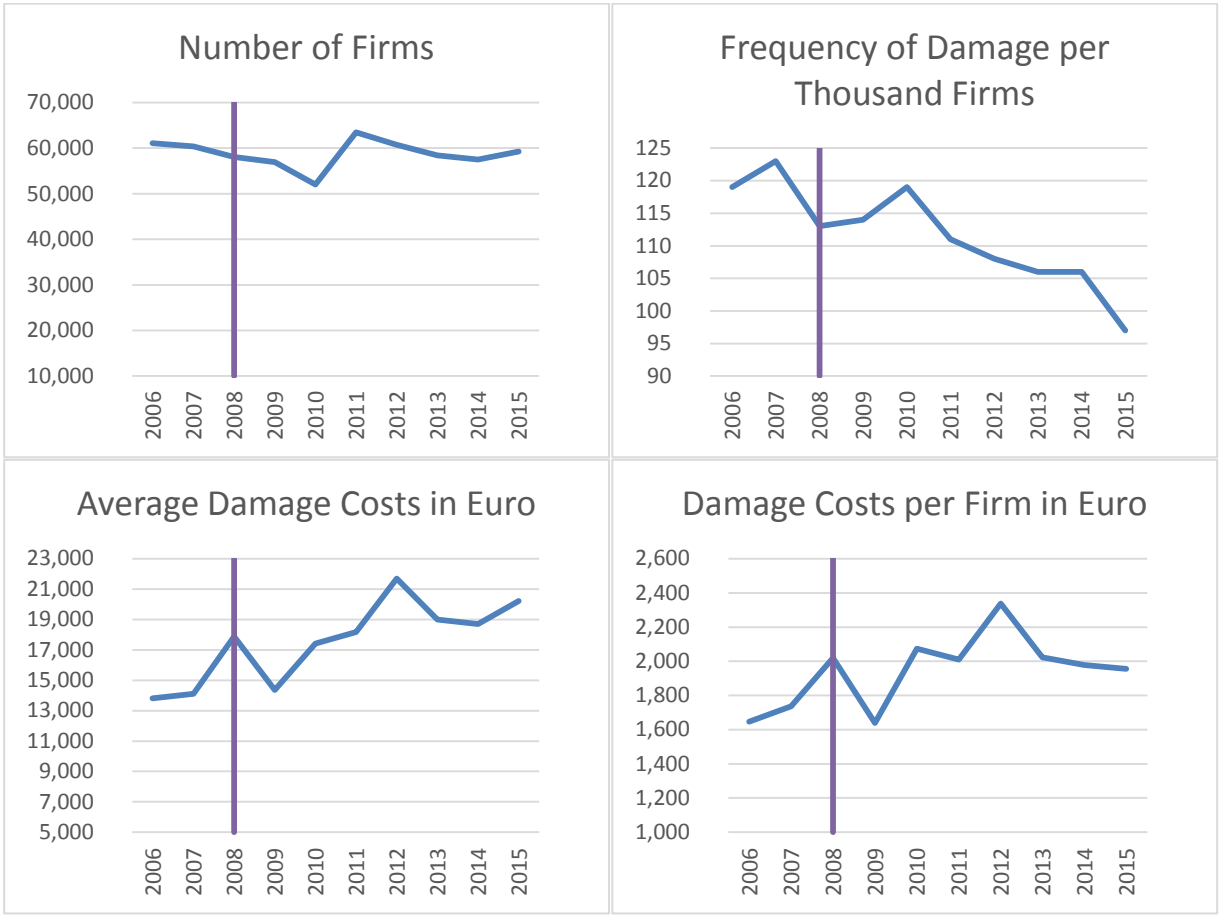
Regulatory Change	Old Version of the VVG	New Version of the VVG
Requirement to provide ex-ante information §19 VVG	Architect required to provide information on all potentially dangerous situations (e.g. number of employees in firm)	Architect only needs to provide information on potentially dangerous situations if explicitly asked for by the insurer
Direct damage claims to the insurer §§ 113, 115 VVG	Building contractor cannot raise direct claim against insurer in case of damage	Introduction of direct claims of third party against insurer
Minimum insurance coverage $114 VVG	No minimum required 0.5 Mio. Euro per year for small and medium-sized enterprises in practice	1 Mio. Euro per year for all cases 250,000 Euro per quarter
Scaling of insurance payments according to degree of negligence that led to violation of obligations §28 VVG	No insurance service if architect violates obligation negligently	No consequences if architect violates obligation in simple negligent way; Reduction of insurance service if architect violates obligation with gross negligence; No insurance service if architect violates obligation wilfully negligently
Acknowledgement or settlement §105 VVG	Acknowledgement or settlement of claims regarding contract partner leads to loss of insurance service	No loss of insurance service if acknowledgement or settlement of claims regarding contract partner
Period of limitation §12 VVG	Two years period of limitation of coverage claims for architects	Three years period of limitation of coverage claims for architects
Transfer claims of cover costs §108 VVG	Transfer of claims from professional indemnity insurance as contract partner of architect prohibited	Transfer of claims from professional indemnity insurance as contract partner of architect allowed and bring up against insurer

Source: Own description.

The market for services of architects and engineers was already highly concentrated before the reform: few firms with large revenues competed with many self-employed at the margin of profitability. The reform was likely to have further increased market concentration as facing higher insurance costs, many firms exited and start-ups were discouraged. If this is the case, then we expect the availability of services to consumers (an important measure of access and thus quality) to have been compromised. On the other hand, it is possible that the withdrawal of low-quality firms from the market as a result of the reforms could have increased the overall quality of services received by the consumer. To assess the effect of this reform on quality is therefore an empirical question.

Unfortunately, we do not have access to insurance premiums to prove this claim. However, if costs of professional indemnity insurance really affected business survival, this will show up in our empirical analysis below. Before we turn to these results, though, we present descriptively the development of some of the determinants of insurance premiums provided by the German Insurance Association.

Figure 3: Damage Frequency and Damage Costs



Source: Own depiction based on data from the German Insurance Association (2006-2015).

The top left graph of Figure 3 shows that in the pre-reform period about 60,000 firms were insured. According to the German Insurance Association, deviations in this number seem to be caused by non-response of some insurers. The top right graph shows that the frequency of damages per thousand firms declined substantially over time by almost 20 per cent from 2006 to 2015. The coincidence with the reform could suggest that it led to this development but this could have been the result of a general trend including, for example, technological improvements. The bottom left graph shows the average damage costs in a given year. The damage costs are measured at the end of the year in which the damage was reported; therefore, the total damage over several years in a specific case could be much higher. This variable increased by 46 per cent from 2006 to 2015. This indicates that costlier damages occurred or, given the evidence that fewer damages were reported, this could be the result of fewer minor damages. The damage cost per firm also increased by less than 20 per cent. In these time series, there are some changes that coincide or follow shortly after the VVG reform; however, it is not clear whether these patterns were caused by the VVG reform.

HOAI-Reforms 2009 and 2013

The German Fee Structure for Architects and Engineers ('Verordnung über die Honorare für Leistungen der Architekten und der Ingenieure', HOAI) is a federal ordinance to regulate the remuneration of services rendered by architects and engineers operating in the construction sector. Based on the Law on the Regulation of Services of Engineers and Architects (Gesetz zur Regelung von Ingenieur- und Architektenleistungen), it is considered a binding price law for all planning and monitoring services offered by architects and engineers and has its early roots in 1871 when the first private fee structure for architects emerged. Before 1977, there were two separate fee scales, binding price ceilings but not price floors for architects, and non-binding recommendations for engineers. The HOAI is justified on the basis of consumer protection, namely that it protects consumers from competition based on prices and forces service providers to focus on competing on quality. This is the main justification for price floors. On the other hand, price ceilings are set to preclude overpricing.

Table 2. Service phases for the service profile "Buildings and Interior" (§34, HOAI 2013)

Phase	Service	in % of total fee	
		Buildings	Interior
1	Basic evaluation	2	2
2	Preliminary planning	7	7
3	Draft planning	15	15
4	Approval planning	3	2
5	Execution planning	25	30
6	Award preparation	10	7
7	Assistance with the award process	4	3
8	Construction supervision and documentation	32	32
9	Project supervision	2	2

Notes: This table lists the nine service phases for the service profile "Buildings and Interior". The last column reports the relative share of the total fee for each phase.

The German Monopolies Commission (2006) emphasizes that the introduction of price floors was a response to the wish of the professional association rather than to actual problems with quality of services, since the actual objective of the HOAI was to use price ceilings to provide incentives for reducing costs of buildings and rents. The obligatory applicability to set prices according to HOAI is in general not profession related, but applies to all services which are explicitly listed in the HOAI. However, for activities in some service phases specific professional degrees are a prerequisite for authorization to provide a service. These broadly include all necessary services pertaining to area planning (landscape, land use), project planning (building and interiors, open air facilities, civil engineering structures, traffic installations) and specialist planning (structural planning, technical equipment). This means that for each service within the work contract, it has to be checked whether it is subject to the HOAI.

All service profiles (like, e.g., the service profile 'Building and Interiors') are further subdivided into up to nine service phases (see Table 2), which follow the object-specific planning process. Each service phase is valued as a fixed percentage of the total fee.[37] Architects usually carry out all nine service phases, but there also exist architectural firms that specialize either in the planning phases one to four or in the construction phase five to nine. An example for the service profile 'Building and Interiors' is given in Table 3. As becomes apparent from the fee structure, the scope for price competition is very limited and restricted to the prices within these fee bands.

Since 1977, the HOAI has been reformed several times with the last major revisions in 1996, 2009 and 2013. The main element of the 2009 reform was a substantial increase in prices of services. Simplification and slight liberalization of the existing regulation scheme were mentioned as further aims. Moreover, this reform implemented the EU services directive of 2006 on the freedom of establishment for suppliers of services. Table 4 gives an overview of the most important changes of the HOAI introduced by the 2009 reform which we describe in detail below.

First, and most importantly, the reform in 2009 led to an increase in all price floors and ceilings. The price increase in 2009 amounted to 10 per cent for most services. Figure 4 shows price indices from 2006 to 2016 for architects, construction engineers, other construction related service providers and other engineers, as described below.

Second, as a reaction to the EU services directive, the geographical scope of the HOAI changed. Therefore, the fee structures started to apply only to services provided by architects and engineers headquartered in Germany. In principle, foreign firms could have offered competitive prices which

[37] Reference for the determination of the total fee, for which the above mentioned price floors and ceilings apply, are either surface units (for area planning) or chargeable costs (for project and specialist planning, calculation is based on DIN 276-1:2008-12).

would, given the advanced communication technologies, not even require physical presence, e.g., for specific services like consultation.

Table 3. Fee structure for 'Buildings and Interior' (HOAI 2013 §35, in €)

Charg. costs	Fee band I		Fee band II		Fee band III		Fee band IV		Fee band V	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
25,000	3,120	3,657	3,657	4,339	4,339	5,412	5,412	6,094	6,094	6,631
35,000	4,217	4,942	4,942	5,865	5,865	7,315	7,315	8,237	8,237	8,962
50,000	5,804	6,801	6,801	8,071	8,071	10,066	10,066	11,336	11,336	12,333
70,000	8,342	9,776	9,776	11,601	11,601	14,469	14,469	16,293	16,293	17,727
100,000	10,790	12,644	12,644	15,005	15,005	18,713	18,713	21,074	21,074	22,928
150,000	15,500	18,164	18,164	21,555	21,555	26,883	26,883	30,274	30,274	32,938
200,000	20,037	23,480	23,480	27,863	27,863	34,751	34,751	39,134	39,134	42,578
300,000	28,750	33,692	33,692	39,981	39,981	49,864	49,864	56,153	56,153	61,095
500,000	45,232	53,006	53,006	62,900	62,900	78,449	78,449	88,343	88,343	96,118
750,000	64,666	75,781	75,781	89,927	89,927	112,156	112,156	126,301	126,301	137,416
…										

Notes: This table presents the total fee structure for the first cost levels for the service profile 'Buildings and Interior'. The table continues to 25 million chargeable costs. For each cost level, five different fee bands exist which apply depending on the complexity of the service provided.

However, as reported by the Federal Chamber of Architects, cross-national provision of services has been rare due to prohibitively high costs of information acquisition. According to the sector study of the Architect's Council of Europe (2017), 12 per cent of architects in EU member states (without Croatia) cited in 2016 local fee scales, 11 per cent professional indemnity insurance and 38 per cent planning or building regulations concerning working in another country. Insufficient language skills were cited by 36 per cent as the main reason not to work in another country. The language concerns likely express similar concerns as those referring to regulation as an impediment, since good command of a language is a prerequisite to understanding the regulations which are often only available in the local language. However, labour and product market factors and – most frequently – practical, relocation or personal concerns are also mentioned as barriers. Therefore, it is no surprise that the number of foreign architectural firms operating in Germany is extremely low.

The provision of services of firms headquartered in Germany to EU member states is virtually non-existent as well, since, as reported by the Federal Chamber of German Architects (2016), 94 percent of surveyed firms in 2015 are not active outside Germany.

Third, some of the engineering services, which have had the status of a service profile, were relabelled and became consulting services for which the binding price scheme does not apply. These engineering services include environmental impact assessment, thermal building physics, sound insulation and room acoustics, soil mechanics, earth and foundation works and technical surveying services. Thus, engineers specialized in these areas might have potentially entered a price competition after the reform.

Fourth, existing building structures are not allowed to be considered in price calculations. Further, changes include the permission to freely agree on hourly rates, model-based salaries that are not connected to actual costs and the possibility of rewarding a reduction in costs with a bonus of up to 20 per cent and penalizing an increase in costs with a malus up to 5 per cent of salary. In addition, the changes with respect to the building structures to be processed have been reversed, resulting in higher chargeable costs and thereby higher fees.

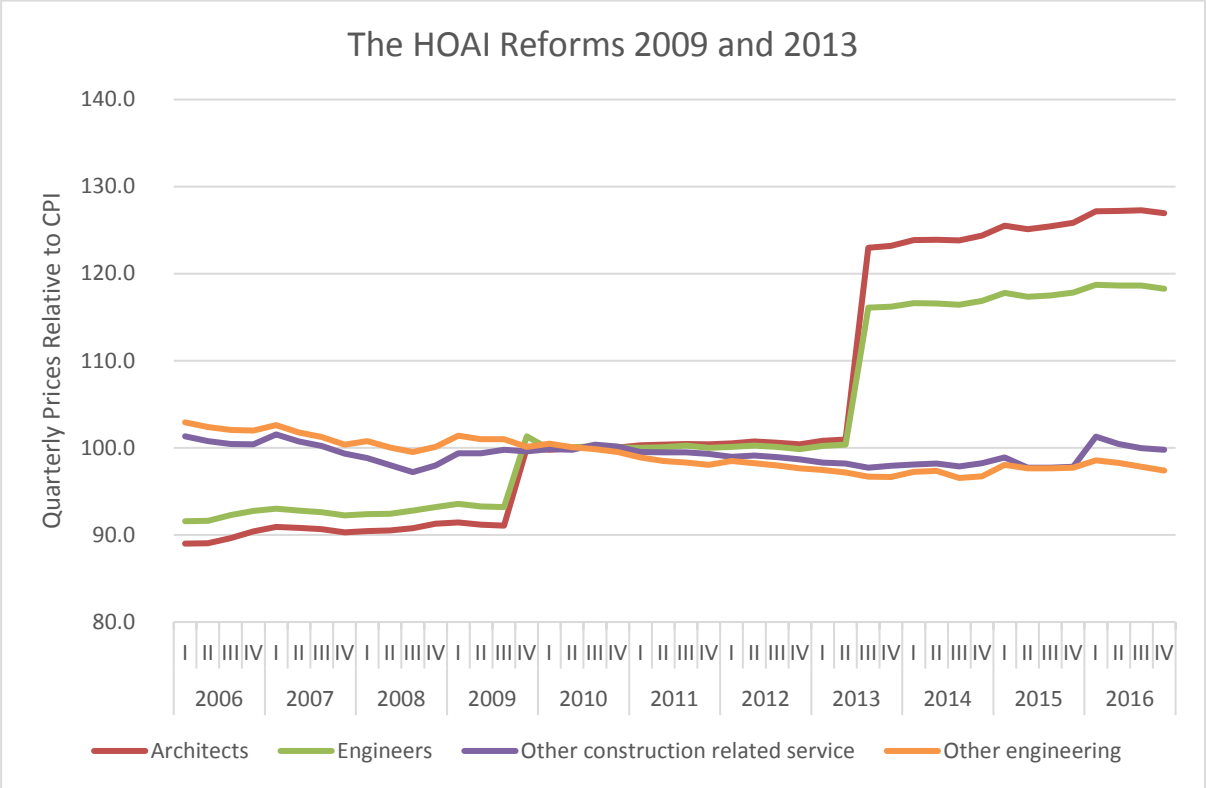
In our empirical analysis, we exploit the exogenous price increase of the HOAI 2009 as a natural experiment and analyse its effect on different quality measures. As apparent from Figure 4, there are other related groups of professions, which are not subject to HOAI. There is no jump in the producer price index for these groups. They will serve as an ideal control group in our setting. This group comprises, for instance, engineers working in the food, mining or machine industries.

Table 4. The Elements of the HOAI-Reform 2009

Regulatory Change	Old Version of the HOAI	New Version of the HOAI
Fee rates	As specified in 1996	Fee rates increased by 10 %
Applicability of HOAI to foreign architects and engineers § 1	The HOAI applied to all architects and engineers.	Foreign architects and engineers are exempt from HOAI and can freely set their prices.
Fees for building structures to be processed § 10	The extent of building structures to be processed are included in the price.	Building structures to be processed are not part of the chargeable costs anymore and therefore not part of the price.
Engineering services subject to HOAI	The following engineering services were covered by the HOAI: environmental impact assessment, thermal building physics, sound insulation and room acoustics, soil mechanics and foundation works and technical surveying services.	These engineering services dropped out of the HOAI.
Fees for consulting services § 3	Prices for consulting services were covered by binding price law.	Prices for consulting services can be freely negotiated.
Bonus and malus fees § 5	In the case of a substantial cost reduction, a bonus fee of up to 20 % could be agreed on. A malus payment does not exist.	In addition to the success fee, a penalty fee up to 5% may be agreed on in writing.
Progress payments § 8	—	Progress payments can be agreed on in advance.
Allowances for modernization and rebuilds § 26	A supplemental allowance amounting from 22% to 33% may be agreed on.	Supplemental allowance for modernizations and rebuilds increases to 80%.

Source: Own description.

Figure 4: Changes to the price level by the HOAI reforms



Source: Own depiction based on data from the German Federal Statistical Office.

To account for inflation, the price indices are presented as ratios to the consumer price index (CPI) provided by the German Federal Statistical Office. The percentage change in the producer price indices over time is the difference on the vertical axis when moving one quarter forward.

As apparent from Figure 4, the binding price increase resulting from the HOAI reform is reflected in the price index. The producer price indices sharply increased by about 10 per cent in the third quarter of 2009 when the HOAI 2009 came into place. Compared to the HOAI 2009, prices were further increased by around 17 per cent on average due to the 2013 reform of the HOAI (see Figure 4). In theory, these changes could have also had positive effects on the quality of services, since architects need to be able to justify the higher prices that they are now allowed to charge.

2.4. Indicators

Measures of Quality

Measuring the quality of services of architects and construction engineers is very difficult because usually it cannot be observed ex ante (like for experience goods) and it is also hard to verify even ex post (like for credence goods). Roman architect Vitruvius named three qualities required of architecture in his De architectura of 15 BC: firmness, usefulness and delight. For contemporary practice, regulation guarantees a minimum level of quality in the form of required education in Germany. In addition to this, Vitruvius's list has been developed further by private international certification agencies (see Draeger 2012 for an overview).

The leading organizations LEED, BREEAM, HQE and DGNB are based in the US, UK, France and Germany respectively but operate internationally.[38] The emphasis of these certificates is on quality in the sense of sustainability and environmental impact; however, assessment of socio-cultural, economical, technical and locational quality by independent auditors are also considered for rating of services on a three to four point scale. We asked the Federal Chamber of German Architects and

[38] LEED is short for Leadership in Energy and Environmental Design, BREEAM for Building Research Establishment Environmental Assessment Methodology, HQE for Haute Qualité Environnementale, and DGNB for Deutsche Gesellschaft für Nachhaltiges Bauen.

the DGNB for access to the data for this study. However, unfortunately, the certification system was not yet widespread enough at the time of the reforms (only 75 buildings were certified by DGNB in November 2009) to allow measuring the effects of the reform but this will certainly be an excellent source of reliable data for future research.

In this study, we use a peer ranking published by BauNetz Media GmbH on its website. It is highly recognized in the profession of architects as a measure of quality and serves as a proxy for aesthetic quality. Many architectural firms signal the quality of their services by referring to their position in this ranking. Universities, for example, also recommend the ranking as a starting point to search for an internship. Of course, one has to account for the fact that this ranking, while it serves as a signalling device for architectural quality in practice, it is selective, since only slightly more than 1 percent of all architectural firms in Germany are included. Given the surprising lack of evidence available from professional associations or other sources, using this ranking as a proxy for service quality is an important first step for the analysis of regulation within the architectural profession, since it shows how to make use of a measure that goes beyond purely 'objective' quality measures.

Measures of Prices, Service Availability, Firm Creation and Destruction, Employment and Training on the Job

Figure 4 in the previous section showed that the reform of the VVG had no impact on prices, while the amendments to the HOAI directly increased prices. We obtained information on prices from the German Statistical Office. The price indices are based on the stipulations of the HOAI, construction sector statistics, national accounts and a survey of 143 firms.

The other outcome variables, i.e. service availability, firm dynamics, labour demand and human capital decisions, are all computed from data of the German microcensus and described below. As a measure of service availability, we use the self-reported indication of working self-employed and compute the annual share of self-employed architects and engineers to 1,000 inhabitants of a federal state within Germany.

Firm creation and destruction is measured using annual entry and exit rates. We calculate entry rates using a binary variable indicating whether a person was self-employed in the respective year and not self-employed the year before. The latter is obtained by explicitly asking for information on employment status 12 months ago. The variable includes being self-employed with and without employees. Similarly, exits are constructed from a binary variable indicating whether a person has exited from self-employment in the respective year. A person who exited was self-employed last year and is not self-employed this year. The previous employment status is obtained by explicitly asking for information on employment status 12 months ago.

Labour demand is measured by the self-reported number of individuals employed in a firm. This variable takes the values 1 to 10 for 1 to 10 employees and then 11 for 11 to 19 employees, 12 for 20 to 49 employees and 13 for 50 or more employees.

To assess the effects of the reforms on training on the job, we look at the extensive as well as the intensive margin. Continued education is thus either a binary variable indicating whether the person has engaged in continued training within the last 12 months or a continuous variable denoting the hours of training.

2.5. Data

Architectural Firm Ranking

We use the architectural firm ranking from 2006 to 2012 published every second month by BauNetz Media GmbH on its website. The ranking is based on publications made within the previous 24 months. The international ranking list, which we use in our analysis, is based on the evaluation of two German (Bauwelt, Detail) and five international professional journals (Architectural Review; a+u; architektur aktuell; Werk, Bauen und Wohnen; domus).

The professional journals are evaluated based on registration of the illustrated editorial contributions and documentations on architectural and urbanistic works of the last 10 years. Moreover, information on object description, object location, firm location, etc. is taken into account. The number of pages of a report is converted into a score, where a one-page illustrated note equals one point, a two-page short presentation two points, a three- to four-page small report three points, a regular report of five to seven pages four points and a large report of eight or more

pages gets five points. The final score is the weighted sum of points, where Bauwelt and Detail are assigned a weighting factor of two, and the other journals a weighting factor of three to take account of the information density of the journal pages.

Architects' Council of Europe

We use data on the number of architects and the market size in each state from data compiled from an online survey for the Architects' Council of Europe.

German Microcensus

The German microcensus (Mikrozensus) is an official representative yearly household survey comparable to the Current Population Survey in the United States. This data set covers approximately 830,000 persons in 370,000 households, which is 1 per cent of all private households in Germany. The microcensus provides information on the population structure as well as on the economic and social living conditions in Germany. Due to the mandatory nature of the German microcensus (based on §13 Mikrozensusgesetz; §15 Bundesstatistikgesetz), the data guarantees a low rate of item non-response, which is a major advantage compared to other surveys. This also ensures that our groups of interest, architects and engineers, are adequately represented.

2.6. Methods

In economic research, the most interesting and challenging part is the identification of cause and effect. In most contexts, policy makers are interested in the (average) causal effect of an intervention or a reform on a certain outcome variable (Angrist and Pischke 2008). In our case, we are interested in the effect of regulation on service quality of architects and engineers in the construction sector. In modern programme evaluation, it is common to think in counterfactuals in order to determine this causal effect.

The question of interest is therefore a hypothetical one: what would have been the outcome in a world without the reform? Transferred to the question at hand, one would like to know the level of service quality if the HOAI reform had not happened. The difference between this hypothetical outcome and the actual outcome would thus represent the causal effect of the reform. Based on Rubin (1974), Holland (1986) has further established the concept of potential outcomes, which is known as the 'Rubin causal model'. Formally, let $D_i = \{0,1\}$ be a binary variable defining the reform status and Y_i be the outcome of interest for unit i. Assuming that each unit is potentially exposable to the treatment (Holland 1986), there are two states of the world and hence two potential outcomes, $Y_i(0)$ and $Y_i(1)$, with the latter being the outcome in the world where unit i is treated. The real-world observed outcome can thus be written as:

$$Y_i = \begin{cases} Y_i(1), \text{ if } D_i = 1 \\ Y_i(0), \text{ if } D_i = 0. \end{cases}$$

This can be restated as

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0)] \times D_i.$$

The difference between the two potential outcomes $\tau_i = [Y_i(1) - Y_i(0)]$ is the causal effect of the reform on the outcome of interest. However, as Germany cannot be in both states of the world simultaneously, the hypothetical outcomes are never observed together, making it impossible to calculate the treatment effect. This is known as the 'Fundamental Problem of Causal Inference'. However, if treatment assignment was randomized, then a simple comparison of averages between treated (Germany) and untreated units (other countries) will identify the causal effect, as treatment would then be independent of any observed or unobserved characteristics.

As policy reforms are usually targeted interventions, treated and untreated units typically differ beyond their treatment status and a comparison of averages between treated and untreated units cannot identify the causal effect. If treatment status is based on unobservable factors as well ('confounded assignment'), OLS regressions or matching methods will fail. Then other methods, like the Difference-in-Differences (DiD) approach or the Synthetic Control Method (SCM), are needed for causal inference. The basic ideas of these two approaches will be explained below.

Difference-in-Differences

The DiD estimator is the most popular and intuitive approach when the treatment varies at a more aggregate level (e.g., at federal state level) and non-random treatment assignment arises from unobserved but fixed variables at the aggregate level (Angrist and Pischke 2008). By exploiting data with a time dimension, the idea is to compare the change in outcome for the treated unit over time with the corresponding change of a control unit, which has not been exposed to the treatment.

Under the crucial assumption that both units would have evolved similarly in absence of the treatment, the difference of these changes captures the causal effect of interest. If assignment to treatment is random, unobserved differences within the two groups cancel by taking first differences and allow to consistently estimate the causal effect.

Synthetic Control

First used by Abadie and Gardeazabal (2003) to estimate the effect of terrorism on economic growth, there is now a wide array of studies estimating treatment effects with SCM. Among other things, SCM has been used to measure the impact of the German reunification on West Germany (Abadie et al. 2014), to assess the effect of economic liberalization (Billmeier and Nannicini 2013), to examine the economic consequences of civil wars (Bove et al. 2014) and to evaluate regional policy programmes (Abadie et al. 2010; Gobillon and Magnac 2013).

Despite the huge content pluralism, those studies have a similar research design, making the use of SCM beneficial as in our case. First, the analyses are based on a small sample over a long time period, including typically aggregated units, i.e. we do not consider individual firm ranks but the distribution across states. Second, there is only one state that receives the treatment and the treatment assignment is assumed to be based on unobservable, time-varying factors. Finally, there are multiple potential control states. The basic setup of SCM is simple, and from a theoretical point of view very close to the potential outcome framework.

The idea is that a weighted average of the available control units, the ‘donor pool’, is able to reproduce the trajectory of the outcome of interest of the treated state in absence of the treatment. The treatment effect of interest can then be calculated by taking the difference between the actual outcome of the treated state in the post-intervention period and the outcome of the synthetic unit built from the control states. Under certain conditions and a large enough number of pre- intervention periods, the SCM estimator is able to identify a causal effect by matching pre-treatment outcomes, even if treatment assignment is based on unobservable, time-varying factors.

2.7. Results

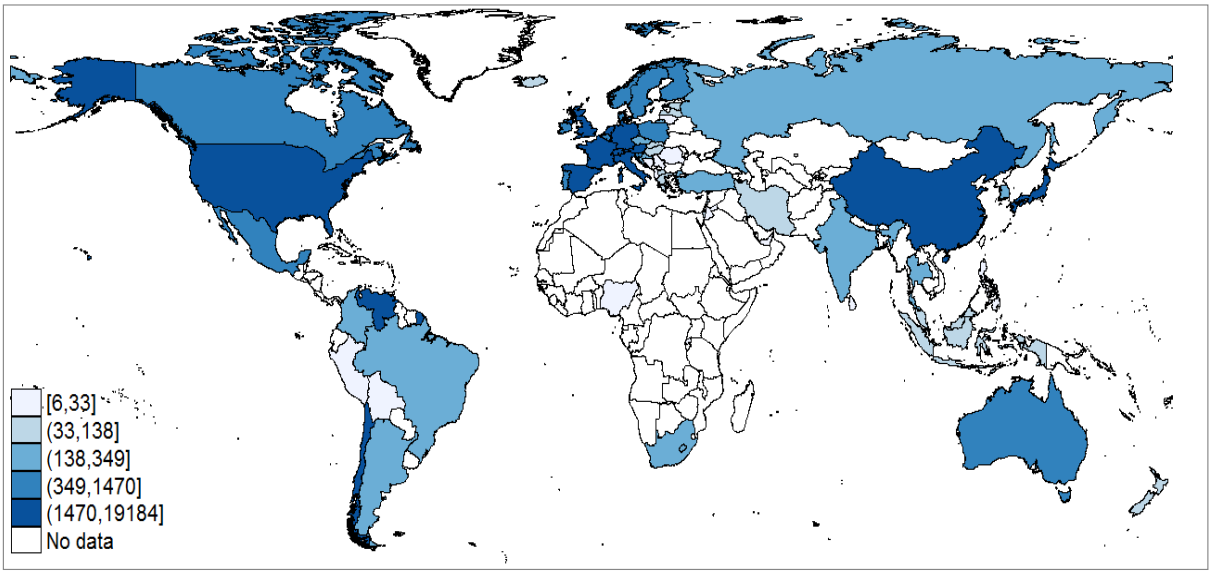
Effects of Increases in Regulated Prices on Service Quality

A major challenge to assess the impact of the reforms on service quality is to obtain reliable and valid measures of quality. We propose to use the peer ratings published by the construction magazine BauNetz since 2006 as one measure of quality.

Figure 5 shows the number of architectural firms within the peer ranking. The ranking includes firms from all over the world, except for most African and some Asian countries. The USA as well as most European countries are well represented with more than 1500 firms within the ranking. Figure 6 shows the average quality of architectural services across the world. Whether the HOAI reform in 2009 had an impact on architectural quality (measured by the scores of the peer ranking) is analysed using the DiD approach and the Synthetic Control method.

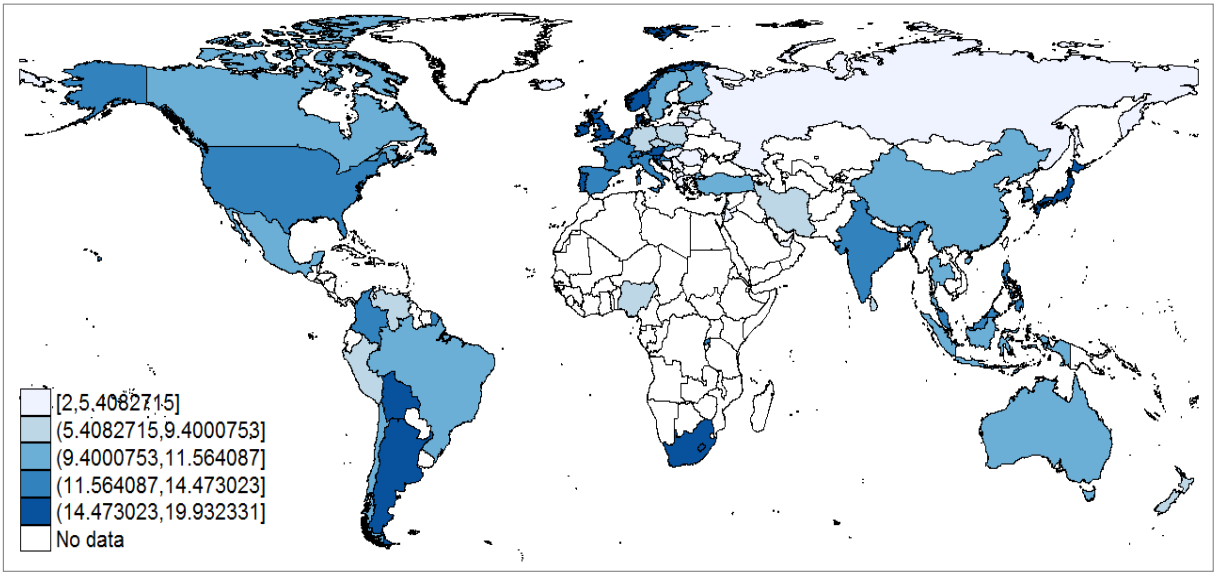
Table 5 shows the results of this pure DiD approach. The table reports the median score in the first row, the mean score (in log) in the second row and the 90 per cent quantile in the third row. All statistics are calculated separately for Germany (GER), the EU (except Germany) and the world (except Germany) in the pre- and post-reform period. If the latter two are suitable control groups, the before–after difference of Germany minus the before–after difference of the control group gives an unbiased estimate of the result of the HOAI reform.

Figure 5: Observations of Firms in the Peer-Ranking



Source: BauNetz Media GmbH: Firm-Ranking 2006-2016. Own Calculations.

Figure 6: Average Score in the Peer-Ranking



Source: BauNetz Media GmbH: Firm-Ranking 2006-2016. Own Calculations.

Table 5. Peer-Ranking Scores Before and After the Reform 2009

		GER	EU	World	Δ to EU	Δ to World
Pre-Reform	Median	1.6	2.2	2.2	-0.6	-0.6
	Mean	1.6	2.2	2.1	-0.6	-0.5
	90th	3.0	3.5	3.3	-0.5	-0.3
Post-Reform	Median	1.4	2.2	2.2	-0.8	-0.8
	Mean	1.6	2.2	2.1	-0.6	-0.5
	90th	2.8	3.4	3.2	-0.6	-0.4
Δ time	Δ Median	-0.2	0.0	0.0	-0.2	-0.2
	Δ Mean	0.0	0.0	0.0	0.0	0.0
	Δ 90th	-0.2	-0.1	-0.1	-0.1	-0.1

Notes: Median above mean above 90% quantile. Lower quantiles do not vary significantly. Scores are in logs.
Source: BauNetz Media GmbH: Firm-Ranking 2006-2012. Own calculations.

Table 5 shows that Germany has an average peer ranking score of 1.6 (in logs) before 2010, which is our pre-intervention period. The median score is 1.6 and the 90 per cent quantile is, of course, higher with a pre-intervention value of 3.0. After the reform, the median and the 90 per cent quantile have decreased, while the average score stays at 1.6. At first glance, the reform seems to have decreased the architectural quality in Germany as the scores decreased after the reform. However, this effect might also be driven by various other factors and could be the result of a general downward trend. To control these effects, we compare these differences for Germany with the respective change in i) the EU and ii) the world.

Compared to Germany, the EU average and the worldwide average score are higher by almost 0.5 points, suggesting that German architects and engineers deliver less quality compared to their European and worldwide colleagues. Thus, a raw comparison between Germany and the rest of the EU (the rest of the world) after the reform would highly overestimate the negative effect of the reform as they already differed in their scores before the HOAI 2009 was introduced. To get a first causal estimate, we therefore take 'the difference of the differences'. More precisely, we subtract the general time trend inferred from the control states from the change in scores over time for Germany. These double differences can be found in the right-hand lower corner of Table 5. Interestingly, the estimates between the two control groups do not differ. For the median, we find a very small negative effect of -0.2 suggesting that the reform had a negative effect on architectural quality. While we find no effect on the average quality score, the effect for the 90 per cent quantile is again negative with an estimate of -0.1. The sign and magnitude of the effect might give a first impression on the quality effect of the HOAI 2009. However, more information on the magnitude of the effect and the statistical significance would be needed to draw any profound conclusion.

In addition, these first results are only credible if the common trend assumption is fulfilled, meaning that Germany follows the same time trend as the EU or the world. As this assumption cannot be directly verified, we go one step further and apply the Synthetic Control Method. The estimation technique relaxes the common trend assumption and is able to control for time-varying, unobserved variables.

The results of the SCM estimation are graphically depicted in Figure 7, which plots the average quality score against time. To analyse the effects of the HOAI 2009, we restrict the sample to the period January 2006 to December 2012. Before discussing the results, Tables 6 and 7 give more information on the output of the estimation procedure. Table 6 lists all countries for which we have full data of the quality scores, the architects to population ratio, population, GDP per capita, market size and government expenditure.

These variables enter the estimation equation and are used to construct the synthetic control unit. Using a complex optimization procedure, the SCM estimation has constructed the synthetic twin of Germany, which is just a weighted average of the countries in the donor pool. In this case, it consists of 6.9 per cent Croatia, 34.4 per cent France, 3.1 per cent Italy, 23.7 per cent Spain, 4.2 per cent Switzerland and 27.7 per cent Turkey. Thus, France, Spain and Turkey are the main ingredients of the synthetic twin.

Table 6. Comparison Germany and its synthetic control: Weights

Country	Weight	Country	Weight
Austria	0.000	Poland	0.000
Belgium	0.000	Portugal	0.000
Croatia	0.069	Slovenia	0.000
Denmark	0.000	Spain	0.237
Finland	0.000	Sweden	0.000
France	0.344	Switzerland	0.042
Ireland	0.000	Turkey	0.277
Italy	0.031	UK	0.000
Netherlands	0.000		

Source: BauNetz Media GmbH: Firm-Ranking 2006-2012, Architects' Council of Europe, World Bank Data, own calculation.

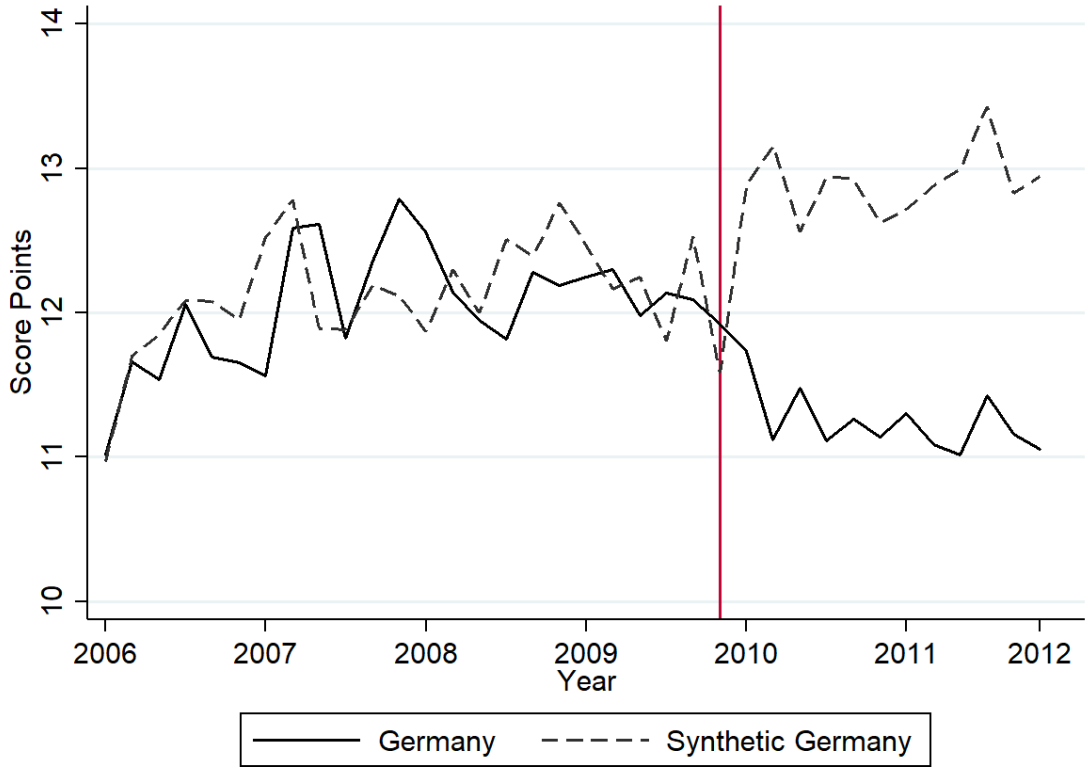
Table 7 further compares the control variables for Germany and its synthetic counterpart. As can be seen from the table, they are both very similar with respect to the pre-intervention outcomes (averages for each pre-intervention year) and the other control variables. This is evidence that the optimization procedure has worked well and the control unit is sufficiently comparable to the treated unit Germany.

Table 7. Comparison Germany and its synthetic control

Predictor	Germany	Synthetic Germany
Average Score in 2006	11.60	11.77
Average Score in 2007	12.29	12.23
Average Score in 2008	12.16	12.31
Average Score Jan to Aug in 2009	12.11	12.13
Architects/population ratio (in 1,000)	1.16	0.66
Population (in log)	18.22	17.64
GDP per capita	39,848	30,319
Market size	212,279	209,336
Government consumption	18.23	18.76

Notes: This table reports the predictor values for Germany and the synthetic control group ("Synthetic Germany"). The synthetic control group is an optimally weighted average of Croatia, France, Italy, Spain, Switzerland and Turkey
Source: BauNetz Media GmbH: Firm-Ranking 2006-2012, Architects' Council of Europe, World Bank Data, own calculation.

Figure 7: Graphical Results Synthetic Control Estimation



Notes: This graph plots the time series from January 2006 to December 2012 of the quality score for Germany and the synthetic control group ("Synthetic Germany"). The synthetic control group is an optimally weighted average of Croatia, France, Italy, Spain, Switzerland and Turkey.
Source: BauNetz Media GmbH: Firm-Ranking 2006-2012. Own Calculations.

The successful construction of the synthetic twin is also graphically represented in Figure 7, which depicts the time series of the outcome of interest (score) for Germany and its control. The red vertical line shows the reform period. The figure clearly visualizes that in the pre-intervention period, the German twin is able to reproduce the trajectory of Germany quite well. Both curves follow the same pre-trend and overlap each other.

However, at the beginning of 2010, the two time series clearly diverge. The difference between the average score of Germany and the score of Synthetic Germany is the estimated treatment effect, which is graphically depicted by the vertical distance between the two curves in Figure 7. While the score for the synthetic control group (dashed) slightly increases over time, indicating a positive time trend in scores, Germany experiences a decline. Thus, the HOAI 2009 seems to have decreased architectural quality. This is in line with the simple DiD results from above. This negative effect even grows stronger over time. Quantitatively, the reform has decreased average scores by about two points, which translates into an effect of minus 18 per cent.

To assess the significance of the results, we assessed several placebo studies, where treatment status is reassigned to all placebo units in the donor pool (see Abadie and Gardeazabal (2003) for technical details). The pseudo p-value from this exercise is 0.18. Although this value does not suggest statistical significance using conventional procedures, it should be noted that this is due to the discrete number of control units in the donor pools. Stated differently, Germany has the third largest effect within all the placebo studies, which would indeed suggest that the effect is significant.

Effects of Increases in Regulated Prices on Service Availability, Firm Creation and Destruction, Employment and Training on the Job

Table 8 reports descriptive evidence on business formation and business survival from self-reported information of entry and exit from self-employment. In addition, the table contains the number of employees as a measure for firm size and the share of professionals who have attended continued education as well as the average number of hours of continued education in the last 12 months. For each variable, we report weighted averages for our three subgroups: architects, engineers in the constructions sector (HOAI Engineers) and all other engineers who are not restricted by the HOAI (Other Engineers).

Table 8. Weighted averages by treatment and control groups in pre- and post 2009 samples

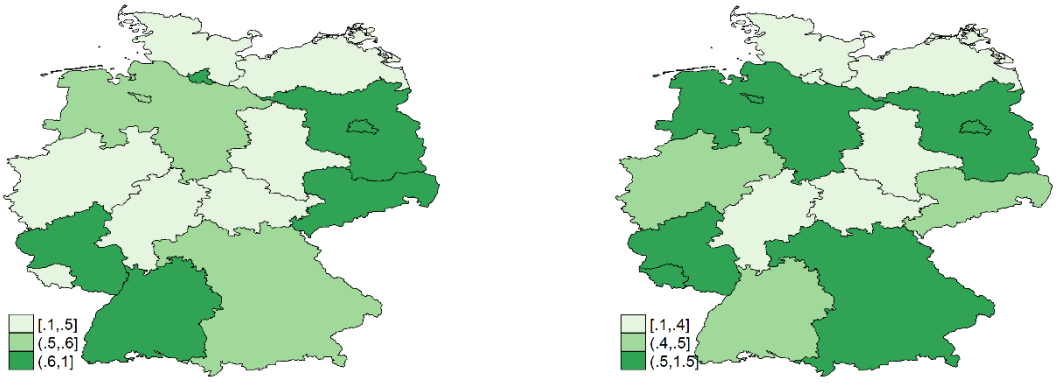
		Architects		HOAI Eng.		Other Eng.		Working Pop.	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post
Self-Employed	%	57.50	53.50	40.80	38.60	7.90	7.60	11.80	11.80
Entry	%	5.30	2.54	1.61	1.34	0.43	0.19	0.59	0.45
Exit	%	1.40	0.86	2.06	0.61	2.36	1.23	1.25	1.31
Employees	#	5.4	5.7	7.1	7.8	11.8	11.9	10.3	10.3
Cont. educ.	%	31.4	30.3	35.8	37.7	33.4	33.3	20.0	21.2
Cont. educ.	h	13.4	14.7	15.6	15.7	15.7	20.3	10.7	13.9

Notes: All numbers are weighted by survey weights provided by the microcensus. Entries and exits are based on a very small number of observations from architects and HOAI engineers.
Source: Own calculations based on the scientific use file of the German microcensus (2006-7;2009-12).

Compared to engineers, the share of self-employed architects is very high. From 2006 to 2009 (denoted as 'Pre'), more than half of all architects were self-employed. After the introduction of the HOAI 2009, this share decreased by four percentage points. The share of self-employed engineers in the construction sector is in general lower than that of architects but still almost four times the corresponding figure for the total working population (last two columns). This share remains constant at somewhat less than 40 per cent after the introduction of the HOAI reform in 2009. All other engineers face a substantially lower and constant self-employment rate of less than 8 per cent.

We moreover investigate whether the reform had an impact on entry and exit rates. The exit rate, for example, is constructed by creating a dummy variable indicating whether a person was an architect/engineer and self-employed in period t-1, but no longer self-employed in the current period t. Entry rates are developed the other way around with being self-employed in t, but not in t-1. Architects had the highest entry rates with 5.30 per cent in the pre-reform period. For all subgroups, we see a decreasing trend in entry rates. The same trend is observed for the exit rates except for the total working population.

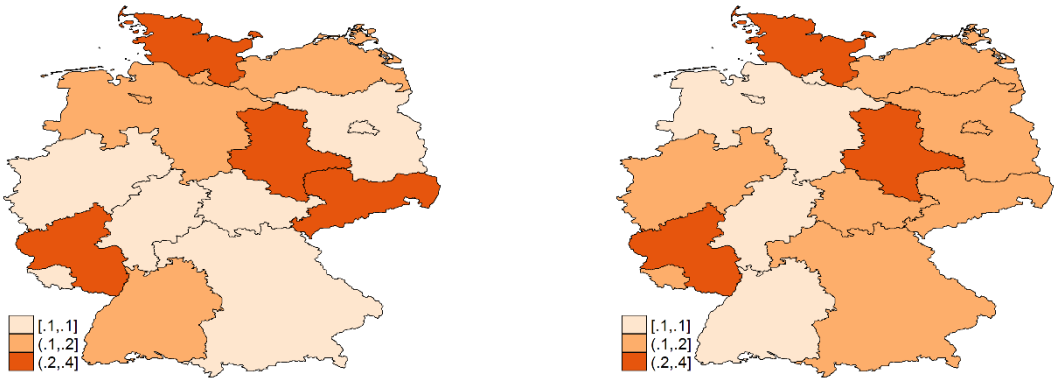
Figure 8: Service Availability of Architects per 1,000 Inhabitants



2006-2007 2009-2012

Source: Own Calculations based on the German Mircocensus.

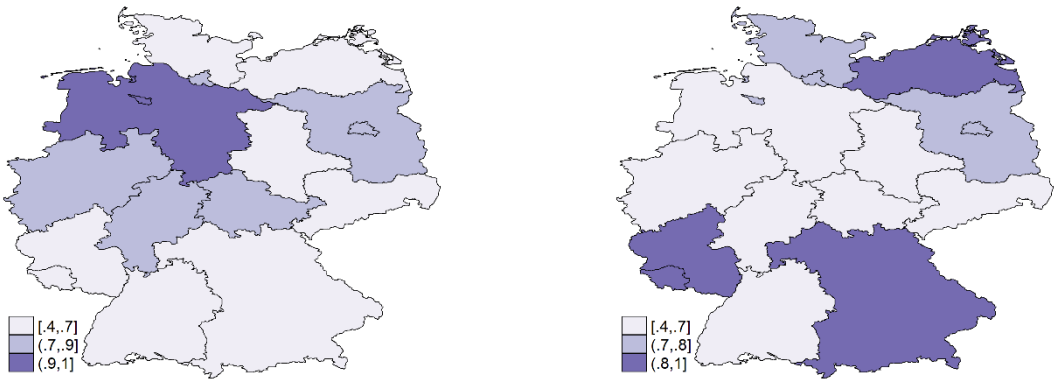
Figure 9: Service Availability of HOAI Engineers per 1,000 Inhabitants



2006-2007 2009-2012

Source: Own Calculations based on the German Mircocensus.

Figure 10: Service Availability of Other Engineers per 1,000 Inhabitants



2006-2007 2009-2012

Source: Own Calculations based on the German Mircocensus.

After 2009, firm size increased by less than one employee for architects and construction engineers, while it stayed constant for the other two subgroups. Interestingly, about one in three architects and engineers take part in regular continuous education. This share increases for all groups. The average number of hours spent on continued education has increased substantially for all groups. Therefore, this seems at first glance not to be an effect of the reform.

Figures 8 to 10 present descriptive evidence for the access to architectural and engineering services from a geographical perspective. Access to services is measured as the share of self-employed architects and engineers to 1,000 inhabitants of a federal state within Germany. The Federal Chamber of German Architects reports in total 134.419 architects for 2018. The number of inhabitants in Germany at the end of 2016 was 82,522,000 persons. The share of architects per 1,000 inhabitants is therefore 1.63. Since about half of all architects are self-employed, the share of self-employed architects among 1,000 inhabitants is about 0.8 implying that 8 self-employed architects serve 10,000 inhabitants. The left-hand map of Figure 8 shows for 2006 and 2007 low service availability (one to five architects per 10,000 inhabitants) in the light colour, medium service availability (five to six architects per 10,000 inhabitants) in a darker shade and high service availability (six to 10 architects per 10,000 inhabitants) in the darkest colour. Generally, the variations across federal states are not very large. Compared to the right-hand map that shows the same information for 2009 to 2012 (though the class definition varies slightly), five federal states remain at the same low level of service availability as in 2006/7.

For HOAI engineers shown in the two maps of Figure 9, the variation across federal states is substantially lower, ranging from one HOAI engineer per 10,000 inhabitants (low availability) and more than one but fewer than two HOAI engineers per 10,000 inhabitants (medium availability) to up to four HOAI engineers per 10,000 inhabitants (high availability). Interestingly, Schleswig-Holstein, Saxony-Anhalt and Mecklenburg-Vorpommern seem to substitute low availability of architects with higher availability of HOAI engineers. Figure 10 shows two maps for availability of services of other engineers. The number of self-employed other engineers is much higher: low availability means four to seven engineers, medium seven to nine (to eight in 2009–12) and high up to 10 engineers per 10,000 inhabitants.

Tables 10 and 11 report the results using the German microcensus to evaluate the VVG and HOAI 2009 reform. Using OLS regression models, we want to identify whether the two reforms had a causal impact on the following outcome variables: the entry and exit rate into and from self-employment, the probability of being self-employed, firm size (measured by the number of employees) and continued education (extensive and intensive margin responses). From the previous theoretical and empirical literature (e.g. Rostam-Afschar and Strohmaier, 2018 and Branstetter et al. 2013), we expect the two reforms to have the following effects on our outcome measures ex ante:

Table 9. Theoretical Predictions

Dependend Variable	VVG 2008	HOAI 2009
Entry Rate	-	+
Exit rate	+	-
Self-employment	-	+
Firm size	-	+
Continued education	+	+

Source: Own considerations.

For all outcomes, the estimation equation reads as follows:

$$y_{ist} = \alpha + \delta\, Treated_i + \beta\, postHOAI_{it} \times Treated + \gamma\, postVVG_{it} \times Treated + X_{ist} + \mu_s + \rho_t + \varepsilon_{ist}\ ,$$

where y_{ist} denotes the respective outcome for individual i in state s in year t. $Treated$ is an indicator that tracks architects and engineers before and after the reforms. This variable is zero for the control group, which consists of all other engineers (as already shown in Figure 4) for which the

HOAI does not apply.[39] The variables $postHOAI_{it}$ and $postVVG_{it}$ are dummy variables which are zero for pre-reform years and one otherwise. Following a difference-in-differences approach, the coefficients β and γ of the respective interactions capture the causal effect of the HOAI and VVG.

In addition, we augment the regression by a full set of control variables X_{ist} including indicator variables for being female, German, married as well as dummies for the educational qualification, age, age squared and net income groups. We also include state and year fixed effects μ_s and ρ_t to absorb time-constant unobserved state-level factors and common shocks. Finally, ε_{ist} denotes an error term. The results for the entry rate into self-employment are reported in Table 10.

To estimate the effects on the decision to become self-employed, we restrict the sample to all individuals who have been in dependent employment 12 months before t. In columns (1) and (2) we solely regress the entry rate on the treated dummy, a post-reform dummy (not reported) and the relevant interaction denoted by $postHOAI \times Treated$ and $postVVG \times Treated$. Interestingly, if we only look at the HOAI reform without taking the VVG reform into account, the HOAI has a negative impact on the entry rate, though not significantly different from zero. However, as soon as we include both reforms simultaneously, the HOAI coefficient turns positive and significant and stays almost constant throughout all specifications.

As predicted by theory, the HOAI with its sharp increase in regulated prices leads to an increase in the probability of entering self-employment as the monetary incentives have increased. Quantitatively, the HOAI reform increased the entry rate by one percentage point. Related to the mean entry rate in 2006[40] of 0.02, this is an increase of 50 per cent. On the other hand, the results show that the VVG reform of 2008 has a strong negative impact on the entry decision. The effect amounts to approximately 0.02 and is very precisely estimated in all specifications. Again, as expected, this reform has led to a drastic decrease in the entry rate.

Table 10. Effect on the Entry Rate

	(1)	(2)	(3)	(4)	(5)
Treated	0.0240***	0.0341***	0.0341***	0.0316***	0.0316***
	(0.0018)	(0.0026)	(0.0026)	(0.0032)	(0.0026)
postHOAI × Treated	-0.0029		0.0066**	0.0071**	0.0072**
	(0.0026)		(0.0032)	(0.0032)	(0.0032)
postVVG × Treated		-0.0154***	-0.0196***	-0.0198***	-0.0199
		(0.0030)	(0.0036)	(0.0036)	(0.0036)
Observations	35,145	35,145	35,145	35,145	35,145
Controls				✓	✓
Year Fixed Effects					✓
State Fixed Effects					✓

Notes: This table shows the results from OLS regressions where the entry rate is the dependent variable. In column (4), we add a full set of control variables, including a dummy for being female, German, married, the number of children, experience, dummies for education and categories of net income (both not reported). In column (5), year and state fixed effects are added. In column (6), we further cluster the standard errors by year, thus allowing the errors within the same year to be correlated. Level of significance: $*$ p < 0.10, $**$ p < 0.05, $***$ p < 0.01.
Source: Own calculations based on the scientific use file of the German microcensus (2006-2012).

[39] We broadly follow the definition of the German Statistical Office to define the treatment group. To be more precise, we decided to combine the German Classification of Occupations 1992 (Codes 609 and 836 for architects and Code 603 for civil engineers) and the German Classification of Economic Activities 2003 and 2008 (Code 742 for WZ 2003 and Code 711 for WZ 2008) to identify our treatment group.

[40] We decided to take 2006 as the baseline comparison year and not the complete pre-reform period due to the potential effect of the VVG.

Table 11. Effect on Exit, Self-Employment, Firm size and Continued Education

	(1)	(2)	(3)	(4)	(5)
	Exit rate	Self-employed	Firm size	Cont. educ. (y/n)	Cont. educ. (in h)
Treated	-0.0105	0.4536***	-5.525***	-0.0019	1.7391
	(0.0066)	(0.0093)	(0.0994)	(0.0151)	(2.8400)
postHOAI × Treated	-0.0094	-0.0200*	0.3662***	-0.0177	-4.1753
	(0.0084)	(0.0115)	(0.1231)	(0.0187)	(2.8400)
postVVG × Treated	0.0097	-0.0222*	0.0432	0.0336	-1.8038
	(0.0092)	(0.0129)	(0.1379)	(0.0209)	(3.3231)
Observations	4,212	35,161	34,936	35,161	33,407
Controls	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓
State Fixed Effects	✓	✓	✓	✓	✓

Notes: This table shows the results from OLS regressions with different dependent variables. All specification include a full set of controls and year and state fixed effects. Standard errors are clustered by year.
Level of significance: $*$ p < 0.10, $**$ p < 0.05, $***$ p < 0.01.
Source: Own calculations based on the scientific use file of the German microcensus (2006-2012).

Column (1) in Table 11 reruns the full specification with all controls and fixed effects for the exit rate. Here, we condition on having been self-employed 12 months before year t to estimate the effects on switching from being self-employed to dependent employment. Although the signs of the coefficients point in the expected direction (based on economic theory), both reform coefficients are statistically insignificant. This could be rationalized by the small number of exits in our sample, as there were only 55 exits in total.

As further apparent from Column (2), the HOAI has decreased the probability of being self-employed with a significant negative effect of 0.02. In line with these results, firm size has increased due to the HOAI reform. On average, firm size increased by 0.4 employees per firm which is an increase of 6 per cent related to the baseline year 2006. This suggests that market concentration seems to have increased due to the reform which might have a negative impact on service quality. With respect to continued education, the regressions are inconclusive and suggest a positive 'extensive effect' (Column (4)) due to the VVG reform, which can be considered marginally significant (p-value=0.109) and a negative, but insignificant, 'intensive' effect (Column (5)). The reason for this is that the so-called "Engineering Card" increased the propensity to engage in education for the control group. It was introduced in 2010 in Germany. This confounding factor makes it very difficult to separately identify the reform effects and therefore these estimates cannot be interpreted as evidence for a decrease in service quality.

2.8. Conclusions

To evaluate the effect of regulation on service quality, two reforms are of special interest regarding architects and civil engineers in Germany. The VVG-Reform 2008 increased professional indemnity insurance requirements to 1 million euros per year which affects mainly small and medium-sized firms that previously contributed 0.5 million euros per year. The HOAI-Reform 2009 increased prices by 10 per cent for architects and by 8 per cent for civil engineers. Four years later, the HOAI reform 2013 increased prices by further 24 per cent for architects and by 17 per cent for civil engineers. However, data to analyse the latter reform are not available yet.

In order to measure the effects of the HOAI 2009 reform, we combine data from different sources from which we generate a synthetic control group against which we compare the average firm rank of a peer ranking for architects. We show that architectural quality measured by the peer ranking score decreased in Germany due to the HOAI 2009 which may give a first hint that the drastic price increase might not have had the intended positive effects on service quality.

Further, we show that the market structure of architects and engineers subject to the HOAI changed in response to the VVG and HOAI reforms. In particular, fewer but larger firms resulted from the reforms that could--with sufficient market power--compromise service quality, though the high rate of self-employment suggests that architects and construction engineers are far from such a situation. The evidence from this study shows that higher insurance costs discouraged new firms from entry, while exits could well have remained constant (though our evidence is weak for this). In line with this, the probability of being self-employed decreased and the results show that most of the former self-employed now work as employees for larger firms. We were not able to identify the effects of the HOAI reform on continued education.

Besides expressing the hope for opportunities to collect and access data of higher quality that will make results of future research clearer, we stress that our results focus only on some dimensions of quality and serve as an example of how to improve measurement of quality and evaluation of reform effects in future research. In our samples there is always some sort of enforcement of minimum quality. Therefore, a prediction of the consequences of deregulating on all guarantees of quality is not within the support of the data and could have much more harmful effects in practice than our linear model suggests.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," Journal of the American Statistical Association, 105, 493–505.

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: a Case Study of the Basque Country," American Economic Review, 93, 113–132.

Abadie, A., A. Diamond, and J. Hainmueller (2014): "Comparative Politics and the Synthetic Control Method," American Journal of Political Science, Wiley Online Library.

Acemoglu, D., and Verdier, T. (1998) "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," Economic Journal, 108, 1381–1403.

Ahlfeldt, G. M. and Holman, N. (2017) "Distinctively Different: A New Approach to Valuing Architectural Amenities," Economic Journal, in press.

Angrist, J. and J. Pischke (2008): Mostly Harmless Econometrics - An Empiricist's Companion, Princeton, NJ: Princeton University Press.

Architect's Council of Europe (2017) "The Architectural Profession in Europe 2016", Sector Study, Architect's Council of Europe.

Atkeson, A., Hellwig, C., and Ordoñez, G. (2015) Optimal Regulation in the Presence of Reputation Concerns, Quarterly Journal of Economics, 130, 415-464.

Billmeier, A. and T. Nannicini (2013): "Assessing Economic Liberalization Episodes: A Synthetic Control Approach," The Review of Economics and Statistics, 95, 983–1001.

Bove, V., L. Eliay, and R. P. Smith (2014): "The Relationship between Panel and Synthetic Control Estimators of the Effect of Civil War," BCAM Working Papers.

Branstetter, L., Lima, F., Taylor, L.J. and Venâncio, A. (2013), "Do Entry Regulations Deter Entrepreneurship and Job Creation? Evidence from Recent Reforms in Portugal", Economic Journal, 124 No. 577.

Darby, M. R., and E. Karni (1973): "Free Competition and the Optimal Amount of Fraud." Journal of Law and Economics, 16(1): 67–88.

Draeger, S. (2012). "Vergleich des Systems des Deutschen Gütesiegels Nachhaltiges Bauen mit internationalen Systemen – Endbericht", Bundesamt für Bauwesen und Raumentwicklung, Aktenzeichen SF – 10.08.17.7-09.15.

European Commission (2016). "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Evaluating National Regulations on Reform Recommendations for Regulation in Professional Services", Brussels 2016, COM(2016)820 final 10.01.2017.

Federal Chamber of German Architects (2016). "Analyse der Büros selbstständig tätiger Mitglieder der Architektenkammern der Länder: Ergebnisse einer Repräsentativbefragung 2016 für das Berichtsjahr 2015", structure study.

Federal Institute for Occupational Safety and Health (2012): "Qualität des Bauens: Eine Studie über den Begriff und die Wahrnehmung von Bauqualität bei privaten und öffentlichen Bauherren und Bauunternehmen," report, 44.

Fuerst, F., McAllister, P. and Murray C. (2012). "Designer Buildings: An Evaluation of the Price Impacts of Signature Architects", Real Estate & Planning Working Papers 2009-10, Henley Business School, Reading University.

Fossen, F., Rees, R., Rostam-Afschar, D., and Steiner, V. (2017) "How do Entrepreneurial Portfolios Respond to Income Taxation?," Hohenheim Discussion Papers in Business, Economics and Social Sciences 12-2017, University of Hohenheim, Faculty of Business, Economics and Social Sciences.

Gat, D. (1998). “Urban Focal Points and Design Quality Influence Rents: The Tel Aviv Office Market”, Journal of Real Estate Research 16(2), 229-247.

German Monopolies Commission (2006). “Mehr Wettbewerb auch im Dienstleistungssektor!“, XVI (2004/2005) of Hauptgutachten der Monopolkommission, chap. VI: Regulierungen der Freien Berufe. Bonn.

Gobillon, L. and T. Magnac (2013): “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls,” IZA Discussion Papers.

Holland, P. W. (1986): “Statistics and Causal Inference,” Journal of the American Statistical Association, 81, 945–960.

Imrie, R. and Street, E. (2009). “Risk, Regulation and the Practices of Architects”, Urban Studies 46, 2555–76.

Krause-Allenstein, F. (2008). “Praxisrelevante Änderungen des neuen Versicherungsvertragsgesetzes für das Bauversicherungsrecht”, NZBau 9(2), 81–144.

Meran, G., and Schwarze, R. (2010). “Can Minimum Prices Assure the Quality of Professional Services?”, European Journal of Law and Economics 30(2), 171–199.

Murphy, K. M., Shleifer, A. and R. W. Vishny (1991) “The Allocation of Talent: Implications for Growth.” Quarterly Journal of Economics 106 (2): 503-530.

Nelson, P. (1970). “Information and Consumer Behavior.” Journal of Political Economy, 78(2): 311–29.

Pesendorfer, W. and A. Wolinsky (2003) “Second Opinions and Price Competition: Inefficiency in the Market for Expert Advice.” Review of Economic Studies, 70(2): 417–37.

Rostam-Afschar, D. (2014). “Entry Regulation and Entrepreneurship: A Natural Experiment in German Craftsmanship”, Empirical Economics 47(3), 1067–1101.

Rostam-Afschar, D. (2015). “Regulatory Effects of the Amendment to the HwO in 2004 in German Craftsmanship”, Economic Analysis for the European Commission, Directorate General Internal Market and Services.

Rostam-Afschar, D. and K. Strohmaier (2018). “Does Regulation Trade-Off Quality against Inequality? The Case of German Architects and Construction Engineers”, Discussion Paper 23-2018, Universität Hohenheim.

Rubin, D. (1974): “Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies,” Journal of Educational Psychology, 66, 688–701.

Schwarze, R. (2007). “Architekturbüros: Ausweitung der Berufshaftpflicht verstärkt Marktkonzentration“, DIW Wochenbericht, DIW Berlin, German Institute for Economic Research, 74(10), 159–161.

Vitruvius, M. “De Architectura Libri Decem”, ed. Krohn, F., Teubner, Leipzig 1912

(This page intentionally left blank)

3. Pharmacists in Italy

3.1. Introduction

The market for pharmacists is heavily regulated in most European countries. This is typically justified by the dangers associated with the improper use of medicines. In Italy, the work of pharmacists and the production, trade and distribution of drugs is heavily regulated. Most drugs must be distributed through an authorized pharmacy, the main exception being over-the-counter (OTC) drugs, which are subject to lighter regulation. Pharmacies must employ licensed pharmacists, subject to educational requirements and behavioural regulations. The number of pharmacies and their location is also subject to strict regulation and it is determined through an administrative procedure (the most important determinant of the number of pharmacies in a given city being the size of the local population). The rules used for determining the number of pharmacies in each city are relatively stable, although significant changes occurred in 1991 and 2012, which provided for increases in the number of pharmacies.

Since most drugs (with the exception of OTC drugs) sold in pharmacies have a regulated price, the key factor affecting the quality of the service is product availability. Hence, in this chapter, we will focus on the impact of regulation on the availability of pharmacies and its potential effects on human health. We use health outcome measures from hospital admission records that can potentially capture the impact of access to medicines (and the other services provided in pharmacies) on the health of the population. We complement this data with information from the National Institute of Statistics (ISTAT) on consumer perception of difficulties accessing pharmacies. Some general data on customers' complaints and satisfaction are also provided.

The data on the exact location and opening date of each pharmacy in Italy are provided by the Italian government. Health indicators are based on administrative data. In fact, the Italian National Health System collects data on all the episodes of care delivered to the residents, irrespective of citizenship, social class and income. Such data are characterized by a high level of standardization and homogeneity across space (different regions and local health units) and time, and are available at population level.

This chapter is organized as follows. In section 3.2, we provide a brief description of the pharmacy market and the literature on the regulation of this specific market. In Section 3.3, we focus on entry regulation in the Italian pharmacy market, describing in detail the mechanism that links the population in each municipality to the number of pharmacies allowed to enter the market. We also describe in detail the data used in the analysis. Section 3.4 provides a description of the empirical methodologies used, and Section 3.5 reports the results. Concluding remarks are presented in Section 3.6.

3.2 The Pharmacy Market

The pharmacy market in Europe is characterized by significant government control on entry (geographical or demographic regulations), scope of activities and profit margins. Over the past two decades, the system began to change due to the introduction of a series of reforms. There is a lack of studies on the impact of such reforms and existing studies tend to suffer from a high degree of heterogeneity across countries in the initial level of regulation and in the type of reform introduced.

Vogler *et al*. (2014) classify countries into two broad groups. The first group includes regulated countries such as Austria, Denmark, Finland and Spain. The second is made up of less regulated countries such as England, Ireland, the Netherlands, Norway and Sweden.[41] In general, regulated countries apply demographic rules (e.g. number of potential consumers for each pharmacy) and geographic criteria (e.g. distance to existing pharmacies) to determine entry into the market. Moreover, they tend to impose restrictions on ownership of pharmacies. Less regulated countries are subject to different types of liberalization.[42]

[41] Unfortunately, Italy is among the countries not included in the analysis.
[42] In England, in 2005, there was a revision of the 'control of entry test' system. Ireland revoked, in 2002, the rules for the opening of new pharmacies and the Netherlands abolished in 1998 the restrictions on the establishment of new

Vogler *et al.* (2014) find that, after deregulation, the number of pharmacies increases and the ratio between the number of inhabitants per community pharmacy drops, especially in Norway and Sweden. These new pharmacies tend to be established mainly in urban areas, with few new pharmacies in rural areas. There are indications of an increased workload for pharmacists in some deregulated countries. Moreover, they note that after liberalization specific stakeholders (e.g. wholesalers) may gain market power and possibly limit competition (Norway), partially offsetting the increase in competition due to the larger number of pharmacies. Anell (2005) finds somewhat similar results in his study of the pharmacy market in Norway and Iceland. The number of pharmacies increases (and prices tend to decrease), but also concentration seems to increase, as a result of horizontal integration of formerly independent pharmacies.

Verboven and Schaumans (2014) study the entrance and the conduct regulation of pharmacies and physicians in Belgium. Belgium is a regulated country, where both professions have regulated fees and mark-ups, and where there is a ban on most types of advertising. Moreover, there are significant restrictions to entry based on geography and size of the local population.

The presence of these kinds of regulations is often justified to ensure a minimum availability of supply in less profitable regions without inducing excessive entry elsewhere. The authors try to evaluate this public interest motivation and develop an econometric model of entry by two types of professions: physicians and pharmacies. They use the models of free entry of Bresnahan and Reiss (1991a) and Mazzeo (2002) to model entry and market structure.

The results show that entry in one profession has a positive effect on the profitability of entry into the other profession, supporting the idea that these professions are strategic complements. Moreover, geographical entry restrictions substantially limit the number of firms. They conclude that the current regime of high regulated mark-ups and restricted entry seems to protect the private interests of existing pharmacies more than the public interests. They also suggest it is possible to induce a large shift in rents to consumers without reducing geographic coverage, combining reductions in geographical and mark-ups restrictions.

The Italian Pharmacy Market

In Italy the production, distribution and pricing of pharmaceuticals are strictly regulated by the Italian Agency for Pharmaceuticals (AIFA). AIFA approves the pharmaceuticals that can be produced, used and marketed in Italy, and authorizes clinical trials. Medicines are grouped into three main classes: Class A are essential medicines that are reimbursable (but require a co-payment, which varies by region) and include those indicated for the treatment of severe illnesses. Class H includes pharmaceutical products delivered only within hospitals. Class C includes non-reimbursable pharmaceutical products (though some regions opt to offer partial reimbursement). Class C pharmaceuticals are divided, in turn, into medicines with compulsory medical prescription and medicines not requiring a compulsory medical prescription. The latter are over-the-counter (OTC) pharmaceuticals for minor, mild or intrinsically transient conditions or symptoms, suitable for self-medication.

Pharmacists are entitled to distribute prescription drugs to consumers. Pharmacists are required to meet educational requirements and to pass a licensing exam. OTC drugs can be distributed to consumers within pharmacies, supermarkets, and other shops.[43] In the latter two cases, a pharmacist must always be present during store opening hours.[44] There is little variability in the type of service provided by Italian pharmacies, since prices are fixed and the training and skills of pharmacists are strictly regulated. Hence, a key indicator of quality is the availability of pharmacies to consumers. This is particularly relevant in smaller towns and villages, particularly in rural areas, where older consumers (or those affected by chronic pathologies) might have substantial mobility problems.

There are few studies on the Italian market for pharmacies. Calzolari *et al*. (2017) investigate how pharmacies adjust prices as a function of the composition of their consumers. They empirically study the pricing strategies of Italian pharmacies in the market for diapers and some hygiene

---

pharmacies. Norway deregulated in 2001 and Sweden, in 2009, ended the monopoly of the state-owned pharmacy company, which owned all pharmacies.

[43] The market was subject to a modest liberalization in 2006, when para-pharmacies and supermarkets were allowed to enter the market and sell OTC drugs. The new regulations were introduced by decree-law (223/2006) at the end of June 2006 and converted into law by the Italian parliament at the beginning of August (L 248/2006).

[44] Attending physicians are not allowed to distribute drugs.

products, whose prices are not regulated. The empirical estimation strategy exploits the demographic rule that links the number of pharmacies to the population size in each city (this approach will be described in detail below). The results show that more competitive markets tend to have significantly lower prices. Mocetti (2016) provides evidence that the possibility of inheriting the ownership of a pharmacy may affect the career choices of pharmacists' children and intergenerational mobility.

The Pharmacy Market and Public Health

Medications can reduce morbidity and mortality, and adherence to therapies is a primary determinant of treatment success. Poor adherence attenuates optimum clinical benefits and therefore may reduce the overall effectiveness of health systems. Unfortunately, medications are not always used appropriately. Inappropriate medication use can cause serious harm, including death, and increased hospitalizations and health care expenditures (WHO 2003).

The treatment of chronic illnesses and control of progression to end-stage disease commonly involves long-term pharmacotherapy. Although these medications are effective in controlling diseases, their full benefits are often not realized because approximately 50 per cent of patients do not take their medications as prescribed (Brown 2011). Within the health care system, pharmacists are uniquely trained to assist patients and other health care providers to deal with issues involving medication management in the community. Many studies have been conducted over the past 40 years to assess potential health effects of pharmacist-provided direct patient care services. There is substantial evidence that pharmacist involvement in direct patient care can have beneficial effects on patient health outcomes (Chisholm-Burns 2010).

A systematic review published by the Cochrane Collaboration to assess the effects of interventions targeting healthcare consumers to promote safe and effective medicines use identified, among the promising interventions to improve adherence, those involving pharmacists in management of medicines, such as medicines reviews (with positive effects on adherence and use, medicines problems management and clinical outcomes) and pharmaceutical care services (consultation between pharmacist and patient to resolve medicines problems, develop a care plan and provide follow-up; with positive effects on adherence and knowledge) (Ryan 2014).

Because community pharmacists are among the most accessible health care providers, clinical services provided by community pharmacists have the potential to have a major impact on patient health outcomes. Community pharmacists are uniquely positioned to help mitigate the discontinuation of medication and improve adherence for patients who initiate the therapy because of their access to prescription refill information and frequent interactions with patients. However, as evidenced by the review of Blalock *et al*. (2013) in the US, the effectiveness of pharmacist-provided direct patient care services delivered in the community setting is more limited than in other settings. Only 50 of the 134 outcomes (37.3 per cent) examined in the 21 articles that Blalock and colleagues reviewed revealed between-group differences or changes over time that were consistent with beneficial intervention effects.

Different levels of evidence have been produced for specific clinical conditions. Type 2 diabetes is a serious and highly prevalent metabolic disorder that imposes unacceptably high costs on many countries around the world. Regardless of the arsenal of therapeutic options currently available to tackle this disease, metabolic control still remains suboptimal. The management of diabetes requires close collaboration between the patient and a multidisciplinary health care team, in which pharmacists may also take part by providing pharmaceutical care programmes. A systematic review of randomized controlled trials provides evidence that pharmacist interventions can have a positive influence on metabolic control, medication adherence and health-related quality of life for patients with type 2 diabetes (Pousinho, 2016).

Cardiovascular diseases (CVD) are one of the main causes of death and the most common diseases managed by health care systems worldwide. Clinical pharmacist interventions in the care of patients with CVD or in the care of those with high risk of CVD have been evaluated in several studies. Koshman *et al.* (2008) conducted a systematic review to measure the clinical benefit of adding a clinical pharmacist to the care of patients with heart failure in both inpatient and outpatient settings worldwide. This review demonstrated an improvement for many outcomes, including mortality and hospitalization. A subsequent review showed that clinical pharmacy services were associated with better improvement in patients' outcomes compared with the control group in coronary heart disease and heart failure and in the prevention of CVD risk factors, whether as part of primary prevention or secondary prevention (Altowaijri 2013). The RxEACH (Alberta Vascular Risk Reduction Community Pharmacy Project) study – a randomized trial conducted in 56 community pharmacies in Alberta – was the first large randomized trial of CVD risk reduction

through the involvement of community pharmacists, demonstrating a significant reduction in risk for CVD events, and suggesting a potential role of community pharmacists at public health level (Tsuyuki 2015).

The transition of patients from primary to secondary care settings (and vice versa) is historically acknowledged as risky. The process is vulnerable to misunderstanding and miscommunication, often leaving the patient, carers and families ill-prepared to manage their care appropriately during the transition home. Community pharmacists can offer accessibility, expertise in therapeutics, face-to-face contact and skills in drug problems and adherence. A systematic review by Nazar *et al.* (2015), including both randomized and non-randomized controlled trials, analysed the literature on interventions that involved community pharmacy post-discharge. Results suggested that community pharmacists should play a role, particularly in the identification and rectification of drug-related problems.

Seasonal influenza is a major cause of excess winter deaths and increased hospital admissions, with consequent relevant economic burden. Although vaccination targets have been set to tackle this international issue, many countries struggle to reach adequate coverage for their at-risk populations using traditional delivery methods. Pharmacists play an important role in influenza vaccination and are an important alternative to traditional settings such as doctors' offices and health clinics. Community pharmacies are convenient and widely accessible and evaluations consistently demonstrate that patients are satisfied with pharmacist-led vaccinations. As observed in a recent review by Kirkdale *et al.* (2017), allowing community pharmacists to administer influenza vaccinations as an alternative option for delivery helps to increase the coverage rate of vaccination and had a positive impact in driving up immunization rates within some European countries. Similar conclusions were obtained in a large retrospective analysis in the US (Gai 2003).

Beside the positive impact on health outcomes, density of community pharmacies is also associated with an increase in drug supply, mainly in the so-called over-the-counter (OTC) medicines. Some OTC medicines may be abused of, with addiction and harmful effects being increasingly recognized. In a review of the literature by Cooper (2013), OTC medicine abuse was identified in many countries and a range of associated problems and harmful effects were identified in three broad categories. First, there were direct harms related to the pharmacological or psychological effects of the drug abuse or misuse. Although there were variations between countries, five key groups emerged: codeine-based (especially compound analgesic) medicines, cough products (particularly dextromethorphan), sedative antihistamines, decongestants and laxatives. Second, there were physiological harmful effects related to the adverse effects of another active ingredient in a compound formulation (e.g. ibuprofen-related gastric bleeding). Both these types of harm led to concerns about overdoses and excessive use of emergency services. Third, there were other negative consequences, such as progression to abuse of other substances, economic costs and effects on personal and social life (Cooper 2013). To the best of our knowledge, the association between the density of community pharmacies and increased hospitalizations due to medicine abuse has not been systematically explored.

3.3 The Reforms

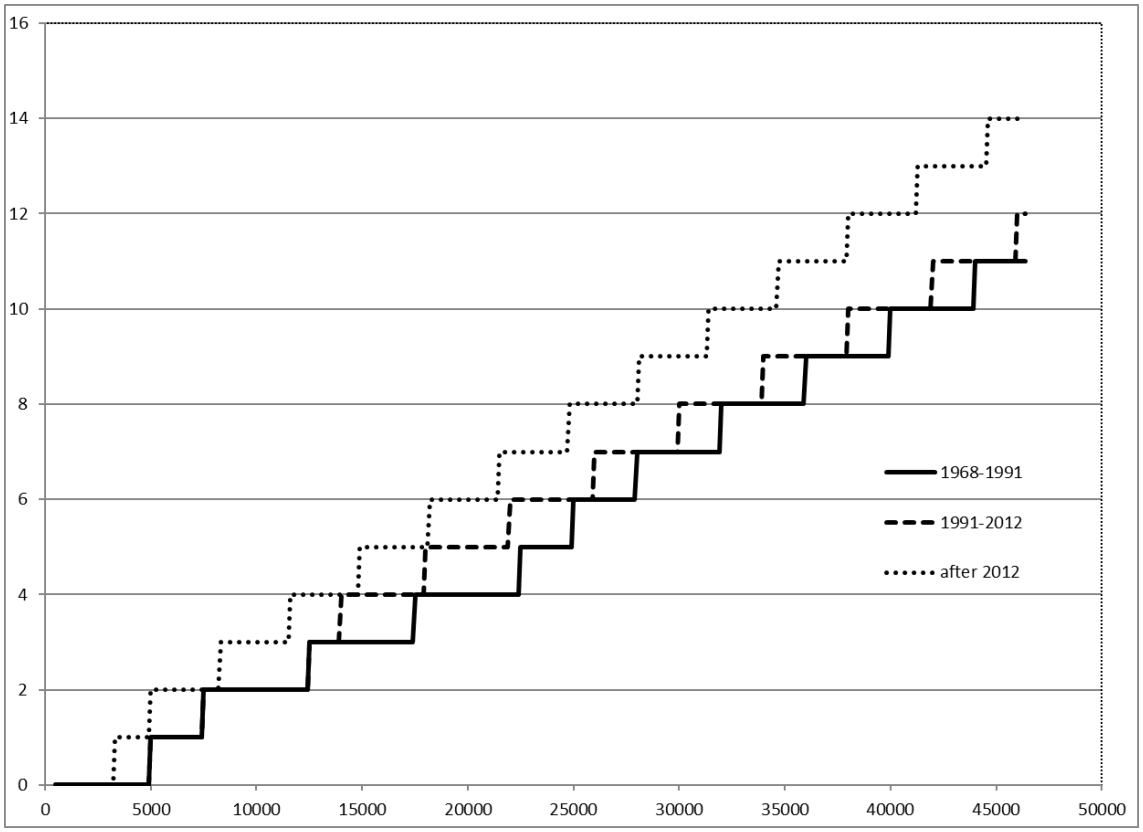
In Italy, the work of pharmacists and the production, trade and distribution of drugs is heavily regulated. Most drugs must be distributed through an authorized pharmacy, the main exception being over-the-counter (OTC) drugs, which are subject to lighter regulation. Pharmacies must employ licensed pharmacists, subject to educational requirements and behavioural regulations. The number of pharmacies and their location is also subject to strict regulation and it is determined through an administrative procedure.

Each of the 20 Italian regional governments determines the number of pharmacies in each municipality and their exact location (minimum distance requirements between pharmacies also apply). The procedure is complex and the process also involves city councils, local health services and the professional associations of pharmacists. The size of the population living in each municipality is the main determinant of the number of pharmacies and the function that maps the local population to the maximum number of pharmacies in each municipality is set by law. Since 1968, three such functions have been used. The first, used between 1968 and 1991, required a maximum of one pharmacy per 5,000 inhabitants in municipalities with a population up to 25,000, and one pharmacy per 4,000 inhabitants in larger towns.[45] Therefore, in small municipalities, (i.e.

[45] Law number 475, 2 April 1968.

with a population below 25,000), the overall number of population is divided by 5,000 to determine the maximum number of pharmacies allowed by law. If after having assigned one pharmacy to each 5,000 inhabitants in a given small town at least 2,500 inhabitants (50 per cent of the population requirement) are left, one more pharmacy is allowed to open. In contrast, for towns with a population larger than 25,000 no additional pharmacy is allowed for any number of inhabitants who remain unassigned after having assigned one pharmacy to each 4,000. This rather complex requirement generates the step function with somewhat irregular steps described in Figure 1.

Figure 1. The number of pharmacies in each municipality according to the demographic rule.



Note: the figure reports the number of pharmacies allowed to operate in each municipality as a function of the local population according to the demographic rule in place in three different periods.

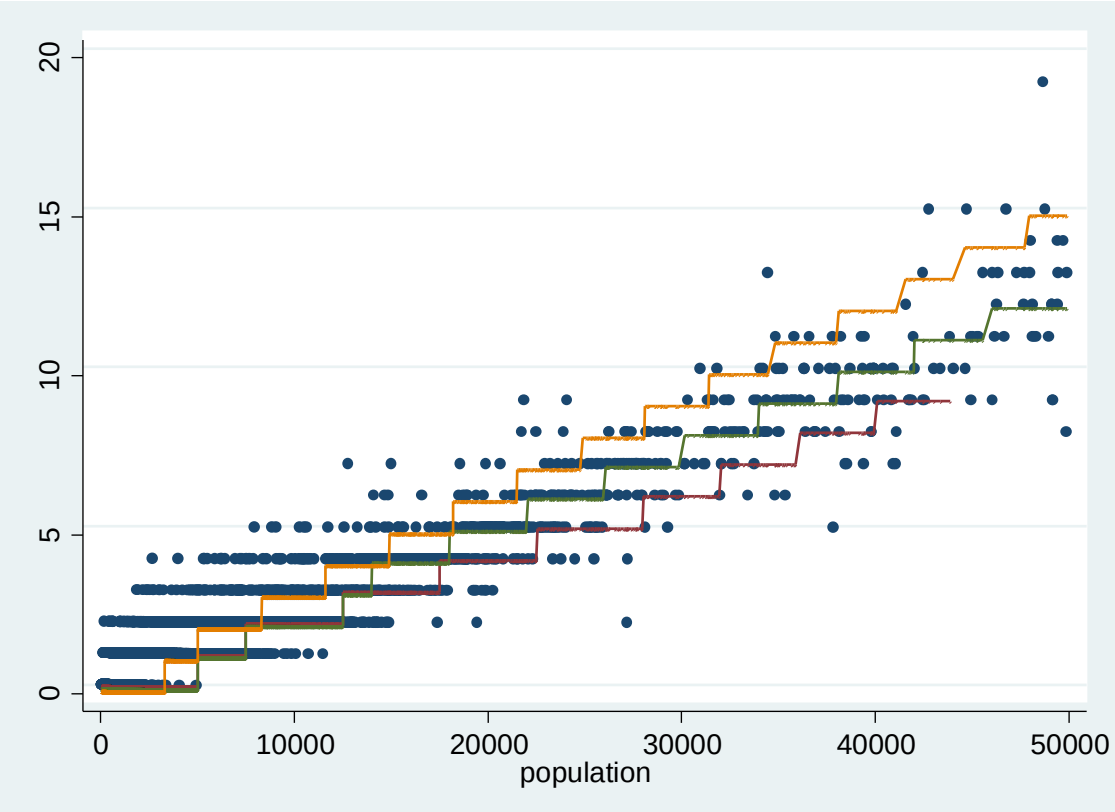
The second function, in place between 1991 and 2012, required one pharmacy per 5,000 inhabitants in municipalities up to 12,500, and one pharmacy per 4,000 in larger towns.[46] The remainder allows for the opening of a new pharmacy only if it is larger than 50 per cent of the population requirement. Broadly speaking, this new function increases the number of pharmacies by decreasing the population threshold (12,500 instead of 25,000) and by not ignoring the remainders in larger towns. Figure 1 compares this new step function with the one in place between 1968 and 1991.

The third function, introduced in 2012, provides for one pharmacy per 3,300 inhabitants.[47] As before, the remainder allows for the opening of a new pharmacy only if it is larger than 50 per cent of the population requirement. This third change generates the third step function described in Figure 1. These new rules further increase the number of pharmacies by increasing the intercept and the slope of the step function.

In principle, the demographic rules described in Figure 1 provide for a simple rule that links local population to the number of pharmacies. However, a number of complications arise in the application of such rules, which make the link between number of pharmacies and population much more complex.

[46] Law number 362, 8 November 1991.
[47] Law number 27, 24 March 2012.

Figure 2. Number of pharmacies and population in Italian municipalities.



Note: the figure shows the number of pharmacies in each Italian municipality and the local population in 2015. The three demographic rules used in the 1968–2016 period are also reported (see Figure 1).

The first issue relates to the municipalities with more pharmacies than the number set by the demographic rule. This may occur as a result of the status quo in 1968 or as a consequence of the shrinking population in some municipalities. In the absence of a clear indication of which pharmacy is supposed to close, and being generally profitable businesses, pharmacies in excess of the demographic rule are often allowed to operate. This allows for municipalities with more pharmacies than the demographic rule would predict.

The second issue relates to small villages, which may not be entitled to a pharmacy according to the demographic rule. In rural areas, the closest pharmacy according to the demographic rule may prove to be too far away to guarantee a minimum availability of medicines. Hence, exceptions to the demographic rule are possible and additional pharmacies can be allowed in small villages (less than 12,500 inhabitants) in rural areas (but subject to minimum distance requirements from other pharmacies).

A third issue arises as a consequence of the complex procedure required to update the list of pharmacies and then to select the pharmacists in charge of each pharmacy (this requires a public competition subject to specific rules). As a result, years of delay are not uncommon. Such an inefficient administrative procedure results in municipalities with fewer pharmacies than the demographic rule would predict. Years after the 2012 reform, only a few regions seem to have completed the procedures to update the list and opened the required new pharmacies.[48] Finally, additional pharmacies may be opened (at most 5 per cent of the total) in locations with large flows of people such as train stations, airports and large shopping centres (but subject to specific minimum distance requirements from other pharmacies).

Origin of the Reforms and Intended Effects

Each change in regulation was motivated by the desire to increase the availability of pharmacies. However, as described above, the impact of the reforms was delayed because of their slow

[48] Michele Bocci, “Il flop delle nuove farmacie”, La Repubblica, 30 May 2016.

implementation. The reforms were implemented by changes in the national law regulating the number of pharmacies. These attempts to increase the availability of pharmacies occurred during a period in which other attempts to liberalise some markets of professional services. These were typically opposed by professional associations and their effects were significantly attenuated with respect to the initial goals. The most significant example is the so-called Bersani reform of 2006, which cancelled price floors (at the time still in place for a number of professions) and lifted the ban on commercial advertising and contingent fees. The market for over-the-counter-drugs was also liberalized, allowing supermarkets to enter into a highly regulated market, in direct competition with pharmacies (see Pagliero 2005 for a description of these reforms and some evidence on their impact).

3.4 Quality Indicators

As discussed above, pharmacists can be considered as primary health care providers and are often asked by patients with chronic disease about their pathologies, medications and other needs. For this reason, pharmacies could represent an important factor contributing to the management of chronic diseases. Based on the organization of the Italian National Health Care system and clinical characteristics, we select four health outcomes of interest: (i) influenza and influenza-associated complications, including pneumonia, (ii) diabetes and its acute complications, (iii) acute coronary syndromes, including acute myocardial infarction, and (iv)upper gastrointestinal bleeding.

Influenza

Influenza is characterized by seasonal outbreaks, at least one during the cold seasons, and its incidence and mortality depend on individual-related factors (e.g. age, comorbidities, immunocompetence), environmental factors (e.g. low temperatures, time course of the outbreak) and medical factors (e.g. antiviral treatment) (Haung 2015). The World Health Organization suggests annual preventive vaccination, especially for high-risk populations (i.e. older people, pregnant women, immunocompromised hosts, residents of nursing homes and other chronic care facilities, patients affected by several chronic cardiovascular and/or respiratory diseases and/or neoplasms, health care providers). In these subgroups of the population, the risk of hospital admissions and death for influenza itself or for some of its complications (e.g. viral pneumonia, bacterial co-infection or respiratory failure) is higher than in the general population.

Before the period of the supposed outbreak, the Italian Ministry of Health promotes the influenza vaccination in order to limit the infection. The Italian Agency for Pharmaceuticals authorizes yearly the vaccine composition to be used for the current season (Gazzetta Ufficiale 2017). The vaccine can be bought in pharmacies with a medical prescription or directly delivered to the frail population by several structures of the National Health Service (Prevention Department Services or general practitioners and paediatricians) free of charge (*www.AIFA.gov.it*). The vaccine bought in the pharmacy by the non-frail population is usually administered by general practitioners.

This explains why influenza was chosen as one of the four outcomes of interest: it is characterized by an acute seasonal outbreak and vaccination, based on work of prevention services, general practitioners/paediatricians and pharmacies, and could have short-term effects on admissions and mortality. It is worth noting that even if influenza and its complications are multifactorial events, individual, environmental and medical determinants may affect the analysis only if they are related to the density of pharmacies. For example, if time trend analyses are conducted within municipalities it is unlikely that changes in individual and environmental factors will overlap with changes in the number of pharmacies to the same extent in all municipalities in Italy.

To identify cases of influenza-related events, we select the following codes of the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9 CM): 480.00, viral pneumonia; 485.00, bronchopneumonia, organism unspecified; 486.00, pneumonia, organism unspecified; 487.1, influenza with other respiratory manifestations; 487.8, influenza with other manifestations; 488.00, influenza due to identified avian influenza virus.

Diabetes

Diabetes is a common chronic syndrome and its prevalence is increasing among all ages. Without correct management, over time diabetes can cause cardiovascular, renal, neural and ocular damage. For this reason, the World Health Organization suggests several preventive strategies but, in case of a confirmed diagnosis, it is essential to obtain a tight control of blood glucose by diet and using pharmacological approaches, with insulin for type 1 and oral medications for type 2 diabetes.

The number and the density of pharmacies can influence both blood glucose level testing and adherence to therapy.

For this condition, we used the following ICD-9 CM codes: 250.1, diabetes with ketoacidosis and 250.2, diabetes with hyperosmolarity. In general, diabetes is not the cause of admission, thus we choose the two most frequent acute complications of diabetes that need urgent management and admission to hospital.

Coronary Syndromes

The coronary syndromes are chronic pathologies with multiple possible acute recurrences. The definition of 'coronary syndrome' is wide and is related to all coronary heart diseases that cause frequent and lethal acute manifestations, such as myocardial infarction and heart failure.

Coronary diseases have a different development than diabetes. Their first manifestation could be more frequently lethal. For this reason, the prevention of these acute events is a cornerstone of the management suggested by the World Health Organization. It is based on diet, physical activity, smoking cessation and blood pressure control. The first three preventive actions are the object of global public health strategies, based on awareness campaigns, but community pharmacies could contribute to the prevention and effective treatment of hypertension.

Many patients are not able to check blood pressure by themselves and automatic devices need frequent calibration. Hence, the possibility to regularly and correctly evaluate blood pressure is a cornerstone of prevention of cardiovascular diseases. Pharmacies could play a role facilitating both the adherence to treatment and blood pressure monitoring. Pharmacies are also needed for the availability of medications for hypercholesterolemia, another risk factor for cardiovascular events. Adherence to lipid-lowering therapies is a well-known problem, and it is about as low as 50 per cent after one year from the first prescription.

For coronary diseases we select the following ICD-9 CM codes: 410.00, acute myocardial infarction and 411.00, other acute and sub-acute forms of ischaemic heart disease.

Upper Gastrointestinal Bleeding

Adherence to therapy is also related to risk of upper gastrointestinal bleeding. Proton pump inhibitors and H2 antagonists are some of the most used drugs worldwide, with a long-term indication. Their importance is related to the prevention of symptoms of gastro-oesophageal diseases and the risk of bleeding due to the use of painkillers. Acetaminophen and non-steroidal anti-inflammatory drugs, the most used painkillers, also without medical indication, may increase the risk of upper gastrointestinal bleeding in the absence of a correct preventive strategy. Access to pharmacies may affect abuse and misuse of non-steroidal anti-inflammatory drugs (thus increasing the risk of upper gastrointestinal bleeding), but, on the other hand, they can also help to increase adherence to therapy with proton pump inhibitors and H2 antagonists.

Cases of upper gastrointestinal bleeding were identified using the following ICD-9 CM codes: 530.2, ulcer of oesophagus; 531.00, acute gastric ulcer with haemorrhage; 532.00, acute duodenal ulcer with haemorrhage; 533.00, acute peptic ulcer of unspecified site with haemorrhage; 534.00, acute gastrojejunal ulcer with haemorrhage; 578.9, haemorrhage of gastrointestinal tract, unspecified.

Other Quality Indicators

We also explor measures of accessibility of pharmacies obtained from a survey of the Italian population conducted by ISTAT. This survey, called 'Aspects of daily life' ('Aspetti della vita quotidiana') involved about 20,000 families to obtain population estimates for several topics on daily life and behaviours. The survey includes a question on perceived difficulty of the household in accessing several services, including pharmacies.[49] Households can report no, some, or many problems/difficulties in accessing services. For the current report the answers 'some' and 'many problems/difficulties' were grouped together.

[49] The Aspects of daily life survey data is accessible through the website of the National Institute of Statistics (ISTAT) at the following address: https://www.istat.it/it/archivio/129959.

We also explor a data set on consumer complaints about the services provided by pharmacies. The data set was compiled and provided by the European Commission, DG Justice. Although this data set is interesting and provides valuable information, it is not well suited for the current analysis for a number of reasons. First, it cannot be matched with municipality, province or regional level data used in this study. Only the total number of complaints at national level is reported. Second, the simple comparison of aggregate data for the periods before and after the reforms is problematic in our setting. This is described in the section on the empirical methodologies used in this study (see 3.6). Third, aggregate complaints data are difficult to use, as the number of complaints depends on the consumer perceived costs and benefits of filing a complaint with the relevant authorities. As such, aggregate data do not allow the control of changes in these variables.

Finally, an overview of customer satisfaction of the pharmacy service in Italy is available from a survey realized by the CFMT (Centro di Formazione Management del Terziario, established by Confcommercio e Manageritalia) from a sample of around 6,000 individuals. The study has the aim of analysing customer satisfaction over several services belonging to different economic sectors in Italy, by means of direct interviews. Satisfaction is described through the following dimensions: accessibility, efficacy, reliability, flexibility, transparency, professional skills, listening, pleasantness and convenience. Being aggregate data at national level, they cannot be matched with municipality, province or regional level data used in this study, and are only used to provide an overview of this aspect of quality.

3.5 Data Sources

The Italian health care system is a regionally based national health service (Servizio Sanitario Nazionale – SSN) that provides universal coverage free of charge at the point of service. Providers are paid for their activities by means of a prospective payment system based on diagnosis-related groups (DRGs). In order to allocate each patient to a specific DRG, the hospital discharge records (HDRs) system has been implemented on a national scale (Ferrè 2014). HDRs are used to collect information about individual patients discharged from hospitals. This is done in observance with the currently applicable Personal Data Protection Code. These records have legal value, and contain general information on the patient, inpatient data and details of implemented therapies/operations, as well as clinical discharge information. HDR collection is mandatory. This includes both ordinary and outpatient admissions (day hospital). HDRs are firstly collected at regional level and then forwarded to the Italian Ministry of Health. In the HDR diagnosis and procedures codes are classified with the International Classification of Disease, 9th revision, Clinical Modification (ICD9-CM).

For the analysis based on the municipality level data (described below), the total number of admissions aggregated by sex, age, class, municipality and calendar year were obtained for the selected groups of causes of hospital admission over the period 2000 to 2015. In relation to the health outcomes of interest described above, we select the following ICD9-CM codes of the main and secondary (up to five) diagnoses:

- Influenza (ICD9-CM codes: 480, 485, 486, 487.1, 487.8);
- Cardiovascular events (ICD9-CM codes: 410–411);
- Diabetes (ICD9-CM codes: 250.10–250.13; 250.20–250.23);
- Upper gastrointestinal tract bleeding (ICD9-CM codes: 530.2, 531–534, 578.9).

For the analyses based on the Italian provinces, data were obtained from the database Health-for-All (HFA). The HFA database, maintained by the World Health Organization, brings together selected indicators on demographics, health status, health determinants and risk factors at national or subnational level.[50] HFA information is obtained from various sources and adaptations are possible at national level. Specifically, the source of data on hospital admissions for selected diseases is based on the flow of HDR data and analysed by ISTAT. ISTAT is also responsible for the collection and coding of mortality data at national level.

[50] The HFA database is accessible through the website of the National Institute of Statistics (ISTAT) at the following address: https://www.istat.it/it/archivio/14562.

Based on availability of data in HFA, we obtained age-standardized, year-, province- and sex-specific mortality and admission rates for:

- Influenza and pneumonia (ICD9-CM codes: 480–487) in individuals aged 65+ years;
- Diabetes (ICD9-CM code: 250) in individuals aged 45+ years;
- Acute myocardial infarction (ICD9-CM code: 410) in individuals aged 65+ years.

The advantage of the HFA data is that they are pre-processed and easy to download for replication purposes. Specifically, data are already aggregated over provinces, include age-standardized rates, and include information on mortality. However, the choice of ICD9-CM codes for the outcome is not flexible and this explains the differences between the health outcome indicators that we used for the analyses at municipality or province level.

HFA mortality data were available up to 2014, while admission data were available up to 2015. To obtain a province-specific measure of mortality and admission rates, we averaged the yearly rates over genders and the five-year periods 2010–2014 for mortality and 2011–2015 for admissions. Analyses were limited to a five-year period because the boundaries and number of Italian provinces have changed over time and a longer time period would have implied a restriction to a smaller number of provinces. In our analysis, we only excluded data for the Sardinia region, which includes a small proportion of the Italian population (less than 3 per cent).

Data from the 'Aspects of daily life' ('Aspetti della vita quotidiana') survey are publicly available for the period 2005 to 2015 and estimates can be obtained at regional level. Data on consumer complaints were provided by the European Commission, DG Justice. Finally, data on customer satisfaction produced by the CFMT were collected through interviews held every two years and results are available for the years 2008, 2010 and 2012 (CFMT 2008; CFMT 2010; CFMT 2012). Interviews had a duration of around 10 minutes and were conducted with the Computer Aided Web Interview (CAWI) method. The sample was representative of the Italian population being stratified by age and geographic area and included service users.

3.6 Empirical Methodology

This section illustrates three different approaches to estimating the impact of the availability of pharmacies on health outcomes of the population. The first is based on the analysis of cross-sectional data on Italian provinces. It is based on the fact that the demographic rule provides for a different number of pharmacies per capita across Italian provinces, as a consequence of the different distribution of the population into small villages and larger towns. This variability in pharmacies per capita across provinces can be used to estimate the effect of interest. The second approach exploits the time series dimension of the data at municipality level and focuses on the impact of the opening of a new pharmacy on health outcomes in the same municipality. In this setting, the opening of new pharmacies is the result of changes in regulation (that typically have a very delayed effect in this market) and changes in the population at city level, which might determine the opening of a new pharmacy according to the demographic rule. The third approach is based on the specific nature of the demographic rule, which provides for sharp steps as illustrated in Figure 1. In this case, the analysis is based on the discontinuities of the demographic rule and it is fundamentally based on cross-sectional variability. Each of the three approaches has its own advantages and disadvantages (discussed in the rest of the chapter) and is complementary to each other.

Analysis Using Province-Level Data

The municipality-specific expected number of pharmacies was estimated on the basis of the 1991 legislation applied to the maximum size of the population of each municipality over the period 1991 to 2009. At least one expected pharmacy was imputed for each municipality irrespective of the population size. The density of observed pharmacies was calculated using the number of pharmacies in each municipality in December 2010. Expected and observed rates of pharmacies were obtained using the populations of 2010 as the denominator.

Table 1. Selected characteristics of the 102 provinces

Variable	min	max	median	mean	std
Median population size of the municipalities	672	19564	3195	4702	4035
Proportion of the population aged at least 19 years holding at least a secondary school degree	33%	54%	40%	40%	3.84%
Proportion of population living in in areas with a low level of urbanization	12%	73%	25%	27%	12.7%
Italian geographical macro-area (1=North, 2= Centre, 3= South)	NA	NA	NA	1.9	0.88

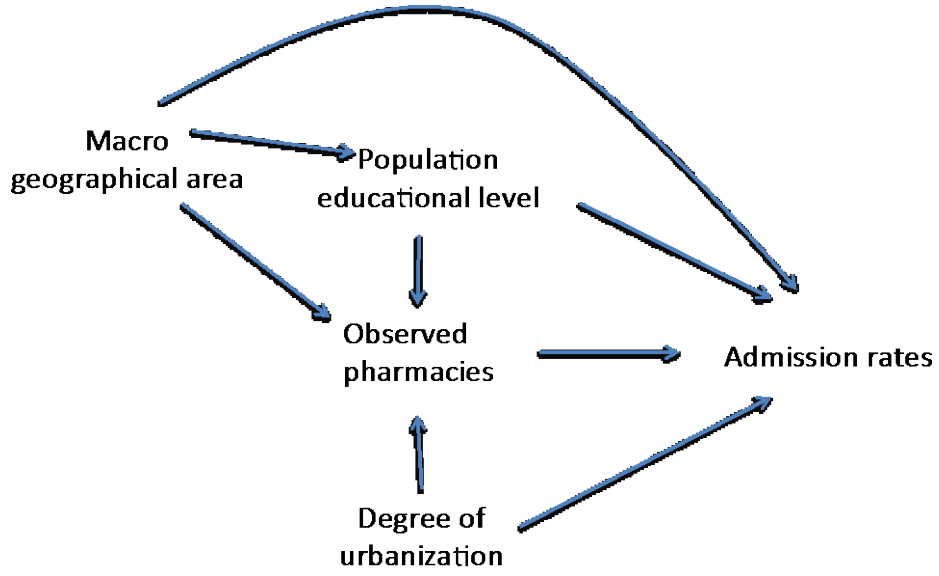
Analyses conducted at province level included 102 observations after the exclusion of provinces of the Sardinia region. Exposure data included the expected and observed densities of pharmacies, while outcome data included mortality and admission rates of influenza/pneumonia, AMI and diabetes. For each province we also obtained data on the median population size of its municipalities, the proportion of population living in areas with a low level of urbanization as measured by the ISTAT at regional level, the Italian macro geographical area (North, Centre, South) and the proportion of the population aged at least 19 years holding at least a secondary school degree as obtained by the 2011 census data. The distribution of these variables over provinces is reported In Table 1.

We first used multivariable linear regression models to estimate the association between the observed density of pharmacies and the selected outcomes, after adjusting for degree of urbanization, population educational level and macro geographical area, as indicated by:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_p x_{ip} + \varepsilon_i \quad (1)$$

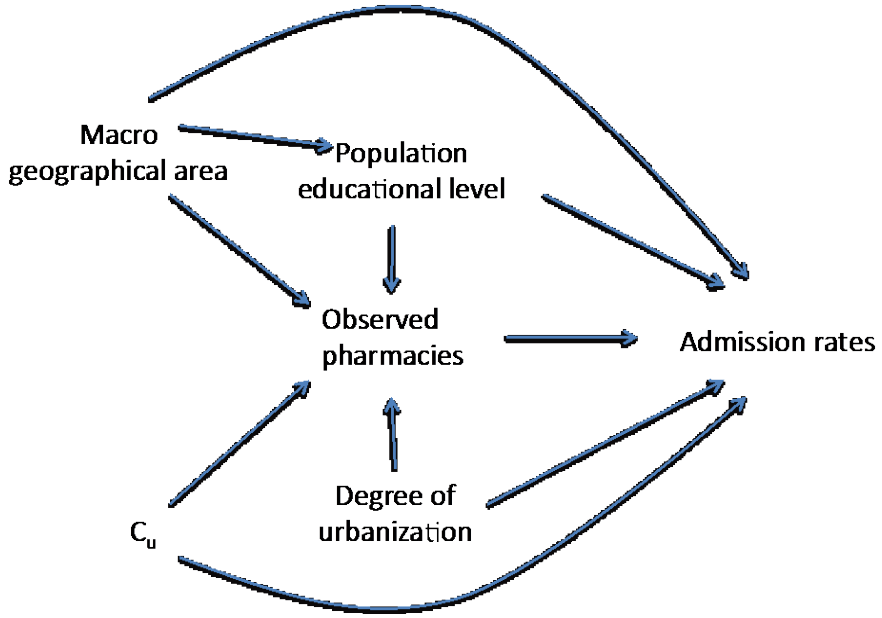
where x_1 is the observed number of pharmacies and x_p is the vector of confounding factors (analyses using Poisson regression provide similar results). The coefficient β_1 has a causal interpretation (effect of an increase of 1 per 10,000 inhabitants in pharmacy density on admission or mortality rates of the outcomes of interest) under the assumption that there is no unmeasured or unknown confounding between the observed density of pharmacies and admission or hospital rates. Using the graphical language of directed acyclic graphs (DAGs), β_1 can be interpreted causally under the assumptions depicted in Figure 3.

Figure 3. Assumed relationship among variables that may lead to an unbiased estimate of the causal effect of the density of pharmacies on the admission rates of the outcomes of interest, using equation (1)



If, as depicted in Figure 4, there is at least one unmeasured or unknown variable that acts as a confounder (i.e. C_u in Figure 4) a causal interpretation of the coefficient β_1 is not possible. The expected density of pharmacies in a given province depends mainly on the distribution of the population size of the municipalities of that province, as the maximum number of pharmacies per inhabitants is regulated at municipality level. If the variable C_u is unrelated to the expected number of pharmacies, the β_1 estimate obtained from the regression model (1) in which x_1 is the expected density of pharmacies instead of the observed density can be interpreted as the causal effect of x_1 on y (Figure 5).

Figure 4. Assumed relationship among variables that hampers a causal interpretation of the estimate of the association between the observed density of pharmacies and admission rates if C_u is unknown or unmeasured



Indeed, under a number of assumptions, the expected number of pharmacies is a so-called instrumental variable (IV) for the effect of the number of pharmacies on the outcome of interest. These assumptions include, for example, the lack of a direct effect of the population size of the municipalities on the admission rates ('exclusion restriction') or the lack of unknown or unmeasured confounders of the effect of population size of municipalities on the outcomes of interest.

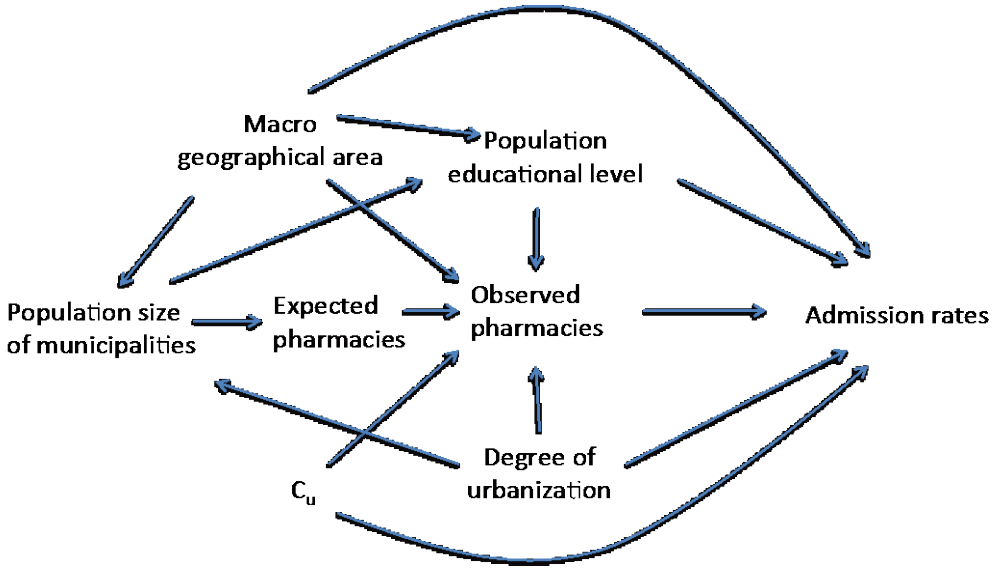
The two-stage least square method (2SLS) is a traditional approach to estimate the IV estimator of the effect of the exposure on the outcome. This approach involves running two models:

$$x_i = \alpha_0 + \alpha_1 z_{i1} + \alpha_p c_{ip} + \varepsilon_i \quad (2)$$

$$y_i = \beta_0 + \beta_{IV} \hat{x}_i + \beta_p c_{ip} + \varepsilon_i \quad (3)$$

where equation (2) estimates the effect of the IV z_1 on the exposure x, after adjusting for the vector of confounders C_p, while model (3) uses the predicted x instead of the observed X to predict the outcome y of interest. In our study, model (2) estimates the effect of the expected density of pharmacies on the observed density, while model (3) estimates the effect of the predicted density of number of pharmacies on the selected outcomes of interest.

Figure 5. Assumed relationship among variables that may lead to a causal interpretation of the estimate of the association between the expected density of pharmacies and admission rates even if C_u is unknown or unmeasured



Before–After Analysis at Municipality Level

Entry regulation in the Italian pharmacy market changed in 1991 and in 2012, when the demographic rule was substantially changed. However, as described above, the application of the reforms was slow and it took years to have a real impact on the market. New pharmacies opened with substantial delay, which varied a lot across regions and municipalities. Hence, we consider variability in the number of pharmacies at municipality level and investigate the correlation between outcome measures and the number of pharmacies in each municipality. The analysis exploits the increase in the number of pharmacies induced by the reforms, but allows for the fact that such reforms had a delayed impact on the number of pharmacies in each municipality.

Consider the following equation describing the relation between health outcome y , the number of pharmacies x and the population p in municipality i and year t :

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 p_{it} + \varepsilon_{it} \qquad (4)$$

OLS estimates of the parameter β_1 provide an estimate of the causal impact of the number of pharmacies under the assumption that $E(\varepsilon|x,p) = 0$, or that there is no correlation between unobserved determinants of health outcomes, ε , and the explanatory variables x . This is a strong assumption as larger cities are likely to have more pharmacies, but they are also likely to have richer and more educated inhabitants. In fact, high wage jobs tend to be concentrated in larger cities rather than in rural areas. Moreover, there might be differences in the availability of health care services between large and small cities, or there might be differences in the way people live in urban and rural areas that affect health outcomes.

While some of these differences are observable, some are likely to be difficult or even impossible to observe and measure. To capture those, we can enrich model (1) by adding a fixed effect τ_i for each municipality i, capturing any unobserved variable that varies across cities but is constant over time. Similarly, one can include in model (1) fixed effects for each year, capturing unobserved variables that vary over time (e.g. a time trend) but not across cities, obtaining the fixed-effects model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 p_{it} + \tau_i + \tau_t + \varepsilon_{it}. \qquad (5)$$

Estimating equation (2) by ordinary least squares exploits the time series variability of the number of pharmacies in each municipality. Informally, it corresponds to estimating the average of hundreds of before–after comparisons that one could perform, studying each case in which a new pharmacy is opened in one city throughout the sample period.

This panel data methodology seems particularly appropriate in estimating the short-run response of health outcomes to changes in the number of pharmacies, since we observe health outcomes for about a decade. In the case of a pharmacy opening in the middle of the sample period, we can observe about five years before and after the event. If the expected lag in response is longer, other estimation methods might be more effective.

Regression Discontinuity Design

The regression discontinuity approach aims at exploiting the arbitrary values taken by a specific step of the demographic rule described in Figure 1. For example, the 7,500 step in the demographic rule remained in place throughout the sample period until 2011. Focusing on the effects of a specific step in the demographic rule might then provide information on the long-term effects of an increase in the availability of pharmacies. Hence, in general, the two methods are complementary and are expected to provide different results.

Consider one specific step in the demographic rule described in Figure 1, for example the 7,500 step, and focus on the municipalities close to the population value that triggers the change in the number of pharmacies. These municipalities are likely to be very similar in all respects, but they do differ because of the step in the demographic rule, which prescribes a different number of pharmacies. On average, municipalities to the right of the threshold will have a higher number of pharmacies. If compliance to the demographic rule is perfect, they will have exactly one more pharmacy. These municipalities may also have different values of the health outcomes that we are considering, as a result of this additional pharmacy. Such differences in health outcomes for municipalities close to the thresholds can then be interpreted as the causal effect of the increase of one unit in the number of pharmacies. This is the main idea behind the so-called sharp regression discontinuity design.[51]

However, in practice, the number of pharmacies varies for many other reasons on top of the demographic rule. Hence, we cannot expect that all the municipalities to the right of the threshold will have exactly one more pharmacy than municipalities to the left of it (see Figure 1). However, municipalities to the right of the threshold will have *on average* a higher number of pharmacies. The so-called fuzzy regression discontinuity design exploits the variability in the average number of pharmacies at the threshold to estimate the impact of the number of pharmacies on health outcomes. This approach leads to an instrumental variable strategy, in which the demographic rule is used as an instrument for the observed number of pharmacies around the threshold.

Consider again the equation describing the relationship between health outcome y, the number of pharmacies x and the population p in equation (1). The regression discontinuity approach exploits the exogenous variability in x (variability in x uncorrelated with ε) generated by the step in the demographic rule. This requires using the predicted number of pharmacies according to the demographic rule, $m(p_{it})$, as an instrument for x_i in a two-stage least squares procedure (Angrist and Lavy 1999). In general, the larger the impact of the function $m(p_i)$ on the average number of pharmacies (the first step in the two-stage procedure), the more precise the estimate of the impact of the number of pharmacies will be. This implies that a very noisy relationship between the demographic rule and the actual number of pharmacies (Figure 2) may provide noisy and inconclusive estimates using the regression discontinuity approach. Whether the correlation

[51] See Angrist and Pischke (2009).

observed in Figure 2 around the step considered is strong enough to provide precise estimates of the parameter of interest is an empirical issue, which we discuss in the next section.

To conclude, the regression discontinuity method provides a different way to estimate the impact of availability of pharmacies on health outcomes. This approach is not based on the systematic comparison of outcomes before and after the opening of a new pharmacy (as in the case of panel data methods), but it relies on the comparison of outcomes for municipalities that are close to a threshold in the demographic rule. Hence, this second approach exploits the specific shape of the demographic rule (a step function), rather than changes in the demographic rule. This difference in methodology also implies that the sample used for the panel data analysis will be different from the sample used for the regression discontinuity analysis.

3.7 Results

The Number of Pharmacies in Italian Municipalities

Our data set includes information on the number, exact location, date of first opening and changes of ownership of pharmacies in each municipality (the data is collected by the Italian government for administrative reasons).[52] The data set includes observations for 7,948 municipalities. We exclude a few municipalities that were created after 2006 or changed name during the same period.[53] The data set provides a cross section (as of January 2017). However, the data set contains enough information to recover the time series of the opening of pharmacies in each municipality since 2006. As the closing of a pharmacy is an extremely rare event, this provides an accurate description of the change in the number of pharmacies throughout the country.

As a result of the demographic rule, the number of pharmacies is linked, although imperfectly, with the local population. This is clearly shown in Figure 2, which reports the number of pharmacies in each Italian municipality and the local population in 2015. Still, there are deviations from the demographic rule, which are particularly frequent in small municipalities (with less than 12,500 inhabitants). For these towns and villages in rural areas, the constraint imposed by the demographic rule is not binding and the number of pharmacies is systematically higher than the values predicted by the demographic rule. However, in larger cities, it is not uncommon to observe fewer pharmacies than the predicted number according to the demographic rule.

Table 2. Summary statistics on the number of pharmacies in Italian municipalities.

Variable	N	Mean	std	p5	p10	p25	p50	p75	p90	p95
Number of pharmacies	7,948	2.22	11.78	0	0	1	1	2	4	6
Inhabitants / number of pharmacies	6,834	2,808	1,697	569	784	1,383	2,585	3,995	5,032	5,845

Note: the table reports the number of observations, the mean, standard deviation (std), and percentiles (5th, 10th, 25th, 75th, and 90th) of the distribution of each variable (2015).

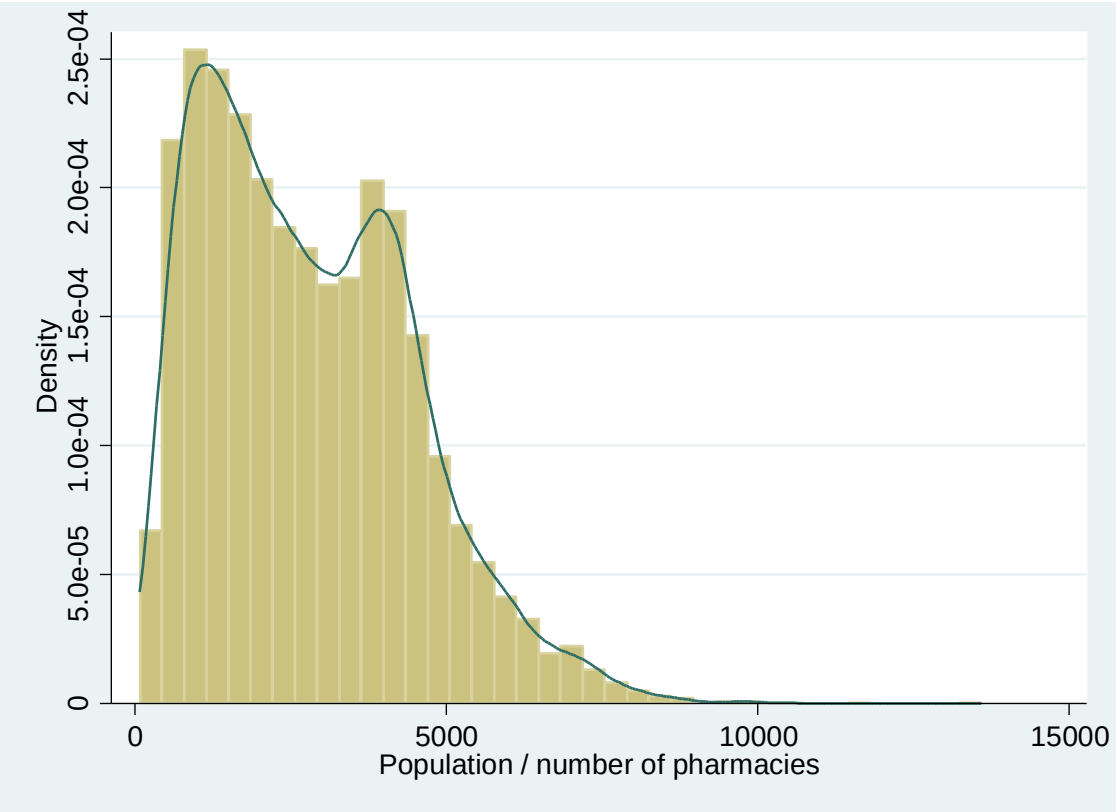
Table 2 provides summary statistics for the same year. On average, municipalities have just 2.2 pharmacies, reflecting the small size of most Italian municipalities. Only 5 per cent of municipalities have six or more pharmacies. About 14 per cent do not have any pharmacy. The number of inhabitants per pharmacy varies across municipalities. The average is 2,808, but there is significant variability. About 50 per cent of municipalities have fewer than 2,585 inhabitants for each

[52] Italian Ministry of Health, http://www.dati.salute.gov.it.
[53] We also exclude temporary points of sale (e.g. open only in summer in touristic areas) or smaller branches of authorized pharmacies (e.g. open in touristic areas and/or in specific periods), which are not subject to the demographic rule.

pharmacy. About 25 per cent have more than 4,000. This illustrates the effects of the delays in implementing the 2012 reform, which set a target of 3,300 inhabitants for each pharmacy. Figure 6 graphically illustrates the heterogeneity across municipalities in the density of pharmacies. The distribution shows that a large proportion of municipalities have too few pharmacies (right tail) or too many pharmacies relative to the demographic rule. While a high density of pharmacies can be explained by the numerous exceptions to the demographic rule described above, a low density of pharmacies in some areas is mainly the result of delays and imperfect implementation.

Table 3 provides additional evidence of these delays. While one would expect a significant increase in the number of new pharmacies in 2012, there is little sign of any systematic change around the reform date. In line with reports in Italian newspapers, only a few regions (Piemonte, Emilia-Romagna, Toscana and Puglia) display significant increases in the number of pharmacies.

Figure 6. Distribution of inhabitants per pharmacy at the municipality level.



Note: The figure displays the distribution of the ratio of the number of inhabitants and the number of pharmacies in each municipality (histogram and kernel density, year 2015).

Table 3. New pharmacies by region and year.

Region	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
Piemonte	0	10	1	1	0	2	20	19	5	49	107
Valle d'Aosta	0	4	0	0	0	0	0	0	1	0	5
Lombardia	3	61	18	16	30	16	14	6	15	9	188
Trentino Alto Adige	0	10	4	3	2	1	1	2	0	1	24
Veneto	5	14	25	5	6	1	7	11	6	2	82
Friuli-Venezia Giulia	1	5	0	0	9	9	2	1	1	1	29
Liguria	0	3	1	0	1	2	2	1	2	0	12
Emilia-Romagna	1	11	7	9	10	3	6	5	4	1	57
Toscana	1	16	4	6	5	3	7	7	2	5	56
Umbria	0	1	1	2	0	1	1	0	0	0	6
Marche	1	3	6	1	2	3	1	2	1	2	22
Lazio	2	25	11	11	9	2	7	5	7	4	83
Abruzzo	0	6	1	1	0	0	2	1	1	0	12
Molise	0	4	1	0	1	0	0	0	0	0	6
Campania	20	51	9	10	13	6	4	3	5	8	129
Puglia	8	28	4	7	4	1	7	3	3	0	65
Basilicata	1	0	0	0	0	1	0	0	0	0	2
Calabria	1	0	0	2	0	1	1	1	0	0	6
Sicilia	0	9	2	1	3	1	28	14	0	4	62
Sardegna	1	2	0	0	19	10	5	4	2	0	43
Total	45	263	95	75	114	63	115	85	55	86	996

Proportion of Families with Difficulties Accessing Pharmacies

The ISTAT data on the 'Aspects of daily life' survey includes about 20,000 families and provides population estimates for several topics on daily life and behaviours. The survey includes a question on perceived difficulty of the family accessing a list of services, including pharmacies.[54] Families may report no, some or a lot of problems/difficulties in accessing services. For the current report the answers 'some' and 'a lot of problems/difficulties' were grouped together.

At national level, the proportion of families who reported at least some difficulties accessing a pharmacy decreased from 23.4 per cent in 2005 to 20.6 per cent in 2015. Region-specific time trends are reported in Figure 7**.** There was a strong heterogeneity among regions, but in most there was a decreasing trend in the prevalence of families reporting difficulties in accessing pharmacies. Table 4 reports the estimated annual changes in prevalence of families reporting difficulties in accessing pharmacies by region. All estimated coefficients were negative; the strongest trend was observed in the Trentino Alto Adige region, with an estimated average decrease of 1.5 percentage points per year.

These trends parallel the trend in the increasing density of pharmacies (see section 3.3). A linear regression analysis of the prevalence of difficulties in accessing pharmacies onto pharmacy density

[54] The Aspects of daily life survey data is accessible through the website of the National Institute of Statistics (ISTAT) at the following address: https://www.istat.it/it/archivio/129959.

estimated a decrease of 11.2 percentage points per an increase in one pharmacy per 10,000 inhabitants (p=0.001), after considering the fixed effect of the regions. These data are, however, not detailed enough to be used to infer the casual link between the two phenomena. For example, the inclusion of calendar time in the model explained almost entirely the association (p-value adjusted for calendar year = 0.84). More broadly, apart from the number of pharmacies and opening times, the accessibility of pharmacies depends on several environmental and individual factors, including for example the transportation service, city planning, age of the population and social welfare services.

Figure 7. Region-specific time trend in proportion of families reporting difficulties in accessing to pharmacies between 2005 and 2015.

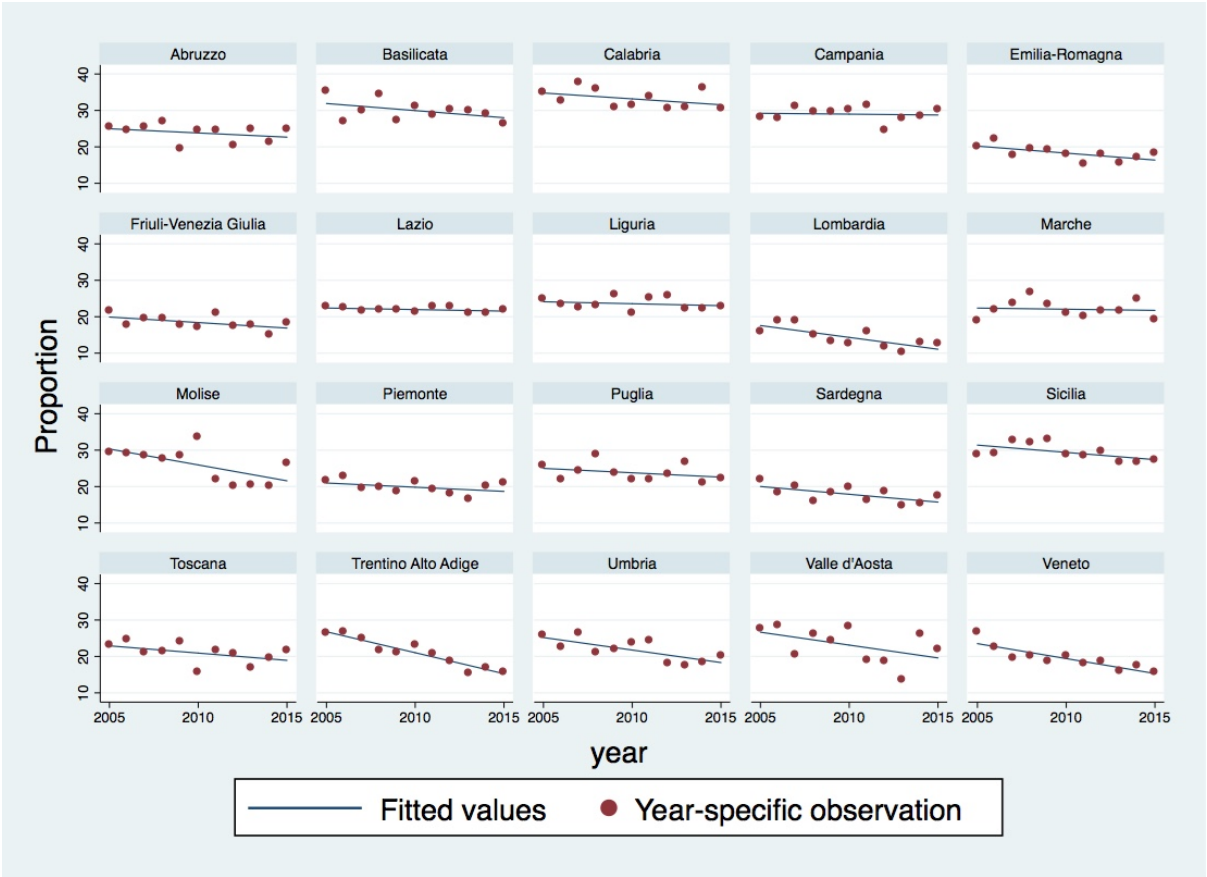


Table 4. Estimated Region-specific annual changes in difficulties in accessing pharmacies. Period: 2005 to 2015.[a]

Region	Annual change	R2	N years	Region	Annual change	R2	N years
Abruzzo	-0.002	0.105	11	Molise	-0.009**	0.400	11
Basilicata	-0.004	0.202	11	Piemonte	-0.002	0.188	11
Calabria	-0.003	0.167	11	Puglia	-0.002	0.110	11
Campania	-0.0005	0.006	11	Sardegna	-0.004**	0.402	11
Emilia-Romagna	-0.004**	0.429	11	Sicilia	-0.004*	0.350	11
Friuli-Venezia Giulia	-0.003*	0.289	11	Toscana	-0.004	0.244	11
Lazio	-0.001	0.174	11	Trentino Alto Adige	-0.015***	0.903	11
Liguria	-0.001	0.005	11	Umbria	-0.007**	0.524	11
Lombardia	-0.007***	0.570	11	Valle d'Aosta	-0.007	0.240	
Marche	-0.001	0.008	11	Veneto	-0.008***	0.760	11

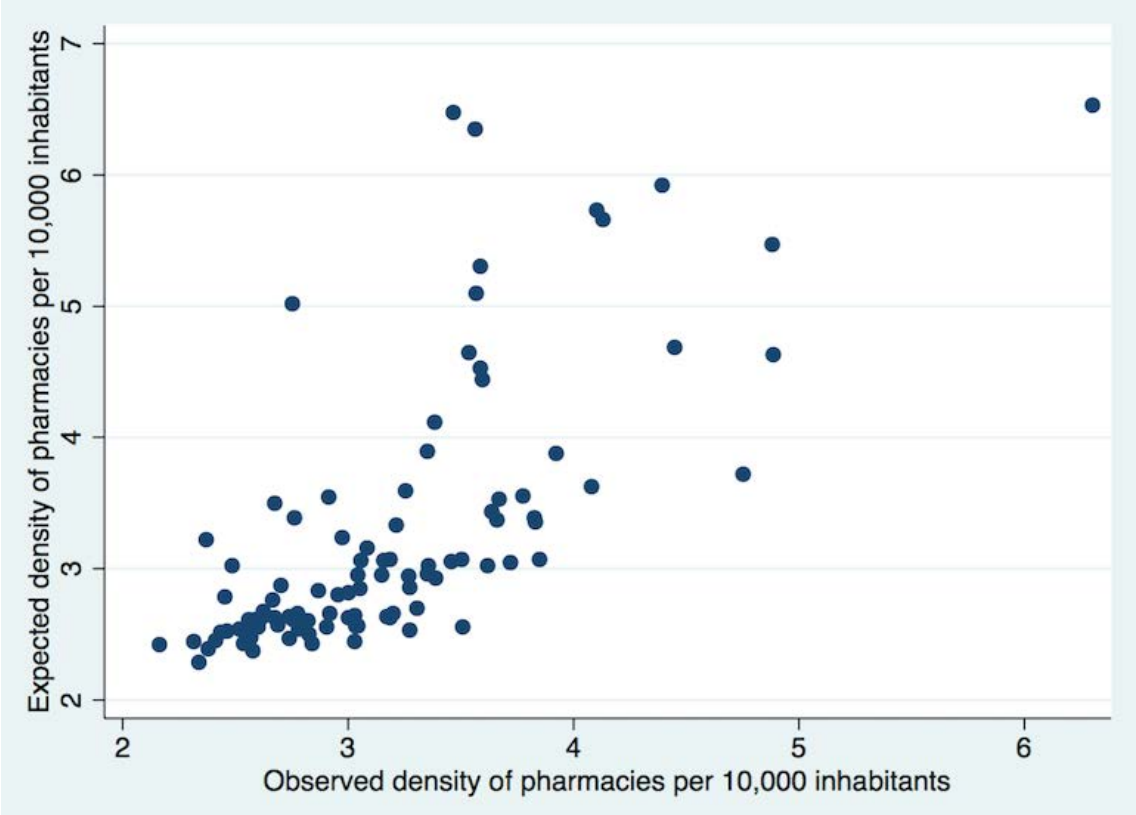
[a] An annual change of -0.002 means an average reduction of 0.2% percentage points each year in the period 2005 to 2015.

*** p<0.01, ** p<0.05, * p<0.1

Results from Data at Province Level

When using the population of 2010 as the denominator, the mean observed density of pharmacies by province was 3.04 per 10,000 inhabitants with a range between 2.16 per 10,000 (Monza) and 6.30 per 10,000 inhabitants (Isernia), while the mean expected density of pharmacies was 2.85 per 10,000 inhabitants with a range between 2.28 per 10,000 (Treviso) and 6.53 per 10,000 inhabitants (Isernia). Thus both densities had a large variation over provinces. Figure 8 shows the correlation between observed and expected densities ($r=0.71$, $R^2=50\%$).

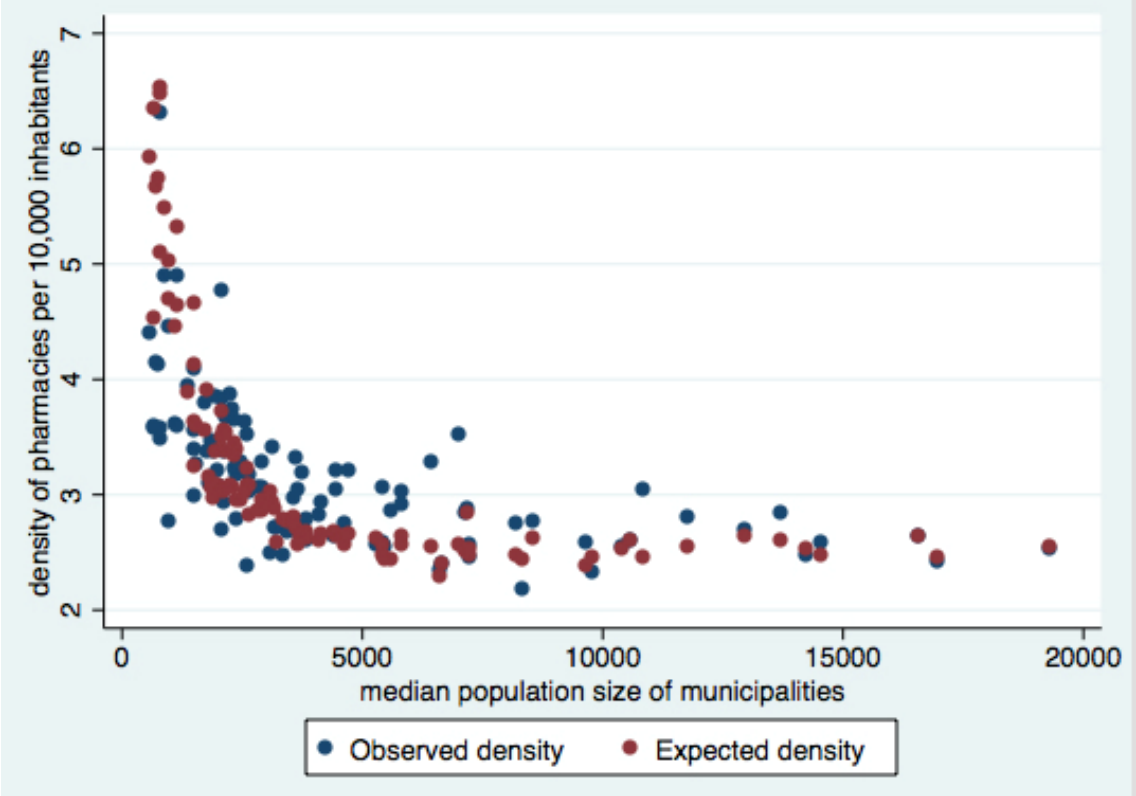
Figure 8. Correlation between the observed and the expected density of pharmacies (per 10,000 inhabitants) by Italian provinces



Both the observed and the expected densities depend on the distribution of the population size of the municipalities within each province. This is depicted in Figure 9, in which observed and expected densities are plotted against the province-specific median population size of municipalities. The median population size explained 31 per cent of the variation in expected pharmacies and 27 per cent of the variation in observed pharmacies.

The age-standardized admission rates per 10,000 inhabitants for the period 2011–2015 ranged between 28.5 (Crotone) and 144.4 (Bolzano) for influenza/pneumonia, between 45.1 (Campobasso) and 120.4 (Fermo) for AMI and between 5.3 (Valle d'Aosta) and 60.5 (Foggia) for diabetes. Maps with province- and gender-specific age-standardized admission rates for the three outcomes of interest are reported in Figure 10.

Figure 9. Province-specific observed and expected densities of pharmacies by median municipality population



While for influenza/pneumonia there was a clear North–South gradient, the rates for the other two outcomes did not follow clear, simple geographical patterns. Table 5 summarizes the results of the regression models carried out to estimate the association between observed/expected densities of pharmacies and age-standardized admission or mortality rates of influenza/pneumonia, AMI and diabetes. Estimates reported in the table should be interpreted as the estimated increase or decrease in admission or mortality rates by a unit increase in expected or observed province density of pharmacies (i.e. an increase of one pharmacy per 10,000 inhabitants). For example, the coefficient of -8.7 in the upper left corner of the table implies an estimated decrease of 8.7 per 10,000 in the age-standardized rate of influenza/pneumonia associated with an increase of 1 per 10,000 inhabitants in the observed number of pharmacies.

The adjusted estimates take into account the possibility that some determinants of the three outcomes of interest (influenza/pneumonia, AMI and diabetes) are also associated with the observed or expected density of pharmacies at province level. For example, provinces located in the North, compared to the rest of the country, have higher rates of influenza/pneumonia and a lower expected density of pharmacies. For both the observed and the expected density of pharmacies, we also conducted a sensitivity analysis restricted to provinces with an overall maximum population size that was always lower than 1,000,000 inhabitants between 1991 and 2009. This analysis aimed at checking that our results were not biased by the contribution of large cities in which the access to pharmacies and their density may be different compared to the rest of the country.

Figure 10. Gender-specific and age-standardized admission rates (per 10,000 inhabitants) for influenza/pneumonia, acute myocardial infarction (AMI) and diabetes, by Italian provinces for the calendar year 2013. Colours indicate quintiles, where the darkest colour corresponds to the highest rate (Source: HFA).

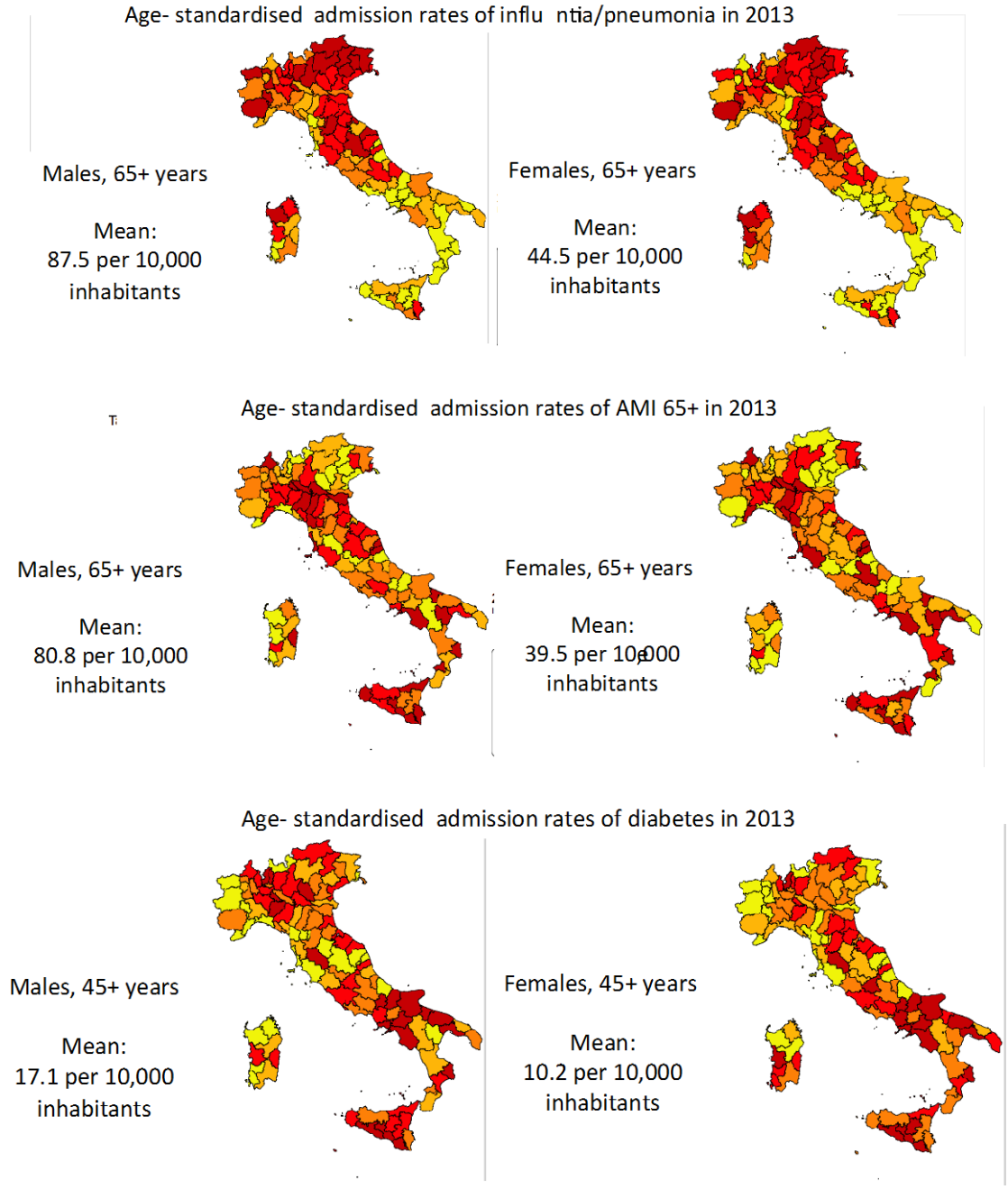


Table 5. Association of observed and expected densities of pharmacies with admission and mortality rates of influenza/pneumonia, acute myocardial infarction and diabetes.^

Density of pharmacies	Influenza/ pneumonia	Acute myocardial infarction	Diabetes	Density of pharmacies	Influenza/ pneumonia	Acute myocardial infarction
	Admission rates	Mortality rates	Admission rates	Mortality rates	Admission rates	Mortality rates
Observed density of pharmacies						
Crude	-8.66***	-0.68*	-0.52	1.24	-0.12	0.14
	(2.96)	(0.38)	(1.95)	(0.98)	(0.99)	(0.77)
	R^2=0.069	R^2=0.031	R^2=0.001	R^2=0.016	R^2<0.001	R^2<0.001
Adjusted coefficient[a]	-8.47***	-0.33	-0.20	1.06	-0.45	-0.33
	(2.53)	(0.24)	(2.08)	(1.01)	(0.99)	(0.39)
	R^2=0.443	R^2=0.669	R^2=0.057	R^2=0.138	R^2=0.165	R^2=0.787
Adjusted and restricted coefficient	-8.52***	-0.29	-0.38	1.11	-0.16	-0.23
	(2.62)	(0.26)	(2.25)	(0.93)	(1.01)	(0.40)
	R^2=0.440	R^2=0.643	R^2=0.052	R^2=0.148	R^2=0.148	R^2=0.764
Expected density of pharmacies on the basis of 1991 legislation						
Crude	0.72	0.42*	-1.92	-0.84	-0.62	-1.08
	(2.02)	(0.25)	(1.26)	(0.64)	(0.65)	(0.49)
	R^2=0.001	R^2=0.028	R^2=0.002	R^2=0.017	R^2=0.009	R^2=0.046
Adjusted coefficient[a]	-2.89	0.02	-1.63	-0.15	-0.31	-0.24
	(1.76)	(0.16)	(1.38)	(0.68)	(0.67)	(0.26)
	R^2=0.395	R^2=0.663	R^2=0.070	R^2=0.129	R^2=0.165	R^2=0.787
Adjusted and restricted coefficient[a]	-2.72	0.02	-1.73	-0.24	-0.33	-0.29
	(1.80)	(0.17)	(1.47)	(0.62)	(-0.66)	(0.26)
	R^2=0.388	R^2=0.638	R^2=0.070	R^2=0.136	R^2=0.150	R^2=0.766

^Based on 102 provinces; 92 provinces in the restricted analyses

[a]Adjusted for proportion of population living in areas with low urbanization, macro geographical area, proportion of the population with high educational level; restricted to provinces with less than 1,000,000 inhabitants (thus exclusion of Roma, Milano, Napoli, Torino, Palermo, Bari, Brescia, Catania, Salerno, Bergamo, Genova, Firenze, Bologna) *** p<0.01, ** p<0.05, * p<0.1

Overall we found no evidence of association between the observed or expected density of pharmacies and admission rates for AMI and diabetes. There was an inverse association between the observed density of pharmacies and admission rates of influenza/pneumonia, although this association was attenuated in the analysis on the expected density of pharmacies. Neither for the observed nor for the expected density of pharmacies did we find evidence of association with mortality for influenza/pneumonia, AMI or diabetes.

Results from the analyses restricted to provinces with a population size lower than 1,000,000 were consistent with the results carried out on all provinces, with no evidence of bias introduced by large cities.

Table 6 summarizes the results on the instrumental variable (IV) analysis. As expected, they are consistent with the results on the expected density of pharmacies reported in Table 5. Overall, the IV analysis magnifies the estimate of the association between the expected density of pharmacies and the outcomes by a quantity that is determined by the strength of the association between the

expected and the observed number of pharmacies. Thus the coefficients reported in Table 6 are larger than those reported in Table 5 for the expected density. In Table 6, there was no association between density of pharmacies and the investigated health indicators, with a possible exception of a positive association between density of pharmacies and admission rates of influenza/pneumonia that had a p-value between 0.05 and 0.10.

The estimates reported in Table 6 rely on a number of (untestable) assumptions, which might well be violated in our analysis. In particular, a casual interpretation is not possible if the distribution of the population within provinces (which is the main determinant of the expected density of pharmacy) is associated, either directly or through shared determinants, with the outcomes of interest, even after adjusting for degree of urbanization, educational level and macro geographical area. Conversely variables that are determinants of the outcomes of interest but are not expected to be associated with the distribution of the population within provinces (e.g. temporal variations in low temperatures in winter for influenza/pneumonia or temporal changes in diet for AMI or diabetes) are not expected to affect the interpretation of our estimates.

Table 6. Results of the analysis using the expected density of pharmacies based on the 1991 legislation as an instrumental variable for the observed number of pharmacies (see methods above). Analysis based on provinces.^

Region	Influenza/pneumonia		Acute myocardial infarction		Diabetes	
	Admission rates	Mortality rates	Admission rates	Mortality rates	Admission rates	Mortality rates
Adjusted coefficient[a]	-5.67*	0.03	-3.19	-0.30	-0.61	-0.47
	(3.34)	(0.32)	(2.76)	(1.33)	(1.31)	(0.51)
	R^2=0.436	R^2=0.661	R^2=0.036	R^2=0.122	R^2=0.165	R^2=0.786
Adjusted and restricted coefficient[a]	-5.34	0.04	-3.39	-0.47	-0.66	-0.56
	(3.41)	(0.34)	(2.93)	(1.22)	(1.30)	(0.51)
	R^2=0.431	R^2=0.637	R^2=0.032	R^2=0.120	R^2=0.145	R^2=0.762

^Based on 102 provinces; 92 provinces in the restricted analyses

[a]Adjusted for proportion of population living in areas with low urbanization, macro geographical area, proportion of the population with high educational level; restricted to provinces with less than 1,000,000 inhabitants (thus exclusion of Roma, Milano, Napoli, Torino, Palermo, Bari, Brescia, Catania, Salerno, Bergamo, Genova, Firenze, Bologna). *** p<0.01, ** p<0.05, * p<0.1

Results from Panel Data and Regression Discontinuity Design

Panel data models are described in section 3.6.2. This approach is based on the systematic comparison of outcomes before and after the opening of new pharmacies at municipality level, after accounting for the potential impact of changes in population, time-invariant city characteristics and year-specific effects. We merge data on the number and location of each pharmacy in each municipality with data from hospital discharges.[55] We exclude municipalities with missing values in health outcomes or population. We then obtain a balanced panel of 7,948 municipalities for 10 years. Table 7 provides summary statistics for this specific sample. The variables bleeding, coronary diseases, diabetes and influenza are the number of cases divided by the population in each city. The number of pharmacies is also divided by the population in each city.

Table 7. Summary statistics for health outcomes and number of pharmacies.

Variable	Obs	Mean	s.d.	Min	Max
bleeding/population	79480	0.000497	0.00079	0	0.022222
coronary diseases/population	79480	0.003167	0.00255	0	0.066667
diabetes/population	79480	6.47E-05	0.000285	0	0.018349
influenza/population	79480	0.002587	0.002322	0	0.054054
Number of pharmacies/population	79480	0.000491	0.000577	0	0.011765
population	79480	7407.929	40788.54	31	2872021

Note: Each observation corresponds to one municipality in one year. The number of hospital admissions and the number of pharmacies is divided by the population in each city in 2011.

Table 8 reports the estimated coefficients, representing the impact of the number of pharmacies, population and year fixed effects on the four health outcomes. The results in Table 8 suggest a non-significant effect of the number of pharmacies on the number of hospital admissions. The point estimate for the impact of number of pharmacies on influenza-related admissions is negative, but the standard error is very large. Also the other estimated coefficients are very noisy.

These non-significant results are likely to be driven by two factors. First, relative to the overall number of municipalities in the country, only a small proportion of cities experienced a change in the number of pharmacies. Hence, the panel data results are based on relatively few observations, while there is a large amount of heterogeneity across municipalities. Second, panel data models can only capture short-run effects, given that municipalities are observed for 10 years and many changes in the number of pharmacies occur in the second half of the sample period (see Table 2). Hence, only very large short-run effects could be detected by panel data models.

55 The number of pharmacies is measured in the beginning of the year.

Table 8. Results from panel data models with fixed effects.

	(1)	(2)	(3)	(4)
Variables	Bleeding	Coronary diseases	Diabetes	Influenza
Number of pharmacies / population	0.0293	-0.175	0.00350	-0.00157
	(0.0479)	(0.113)	(0.0107)	(0.123)
Population	1.61e-09***	3.47e-09***	3.80e-10***	1.73e-09***
	(5.36e-10)	(1.19e-09)	(1.40e-10)	(5.80e-10)
City f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
Observations	79,480	79,480	79,480	79,480
R-squared	0.016	0.010	0.004	0.011
Number of cities	7,948	7,948	7,948	7,948

Note: The table reports the coefficients of a linear regression model (OLS). Each observation corresponds to one city in one city in one year. The number of hospital admissions and the number of pharmacies are divided by the population in each city in 2011. City and year-specific fixed effects are included in each column. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The regression discontinuity approach provides a different way to explore the potential impact of the availability of pharmacies on the health outcomes. First, we focus on one specific step in the demographic rule (at 7,500 inhabitants) for the 2005–2011 period. This choice is motivated by the fact that the step has not changed since the introduction of this type of regulation in the 1960s, which allows us to explore the long-run effects of the regulation. Moreover, most Italian municipalities are small, which implies that there are many observations around the threshold, which is a necessary condition for using this method. We select 1,214 municipalities with a population between 5,000 and 10,000, creating a symmetric window (+/- 2,500) around the step.[56] This is a relatively homogeneous sample of small towns and villages. On average, municipalities in this sample have 1.7 pharmacies. Hence, the opening of one pharmacy can potentially make a difference for consumers.

Table 9 reports the average number of admissions and the average number of pharmacies in the sample. Figure 11 reports the average number of pharmacies for municipalities of different sizes. Each dot in the figure corresponds to the average number of pharmacies in a specific interval of 100 between 5,000 and 10,000 inhabitants. The blue lines are the fitted values of a linear regression on each side of the 7,500 threshold (vertical red line), which corresponds to an increase from one to two pharmacies according to the demographic rule. The discontinuity at the threshold in the blue line corresponds to a 0.28 increase in the number of pharmacies, which is statistically significant at the 5 per cent confidence level. This is in line with the fact that regulation induces an increase in the number of pharmacies at the threshold, although many other variables may also influence the observed number of pharmacies.

[56] In the regression discontinuity analysis we use the maximum historical population method proposed by Calzolari et al. (2013). In practice, the discontinuity is based on the maximum population in the 1971–2001 period instead of current population to account for the fact that the number of pharmacies does not adjust downwards when the population shrinks. See Calzolari et al. (2013) for a detailed discussion of this method.

Table 9. Summary statistics for health outcomes and number of pharmacies (regression discontinuity design).

Variable	Obs	Mean	S.D	Min	Max
Bleeding	8,498	3.52	2.46	0	26
Coronary diseases	8,498	21.28	10.02	0	80
Diabetes	8,498	0.49	1.16	0	45
Influenza	8,498	15.29	9.43	0	77
Number of pharmacies	8498	1.78795	0.679903	0	5
Number of pharmacies according to the demographic rule	8498	1.350906	0.477282	1	2

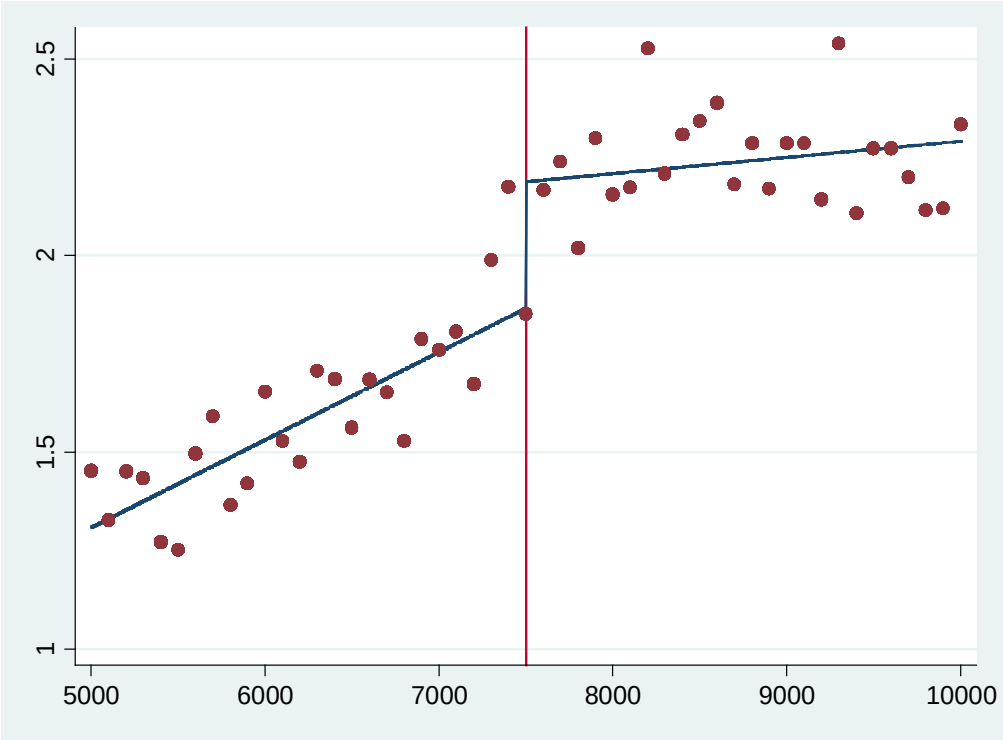
Note: Each observation corresponds to one city in one municipality in one year in the 2005-2011 period.

Figure 12 shows the corresponding results for the average number of influenza admissions. In this case, there is a negative jump at the threshold of 2.1 cases (statistically significant at 5 per cent confidence level). Hence, Figures 11 and 12 suggest that the increase in the number of pharmacies caused by the threshold is associated with a significant decrease in the number of influenza admissions. The magnitude of the effect is sizeable. An increase in the average number of pharmacies of 0.28 (as in Figure 11) corresponds to a drop of about 2.1 in the average number of hospital admissions. Hence, a 16 per cent increase in the number of pharmacies corresponds to a 13 per cent reduction in the average number of influenza-related admissions.

Table 10 combines the results described in Figures 11 and 12. It reports the results of IV regressions (2SLS) of the number of hospital admissions on the number of pharmacies in each municipality, controlling for region-specific fixed effects. The instrument is the predicted number of pharmacies according to the demographic rule. Hence, it is equal to one before the 7,500 threshold and two after. The first stage regression coefficients are highly significant, reflecting the significant jump at the threshold described in Figure 11. The estimated coefficients are not significantly different from zero when the dependent variable is the number of admissions for bleeding, coronary diseases and diabetes. However, the estimated impact of the number of pharmacies on the number of admissions for influenza is positive and statistically different from zero. The magnitude of the coefficient corresponds to an impact of the number of pharmacies on the number of influenza admissions similar to that computed from the results in Figures 11 and 12.

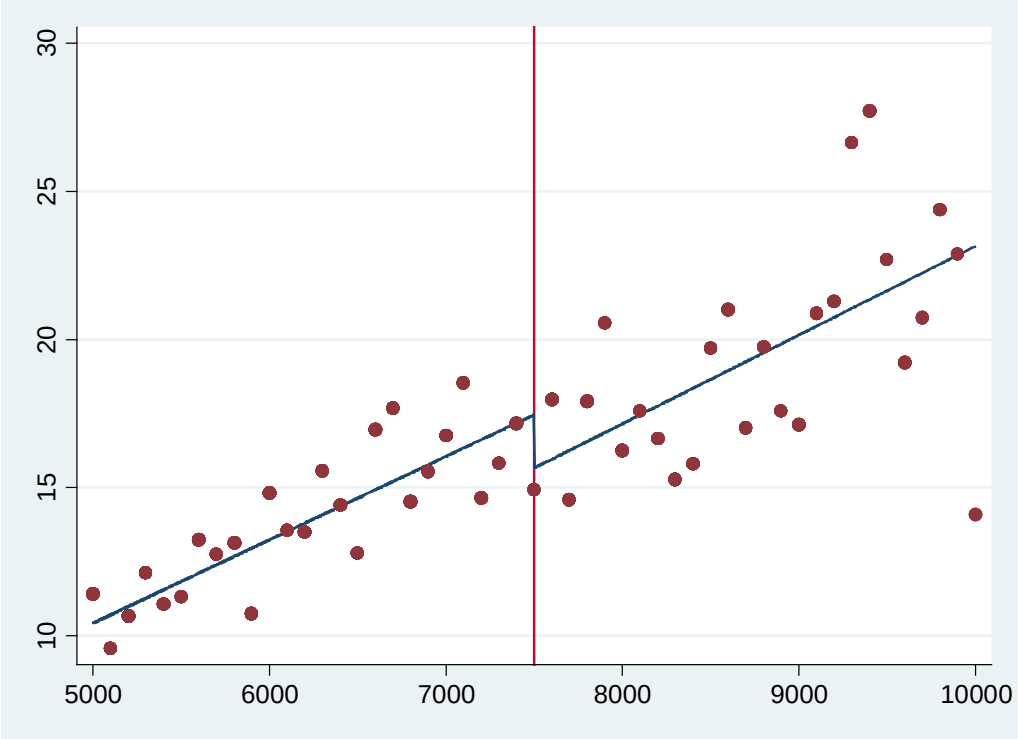
Overall, the panel data and the regression discontinuity results suggest the existence of a negative long-run effect of the availability of pharmacies on the number of hospital admissions related to influenza. The short-run effects might be too small to be detected with the available data. The regression discontinuity design focuses instead on a subsample of small municipalities in which about half are above and below the threshold. Moreover, the threshold remained in place for a long time and provides for a relatively large increase in the availability of pharmacies. Hence, in this sample, we can detect a significant impact of the availability of pharmacies on one health outcome.

Figure 11. Average number of pharmacies and historical population.



Note: the figure reports the average number of pharmacies in municipalities of different sizes according to historical population (intervals of 100 inhabitants). The blue lines are the fitted values of a linear regression of number of pharmacies on population on each side of the 7,500 threshold, which corresponds to an increase from one to two pharmacies according to the demographic rule in place until 2011.

Figure 12. Average number of influenza admissions and historical population.



Note: the figure reports the average number of influenza admissions in municipalities of different sizes according to historical population (intervals of 100 inhabitants). The blue lines are the fitted values of a linear regression of number of admissions on population on each side of the 7,500 thresholds, which corresponds to an increase from one to two pharmacies according to the demographic rule in place until 2011.

Table 10. IV regression discontinuity results.

	(1)	(2)	(3)	(4)
Variables	Bleeding	Coronary diseases	Diabetes	Influenza
Number of pharmacies	-0.578	-0.971	-0.309	-8.823**
	(0.685)	(3.230)	(0.251)	(4.203)
Region f.e.	✓	✓	✓	✓
Observations	8,498	8,498	8,498	8,498
R-squared	0.073	0.274	0.024	0.005
Number of cities	1,214	1,214	1,214	1,214

Note: The table reports the coefficients of a linear IV regression model (2SLS), where the instrument is the predicted number of pharmacies according to the demographic rule (equal to one before the 7,500 threshold and equal to two after). The sample includes municipalities with population between 5,000 and 10,000. Standard errors (clustered by municipality) are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

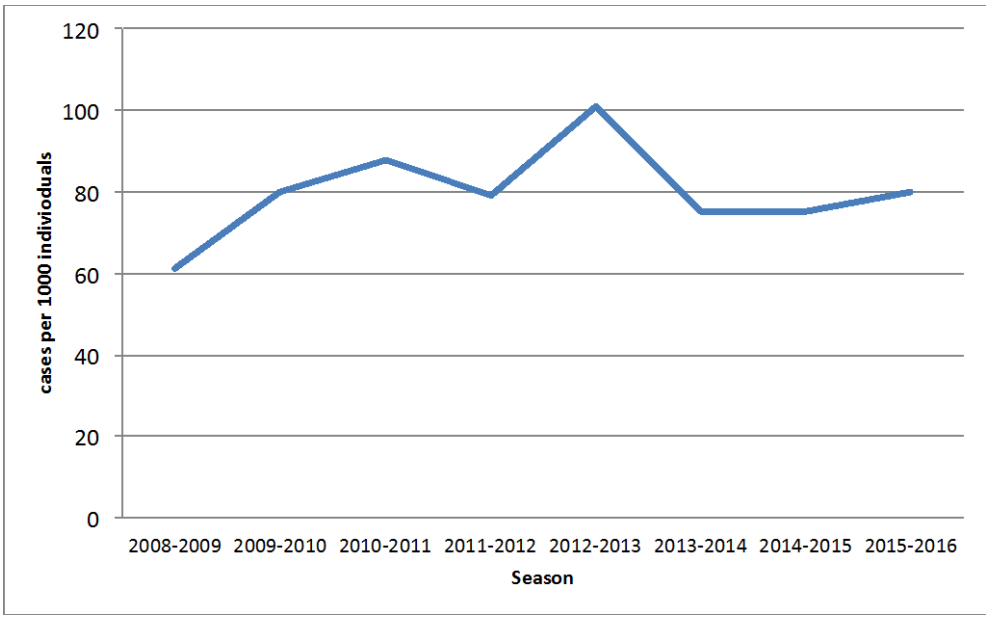
The Role of other Determinants of Influenza

Influenza outbreaks are determined by several factors, including the seasonal viral type and its intrinsic virulence, the proportion of susceptible individuals in the population, the geographical location (see for example the strong North–South gradient in Italy depicted in Figure 10) and individual characteristics (e.g. age, immunocompetence and comorbidities). We discuss here whether these determinants could bias the results reported in sections 3.7.3 (analysis at province level) and 3.7.4 (analysis at the level of municipalities).

The analysis based on provinces drew on geographical variations in density of pharmacies and incidence of influenza/pneumonia, rather than on temporal changes. Even if we adjusted for the three Italian macro geographical areas, we cannot exclude a residual effect of geographical variations in influenza outbreaks. This is one of the reasons why we concluded that a causal interpretation of the analyses based on provinces should be reached with much caution (see the last paragraph of section 3.7.3).

The before–after analysis exploited the temporal variation in density of pharmacies over municipalities in Italy. This analysis can thus be affected by temporal variations in influenza outbreak if, in most municipalities, temporal variations in influenza parallel temporal changes in the number of pharmacies. However, results including year fixed effects control for year to year variability in the intensity of the influenza outbreak. Figure 13 reports the trend in incidence of influenza in Italy from 2008 to 2016 for all age groups. Data are obtained from the surveillance system of the Istituto Superiore di Sanità (Italian National Institute of Health). Incidence peaked in 2011–2012 and the shape of the trend does not match with the time trends in number of pharmacies reported in section 3.7.1.

Figure 13. Incidence of influenza in Italy from 2008 to 2016.

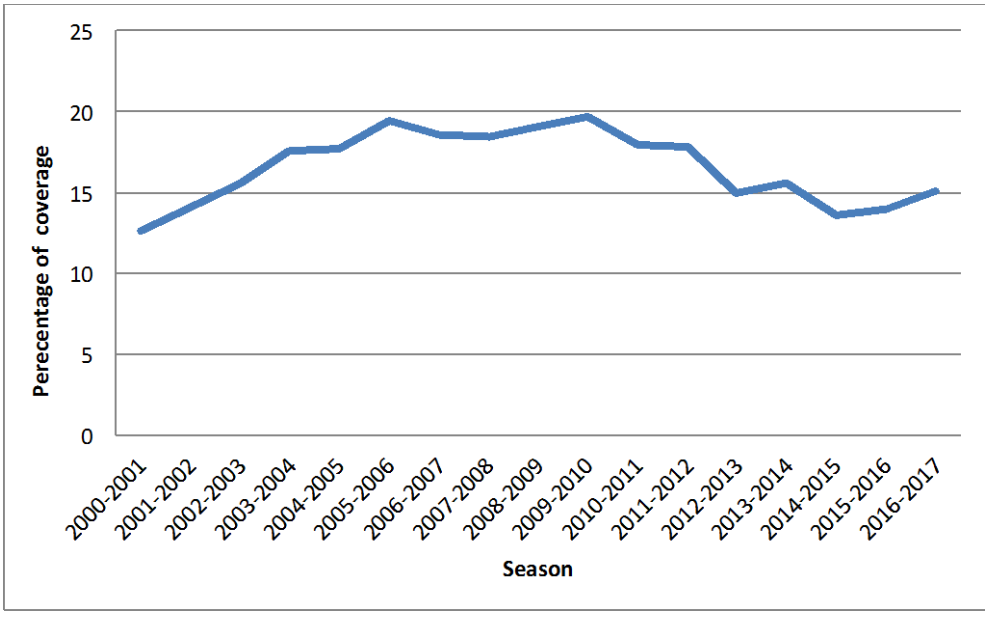


Source: Istituto Superiore di Sanità http://www.epicentro.iss.it.

The regression discontinuity analysis compares municipalities that are likely to be very similar in all respects, but have small differences in population size that, through the demographic rule, prescribe a different number of pharmacies. Thus, this analysis is unlikely to be affected by either geographical or temporal trends in influenza outbreaks. The consistency in the results of the different approaches also indicates that geographical and temporal trends in influenza outbreaks are unlikely to explain our findings.

Population coverage of the influenza vaccine is a key determinant of the magnitude of influenza outbreaks as well as the hospitalization rates for influenza complications. Indeed, pharmacies may affect influenza mainly through promoting and favouring vaccination in the population and in high-risk groups. Thus vaccination coverage is a mediator in all our analyses and should not be controlled for statistically. For the sake of completeness, we report in Figure 14 the time trends in influenza vaccination rates in Italy between 2000 and 2016. The source of data is the Italian Ministry of Health. Unfortunately, coverage is affected by several determinants. Hence, this data does not allow to infer the role of pharmacies in shaping the reported trends.

Figure 14. Influenza vaccination coverage in Italy between 2000 and 2016.

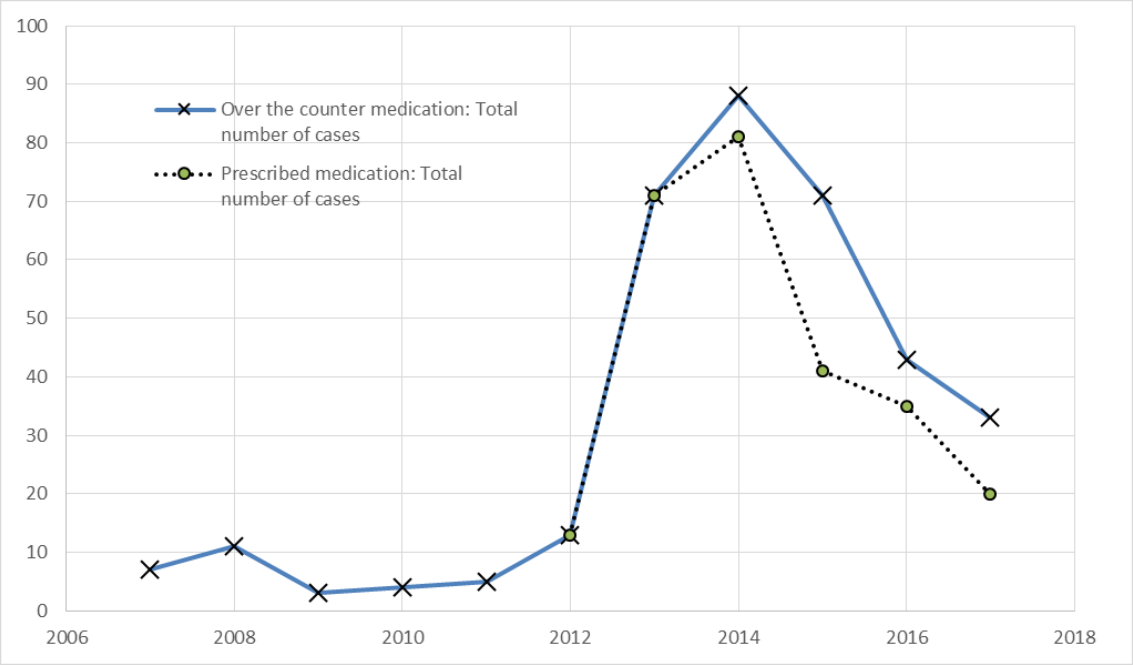


Source: Italian Ministry of Health,
http://www.salute.gov.it/portale/documentazione/p6_2_8_3_1.jsp?lingua=italiano&id=19

Impact on Number of Complaints, Employment and Prices

As discussed in section 3.4, the data on the number of complaints is collected only at national level and cannot be matched with either province or municipality data. The data provides the number of consumer complaints regarding over-the-counter medication and prescribed medication. Complaints refer to the delivery of the good and services, the quality of the good and services, unfair commercial practices and other issues. There is an increase in the number of complaints, which peak in 2013 and then decline (Figure 15). However, this can be due to changes in the perceived cost of complaining and measurement error, which are very difficult to assess with the available data. Indeed, the number of complaints is rather small, given the size of the market, and these changes cannot be interpreted as an indication of real changes in the average quality of the services provided.

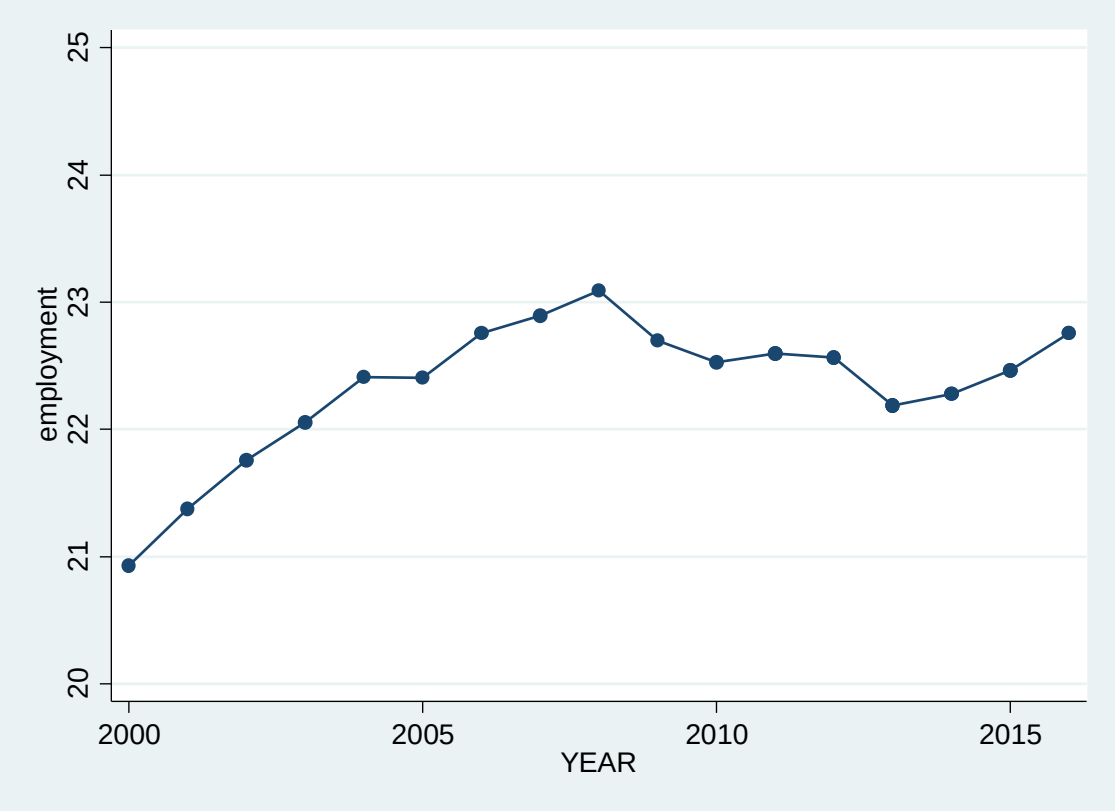
Figure 15. Total number of complaints by year.



The increase in the number of pharmacies could have an impact on employment in the pharmacy sector. Since new pharmacists are required to staff each new pharmacy, we expect an increase in pharmacies to lead to an increase in employment. Still, other effects might occur, as pharmacies compete more intensely at local level. This implies that employment might increase more or less proportionally to the increase in the number of pharmacies.

Given the data used in this study, we cannot estimate the impact of an increase in the number of pharmacies on employment. We are not aware of any data set that could be matched with our current data to address this issue. The available data from the EULF provides some estimates at national level for employment in 'other health professions' (ISCO code 226), which includes pharmacists. Estimates are not reliable due to the small number of survey respondents for this specific occupational category. Figure 16 shows the aggregate trend in employment in this group. Using a different data set on a different (but related) market, Pagliero (2015) studies the impact of the introduction of para pharmacies in Italy and provides some evidence that it increased the demand for pharmacists.

Figure 16. Employment in 'other health professions' (EULF, ISCO code 226).



Note: Figure shows the estimated employment in other health professions (ISCO code 226) in Italy (data from the EULFS). Estimates are not reliable due to the small number of observations.

Finally, the impact of an increase in the number of pharmacies on prices of medicines subject to prescription is likely to be negligible, as the price of medicines is regulated at national level. However, more competition in the pharmacy market might realistically decrease the price of OTC drugs and other products sold in pharmacies. Calzolari *et al.* (2013) provide evidence of this type of effect. However, this requires very specific data sets on specific products and regions, which are generally proprietary and cannot be easily accessed.

Customer Satisfaction

Like data on the number of complaints, customer satisfaction data have been collected only at national level and cannot be matched with either province or municipality data. They are here reported to provide a general overview.

Customer assessments were expressed on a scale from 1 (complete dissatisfaction) to 10 (complete satisfaction) and then transformed into a 0–100 index. Minimum threshold scores were: 50 'Critical neutrality'; 60 'Satisfaction'; 70 'Full satisfaction'; 75 'Excellence'.

In terms of scale of satisfaction, pharmacies had the highest score: above 75 in all the four two-year surveys. Pharmacies were mainly appreciated for the calm and quiet setting and personnel. Other relevant aspects affecting customer perception are: high level of professional skills, flexible opening times, listening availability, professional advice, and a wide range of products.

In 2012 the score was 73.5. If we consider that the score was 76.9 in 2008, 75.5 in 2010 and 73.5 in 2012, the sector is suffering a slow reduction in customer satisfaction, despite the improvement in accessibility observed by the ISTAT survey (see section 3.7.2).

3.8 Conclusion

This case study examined the impact of regulation on the availability of pharmacies in Italy and its potential effects on human health. We used health outcome measures from hospital admission records that can potentially capture the impact of access to medicines (and the other services provided in pharmacies) on the health of the population. We complement this data with information from the National Institute of Statistics (ISTAT) on consumer perception of difficulties accessing pharmacies. Some general data on customers' complaints and satisfaction were also discussed.

The results from the analysis at province level support the hypothesis of an inverse association between density of pharmacies and admission rates of influenza. The instrumental variable regressions are also consistent with this hypothesis. However, these results are based on data at province level, which implies that a significant amount of variability is ignored in the aggregation of the data.

Results using municipality level data provide complementary evidence. Panel data models do not detect a significant effect of the availability of pharmacies on health outcomes. However, results from the regression discontinuity design show a negative impact of availability of pharmacies on the number of hospital admissions related to influenza. These results are in line with the results obtained from the analysis of data at province level and suggest a long-run effect of pharmacy availability on consumer health.

While these are promising results, they should be interpreted with care. While short-run effects are probably small, longer time series could be used in the future to detect the dynamic effects of the availability of pharmacies. Regression discontinuity designs can capture the long-run effects, but they provide results that are difficult to extrapolate out of the sample, as these techniques focus on a specific sample. Overall, the results provide the first evidence of how regulation of the market for pharmacies might affect service availability (a crucial determinant of quality in this market) and consumer health.

References

Altowaijri, A., Phillips C. and Fitzsimmons D. (2013). Systematic review of the clinical and economic effectiveness of clinical pharmacist intervention in secondary prevention of cardiovascular disease. Journal of Managed Care Pharmacy. 19(5):408-16.

Anell, A. (2005). Deregulating the pharmacy market: the case of Iceland and Norway. Health Policy 75,1: 9-17.

Berbatis C., Sunderland V. and Bulsara M. (2003). The services provided by community pharmacists to prevent, minimise and manage drug misuse: An international perspective. In: Drug Misuse and Community Pharmacy. By Sheridan J and Starng J., (eds.) Taylor & Francis.

Blalock S., Roberts A., Lauffenburger J., Thompson T and O'Connor SK (2013). The effect of community pharmacy-based interventions on patient health outcomes: a systematic review. Medical Care Research and Review. 70(3):235-66.

Brown M. and Bussell J. (2011). Medication adherence: WHO cares? Mayo Clinic Proceedings. 86(4):304-14.

Calzolari, G., Ichino, A., Manaresi, F., and Nellas, V. (2013). When the baby cries at night: inelastic buyers in non-competitive markets. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2281909

CFMT - Centro di Formazione Management del Terziario. Italiani Rassegnati? La Customer Satisfaction nei servizi in Italia 2010. Milano – Roma, Novembre 2010. Available from: *https://www.federfarma.it/Documenti/Indagini/RAPPORTO_RICERCA_CUSTOMER2010.aspx*

CFMT - Centro di Formazione Management del Terziario. La customer satisfaction nei servizi in Italia 2008. https://www.federfarma.it/Documenti/Indagini/CFMT2008.aspx

CFMT - Centro di Formazione Management del Terziario. Italiani in Evoluzione. La Customer Satisfaction nei servizi in Italia. Italiani in Evoluzione. La Customer Satisfaction nei servizi in Italia 2012. Milano – Roma, Novembre 2012. *https://www.federfarma.it/Documenti/Indagini/Customer_Indagine_CFMT_2012.aspx*

Chisholm-Burns, M. A., Lee, J. K., Spivey, C. A., Slack, M., Herrier, R. N., Hall-Lipsy, E., and Kramer, S. S. (2010). US pharmacists' effect as team members on patient care: systematic review and meta-analyses. *Medical care*, 923-933.

Cooper, R. J. (2013). Over-the-counter medicine abuse–a review of the literature. *Journal of substance use*, *18*(2), 82-107.

Ferré, F., de Belvis, A. G., Valerio, L., Longhi, S., Lazzari, A., Fattore, G., and Maresso, A. (2014). Health systems in transition. *Health*, *16*(4).

Gai, Y., and Feng, L. (2017). Relationship between pharmacist density and adult influenza vaccination after controlling for individual and neighbourhood effects. Journal of the American Pharmacists Association, 57(4), 474-482.

Huang WT, Chang CH, Hsu YF and Chuang JH. (2015). Prognostic factors for mortality in patients hospitalized with influenza complications, in Taiwan. International Health. 7, 73-5.

Kirkdale, C. L., Nebout, G., Megerlin, F., and Thornley, T. (2017, January). Benefits of pharmacist-led flu vaccination services in community pharmacy. In *Annales pharmaceutiques francaises* (Vol. 75, No. 1, pp. 3-8). Elsevier Masson.

Koshman, S. L., Charrois, T. L., Simpson, S. H., McAlister, F. A., and Tsuyuki, R. T. (2008). Pharmacist care of patients with heart failure: a systematic review of randomized trials. *Archives of Internal Medicine*, *168*(7), 687-694.

Mocetti, Sauro. (2016). Dynasties in professions and the role of rents and regulation: Evidence from Italian pharmacies. Journal of Public Economics 133, 1-10.

Nazar, H., Nazar, Z., Portlock, J., Todd, A., and Slight, S. P. (2015). A systematic review of the role of community pharmacies in improving the transition from secondary to primary care. *British journal of clinical pharmacology*, *80*(5), 936-948.

O'reilly, C. L., Wong, E., and Chen, T. F. (2015). A feasibility study of community pharmacists performing depression screening services. *Research in Social and Administrative Pharmacy*, *11*(3), 364-381.

Pagliero, M. (2015). The Effects of Recent Reforms Liberalizing Regulated Professions in Italy, Report to the European Commission, available from: *http://ec.europa.eu/growth/content/effects-reforms-regulatory-requirements-access-professions-country-based-case-studies-0_en .*

Pousinho, S., Morgado, M., Falcão, A., and Alves, G. (2016). Pharmacist interventions in the management of type 2 diabetes mellitus: a systematic review of randomized controlled trials. *Journal of managed care & specialty pharmacy*, *22*(5), 493-515.

Rubio-Valera, M., Chen, T. F., and O'Reilly, C. L. (2014). New roles for pharmacists in community mental health care: a narrative review. *International journal of environmental research and public health*, *11*(10), 10967-10990.

Ryan, R., Santesso, N., Lowe, D., Hill, S., Grimshaw, J. M., Prictor, M. and Taylor, M. (2014). Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. *Cochrane Database of Systematic Reviews*, (4).

Sabaté E. (2003) Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization

Schaumans, C., and Verboven, F. (2008). Entry and regulation: evidence from health care professions. *The Rand journal of economics*, *39*(4), 949-972.

Tsuyuki, R. T., Al Hamarneh, Y. N., Jones, C. A., and Hemmelgarn, B. R. (2016). The effectiveness of pharmacist interventions on cardiovascular risk: the multicenter randomized controlled RxEACH trial. *Journal of the American College of Cardiology*, *67*(24), 2846-2854.

Vogler, S., Habimana, K., and Arts, D. (2014). Does deregulation in community pharmacy impact accessibility of medicines, quality of pharmacy services and costs? Evidence from nine European countries. *Health policy*, *117*(3), 311

(This page intentionally left blank)

4 Tourist Guides in Greece

4.1 Introduction

In this study, we examine the impact of reforms to the regulatory framework governing entry to the tourist guide profession in Greece on the quality of services provided. The reforms took the form of opening up the qualifying routes to access the profession and the simplification of the administrative procedures to obtain the relevant identity card (also referred to as the tourist guide licence). The overall aim of this study is to examine how the reform process affected the supply and quality of tourist guides in the labour market, as well as the quality of the service offered by tourist guides based on customer evaluations.

Our analysis draws on two data sources. First, data on the labour supply of tourist guides is extracted from the Greek Labour Force Survey (LFS) which is collected by the Hellenic Statistical Authority (ELSTAT). The second data source contains individual level data on customers' evaluations and tourist guides' characteristics and is compiled from an internet booking website for tourist guides. Using appropriate statistical tests we examine the quality of services provided by tourist guides before and after the reform. Moreover, utilizing suitable multivariate models, we are able to quantify the *ceteris paribus* magnitude of any change in the quality indicators.

The rest of this chapter is organized as follows. In section 2, we provide a detailed description of the legislative framework of the tourist guides profession, paying special attention to the necessary qualifications to enter this profession and how these have recently changed. In Section 3, we describe in detail all the statistical data used in the analysis and in the following section we discuss the quality indicators we used based on the available data. Section 4 discusses the quality indicators and section 5 provides an outline of the empirical methodologies used, while section 6 reports the results. Conclusions and policy implications are discussed in the final section.

4.2 Institutional Background

Examining the evolution of the quality of the tourist guides profession in Greece is important given the current economic and social context. The economic importance of the tourism industry in Greece is high and has been rising from a 15.6 per cent contribution to GDP in 2010 to 20 per cent in 2017. Employment in the tourist sector is also sizeable, accounting for 23 per cent of total employment. The sector has displayed substantial resilience during the current economic crisis, as revenues from tourism declined at a lower rate compared to other industries and were the first to make a recovery. The future prospects for the Greek tourism industry, at least in the short to medium term, seem relatively good, especially as the nearby holiday destinations continue to suffer from instability and a climate of insecurity due to terrorism. Tourist guides, despite not being the largest occupational group in the industry, play a role in shaping visitors' experiences of the country.

Tourist guides in Greece are licensed by the appropriate state authority and have exclusive legal rights to perform guided tours of archaeological sites. This is distinct from a tour manager or tour escort who commonly manage an itinerary of activities on behalf of a tour operator, ensure the programme is carried out and provide local practical information. Tour managers and escorts may or may not be tourist guides. According to the Greek legislative framework, while any person can perform city tours or show visitors around various landmarks, only licensed tourist guides are allowed to enter archaeological sites and deliver guided tours. Being caught by the authorities providing such tours without the relevant badge (i.e. licence) carries severe financial penalties.

Regulatory Reforms

Until recently the legislative framework governing entry into the tourist guide profession in Greece was Law 710/1977, under which a mandatory administrative licence and a professional ID were issued by the Hellenic Tourism Organisation (EOT) only for those who had graduated from the tourist guides schools of the Organization for Tourism Education and Training (OTEK).[57] The ID was annual and, as long as the holder had fulfilled his financial obligations arising from his social

[57] Candidates should also have fulfilled their military obligations or have been legally exempted from them and should not have been irrevocably convicted of any major offence.

security, it was renewed. To gain entry to OTEK's schools for tourist guides one must at least hold a degree of upper secondary education (lyceum), arguably a relatively low educational hurdle. The candidates were also examined in a foreign language, where they were required to achieve a score of at least 16 out of 20. Upon successful completion, they were subsequently tested in three modules (history, geography and writing) as a means of selection. Usually these schools accepted about 40 new students every year. Schools in Athens and Thessaloniki operated on a permanent basis, while occasionally and subject to demand and availability of funding, schools in other locations such as in Rhodes and Crete would also run. Course duration in these schools was five semesters, and included sitting exams and submitting a dissertation. Upon graduation, individuals obtained the right to become a 'licensed tourist guide', but the actual commencement of activities procedure was administratively cumbersome. This system of entry to the occupation was operated for a number of years; however, in 2010 OTEK's schools for tourist guides stopped accepting students due to lack of funding. As a result, between 2010 and 2013, no new tourist guides could enter the labour market.

In 2011, Omnibus Law 3919/2011 abolished a number of restrictions regarding entry and conduct within a number of professions in Greece, including tourist guides. First, the licencing procedure was replaced by a simpler and faster announcement of 'activity commencement procedure'. According to the new provisions, once all the necessary documents[58] are submitted to the Ministry of Tourism, the applicant must be notified within 10 days of the result of his/her application. If the ministry does not respond within a 10-day period, then the applicant automatically assumes that his/her application has been approved and can start practising the profession of tourist guide. If successful, the applicant is included in the tourist guides' registry kept by the ministry and can obtain the professional ID (badge)[59].

The second amendment affected the educational requirements of prospective practitioners. In particular, a review by the Hellenic Competition Committee concluded that the requirement to be a graduate from OTEK's tourist guides schools did not meet the criteria of proportionality as it excluded graduates from other schools with similarly specialized knowledge, such as archaeologists and historians. As such, Law 4093/2012, Law 4111/2013, Law 4152/2013 and Ministerial Decision 16368/25-7-2013 established the creation of intensive tourist guide training programmes. These were to be held by public universities, they would last for two months and eligible to attend were graduates of Archaeology, History, History and Archaeology, History of Archaeology and Social Anthropology, History of Archaeology, Management of Cultural Goods and History and Ethnology. The training programmes include nine modules and involve written and oral exams. Upon successful completion of the programme, candidates become 'fully licensed tourist guides'.

Therefore, under the new regime, the routes to access the occupation were enlarged and the educational requirements extended to include individuals from other educational backgrounds. It should be noted, however, that under the old regime secondary education was sufficient to apply to the OTEK's schools for tourist guides, while under the new regime only those who had graduated from specific university departments could attend the intensive training programmes. Furthermore, from 2010 until 2017, when OTEK's schools were not operating, entry into the profession of tourist guide was limited only to graduates of tertiary education from specific academic disciplines. Table 1 summarizes the legislative framework and reforms between 1977 and 2017.

[58] According to Greek legislation, in order to get a professional ID and be allowed to practice the profession of tourist guide the applicant must: I) For Greek citizens and nationals of member states of the European Union, hold a diploma from the Tourist Education and Training Organization (OTEK) Guides School. II) For nationals of member states of the European Union, hold a post-secondary degree which is required by the competent authority in their member state of origin for the purpose of pursuing the profession of tourist guide and: A) Have sufficient knowledge of Greek culture, Greek history and archaeology, as evidenced by the successful attendance of relevant courses within a post-secondary degree. B) Have practiced educational excursions to archaeological sites, museums and historical monuments of Greece. C) Candidates of EU member states outside Greece and foreign nationals of Greek origin should hold a Greek Language Certificate at Level C from the Greek Language Centre or from the Modern Greek Language Teaching School. D) Have fulfilled their military obligations or have been legally exempted from them. Military obligations are not required for nationals of a member state of the European Union for whom such an obstacle is not foreseen in their country. E) Have not been irrevocably convicted of a criminal offence

[59] The ID costs 20 euros which is paid once on initial issue.

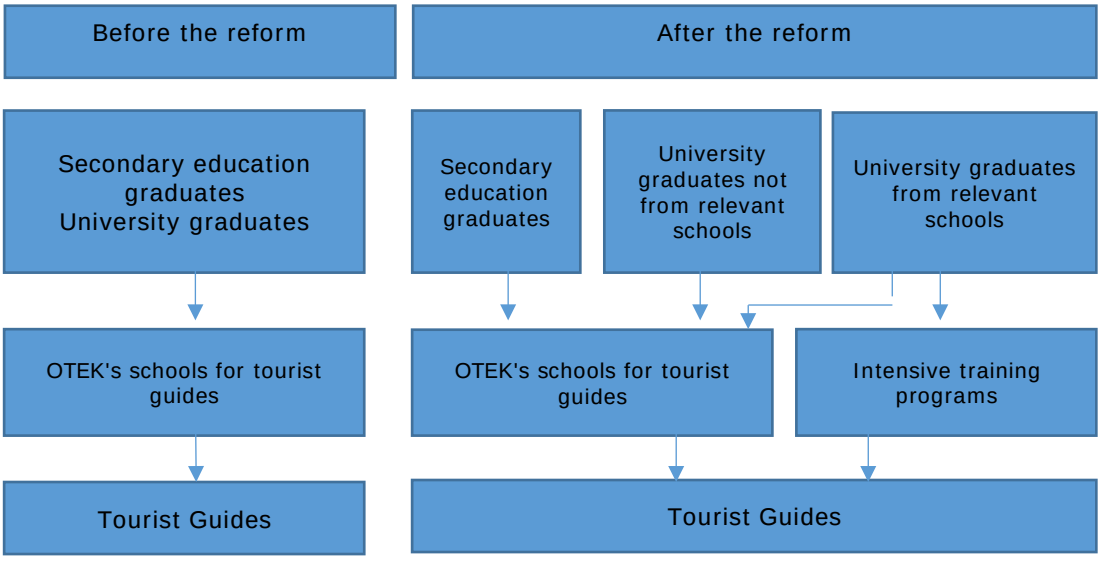
Table 1. Summary of reforms and relevant legislation

Date	New legislation	Description / Reform	Additional information
1977	Law 710/1977	Legal framework defining the process of entry of tourist guides into the labour market, the general rules for the exercise of the profession and related details.	A mandatory administrative licence and a professional ID were issued by the Hellenic Tourism Organisation (EOT) only for those having graduated from OTEK's schools for tourist guides. Candidates to OTEK's schools for tourist guides required to hold at least a degree of upper secondary education (lyceum). For candidates to be accepted to study, the selection process involved written and oral exams.
2010			OTEK's schools for tourist guides stopped accepting students due to lack of funding. No new tourist guides could enter into the labour market.
2011	Law 3919/2011	Abolished a number of restrictions regarding entry and conduct of a number of professions in Greece.	To become a tourist guide one needed to graduate from OTEK's Tourist Guides schools and get an administrative license. The Law simplified the licensing process.
2012	Hellenic Competition Committee (Opinion no18/VI/21.1.2012)	Hellenic Competition Committee under Law 3919/2011 issued a series of opinions on the legislative changes that must be made to various professions in order to liberalize them.	Argued in favour of reviewing the necessary qualifications for practicing the profession of tourist guide. Allow graduates with relevant degrees, like archaeologists and historians, to become tourist guides but not through the two-year programme of OTEK's schools for tourist guides.

2012	Law 4093/2012	Established the creation of two-month intensive tourist guide training programmes by public universities for graduates of specific university departments. Those successfully completing the program would satisfy the educational requirements to become licensed tourist guides.	With this reform new tourist guides can enter into the labour market. However, only graduates holding specific degrees can become tourist guides, suggesting a partial opening of the profession.
2013	Law 4111/2013		
2013	Law 4152/2013		
2013	Ministerial Decision 16368/25-7-2013		
2017		OTEK's school for tourist guides in Athens reopened and new programmes were launched.	

A more comprehensive liberalization of the profession did not come into force until 2017 when both the OTEK's schools and the intensive training university programmes routes became simultaneously available (see Figure 1).

Figure 1: Routes to the tourist guide profession before and after the legislative reform



4.3 Data sources

This section outlines the data sets used to identify the impact of the reforms for tourist guides on the quality of services delivered. For each data source we describe the available quality indicators/proxies and composite indices we construct. We draw on two main data sets: (i) the Labour Force Survey (LFS) of the Hellenic Statistical Authority (ELSTAT) and (ii) a data set compiled using the information on an online platform used to view the profiles of tourist guides and book their services.

Labour Force Survey (LFS)

The first data set used in the analysis is the Greek Labour Force Survey (LFS). LFS is coordinated by Eurostat and collected at a national level every quarter from the ELSTAT. LFS is the largest household survey in Greece responsible for collecting detailed information on the labour market characteristics of the Greek workforce. Individuals are asked about their employment status, occupation and sector of activity. We use this information to identify those in the tourist guide occupation.[60] We supplement this with data relating to their employment profiles to construct quality proxies before and after the reform.

One limitation pertaining to the Greek LFS is that it does not collect very detailed information regarding the classification of occupations. The most detailed occupation classification reported in LFS is ISCO-08 at the 3-digit level and not at the 4-digit level. Unfortunately, in Greece there is no data available containing information regarding occupations at a 4-digit level, not even in the population census. Athanassiou *et al*. (2015), when examining the effect of regulatory reforms on employment, overcome this by using data from alternative sources and their results remain the same, thus suggesting that the LFS, even though not so detailed, performs well in capturing detailed occupational trends. Moreover, during our conversations with ELSTAT officials we were assured (on the basis of the parameters we set for the definition of tourist guides) that they are more than 90 per cent of the cases in our sample.[61]

Data on Customers' Evaluations

The second aim of this study is to examine whether tourists assess the services provided by tourist guides who have attended the two-month intensive programmes differently than those offered by tourist guides who have graduated from OTEK's schools for tourist guides. In order to do this, one needs access to customer data that records, directly or indirectly, the quality of the provided services by tourist guides. As we have already mentioned, to our knowledge, there is no such official data source. Consequently, we will use consumer ratings and other proxies from an online platform through which a tourist can book the services of licensed tourist guides.

More specifically, we compile a data set using information from the website *https://tourguides.viator.com/*.[62] This website contains lists of tourist guides and is used internationally by tourists from all over the world to book tours in various countries, including Greece[63]. Included in this platform are the ratings from previous users (score 1–5), the number of tours booked via this website per guide and the total number of reviews he or she has received. Moreover, through each tourist guide's personal description, we can extract information about their gender, age and professional experience. Furthermore, available information includes the place where they operate, the languages they use in their tours and the prices they charge. Importantly,

[60] The identification closely follows the definition of the tender. In particular tourist guides are those working in sector 79 (NACE Rev.2) and in occupations 5113, 4221 (ISCO-08).
[61] ELSTAT can perform a check since in the primary data they have, the type of occupation is expressed in words and not in a classification.
[62] Unfortunately, there is no other data source providing information on customers' evaluations that allows us to distinguish between tourist guides that have graduated from OTEK's schools and those from the intensive training programmes. In fact some other popular booking websites include, among others, individuals who perform tour guides without the proper licence.
[63] Based on the Ministry of Tourism, the tourist guides in the registry before the reform were around 1,900 and in the dataset we constructed from this website they are 79 (i.e. approximately 4%). The new entrants undertaking the two-month intensive training programs as a result of the reform were 173 and in our dataset we have 27 or 15.6%. While the latter is satisfactory, the former is relatively low. However, as we discuss in our analysis, there are no exits from the registry of tourist guides since the relevant legislation allows for grandfathering (i.e. existing tourist guides can keep their professional ID and work as long as they wish). Thus, we believe that the actual number of tourist guides under the old regime is clearly lower than 1,900 and therefore their percentage in our sample is higher and therefore closer to satisfactory levels of sample representativeness. To make our sample more representative of the population we attempted constructing weights but unfortunately, we had limited information for the stratification process.

based on their professional identity card, which is also available on the website, it is possible to distinguish between the tourist guides who graduated from OTEK's schools for tourist guides (old regime) and those who have attended the two-month intensive programmes (new regime). Utilizing all this information we construct quality proxies and test whether they differ with respect to when and how tourist guides entered their profession.

4.4 Quality Indicators

Labour Force Survey

The LFS has available information which can be used as proxies for the quality of services provided. Defining variables that capture the effect of quality is not an easy task. Proponents of regulation argue that the quality of services is closely related to the level of education in that the higher the educational attainment the better the quality of the practitioner. Given that the higher level of education was a key pillar of the regulatory reform, we use this as an input proxy for quality (treat it as an ordinal variable). A similar approach to measure quality service, although limited to descriptive statistics and a graphical illustration, was used by Athanassiou *et al.* (2015). Here we extend this methodology using more detailed educational levels (seven categories), as well as more elaborate estimation techniques. Moreover, we extend our analysis by utilizing additional information related to the nature and characteristics of the tourist guide profession, which we consider is related, directly or indirectly, to the quality of services provided (Table 2).

Table 2. Proxies for quality of provided services from the LFS

Variable	Link to Quality	Relationship to Quality
Level of education (input proxy)	Tourist guides with tertiary education are expected to provide higher quality services compared to those with lower levels of educational attainment.	+
Had formal training or not (input proxy)	Individuals in receipt of training have acquired more skills and are expected offer better quality services. Since the training is formal in nature, we expect it to directly relate to one's occupation (i.e. offering tourist guide services)	+
Is looking for a another job	We would expect tourist guides of inferior quality to be looking for another job as they are being forced out of the market by their better quality counterparts who are able to secure more bookings.	-

Additionally, we use information on various labour market indicators in the LFS to control for observed heterogeneity and thus isolate, to the extent that this is possible, the effect of the reform (Table 3).

Table 3. Control Variables from the LFS

Category	Variables
Demographic characteristics	Gender; ethnicity; marital status; urbanity; region of residence; number of children aged 0-4; number of children aged 5-14; household composition[64]; labour market status of other family members[65]
Human capital	Age; qualifications; previous labour market status
Job characteristics	Position in the labour market; tenure; full/part-time; permanent/temporary contract; size of local unit of work; insured in social security
Survey characteristics	Year indicators or time trend; quarter indicators
Other	Indicators for the announcement and implementation of reform

Quality Indicators from Customers' Evaluations

Utilizing the available information on the viator website we construct the following proxies of quality of the services provided. First, the number of foreign languages each tourist guide speaks. Since the majority of their services are provided to tourists from abroad it is reasonable to assume that the quality of a tour guide *ceteris paribus* will be higher if this is performed in the mother tongue of the tourist rather than in another widely spoken language (e.g. English, French, German, etc.). To measure the number of foreign languages, that is any other language than Greek, we total the number of languages that are recorded in their professional ID and in some cases the number of other foreign languages they might use but are not officially recognized.

The second proxy is whether the tourist guides speak any 'uncommon' foreign languages. One argument put forward by critics of the new regulatory regime is that the new tourist guides speak mainly English and French, while tourist guides of the old regime were competent in a wider range of foreign languages. To test this, we check if the probability of speaking an 'uncommon' language differs between the two regimes. We use two definitions for an uncommon language. Initially any language other than English, French, German, Italian and Spanish is considered uncommon. In the second definition, we add Russian to the list of uncommon languages. Next, we use the ratings that customers gave to tourist guides for their services. It is reasonable to assume that tourist guides with higher ratings offer better quality services. As long as these do not differ between the two regimes then the liberalization process did not affect the service quality. However, since any differences can be due to other reasons, we also control for other observable characteristics and, importantly, experience. The next two indicators are the number of reviews each tourist guide has received and the number of bookings via this online platform. Both are expected to be positively correlated with quality.[66] Finally, because consumers can be biased when drafting reviews in the sense that they tend to write reviews when they have a negative experience or a very positive one but rarely do it for the intermediate ones, we create a set of composite indices using the information on ratings, the number of reviews and the number of bookings. To construct these indices we follow the next steps as depicted in Figure 2:

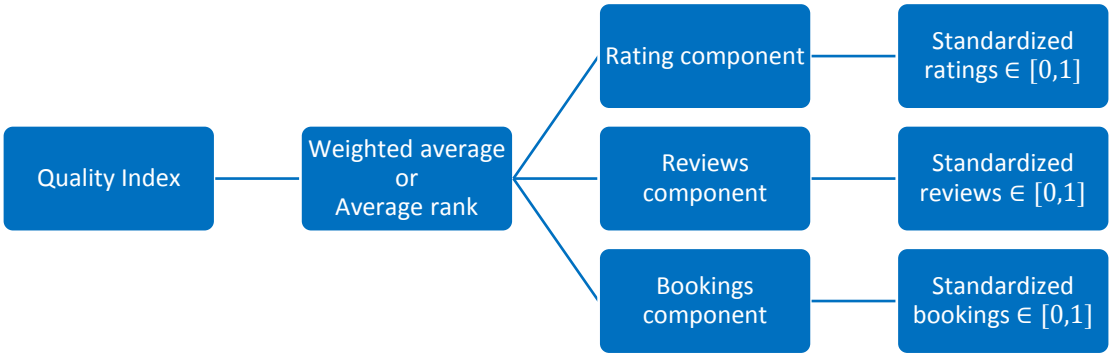
1. We standardize the variables. Since the tree components of the index have different numerical scales (we need to compare apples with apples), we have standardized them to a range between 0 and 1.
2. We assign importance weights to each component.
3. Finally, we estimate a weighted average as in equation (1).

[64] This comprises eight indicator variables representing different combinations of household characteristics depending on whether there are children, a spouse or other family members.
[65] This includes four variables of the number of employed, unemployed, pensioners and inactive in the household respectively.
[66] Interestingly, the number of reviews and the rating score are positively highly correlated in our data set and we also find that the majority of reviews contain positive comments. We are mindful of this and we take it into account in our evaluation of the findings.

$$QI_i = \frac{w_{1i}Rating + w_{2i}\,\mathrm{Re}views + w_{3i}Bookings}{w_{1i} + w_{2i} + w_{3i}} \quad (1)$$

We also estimate the average rank each tourist guide receives in the three components

Figure 2. Quality index construction process



To check whether our results are sensitive to the chosen importance weights we estimate a range of quality indices using various weight combinations as described in Table 4.

Table 4. Quality Index (QI) weighting scheme

Quality Index (QIi)	Relative importance weight	Quality Index (QIi)	Relative Importance weight	Region	Annual change	R2	N years
i	Rating	Number of reviews	Number of bookings	i	Rating	Number of reviews	Number of bookings
0	33.3%	33.3%	33.3%	7	15%	70%	15%
1	40%	30%	30%	8	5%	90%	5%
2	50%	25%	25%	9	30%	30%	40%
3	70%	15%	15%	10	25%	25%	50%
4	90%	5%	5%	11	15%	15%	70%
5	30%	40%	30%	12	5%	5%	90%
6	25%	50%	25%				

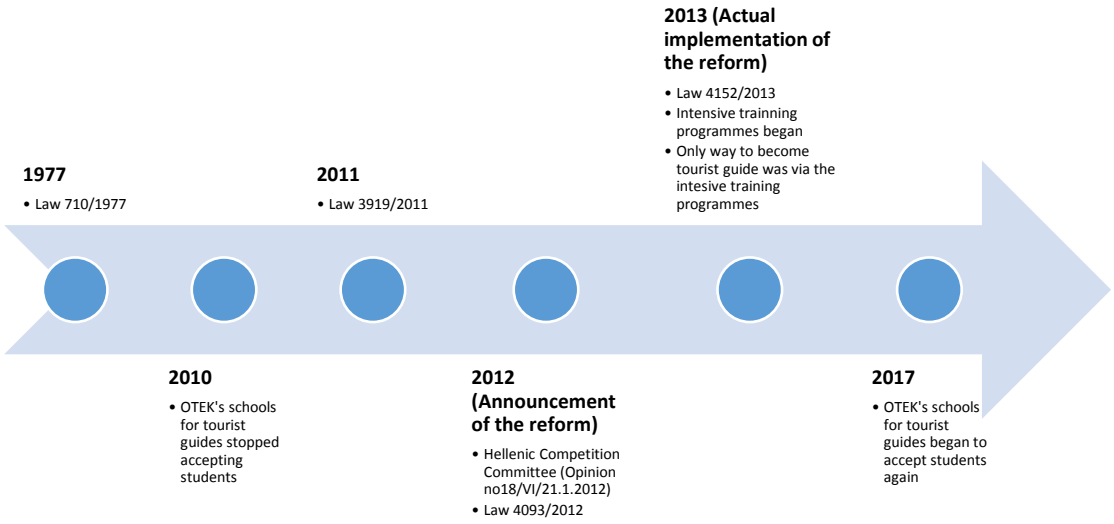
Given data availability limitations, the aforementioned indicators are the best instruments in our disposal. It would be very useful to have data on complaints and tips as proxies for customer satisfaction, but unfortunately, we have no way of accessing this type of data for this study. We recognize that the indicators we use have some limitations; however, in the absence of better data

sources, we proceed with these caveats in mind and try to minimize these effects by using alternative measures and multiple statistical methodologies.

4.5 Methodology

In this section, the methodological approaches used to quantify the effect the regulatory reforms have had on the quality of services provided by tourist guides are presented. The adopted methodology depends strongly on the available data and the (statistical) nature of the variables used.[67] As mentioned earlier while the legal reform was initiated with Law 3919/2011, for its full implementation a series of subsequent relevant laws as well as ministerial decisions were necessary (see Figure 3). Following these legislative reforms, the first intensive training programme for university graduates was completed in the summer of 2013. It is reasonable to accept that this is the exact point in time when the reform actually came into force. Therefore, we can identify the effect the reform had on paper, that is since it was initially announced and before being fully implemented, and the effect it has had after its actual implementation. Consequently, two indicator variables marking the two periods – d_1 for the period starting the first quarter of 2012, when the Hellenic Competition Committee argued in favour of the intensive training programmes, and d for the period since the third quarter of 2013, when the first graduates from the intensive training programmes entered the labour market – are defined[68].

Figure 3. Summary of reforms and relevant legislation



[67] For instance, to examine how the level of certain variables has changed since the reform different tests will be performed depending on whether the examined variable is continuous, binary or ordinal.

[68] In the interest of clarity regarding the choice of periods to study: In 2011 an omnibus law (Law 3919/2011) declared that restrictions regarding entry in and conduct of *all* occupations should be abolished. However, since this was an omnibus law, it did not specify which particular occupations would be affected and additional occupation-specific laws and directions were anticipated. Moreover, Law 3919 stated that the Hellenic Competition Committee should make the final decision regarding which occupations will be affected and specific restrictions to be abolished, and that such decisions would be binding. As such, in 2011, while changes in entry restrictions to professions were expected, their precise nature was not known. It was not until 2012 when the Hellenic Competition Committee made specific reference to the restrictions applied to tourist guides that it became clear that this particular occupation would be affected, while in 2012 the changes were formally announced (Law 4093). We refer to this as the 'announcement period'. The 'actual implementation' period starts mid-2013when the first intensive training programme was completed.

From the LFS certain proxies for the quality of services provided by tourist guides are identified. Using these we can perform appropriate statistical tests to examine whether they have changed before and after the legal reforms. These will be performed for both time periods, d_1 (period after the announcement of the reform) and d (period after the first actual implementation of the reform). Since the proxies are not continuous variables,[69] the usual t-test is not an appropriate way to determine if there are statistically significant differences before and after the reform. The appropriate tests for the selected variables are:

1. Two-sample Wilcoxon rank-sum (Mann–Whitney) test
2. Kruskal–Wallis equality-of-populations rank test
3. Pearson chi^2 test
4. Fisher's exact test

Any difference in the quality indicators before and after the reform as they are estimated by the statistical tests might also be a result of heterogeneity. To account for observed heterogeneity after the statistical tests a regression analysis is conducted, taking into account the aforementioned control variables (see Table 3). The estimated equation takes the following general form:

$$g\left\{\mathrm{E}(Y_{quality} \mid x, d)\right\} = \beta_0 + x_i'\beta + \gamma d_{ij} + \varepsilon_i \qquad (2)$$

where $Y_{quality}$ is any of the indicators in Table 2, the vector x_i depending on the model contains variables from Table 3 and ε_i is an error term with the usual properties. d_{ij} is an indicator variable presenting the initial announcement of the reform (j=1) and/or the actual implementation of the reform.[70] g*()* is a link function depending on the quality indicator used.[71]

A variation of equation (2) is used to model the effect of the reform on service quality when data from customers' evaluations are used. In particular, we use a generalized linear model (GLM)[72] to test whether there is a statistically significant difference between graduates from OTEK's tourist guides schools and those from the intensive training programmes on the level of the quality indicators used. Thus, we estimate:

$$g\left\{\mathrm{E}(Y_i \mid x_i, \mathrm{z}_i)\right\} = \beta_0 + x_i'\beta + \gamma z_i + \varepsilon_i \qquad (3)$$

Here we use the quality indicators described above as our dependent variable. Since the data allows us to distinguish whether each tourist guide belongs to the previous regime (graduate from OTEK's schools for tourist guides) or to the new regime (graduate from an intensive training programme), we add a dummy variable z taking the value 1 for the new regime to capture the effect of the reform on the quality indicator. As before, because the effect might also be the result of individual heterogeneity we control for observed heterogeneity. It is crucial to mention that both equations (2) and (3) are non-linear. Thus, the coefficients cannot be interpreted in a similar way as in a simple regression, i.e. the coefficient is not equal to the magnitude of the effect. Thus, we need to estimate the so-called marginal effects, or else we need to see how much $g\left\{\mathrm{E}(Y_i)\right\}$ changes if a dummy variable changes from 0 to 1.[73]

Finally, because our dependent variable in equation (2) has many zeros, to check if our results are sensitive to the choice of the econometric model, we also estimate a two-part model. In these models, there are two separate independent parts. The first part models the probability that the dependent variable will be greater than zero using a binary choice model. The second part models the distribution of the dependent variable conditional that it is greater than zero using (generalized) linear models.

---

[69] In particular the level of education is an ordinal variable, while the rest are binary variables.

[70] When we examine the effect of the announcement (d_1), 2011 is the before period and the subsequent years until 202013q3 define the after period. When we examine the effect of the actual implementation of the law (d) the period between 2012 and the third quarter of 2013 is the before period and subsequent quarters until the end of 2016 is the after period. We estimate these in separate models but also jointly in a single model.

[71] For the level of education $g()$ is an ordered probit and for the rest proxies it corresponds to a probit.

[72] The choice of a GLM stems from the fact that the dependent variable is bounded between 0 and 1. In particular we use a logit link function.

[73] For continuous variables we need to estimate the first derivative of each equation. However, the main variables of interest (d_{ij} and z_i) are all dummy variables.

4.6 Results

Results from the Labour Force Survey

We begin the analysis by looking at changes in the proxies before and after the reform. Because there was a significant time lag between the initial announcement of the reform and its actual implementation, it is interesting to identify the effect the reform had on paper and the effect of its actual implementation. To do this we distinguish two time points. The first is the announcement of the reform (i.e. when Law 3919/2011 a all its implementing laws were voted for and an Opinion by the Hellenic Competition Committee was issued [74]), and the second is when the reform had an actual effect on tourist guides (i.e. when the first tourist guides under the new regime were able to enter the market[75]). Two measures of 'correlation' between the quality proxies and the two examined periods are used, namely Pearson's x^2 and Fischer's exact test.[76] As can be seen from Tables 5 and 6, Law 3919/2011 and all its implementing measures had an effect on the decision of tourist guides to follow a training programme and on their probability of looking for another job. We find that the initial announcement of the law had no effect on the level of education. However, the actual implementation of the reform had a highly significant effect both on the level of education and on the decision to follow a training programme.[77]

Table 5. Pearson's χ2 test

	Effect of announcement		Effect of actual implementation	
	Test statistic	P-value	Test statistic	P-value
Level of education	3.138	0.679	25.907	0.000
Training	3.052	0.081	3.669	0.055
Looking for another job	2.750	0.097	0.209	0.648

Source: Labour Force Survey. Author's own calculations; Highligted cells indicate statistical significance at 10% level.

Table 6. Fischer's exact test

	Effect of announcement		Effect of actual implementation	
	Test statistic†	P-value	Test statistic†	P-value
Level of education		0.723		0.000
Training		0.122		0.074
Looking for another job		0.121		1.000

Source: Labour Force Survey. Author's own calculations Highligted cells indicate statistical significance at 10% level. † Unlike most statistical tests, Fisher's exact test does not estimate a test statistic; instead, it calculates the p-value.

Moreover, two non-parametric tests analogue to the *t*-test are estimated, which are appropriate when the dependent variable is not normally distributed. Namely we run the Kruskal–Wallis (H)

[74] Law 3919 was voted in 2011. However, in order to be formally implemented a number of Ministerial Decisions and other legal documents were necessary. This process was finalized in 2012. So there is an overlap with the Opinion of the Hellenic Competition Committee. Moreover, the involvement of the Competition Committee was foreseen in Law 3919 since the latter was a general omnibus law. Thus, to check the effect of the announcement we use 2011 as the before period and the time between 2012 and the actual implementation of the law, that is 2013q3 as the after the announcement period.

[75] The first tourist guides under the new regime graduated from the new intensive training programs in the summer of 2013. Thus, in order to test the effect of the actual implementation of the reform we define as the before period the time between 2012 until 2013q3 and as the actual implementation period the time between 2013q4 until the end of 2016.

[76] Both tests examine the hypothesis that the rows and columns in a two-way table are independent. We use Fischer's exact test because in some cases the observations we have are very small and results from Pearson's x^2 may not be so reliable.

[77] It is reassuring that both tests provide, at least for the effect of the actual implementation of the law, similar results (see Table 5 and Table 6).

test (Table 7) and the Wilcoxon–Mann–Whitney (U) test (Table 8).[78] The Kruskal–Wallis (H) test was conducted to determine if the examined quality proxies differ between two time periods, i.e. before and after the reform. The H test shows that the announcement of the reform is associated with a statistically significant difference in the probability of a tourist guide looking for another job and in the probability of undergoing a training programme. Moreover, the actual implementation of the reform had a statistical significant effect only in the probability of undergoing a training programme. This is similar to what we found with Pearson's χ^2 tests.

Table 7. Kruskal–Wallis H test

	Effect of announcement		Effect of actual implementation	
	Test statistic	P-value	Test statistic	P-value
Level of education	0.986	0.321	1.484	0.223
Training	3.042	0.081	3.663	0.056
Looking for another job	2.741	0.098	0.206	0.650

Source: Labour Force Survey. Author's own calculations; Highligted cells indicate statistical significance at 10% level

Table 8. Wilcoxon–Mann–Whitney (U) test

	Effect of announcement		Effect of actual implementation	
	Test statistic	P-value	Test statistic	P-value
Level of education	0.993	0.321	-1.218	0.223
Training	1.744	0.081	-1.914	0.056
Looking for another job	1.656	0.098	-0.456	0.648

Source: Labour Force Survey. Author's own calculations; Highligted cells indicate statistical significance at 10% level.

To sum up, three out of four tests confirm that the voting of Law3919/2011 and all its implementing measures by the start of 2012 had an effect on the probability of looking for another job, while two out of four tests indicate a relationship between the voting of the law and the tourist guides' decision to follow a training programme. Moreover, tourist guides' level of education as well as their decision to undergo a training programme seem to be affected by the actual implementation of the law. Therefore, the voting of the law (in 2011) and all its implementing measures by the start of 2012 had an impact on different quality proxies than the actual implementation (in 2013q3). To some extent, this is to be expected since given the threat of new entrants in the market the 'insiders' are likely to try and maintain their market position by improving the quality of the services they offer. Nevertheless, we interpret these tests with caution, as other reasons (ongoing recessions, observed and unobserved heterogeneity, etc.) might drive the observed changes. Consequently, we supplement this with multivariate regression estimates.

We proceed with the regression analysis by estimating equation (2). Three sets of estimates are presented depending on the date used to determine the beginning of the reform.[79] Initially we run separate models using the two time points discussed before (see Table 9) and next we proceed by including both time points in the estimated model (see Table 10).[80] Keep in mind that the coefficient in Table 9 compares each period with the base period in each occasion as described before, while in Table 10 the base period is common and is 2011.

It is interesting that the previous bivariate analysis did not reveal any relationship between the level of education and the announcement of the reform. However, the multivariate analysis shows that the reform, regardless of when its beginning is set, had a positive effect on the level of

[78] These tests are very similar, but we use both for robustness purposes. Since they provide similar results, we describe results from the former.

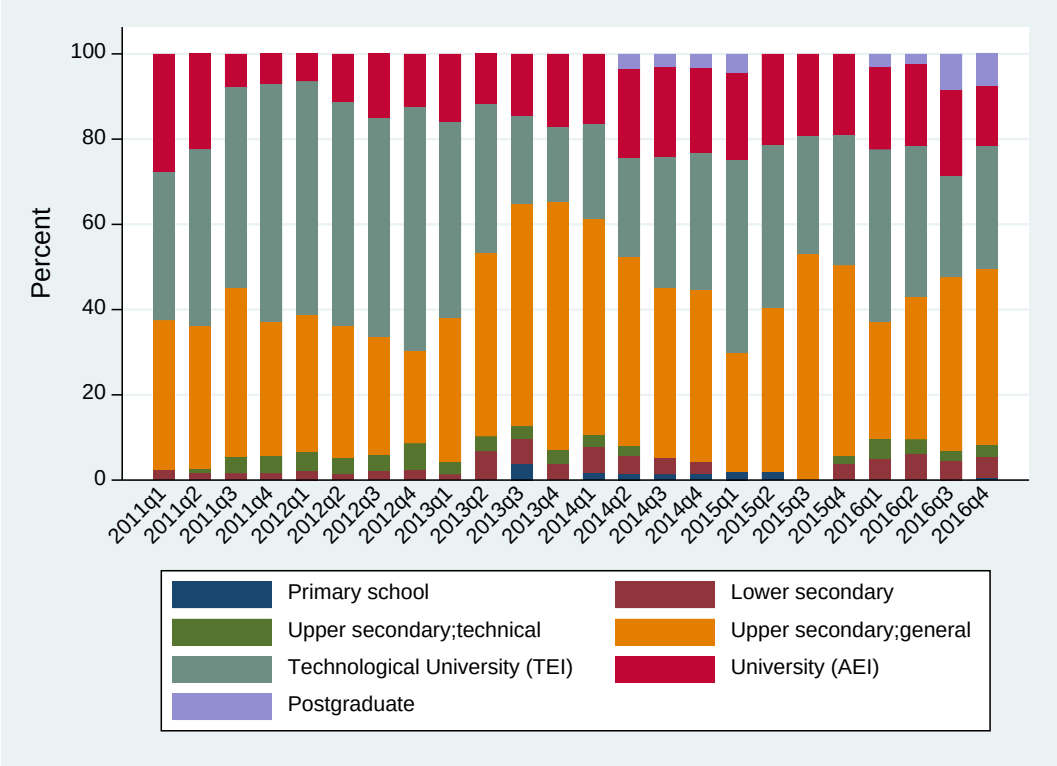
[79] As in statistical tests we examine the effect from the voting of the omnibus law that set the beginning of deregulation reform and from the actual implementation of the reform, which is when the first tourist guides under the new regime could obtain a licence and legally start working. In some instances because our dependent binary variable had very small variability the estimated models did not converge.

[80] A full set of the results can be found in the Appendix.

education, i.e. tourist guides' level of education increased after the reform. The effect is larger after the general announcement of the reform by Law 3919 in 2012 ($\hat{\gamma}$=0.404) compared to the actual implementation of the reform ($\hat{\gamma}$=0.047). This suggests that the qualifications of tourist guides after the reform did increase with a decreasing effect. Since the rationale of the reform was (among other things) to improve service quality, we find that the reform had a small but positive effect on the quality of services provided.

Surprisingly, due to the small variability of the dependent variable, the estimated models for the probability of attending a training programme did not converge when we examined each period separately.

Figure 4. Distribution of tourist guides' educational level %



Source: Labour Force Survey. Author's own calculations.

Following the analysis of human capital accumulation, the next proxy used has to do with whether the tourist guide is looking for another job. An individual might start looking for a second job to compensate income losses, either because wages have fallen or working hours have decreased. However, one cannot rule out that individuals might want to obtain a 'job portfolio'. Under this theory tourist guides may prefer diversity in job tasks and thus increase their overall utility (see Renna and Oaxaca (2006) who use personal preferences for job differentiation to develop a job portfolio model). On the other hand, a second job might be considered a form of insurance due to the overall crisis or the legal reforms affecting the primary job. Alternatively, tourist guides might want to switch occupations or employers due to a poor match and use the second job as a stepping stone (keep in mind that a lot of tourist guides are university graduates from various scientific backgrounds). Regression analysis suggests that the probability of looking for a second job has fallen since both the announcement and the actual implementation of the law, which is likely to be indicative of better matching.

Table 9. Estimates for quality proxies by announcement and implementation (separate models)

	Level of education	Training	Looking for a second job
	(+)	(+)	(-)
Effect of announcement	0.404***	†	-19.810***
	(0.025)		(1.218)
Pseudo R^2	0.3354		0.7201
Log likelihood	-49344.1		-1328.8
Sample size	293		293
Effect of actual implementation	0.047***	†	-7.946***
	(0.019)		(0.613)
Pseudo R^2	0.2254		0.9358
Log likelihood	-147150.1		-442.7
Sample size	596		596

Source: Labour Force Survey. Author's own calculations. † Model did not converge; * p<.1; ** p<.05; *** p<.01

Table 10 presents estimates for these quality proxies but now indicators of the effect for both the announcement and the actual implementation of the reform are included simultaneously and compared to the same base year, namely 2011. Here again we find that the reform had a positive effect on the level of education, which now as expected is larger for the actual implementation of the reform.

The effect on the probability of training can be estimated only for the actual implementation of the reform and its impact is negative. The size of the decrease is statistically significant and the magnitude of the coefficient quite large. This finding seems surprising, as we would expect that insiders would try to establish their position in the market and in doing so, they would try to acquire as many qualifications as possible. On the other hand, the financial constraints due to the economic crisis might have affected negatively both the attendance as well as the organization of training seminars. The model examining if one is looking for a second job indicates that the reform, regardless whether we examine its announcement or its actual implementation, had a negative impact on the likelihood of looking for another job, possibly indicating better job matching in the labour market.

Table 10. Estimates for quality proxies by announcement and implementation (joint model)

Region	Effect of announce-ment	Effect of actual implementation	Pseudo R^2	Log likelihood	Sample size
Level of education (+)	0.064***	0.117***	0.1897	-181118.5	742
	(0.019)	(0.027)			
Training (+)		-6.605***	0.667	-4707.9	591
		(0.377)			
Looking for a second job (-)	-4.772***	-6.961***	0.7803	-2329.0	738
	(0.160)	(0.246)			

Source: Labour Force Survey. Author's own calculations. * p<.1; ** p<.05; *** p<.01

To sum up, we use data from the LFS to quantify the effect the latest reform on the licensing requirements and procedure of tourist guides in Greece had on the quality of the services they provide. Results on the effect of the reform on quality are mixed. There are indications of better quality when using the level of education and the probability of looking for a second job. However, when examining the probability of undergoing training, the results suggest a deterioration in quality.

Results from Customers' Evaluations

We begin our analysis by examining the total number of foreign languages spoken by tourist guides and whether they speak an 'uncommon' foreign language. With regards to the former, results are presented in Table 11. At the top of the table we show the estimated coefficients and at the bottom the marginal effects of being a tourist guide under the new regime on the probability of speaking one to four foreign languages. Overall it seems that there is no significant evidence that tourist guides in the new regime are better or worse than those from the old regime. This evidence holds both in a simple model without covariates and in a richer model with covariates to capture, to the degree that this is possible, any observed heterogeneity. Because these models do not have such good explanatory power (pseudo R^2 is lower than 10 per cent), we draw our conclusions with caution. Further, this model does not distinguish which languages are spoken. For instance, two tourist guides might both speak two foreign languages, but if one speaks a less commonly used one it is reasonable to assume that he/she can provide more specialized services and attract more business.

Table 11. Total number of foreign languages spoken

	Total number of foreign languages spoken	Total number of foreign languages spoken[1]
New regime	-0.680	1.097
	(-1.61)	(1.32)
Pseudo R^2	0.0105	0.0964
Log likelihood	-124.8	-113.9
Number of observations	106	106
Marginal effect of new regime tourist guide	-0.003*	0.289
Probability of speaking 1 language	0.147*	-0.202
	(1.67)	(-1.35)
Probability of speaking 2 languages	-0.0326	0.0376
	(-1.21)	(1.00)
Probability of speaking 3 languages	-0.0725	0.0995
	(-1.57)	(1.31)
Probability of speaking 4 languages	-0.0420	0.0652
	(-1.40)	(1.24)

source: Data from customers' evaluations. Author's own calculations. 1 Other controls: Gender, occupational experience, place of residence; t statistics in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01

We proceed to examine whether the probability of speaking an uncommon foreign language differs between the two regimes. We use two definitions for an uncommon language. In the first, any language other than English, French, German, Italian and Spanish is considered rare. In the second, we include Russian in the list of common languages. Results are presented in Tables 12 and 13 respectively. Overall, we see that tourist guides under the new regime are less likely to speak an uncommon language, regardless of the definition used. In particular for the first definition the probability is lower by 19.4 per cent to 22.5 per cent, while for the second definition the probability, as expected since we narrowed our definition, increases marginally to -21.4 per cent. Our result is quite robust and both our models seem to have a reasonable fit. Thus, we can conclude that quality (as measured by the variety of foreign languages used) has decreased as a result of the liberalization reform.

Table 12. Probability of speaking a rare language (definition 1)

	Coefficient	Marginal effect	Coefficient[1]	Marginal effect[1]
New regime	-1.305*	-0.194*	-1.983	-0.225*
	(-1.67)	(-1.69)	(-1.64)	(-1.68)
Constant term	-1.221***		-3.216**	
	(-4.55)		(-2.30)	
Pseudo R^2	0.0353		0.237	
Log likelihood	-49.53		-39.16	
Number of observations	106		106	

Source: Data from customers' evaluations. Author's own calculations. 1 Other controls: Gender, occupational experience, place of residence, total number of foreign languages spoken; t statistics in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01

Table 13. Probability of speaking a rare language (definition 2)

	Coefficient	Marginal effect	Coefficient[1]	Marginal effect[1]
New regime	-1.633	-0.182	-2.732*	-0.214*
	(-1.54)	(-1.51)	(-1.66)	(-1.70)
Constant term	-1.625***		-3.562**	
	(-5.35)		(-2.12)	
Pseudo R^2	0.0428		0.325	
Log likelihood	-39.60		-27.94	
Number of observations	106		106	

Source: Data from customers' evaluations. Author's own calculations. [1] Other controls: Gender, occupational experience, place of residence, total number of foreign languages spoken; *t* statistics in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01

The next step in our analysis involves utilizing information from customers' evaluations. The ratings customers gave after using a tourist guide are presented on a scale of one to five (five indicating the highest level of satisfaction). The second piece of information involves the total number of reviews that a tourist guide has received. Although the higher number of reviews does not necessarily imply better quality service, it turns out that there is a high positive correlation between the number of reviews and the rating score. Finally, the third proxy of consumers' satisfaction used is the total number of bookings for each tourist guide.[81] Following the same methodology, we regress these indicators of customer satisfaction on a dummy independent variable indicating whether the tourist guide belongs to the old or to the new regulatory regime (Table 14). We find that tourist guides regulated by the new legislation offer lower quality services. In all three examined proxies of consumer satisfaction, being a new tourist guide exerts a negative and statistically significant impact upon these proxies. In particular as the marginal effects estimates indicate, being a new tourist guide reduces the consumer rating by almost 25 per cent, while the negative impact upon the number of reviews as well as upon the number of bookings is lower, around minus 5 per cent. We consider our results to be fairly robust as we have been able to control for all relevant variables such as professional experience, gender, place of residence, number of foreign languages spoken and prices charged.

[81] Since the same scale was not used for these indicators, they were standardized to range from zero to one.

Table 14. Effect of reform on customers' evaluations

	Ratings	Number of reviews	Number of bookings
	Coefficient		
New regime	-1.156**	-1.654**	-1.453*
	(-1.99)	(-2.46)	(-1.86)
	Marginal effect		
	-0.247**	-0.0586***	-0.0538**
	(-2.19)	(-2.88)	(-2.33)
Number of observations	106	106	106

Source: Data from customers' evaluations. Author's own calculations. Note: t statistics in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01

While the results of Table 14 are robust, they do have the limitation of analysing separate indices of service quality, which should be approached by a single variable. To this end, we calculate the average rank for each tourist guide of the above three proxies and use it as the dependent variable. In order to test whether our results are robust and not sensitive to the chosen weight we have run regressions for the alternative weights as they are presented in Table 15.

Table 15. Effect of reform on composite indices by customers' evaluations

	Coefficient	Standard error	Marginal effect	Standard error	Number of observations
Average Rank	-1.091**	(-2.38)	-0.219***	(-2.70)	106
QI_0	-0.833**	(-2.19)	-0.114**	(-2.53)	106
QI_1	-0.832**	(-2.16)	-0.127**	(-2.47)	106
QI_2	-0.848**	(-2.11)	-0.146**	(-2.40)	106
QI_3	-0.924**	(-2.05)	-0.185**	(-2.29)	106
QI_4	-1.059**	(-2.01)	-0.225**	(-2.22)	106
QI_5	-0.840**	(-2.22)	-0.108**	(-2.57)	106
QI_6	-0.857**	(-2.27)	-0.0996***	(-2.64)	106
QI_7	-0.943**	(-2.39)	-0.0822***	(-2.78)	106
QI_8	-1.218**	(-2.52)	-0.0659***	(-2.88)	106
QI_9	-0.836**	(-2.20)	-0.108**	(-2.55)	106
QI_{10}	-0.848**	(-2.22)	-0.0986***	(-2.58)	106
QI_{11}	-0.913**	(-2.23)	-0.0799***	(-2.61)	106
QI_{12}	-1.128**	(-2.10)	-0.0621**	(-2.48)	106

Source: Data from customers' evaluations. Author's own calculations. Note: t statistics in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01

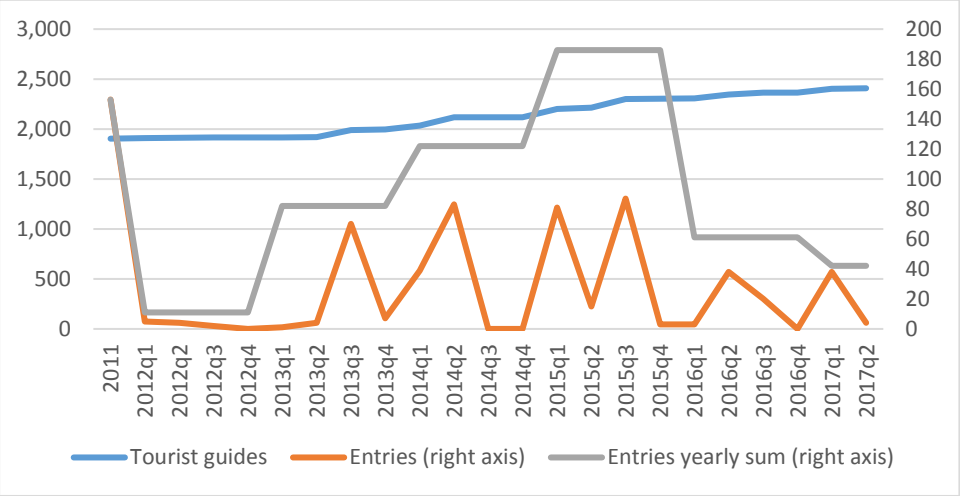
Table 15 shows that approaching the service quality by a single index, namely the average rank of the used variables, is also affected negatively by those in the new institutional regime. It is interesting to note that the size of the marginal effect in this case is rather close to those of Table 14 (that is around 22 per cent). When our dependent variable is approached by the alternative weights (Table 15), it is also always negatively and significantly affected by those entering under the new legislation. In particular this negative impact varies from -6 per cent to -22.5 per cent along the various weight combinations. These findings enhance those of Table 14, confirming the negative impact of the new regime on the quality of service as measured by customer evaluations.

Effect on Stock and Flows in the Profession

We extend our analysis to estimate the effect of the reform on stock and flows in the tourist guide profession. Administrative data from the Ministry of Tourism provide relevant information that is summarized in Figure 5. It appears that the total number of tourist guides was constantly around 1,910 persons for the years until the second semester of 2013. Since then, there has been a consistent increase of registered tourist guides and in the second quarter of 2017 they reached 2,408 persons, which is equivalent to a total increase of 26 per cent. This increase should be considered as one of the biggest increases of professional employment during the great economic crisis that has hit Greece in recent years.

The increase in the stock of tourist guides reflects mainly entries of new tourist guides into the profession, as there were no exits from the relevant registry held at the Ministry of Tourism. The lack of exits from the registry is also explained by the relevant legislation, which allows for grandfathering (i.e. existing tourist guides can keep their professional ID and work as long as they wish). The new entries to the tourist guides profession are concentrated in the years 2013–2015 when around 130 per year were newly registered. These entries come only from those completing the aforementioned intensive training programmes, as the other route into the profession of tourist guides (i.e. OTEK's schools of tourist guides) has not been operational in the period 2010 to 2017.

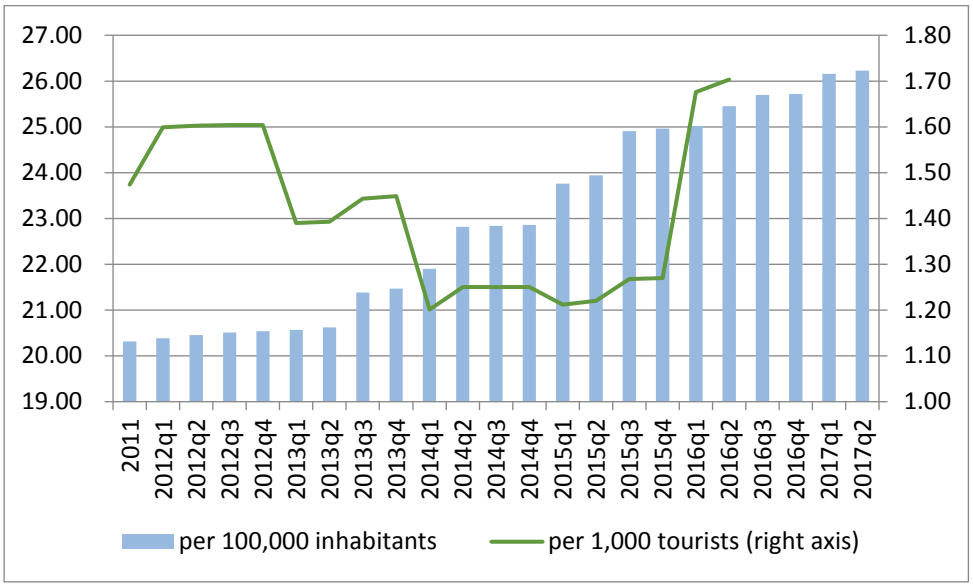
Figure 5. Tourist guides, stock and inflow



Source: Ministry of Tourism

The observed increase in the number of tourist guides in recent years can be better understood when it is standardized (i.e. expressed relative to the number of tourists they are supposed to serve or by the country's population). Figure 6 presents these standardizations and shows that in the years 2013–2015 there were roughly 1.2 tourist guides per 1000 tourists, and this figure increases to 1.7 towards the end of 2015 (i.e. when we expect the first wave of university graduates to have entered the market).

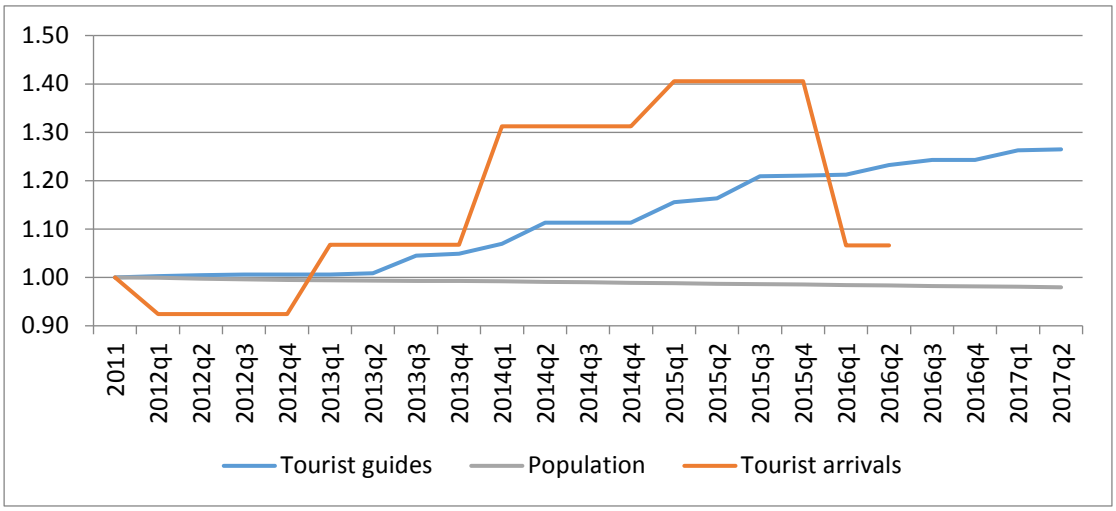
Figure 6. Tourist guides as per inhabitants and tourists



Source: Ministry of Tourism, ELSTAT

Another aspect of the evolution of tourist guides and tourist arrivals is presented in Figure 7 where these variables are expressed as indices with base 1. It is clear that tourist arrivals in the years 2012–2015 were much higher than the number of tourist guides, indicating that there was room in the market to increase the supply of tourist guides.

Figure 7. Evolution of tourist guides, population and tourist arrivals



Source: Ministry of Tourism, ELSTAT, Bank of Greece

4.7 Conclusions

This study examined how the reforms to the profession of tourist guide affected the supply and quality of tourist guides in the labour market, as well as the quality of the service offered by tourist guides based on customer evaluations. Prior to the reform, in order to become a tourist guide in Greece it was necessary to successfully complete a five-semester programme provided only in a specific post-secondary public school (known as OTEK). Minimum entry requirements to the programme were set at secondary education level and excellent knowledge of at least one foreign language. While there were no fees, admission levels to this programme were in practice restricted

to no more than 40 students per year[82] and written exams were used to select the candidates. Under the new legal framework, public universities were allowed to run two-month intensive training programmes addressed to graduates of specific departments related to the tourist guide profession. Attending these programmes became subject to a fee and successful applicants became fully licensed tourist guides. In addition the new procedure of activity commencement is simpler and faster. Once all the necessary documents are submitted to the Ministry of Tourism, the applicant must be notified within 10 days of the result of his/her application, otherwise he/she automatically assumes that his/her application has been approved and can start practising the profession of tourist guide.

One major challenge for a study of this kind relates to the availability of reliable data measuring the level of service quality, while enabling us to distinguish between tourist guides who had graduated from OTEK's schools (old regime) and those from the intensive training programmes (new regime). To identify the effects this reform process we used data extracted from the Greek LFS and customer evaluations data available from an online tourist guide booking platform. Using these sources, we constructed a series of proxies for service quality and estimated alternative statistical models to check the robustness of our results. We recognize that our proxies are not flawless, but in the absence of better alternatives and combined with the use of robust estimation methods, they are at the very least informative.

The LFS analysis has shown that liberalization of the tourist guide profession has increased the stock of human capital in the market (a measure of input quality). In particular, after the reform the educational level of tourist guides has increased substantially, which is most likely related to the fact that university graduates can now enter the profession in an easier and faster way. It is also positive to find that the search for a second job has declined, an indication that tourist guides after the reform are of good quality and therefore are not being forced out of the market. However, we do not find enhanced levels of formal training among tourist guides after the reform (an additional indicator of human capital stock).

To gauge the effect of the reform on the quality of service offered as evaluated by consumers, we compiled a data set from customers' evaluations from an online booking platform and matched it to the characteristics of the tourist guides. Our data source enabled us to distinguish between the tourist guides that qualified under the old regime and those entering the market after the reform. Overall, we find that tourist guides entering the profession under the new regime speak as many foreign languages as those under the previous regime. However, we find that new tourist guides have a significantly lower probability of speaking 'uncommon' languages. Our results also suggest that tourists evaluate graduates from the two-month intensive training programmes with a lower score, as it is approximated by the value of the assessment ratings, the number of bookings, the number of reviews and a composite index of these. This finding is robust as it is not sensitive to the choice of quality proxy utilized or the statistical model used.

Finally, we examine the effect of the reform on the stock and flow of practitioners in the market (an indicator of service availability). Against the backdrop of a period with a stable number of tourist guides in the labour market and falling employment during the economic crisis that has hit Greece in recent years, post-reform we observe a substantial increase of around 26 per cent between 2010 and 2017. Nevertheless, this increase is still smaller than the rise in the number of tourist arrivals, suggesting that there is room in the market to increase the supply of tourist guides.

Overall, it transpires that the reform has had a positive effect on service availability indicators (stock and flow of tourist guides), which was much needed in a context of increasing tourism in Greece. However, although the reform opened up the entry route to more and better qualified practitioners (as measured by education levels and attachment to the job), we do not find any evidence that new entrants are of better quality as judged by customer evaluations. One interpretation is that the educational requirements prior to the reform (i.e. completion of secondary education) coupled with the specialized training offered by the OTEK schools were sufficient to ensure high-quality service. Or else that higher educational credentials of a more general nature (in the form of university education) even when combined with a short specialized course do not translate to better quality.

[82] The number of students each year was decided by the Ministry of Tourism and Finance and usually it was set at around a maximum of 40.

References

Altman, D. G. (1991). Practical Statistics for Medical Research. London: Chapman & Hall/CRC.

Athanassiou, E., Kanellopoulos, N., Karagiannis, R. and A. Kotsi, (2014). The Effects of Liberalization of Professional Requirements in Greece, Report prepared for the European Commission, available from: https://ec.europa.eu/growth/content/effects-reforms-regulatory-requirements-access-professions-country-based-case-studies-0_en

Conover, W. J. (1999). Practical Nonparametric Statistics. 3rd ed. New York: Wiley.

Finney, D. J. (1948). The Fisher–Yates test of significance in 2x2 contingency tables. Biometrika 35: 145–156.

Fisher, R. A. (1935). The logic of inductive inference. Journal of the Royal Statistical Society 98: 39–82.

Kruskal, W. H., and W. A. Wallis. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47: 583–621.

Kruskal, W. H., and W. A. Wallis. (1953). Errata: Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 48: 907–911.

Mann, H. B., and D. R. Whitney. (1947). On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18: 50–60.

Renna F, Oaxaca RL (2006) The economics of dual job holding: a job portfolio model of labor supply, IZA Discussion Paper No. 1915

Riffenburgh, R. H. (2012). Statistics in Medicine. 3rd ed. San Diego, CA: Academic Press.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. Biometrics 1: 80–83.

Appendix A

Wilcoxon rank-sum test, which is also known as the Mann–Whitney two-sample statistic (Wilcoxon 1945; Mann and Whitney 1947) tests the hypothesis that two independent samples (in our case the level of quality proxies before and after the reform data) are from populations with the same distribution. Wilcoxon's test T statistic is the sum of the ranks for the observations in the first sample, i.e. before the reform

$$T = \sum_{i=1}^{n_1} R_{1i} \tag{1}$$

And Mann and Whitney's U statistic, which is based on Wilcoxon's test T, is the number of pairs $(x_{1i}; x_{2j})$ such that $x_{1i} > x_{2j}$

$$U = T - \frac{n_1(n_1+1)}{2} \tag{2}$$

The Kruskal–Wallis test (Kruskal and Wallis 1952, 1953; Altman [1991, 213–215]; Conover [1999, 288–297]; Riffenburgh [2012, sec. 11.6]) is a multiple-sample generalization of the aforementioned two-sample Wilcoxon rank-sum test. Kruskal–Wallis tests, such as the Wilcoxon test, test the hypothesis that the two samples (before and after the reform) are from the same population. The Kruskal–Wallis test H is defined as:

$$H = \frac{1}{S^2}\left\{\sum_{j=1}^{m}\frac{R_j^2}{n_j} - \frac{n(n+1)^2}{4}\right\} \tag{3}$$ and

$$S^2 = \frac{1}{n-1}\left\{\sum_{ALL} R(x_{ji})^2 - \frac{n(n+1)^2}{4}\right\} \tag{4}$$

The test is found to follow an χ^2 distribution with m-1 degrees of freedom.

The next two tests must be considered in the context of a two-way table, where rows represent the level of the quality proxy and columns divide this value to before and after the reform period. In particular we denote nij the number of observations in the ith row and the jth column of a table.

Pearson's χ^2 test examines the hypothesis that the rows and columns in a two-way table are independent. For a population sample n_{ij} let $i = 1, \ldots, I$ and $j = 1, \ldots, J$ be the number of observations in the ith and jth column of a 2x2 matrix of two variables. The Pearson χ^2 statistic with $(I-1)(J-1)$ degrees of freedom is

$$x^2 = \sum_i \sum_j \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \tag{5}$$

Where $m_{ij} = (\sum_{j=1}^{J} n_{ij} \sum_{i=1}^{I} n_{ij})/n$.

Fisher's exact test (Fisher 1935 and Finney 1948) estimates the probability of observing a table in which the assumption of association is the same as the assumption of no association. This is done following an iteration process holding row and column marginals fixed. Therefore, if A is a set of all tables with the same marginals as the observed, the estimated probability is defined as:

$$P = \sum_{T \in A} \Pr(T) \tag{6}$$

Appendix B

Table A.1 Estimates the effect of educational level

	Model 1	Model 2	Model 3
Unemployed one year ago	0.402***	0.666***	0.601***
	(0.041)	(0.018)	(0.016)
Inactive one year ago	-5.058***	-2.621***	-2.691***
	(0.065)	(0.032)	(0.030)
Female	-0.672***	0.277***	0.123***
	(0.016)	(0.008)	(0.008)
European	1.726***	0.114***	0.257***
	(0.043)	(0.020)	(0.017)
Third country	0.251***	1.450***	1.425***
	(0.059)	(0.018)	(0.018)
Married	0.377***	0.462***	0.385***
	(0.028)	(0.013)	(0.012)
Urban	-0.176***	0.454***	0.438***
	(0.021)	(0.013)	(0.011)
Central Macedonia	-1.404***	-0.556***	-0.733***
	(0.069)	(0.031)	(0.029)
West Macedonia	1.958***	-10.743	-4.522***
	(0.117)	(98.021)	(0.050)
Epirus	-1.049***	-1.658***	-1.387***
	(0.075)	(0.038)	(0.034)
Thessaly	-1.315***	-1.245***	-1.338***
	(0.102)	(0.037)	(0.035)
Ionian Island	-3.468***	-1.195***	-1.413***
	(0.074)	(0.033)	(0.030)
West Greece	-0.487***	0.474***	0.164***
	(0.086)	(0.035)	(0.032)
Sterea Ellada	-4.864***	-2.277***	-2.387***
	(0.092)	(0.044)	(0.040)
Attica	-0.652***	-0.883***	-0.951***
	(0.066)	(0.029)	(0.027)
Peloponnese	-2.908***	-1.153***	-1.198***
	(0.096)	(0.039)	(0.037)
North Aegean	-2.804***	-0.325***	-0.859***
	(0.076)	(0.039)	(0.035)
South Aegean	-2.248***	-1.854***	-1.680***
	(0.068)	(0.033)	(0.029)
Crete	-1.058***	-0.531***	-0.677***
	(0.067)	(0.030)	(0.028)

Number of children age 0-4	-1.479***	0.607***	0.207***
	(0.027)	(0.014)	(0.012)
Number of children age 5-14	0.293***	0.211***	0.373***
	(0.023)	(0.012)	(0.011)
3.household type	-0.063**	-0.448***	-0.563***
	(0.030)	(0.018)	(0.015)
4.household type	-0.094**	-0.799***	-0.445***
	(0.046)	(0.028)	(0.024)
5.household type	0.710***	1.569***	1.196***
	(0.031)	(0.016)	(0.015)
6.household type	-0.517***	-0.494***	-0.518***
	(0.034)	(0.018)	(0.016)
7.household type	-0.488***	0.224***	0.214***
	(0.057)	(0.023)	(0.022)
Age	-0.485***	-0.211***	-0.224***
	(0.006)	(0.004)	(0.003)
Age square	0.005***	0.002***	0.003***
	(0.000)	(0.000)	(0.000)
Part-time	1.313***	0.519***	0.653***
	(0.029)	(0.012)	(0.011)
Self-employed	1.142***	0.320***	0.063***
	(0.049)	(0.020)	(0.018)
Employee	2.371***	-0.580***	-0.549***
	(0.047)	(0.020)	(0.018)
Assist family business	1.150***	-0.701***	-0.762***
	(0.064)	(0.024)	(0.022)
Permanent contract	-0.130***	0.186***	0.147***
	(0.021)	(0.012)	(0.010)
Tenure	-0.025***	-0.031***	-0.037***
	(0.001)	(0.001)	(0.000)
Small	-0.744***	0.039***	0.142***
	(0.026)	(0.011)	(0.011)
Insured	0.000	0.000	-0.053
	(.)	(.)	(0.054)
Number of employed in the household	0.032*	0.745***	0.630***
	(0.018)	(0.009)	(0.008)
Number of pensioners in the household	-0.054**	0.984***	0.680***
	(0.023)	(0.014)	(0.011)
Number of unemployed in the household	1.441***	0.679***	0.613***
	(0.030)	(0.011)	(0.010)
Number of inactive in the household	-0.184***	0.424***	0.208***
	(0.021)	(0.012)	(0.010)

2012	-0.594***	0.000	-0.266***
	(0.023)	(.)	(0.018)
2013	-1.199***	-0.301***	-0.553***
	(0.030)	(0.014)	(0.023)
2014		-0.043**	-0.346***
		(0.022)	(0.029)
2015		-0.066***	-0.371***
		(0.022)	(0.029)
2016		-0.093***	-0.370***
		(0.022)	(0.029)
Q2: April-June	-0.291***	-0.036***	-0.051***
	(0.016)	(0.010)	(0.009)
Q3: July-September	-0.398***	-0.018*	-0.069***
	(0.016)	(0.009)	(0.008)
Q4: October-December	-0.462***	-0.044***	-0.095***
	(0.019)	(0.010)	(0.009)
1.d1	0.404***		0.064***
	(0.025)		(0.019)
1.d		0.047**	0.117***
		(0.019)	(0.027)
Cut 1	-15.108***	-7.812***	-8.561***
	(0.160)	(0.092)	(0.090)
Cut 2	-14.110***	-6.531***	-7.463***
	(0.160)	(0.090)	(0.089)
Cut 3	-13.593***	-6.219***	-7.155***
	(0.160)	(0.090)	(0.089)
Cut 4	-11.481***	-4.227***	-5.293***
	(0.158)	(0.089)	(0.089)
Cut 5	-9.382***	-2.918***	-3.956***
	(0.155)	(0.089)	(0.088)
Cut 6		-1.371***	-2.400***
		(0.090)	(0.089)
Pseudo R2	0.3354	0.2254	0.1897
Log likelihood	-49344.1	-147150.1	-181118.5
Sample size	293	596	742

Table A. 2 Estimates the probability of training programme

	Model 1	Model 2	Model 3
Worked last year	0.278***	-0.504***	
	(0.088)	(0.069)	
Female	-2.216***	-1.446***	
	(0.104)	(0.064)	
Married	2.318***	1.252***	
	(0.113)	(0.068)	
Number of children in household	2.077***	1.164***	
	(0.110)	(0.058)	
Age	-2.679***	-1.335***	-0.154***
	(0.099)	(0.040)	(0.003)
Age square	0.024***	0.010***	
	(0.001)	(0.000)	
Years of education	-0.747***	-0.399***	-0.286***
	(0.044)	(0.019)	(0.007)
Permanent contract	6.454***	2.076***	-0.340***
	(0.235)	(0.091)	(0.024)
Tenure	0.510***	0.385***	0.121***
	(0.023)	(0.013)	(0.003)
Small	-4.086***	-1.907***	-0.470***
	(0.162)	(0.082)	(0.030)
Number of employed in the household	-2.256***	-0.248***	0.564***
	(0.093)	(0.041)	(0.018)
Number of pensioners in the household	0.035	0.553***	0.731***
	(0.054)	(0.041)	(0.018)
trend	8.939***	1.761***	0.201***
	(0.382)	(0.088)	(0.008)
Q2: April-June	0.049	0.457***	0.097***
	(0.067)	(0.057)	(0.033)
Q3: July-September	0.510***	0.815***	0.276***
	(0.071)	(0.059)	(0.032)
Q4: October-December	0.523***	0.844***	0.068**
	(0.073)	(0.062)	(0.034)
1.d		-6.605***	
		(0.377)	
Constant	6.068***	26.404***	4.619***
	(0.571)	(0.804)	(0.113)
Pseudo R2	0.6699	0.6674	0.4277
Log likelihood	-3631.0	-4707.9	-8482.5
Sample size	445	591	738

Table A. 3 Estimates the probability of looking for a second job

	Model 1	Model 2	Model 3
Worked last year	-16.066***		-4.633***
	(0.888)		(0.141)
Female	10.786	1.972***	6.234***
	(100.819)	(0.419)	(0.219)
Married	4.358***	-1.050***	1.164***
	(0.398)	(0.274)	(0.094)
Urban	-14.647***	-4.623***	-2.713***
	(0.972)	(0.273)	(0.116)
Number of children in household	12.676***	6.117***	2.549***
	(0.784)	(0.319)	(0.091)
Tenure	0.021	0.054***	0.156***
	(0.013)	(0.016)	(0.005)
Age	8.458***	-0.958***	1.536***
	(0.423)	(0.075)	(0.054)
Age square	-0.090***	0.015***	-0.016***
	(0.004)	(0.001)	(0.001)
Years of education	2.625***		0.893***
	(0.146)		(0.030)
Number of employed in the household	-23.589***	-2.820***	-5.343***
	(1.552)	(0.511)	(0.183)
Number of pensioners in the household	6.680***		1.352***
	(0.386)		(0.051)
Number of unemployed in the household	-5.964		3.449***
	(100.820)		(0.104)
trend	0.890***	0.639***	0.460***
	(0.128)	(0.209)	(0.050)
1.d		-7.946***	-4.772***
		(0.613)	(0.160)
1.d1	-19.810***		-6.961***
	(1.218)		(0.246)
Constant	-231.402**	-1.546	-53.470***
	(101.490)	(1.537)	(1.691)
Pseudo R2	0.7201	0.9358	0.7803
Log likelihood	-1328.8	-442.7	-2329.0
Sample size	293	596	738

Table A. 4 Estimates the probability of a full-time job

	Model 1	Model 2	Model 3
Unemployed one year ago	-1.961***	0.014	-0.077***
	(0.130)	(0.029)	(0.028)
Inactive one year ago	0.000	-2.524***	-2.229***
	(.)	(0.044)	(0.040)
Female	-6.644***	-1.710***	-1.483***
	(0.159)	(0.023)	(0.020)
European	0.000	0.433***	0.882***
	(.)	(0.047)	(0.045)
Third country	0.000	-0.499***	-0.447***
	(.)	(0.027)	(0.026)
Married	-1.062***	-1.004***	-0.902***
	(0.084)	(0.019)	(0.017)
Urban	4.316***	0.308***	0.415***
	(0.121)	(0.016)	(0.015)
Number of children age 0-4	0.000	-0.479***	-0.262***
	(.)	(0.024)	(0.022)
Number of children age 5-14	-0.934***	-0.354***	-0.243***
	(0.060)	(0.018)	(0.016)
Age	0.472***	0.234***	0.157***
	(0.037)	(0.006)	(0.006)
Age square	-0.008***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.000)
Years of education	-0.721***	-0.243***	-0.250***
	(0.022)	(0.004)	(0.004)
Permanent contract	-0.772***	0.558***	0.353***
	(0.057)	(0.013)	(0.012)
Tenure	-0.130***	0.027***	0.018***
	(0.004)	(0.001)	(0.001)
Small	5.467***	0.234***	0.690***
	(0.124)	(0.023)	(0.019)
Number of employed in the household	-0.900***	-0.211***	-0.138***
	(0.059)	(0.012)	(0.011)
Number of pensioners in the household	0.146**	-0.419***	-0.234***
	(0.072)	(0.016)	(0.015)
Number of unemployed in the household	0.000	-0.180***	-0.114***
	(.)	(0.019)	(0.017)
2012	-2.267***	0.000	-0.700***
	(0.086)	(.)	(0.036)
2013	-3.653***	-0.070***	-0.881***
	(0.128)	(0.026)	(0.045)

2014		0.228***	-0.795***
		(0.041)	(0.055)
2015		0.380***	-0.655***
		(0.042)	(0.055)
2016		0.919***	-0.135**
		(0.041)	(0.055)
Q2: April-June	0.231***	0.448***	0.345***
	(0.053)	(0.018)	(0.017)
Q3: July-September	0.193***	0.770***	0.637***
	(0.053)	(0.019)	(0.017)
Q4: October-December	-1.266***	0.449***	0.190***
	(0.067)	(0.019)	(0.017)
1.d1	0.198**		-0.339***
	(0.083)		(0.035)
1.d		-0.229***	-0.426***
		(0.035)	(0.050)
Contant	12.493***	1.565***	3.727***
	(0.695)	(0.147)	(0.143)
Pseudo R2	0.7600	0.3514	0.3260
Log likelihood	-4059.5	-28652.8	-33484.6
Sample size	228	592	738

Table A.5 Estimates the probability of holding a permanent job

	Model 1	Model 2	Model 3
Unemployed one year ago	0.736***	-0.302***	-0.164***
	(0.047)	(0.019)	(0.018)
Inactive one year ago	0.000	-0.315***	-0.799***
	(.)	(0.035)	(0.031)
Female	-0.170***	0.191***	0.151***
	(0.016)	(0.009)	(0.008)
European	-1.547***	-0.864***	-1.091***
	(0.051)	(0.023)	(0.020)
Third country	0.000	-0.320***	-0.276***
	(.)	(0.020)	(0.020)
Married	0.057***	-0.235***	-0.118***
	(0.020)	(0.010)	(0.009)
Urban	0.783***	0.784***	0.688***
	(0.019)	(0.011)	(0.010)
Number of children age 0-4	0.000	0.715***	0.765***
	(.)	(0.017)	(0.015)
Number of children age 5-14	-0.296***	0.251***	0.263***
	(0.023)	(0.012)	(0.011)
Age	0.095***	0.074***	0.050***
	(0.006)	(0.003)	(0.003)
Age square	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Years of education	0.093***	-0.026***	-0.017***
	(0.005)	(0.002)	(0.002)
Part-time	-0.375***	-0.762***	-0.560***
	(0.028)	(0.014)	(0.013)
Tenure	0.026***	-0.005***	0.004***
	(0.001)	(0.001)	(0.000)
Small	-0.593***	-0.395***	-0.519***
	(0.032)	(0.014)	(0.013)
Number of employed in the household	0.101***	0.143***	0.024***
	(0.014)	(0.007)	(0.007)
Number of pensioners in the household	-0.004	0.619***	0.425***
	(0.014)	(0.011)	(0.009)
Number of unemployed in the household	0.563***	0.212***	0.128***
	(0.031)	(0.010)	(0.009)
Number of inactive in the household	-0.275***	0.562***	0.369***
	(0.015)	(0.010)	(0.009)
2012	0.622***	0.000	0.395***
	(0.030)	(.)	(0.025)

2013	0.632***	-0.282***	0.158***
	(0.038)	(0.016)	(0.030)
2014		-0.424***	0.101***
		(0.028)	(0.038)
2015		-0.520***	-0.014
		(0.028)	(0.038)
2016		-0.281***	0.202***
		(0.028)	(0.038)
Q2: April-June	0.201***	-0.164***	-0.114***
	(0.020)	(0.012)	(0.011)
Q3: July-September	0.311***	-0.196***	-0.120***
	(0.020)	(0.011)	(0.010)
Q4: October-December	0.676***	0.012	0.136***
	(0.024)	(0.012)	(0.011)
1.d1	-0.801***		-0.218***
	(0.033)		(0.026)
1.d		0.014	-0.253***
		(0.024)	(0.036)
Constant	-1.966***	-0.838***	-0.349***
	(0.169)	(0.084)	(0.077)
Pseudo R^2	0.2039	0.1289	0.1291
Log likelihood	-25960.8	-77368.6	-91376.4
Sample size	255	592	738

(This page intentionally left blank)

5. Driving Instructors in the UK

5.1 Introduction

It was not until the introduction of the driving test in 1935 that any serious attempt was made to coordinate the ways in which people learned to drive. Even then, however, the state did not see the need for a parallel system to formalize eligibility to teach individuals driving skills. So for many years there were no legal controls over the training and qualifications of those who taught driving. The first initiative came from the Motor Schools Association and the Royal Automobile Club (both independent associations formed by driving schools owners) who also in 1935 produced their own independent registers of qualified driving instructors and schools. Each organization outlined its own version of the basic skills that were needed to join the register. The register of approved driving instructors (ADIs) was finally approved by Parliament in 1964 but remained voluntary until 1970. During that time, on various occasions the state attempted to persuade the industry to formalize the standards and put in controls, without success. However, it was not until the driving lessons industry grew substantially that driving instructor training was formalized and membership of the register of approved driving instructors became compulsory.

In practice, this meant that any person giving paid instruction in the driving of a motor car whose name is not on the register is guilty of an offence. As such, the evolution of regulation in the occupation can be understood as a shift from voluntary membership of a register belonging to a professional body to legally enforced state regulation in the form of specific training requirements to practise (as evidenced by one's membership of the Approved Driving Instructor register or ownership of a trainee licence). As we shall show below, however, in reality a dual market of fully trained and trainee driving instructors is in operation. From the learner's perspective, it is worth pointing out that driving lessons can be provided by driving instructors or any other person that has a valid driving license (e.g. family or friend). Therefore, there is no legal requirement to receive paid tuition from a driving instructor when learning to drive.

More recently, various changes have taken place with regards to requirements for regular periodic assessments that instructors have to undergo as well as proposals to reform the occupation in terms of training requirements to practise. In this study we assess the impact of these proposals on outcome and value-added types of indicators.

5.2 Regulatory Context

The Driver and Vehicle Standards Agency (DVSA) is the regulatory body that manages the register. The regulatory framework also covers the qualification process that prospective practitioners must pass to be included on the register. In order to be allowed on to the register applicants must first meet 'fit and proper' conditions. This includes passing a criminal record check, followed by success in a three-part test, administered by the DVSA. This test consists of:

- Part 1: a computer-based driving/instructional theory and hazard perception test
- Part 2: a practical driving ability test
- Part 3: a practical instructional ability test

Once a candidate passes the first two tests, they can then apply to the regulatory body for a trainee licence which allows them to give paid driving instruction without supervision as a means to gain practical experience in giving on-road instruction before taking the Part 3 qualifying test. The cost of taking all the required tests, obtaining a trainee licence and joining the ADI Register is approximately £750, of which about £111 is towards the cost of the Part 3 test (i.e. the cost of moving from a trainee to an ADI). Registration is for four years, after which there is a renewal application. The total fee covers the issue of the ADI licence and other administration costs, the cost of the check test and a further Disclosure and Barring Service (DBS) check prior to renewal of their licence at the end of the four-year period.

Only those on the register (ADIs), or potential ADIs that have been granted a trainee licence by the regulator (hereafter "trainee instructor"), could give in-car driving instruction for money. In practice, the profession of driving instructor in the UK operates as a dual market of certified ADIs (fully qualified to provide tuition) and non-fully qualified (trainee) driving instructors also legally allowed to give tuition. The difference between the two is that the former have passed Part 3 of the test, whereas the latter have only passed Parts 1 and 2. Therefore, in theory, it is left to the consumer to choose between the two service providers. Although instructors are legally required to inform the students of their status (through displaying their badge when they are giving tuition),

anecdotal evidence collected by the Driver and Vehicle Licensing Agency (DVLA) showed that learners tend not to be aware of whether they are receiving tuition by a fully approved (ADI) or a trainee instructor, and the tuition fee they were paying was often the same regardless of the instructor's status.

From a comparative perspective, the profession of driving instructor is regulated to some extent in all 28 member states, albeit to various degrees. With regards to the scope of reserved activities we find a distinction with regards to whether the instructor can offer both theoretical and practical training, with some member states such as Austria and Italy having two different occupations and different training requirements for each (EU Report 2016). In the UK, driving instructors are responsible for teaching students the practical side of driving and their input in the theoretical part of the test is minimal, if any, and only in the form of clarification. In comparison to other member states, the UK has arguably one of the least restrictive regimes for entry to the occupation. As shown above, while stipulations exist regarding the type of training required, the corresponding exams and subsequent testing of competence, there are no preconditions regarding primary or secondary education, the length of required training and mandatory traineeship. This is in sharp contrast with some European counterparts where it is not uncommon to find requirements for post-secondary education (e.g. Finland, Greece and Romania) and lengthy training provisions (e.g. 300 hrs in Belgium; 630 hrs in France; two years in Estonia, Greece and Ireland). Turning to the justification offered for regulation, road safety and consumer protection are put forward as the overriding reasons for the regulation of the occupation by the majority of member states including the UK. Interestingly, from a comparative perspective, the UK in the period 2010–2013 had the second lowest number of fatalities in the EU after Sweden (CARE 2014). To the extent that the level of regulation of driving instructors has (among other things) some causal impact on car accidents, then it would seem that the level of regulation prior to the reform was satisfying the road safety dimension of its existence.

5.3 The Reform Proposals

In 2013, the government began a review of the existing regulatory framework. Concerns were raised regarding the extent to which after becoming qualified as an ADI, instructors started deviating from the guidance provided by the Driver and Vehicles Standards Agency (DVSA). For this reason, a new 'standards check' was introduced in 2014 which requires ADIs to undergo further assessment of their instructional methods a least once during the four-year period that their licence is valid, otherwise risking being struck off the register. The process involves an assessor observing the ADI giving tuition and providing a grade along a list of criteria. The grade varies from A (indicating a high standard of driving instruction) to B (indicating a sufficient level of competence) and finally FAIL (indicating unsatisfactory driving instruction performance). The new system replaces the older version of a check through 'role-play' between examiner and instructor, and was largely seen as a tightening of the right-to-practice regulations in the industry.

Further concerns were raised regarding the low completion rates among trainee driving instructors as evidenced by the large proportion of individuals who had completed the minimum training required by the regulations to provide tuition as a trainee instructor (i.e. Parts 1 and 2) had not progressed to complete their training (i.e. Part 3) and become ADIs. There was also evidence of high rates of failure among those trainee instructors in the Part 3 test on instructional ability, despite multiple attempts. As a result, the market was populated by a mix of fully trained instructors and trainee ones (many with questionable ability to ever qualify), both of which were allowed to offer identical services and often at a similar price to consumers. Overall, this regime was seen to be compromising the quality of tuition received by prospective drivers, not incentivising trainee driving instructors to fully qualify, and not providing transparency to the public, thus driving down standards in the market. As a result, the government put forward proposals to reform the trainee licence scheme. In addition to improvements in the qualification tests, the proposals stipulated that trainee licence holders cannot provide driving lessons for money unless they are accompanied by a fully licensed ADI for at least part of the duration of the total tuition. As a result of these proposals, the regulation regime which entered into force in April 2014 by the Motor Cars (Driving Instruction) Amendment Regulations tightened the educational requirements to perform the activity of providing paid tuition.

5.4 Literature Review

There is a marked absence of studies on the regulation of driving instructors in general and on its impact on quality in particular. One exception is a study on the deregulation of Portuguese driving schools in the late 1990s (Seim and Vitorino 2011). Prior to the reform, the operation of driving schools was heavily regulated by the state and took the form of restrictions in the issuing of licences based on the population size and minimum distance requirements from competitors, as well as ceilings on fees. These strict requirements were partially lifted in 1998, which the preliminary results show had a direct impact on service availability as evidenced by a 116 per cent increase in the number of driving schools by 2010. Using student performance in driving exams the authors are currently working on estimating the effect of the reform on quality of service provision. Avrillier et al. (2010) explore the effect of a demand shock (in the form of the abolition of compulsory conscription in France, which automatically provided young French men with a driving licence) on the heavily regulated French driving school industry. Their results demonstrate that due to the strict entry restrictions (long and costly to become an instructor) the industry could not respond by quickly increasing supply (as it would have done if it was competitive) and therefore service availability suffered. The authors also find an increase in the price of driving lessons (even when the industry finally started growing) and a reduction in the number of individuals getting their licences (which can either be attributed to better screening of low-skilled students or to a deterioration of the quality of tuition offered by driving instructors).

Turning to the wider literature, we can find some studies that parallel the rationale and quality measure proposed here. In particular, several studies have looked at the relationship between different regulatory regimes within the teaching profession and their effect on quality. While these studies have not used reform proposals to capture the effect of interest, they are of relevance not least due to their use of a generally recognized measure of 'quality' in education, namely student test scores. Overall, we deem these to be of interest given that driving instructors are also engaged in teaching and therefore are analogous to the profession of teachers. An early study by Kleiner and Petree (1988) examines the impact of variations in the restrictiveness of licensing on educational performance using state-level data for the 1972–1982 period. Their measures of performance include the average level of the SAT and ACT standardized test scores and the proportion of students who graduate high school; their results are mixed and dependent on the estimation technique. With regards to the standardized scores, a positive relationship is found between regulation and standardized test scores, while the reverse is true for graduation rates. In a similar vein, Angrist and Guryan (2004) studied the relationship between testing prospective teachers (to certify they meet minimum standards) and subsequent performance by students. Their evidence is also mixed in that while students of accredited teachers do better in the Praxis test, teacher testing has a negative effect on the average SAT score of a teacher's undergraduate institution. Kane et al. (2005) using panel data also find little difference in student achievement between students taught by licensed, unlicensed and certified teachers in the same schools in New York, while Kane and Staiger (2005) using a quality measure based on improvements in student achievement again find no evidence that licensed teachers are more effective.

5.5 Data

The Department of Transport (DoT) and the Driver and Vehicle Licensing Agency (DVLA) produce aggregated data on driving test pass rates for learner drivers and instructors publicly available on the UK government's website in a series of Excel files (see Appendix A). The raw files are not organized in a consistent manner in two main ways. First, the geographical unit. Some are organized by Driving Test Centre (DTCs, a local area unit), while others by postcode area where the exam took place (a slightly higher-level geographical unit). While DTCs and postcode areas may coincide, the match is not always perfect. Second, the time unit. Most files contain data by month, but others are annual. There is also some missing data in the files. There are two main sources of missing data. First, differences between files in when the data runs from and ends. For example, while the data on the driver pass rates across geographical units runs from April 2007, data on the proportion of qualified/trainee instructors within each unit only runs from March 2012. Second, a small but not insignificant number of DTCs close over time, meaning we potentially lose some panels before the reform. To remedy this, we aggregate the DTC geographical units up to postcode level, maximizing the number of observations within panels. We are thus left with a data set of observations (time points) nested within postcode areas. The exact sample sizes depend on which variables are included in the models. We supplement this data source with historical data on employment and cost of driving lessons from the UK Labour Force Survey and the Office for National Statistics.

Measures of Quality

Within these files, we extract four measures of quality, relating to both quality of instructor and quality of service. First, driving instructor quality within postcode-time cells is proxied by driving instructor practical test pass rates (Part 2 test). Second, quality of service is proxied by practical

test pass rates of learner drivers. This is supplemented with two further indicators of service quality: practical test pass rates at first attempt and practical test pass rates at the first attempt with zero faults (a perfect test score).

Our aims in this study are as follows. First, we would like to establish whether the reforms affected the quality of instructors as measured by their success in the Part 2 test. The reforms would have had a positive effect on the quality of instructors if it deterred those with sub-standard skills from entering the occupation. As such, we would expect to see an increase in the pass rates of instructors in the Part 2 test. Second, using the three indicators of learner outcomes, we explore whether the quality of service offered changed during the period of observation. Our analytical approach follows similar studies on the impact of licensing on teachers (see the literature review section) where teacher quality is derived from variables relating to student performance in relevant tests. Keeping in mind the caveats of this approach (e.g. mediating factors such as student ability), if the quality of instructors improved after the reforms, then we would expect these indicators to also improve. Third, we compare whether the reforms affected the quality of learner drivers of cars vis-à-vis their motorcycle counterparts (where no reforms were proposed) as measured by the practical test pass rates for each group. If the effect was positive, then we would expect to find no difference between the two observation periods in the pass rates of motorcycle learners. The quality measures adopted to achieve these three aims can be understood as 'measures of outcomes' specific to the occupation.

Fourth, we examine whether the reform proposals discouraged sub-standard trainee instructors from entering the occupation or/and encouraged them to exit it, as well as whether it incentivized trainees to take the tests so that they can fully qualify as fully licensed trainers (i.e. to switch from trainees to ADIs). We supplement this analysis with (a) labour force survey data and explore changes in total employment during this period within this occupational group[83] and (b) changes in the price of driving lessons. We take all these measures to serve as more indirect indicators of quality and we group them under the 'value-added' indicator of service availability and access.

Our fifth and final test takes a different approach to quality by looking at post-qualification outcomes, namely road accidents. Here we can derive one indicator from the available data (namely the number of road accidents within different postcodes) and assume that if quality of driving instruction improved then we would expect to see a fall in all the accident rates. We treat this final part of our analysis with extreme caution given the difficulty in isolating the effects of regulation from the variety of factors that can affect accident rates in a given area, a theme that we will return to in the conclusion. As before, the quality measures used here represent 'measures of outcomes' specific to the occupation.

Overall, although not perfect, we believe our approach explores a plethora of quality indicators that taken together provide a comprehensive test of how the reform proposals might have affected these outcomes.

5.6 Methodology

The analysis proceeds in five steps. First, we examine descriptive trends in the numbers of trainees and fully qualified instructors over time and descriptive trends in several quality indicators, paying particular attention to trends before and after the reform. Second, we switch to a fixed-effects estimation approach to more robustly parse out whether the regulatory changes had any detectable effects on measures of quality of instructors and service. We do this by estimating fixed-effects linear regressions with postcode area fixed-effects. That is, all time-invariant differences across postcode areas are controlled. To identify the differences in quality before and after the reform proposals, we create a dummy variable coded 1 if an observation was post-April 2014 and 0 if the observation occurred prior to this time. The coefficient on this variable shows the average difference in quality (or whatever outcome) between the two time periods within each postcode area, which may be (cautiously) interpreted as the effect of the reform. This model is extended by including a linear time trend to take into account general trends common across postcode areas in the outcome of interest (e.g. the general rise in practical pass rates over time). In other specifications, we extend the model by including month and year fixed-effects (individually and together). The results are almost identical to the ones reported, but introduced high levels of collinearity, so we do not report them.

[83] Analysing trends in enterprise creation and exit rates would not be appropriate in the case of driving instructors as, according to the Labour Force Survey, 90 per cent of driving instructors are self-employed.

In the third step, we extend the fixed-effects models used in the second step by examining variation in trends pre- and post-reform within postcode areas by whether postcode areas fell into the top, middle or bottom tercile in terms of the share of instructors who were trainees prior to the reform. This is done by interacting the reform dummy with dummies indicating which share of trainee tercile a postcode area is in. Since this is time-invariant, no main effect is estimated. However, the interaction informs us on how trends before and after the reform vary according to initial levels of the share of trainees. This is a particularly informative analysis as we expect the reform to affect both instructor and learner outcomes through its effect on the ratio of trainees to fully qualified instructors. We present the results of this part of the analysis in graphs as opposed to presenting large tables of numbers.

In the fourth step, we take a difference-in-differences (DiD) approach focusing on a single indicator of quality: practical test pass rates. DiD estimates the effect of a 'treatment' on a 'treatment group' relative to a 'control group'. In our case, the treatment is the reform proposals that entered into force in April 2014 that only affected the tuition of learner drivers: the treatment group. We use learner motorcyclists as a control group for the reform which should have in theory only affected the quality of tuition received by learner drivers (no reforms in the professional regulation status of motorcycle riding instructors occurred over the period). For this part of the analysis, we pool data on practical test pass rates of drivers and motorcyclists together at the postcode area by quarter (since motorcycle pass rate data is only available on a quarterly basis rather than monthly). We restrict the analysis to the period from 2009 (since motorcycle practical test pass rates are only available from 2009, whereas driver practical pass rates are available from 2007). We control for postcode area fixed effects. The DiD coefficient estimates the average change before and after the reform in practical test pass rates for learner drivers within postcode areas compared to the average change before and after the reform for learner motorcyclists within postcode areas. For this part of the analyses we only focus on practical test pass rates across postcode areas as this is the only quality measure we have that is comparable across learner drivers and learner motorcyclists.

In the fifth and final step, we shift to much broader measures of quality, exploring trends in accident rates before and after the reform. As with much of the other analysis, our focus is on differential trends before and after the reform according to pre-reform shares of trainee instructors. This analysis is much more speculative than our other analyses given that learner and newly qualified drivers make up a small fraction of overall traffic. Moreover, there are plenty of confounding and complicating factors such as accidents potentially happening far away from where a learner driver took their lessons, not to mention differences in rural and urban areas, weather and road conditions, demography, whether the accident involved a car or other mode of vehicle (e.g. lorry, motorcycle), etc. Nonetheless, we still present this as a final piece of evidence, albeit with these caveats in mind.

5.7 Results

Descriptive Trends

We begin by describing key aggregate trends in the data, beginning with examining the number of instructors in Figure 1. As can be seen, there has been a sharp decline in the number of trainee instructors and a smaller decline of qualified instructors (ADIs) over time. Overall, the number of trainees is much smaller than the number of qualified instructors. The proportion of instructors that are trainees has fallen relatively dramatically from around 12 per cent in March 2012 to around 3 per cent by March 2017. Examining the timing of the fall, it appears there were some anticipatory effects of the reform proposals (which became operational in April 2014). It is possible, for example, that trainee instructors who did not feel competent enough to pass the Part 3 exam and qualify as ADIs exited the market, leaving only the more competent ones active. Unfortunately the data does not go further back in time, so we are unable to provide a longer-term picture to put this fall into context. Nonetheless, on the surface, it appears the reform proposals have had an effect on the composition of instructors which translates to fewer driving instructors in the market. This trend is also confirmed when looking at Labour Force Survey data for the occupational code 'driving instructors' (SOC code 8215). As can be seen in Figure 2, the proportion of self-employed driving instructors in the labour market (i.e. the majority of individuals in this occupational group) has been declining in anticipation of the reform proposals and after, while aggregate self-employment has been steady during the same period.

To get a handle on the extent to which the reform may have affected the distribution of the share of trainees across postcode areas, Figure 3 plots the share of trainee instructors by whether postcode areas fell in the bottom, middle or top third of the distribution of shares in trainee instructors for the period prior to the reform. As Figure 3 makes clear, there were obvious disparities in the share of instructors who were trainees between postcodes prior to the reform. In the period following the reform, the disparities in the share of instructors who are trainees across postcode areas all but closed. While the average share of instructors who were trainees across postcode areas fell from 6.8 per cent prior to the reform to 2.0 per cent afterwards, the standard deviation also fell from 14.1 per cent to 3.4 per cent.

Turning to trends in the indicators of instructor quality, there was a four percentage point increase in the pass rate of trainee instructors in Part 2 (the driving practical) between March 2012 and March 2017, indicating that the average quality of instructors (at least new entrants and as judged by their success in this exam) has been getting better over time (Figure 4). The rate of upward trend in pass rates seemed to increase after April 2014 (see the fitted values which fit a quadratic trend term). One possible explanation of this is that in anticipation of the reform which would make entry and operation in the market more stringent, the low-quality prospective driving instructors were discouraged from entering the occupation. Overall, we can certainly rule out explanations relating to the examination becoming easier or simpler, or explanations relating to any improvements in the training and support prospective driving instructors received between the two periods, as we know that the procedure and requirements remained the same.

In Figure 5, we examine trends in instructor pass rates by whether postcode areas fall into the bottom, middle or top third of the distribution of shares in trainee instructors for the period prior to the reform. If the reform increased the average quality of trainee instructors (as measured by pass rates on the instructor practical tests), then we might see steeper increases in pass rates in those areas that had a higher share of trainee instructors in the first place (as their higher share in those areas would translate to higher pass rates). While we find definite differences in pass rates prior to the reform proposals, with pass rates being the lowest in areas with the highest share of trainee instructors, we find that pass rates seem to improve the most in areas that had lower shares of trainee instructors to begin with.

Turning to our proxies for service quality for learners, we can see general upward trends in total practical test pass rates (Figure 6) – the part of the qualification process that is affected by driving instructors. After the reform, trends in pass rates slowed and levelled out. By contrast, theory tests have been steadily falling, likely partly because of stricter procedures. Since the theory test must be taken before the practical test, one possible reason for the upward trend in practical pass rates might be due to elimination of very low ability would-be drivers who cannot pass the theory test as the procedures became stricter. In Figure 7, we explore two more fine-grained measures of performance in the practical test: pass rates at first attempt (as learners can take the test multiple times if they fail) and zero faults pass rates at first attempt (obtaining a perfect score on their first attempt). Here too, we find similar trends as with the overall pass rates for the practical test: a general upward trend that levels out after the reform proposals. On the face of it, then, it appears the reform has had no visible effect on the quality of the service as measured by performance on the practical test by learners.

Finally, we explore whether trends in practical test pass rates over time vary according to the share of instructors who were trainees prior to the reform (Figure 8). While we find clear differences in pass rates according to the share of instructors who were trainees, with postcodes in the bottom third having significantly lower pass rates than all of the postcodes, we find broadly similar over time trends as in Figure 5. Thus even in this more fine-grained approach, we find little evidence that the reform resulted in higher pass rates for learners, even though we know from Figure 2 that the share of trainee instructors fell.

Figure 1. Number of qualified and trainee driving instructors over time

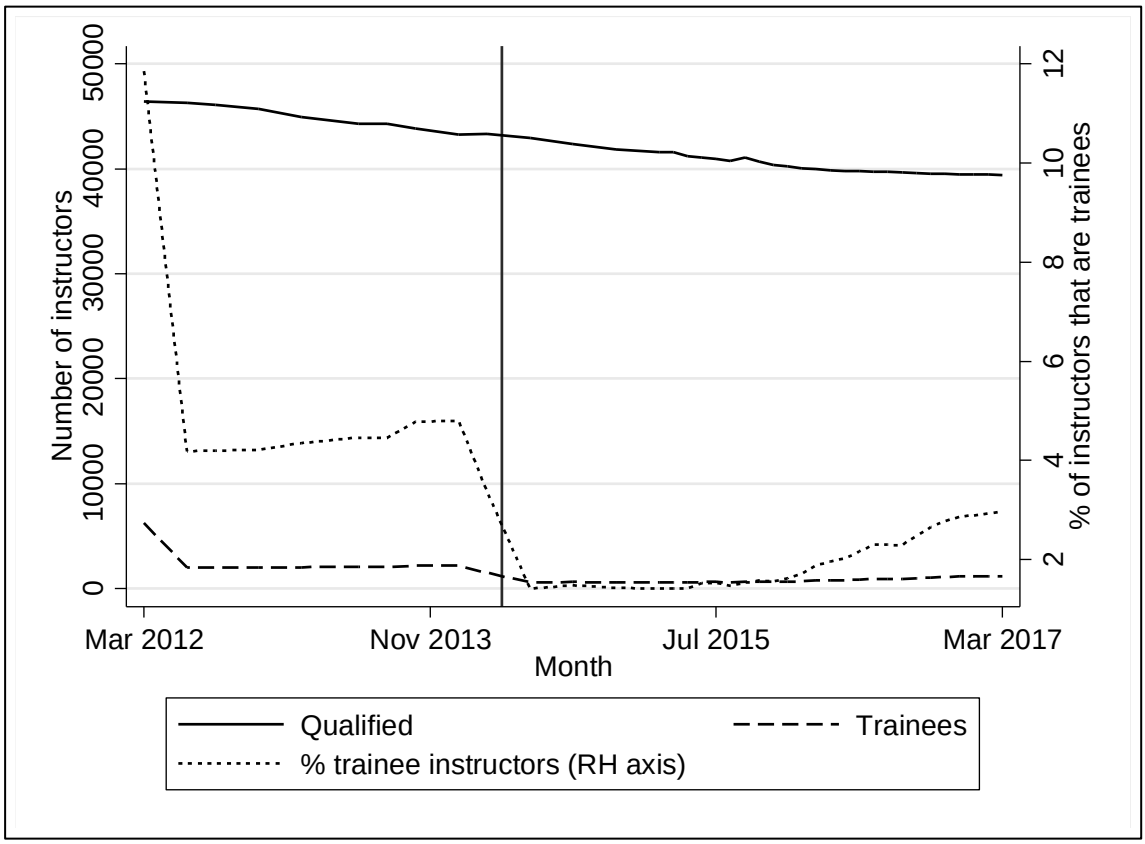


Figure 2: Driving Instructors in the labour market over time (Labour Force Survey Data)

Figure 3. Share of trainee driving instructors by pre-reform trainee instructor share tercile

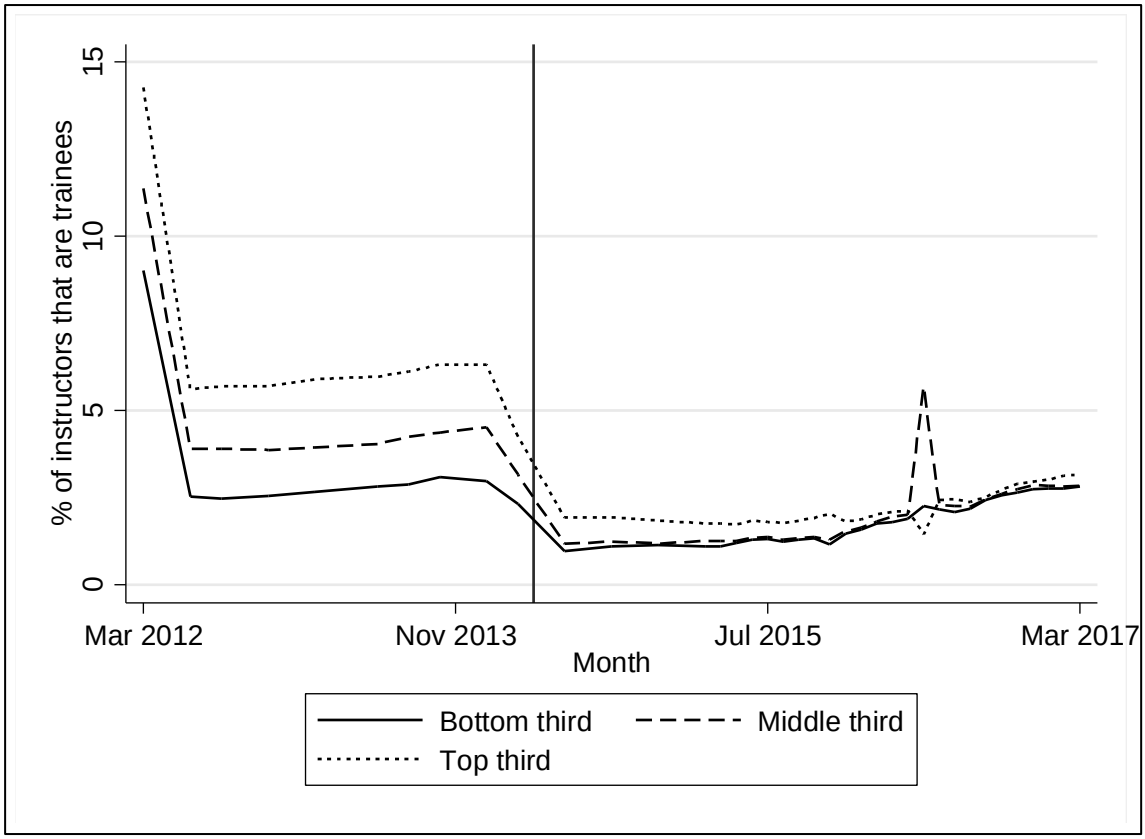


Figure 4. Instructor practical test pass rates over time

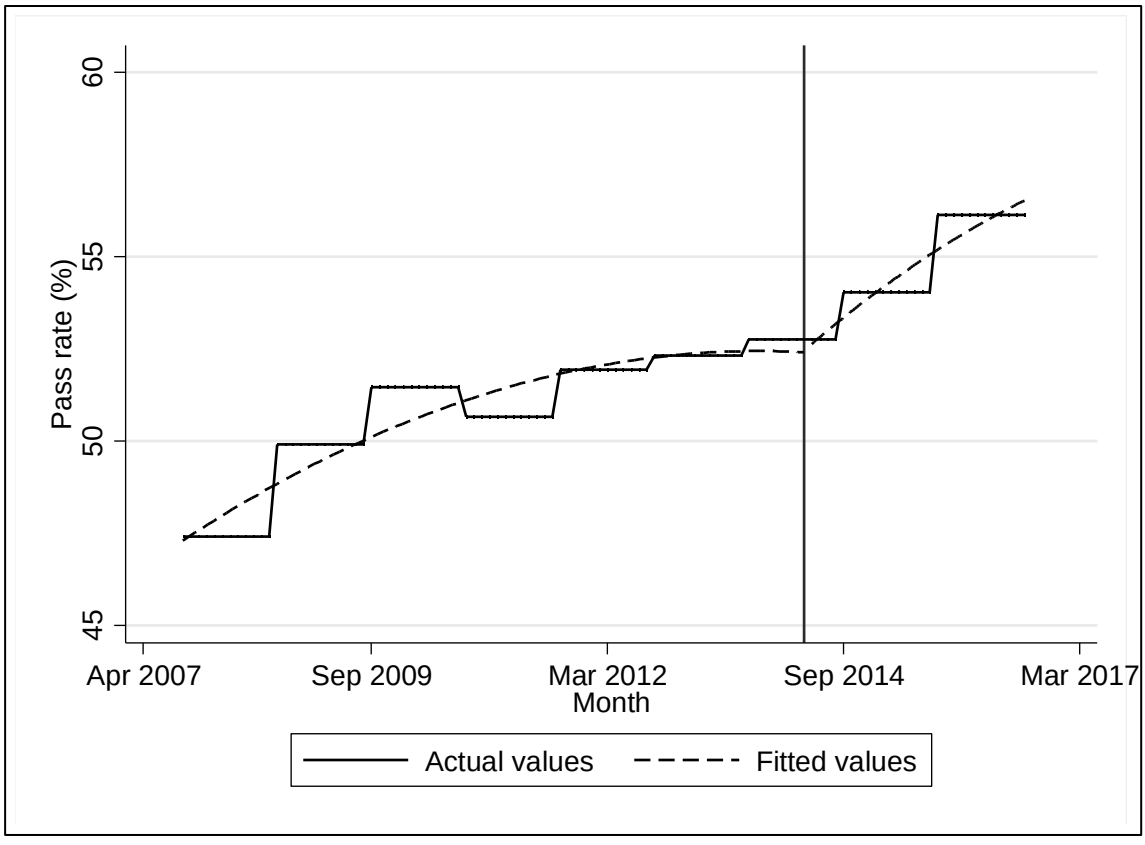


Figure 5. Instructor practical test pass rates by pre-reform trainee instructor share tercile

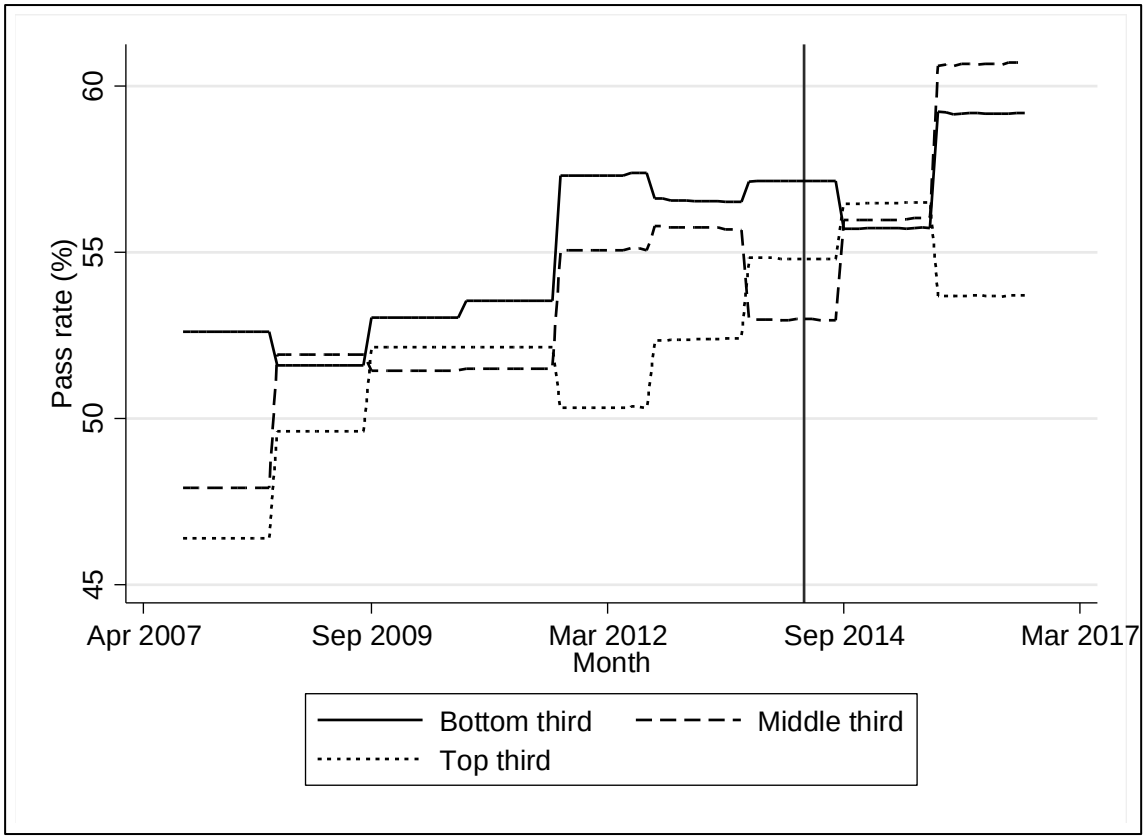


Figure 6. Learner practical and theory test pass rates

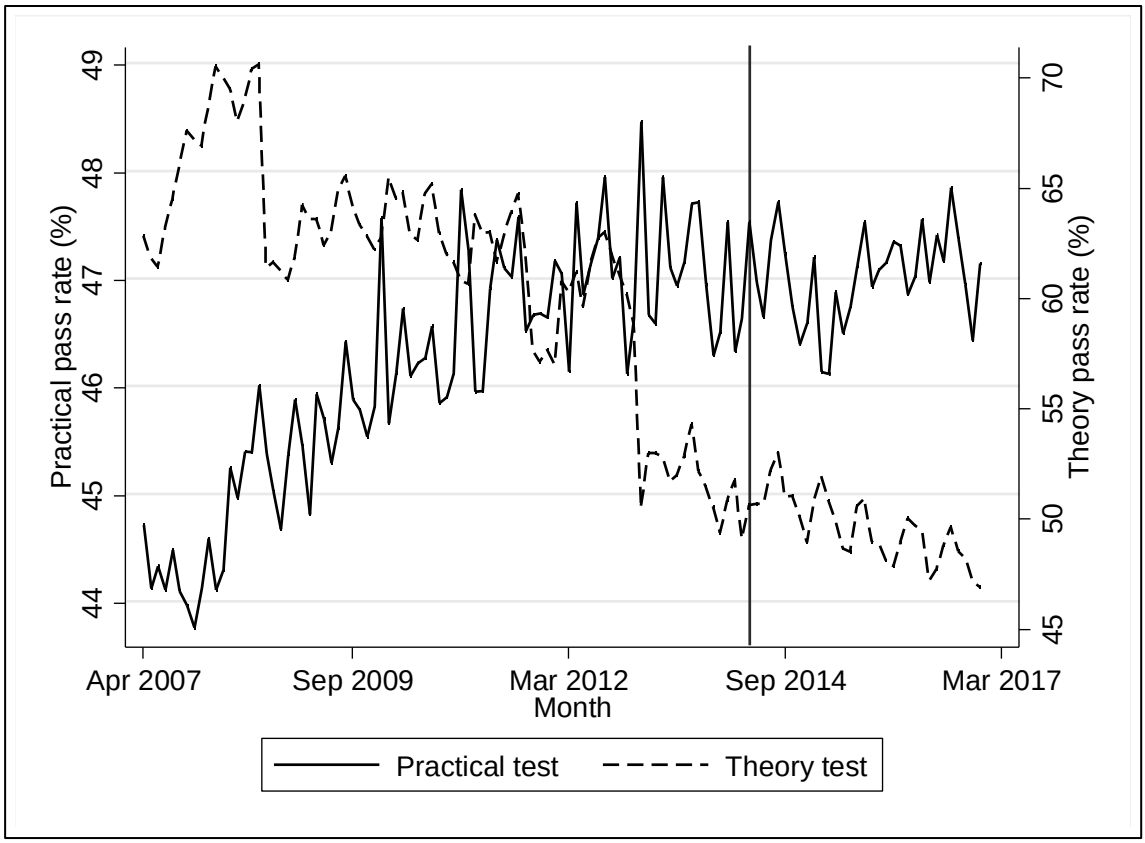


Figure 7. Learner practical test pass rates – first attempt and zero faults

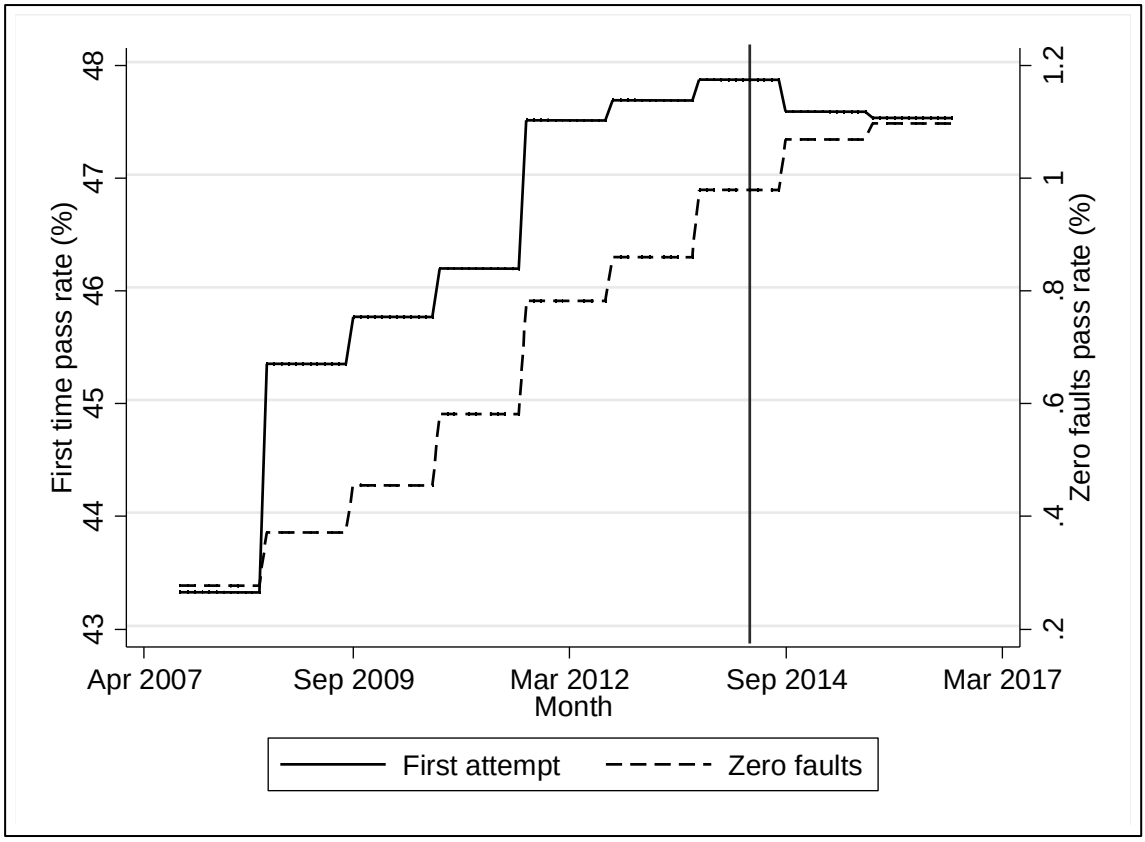
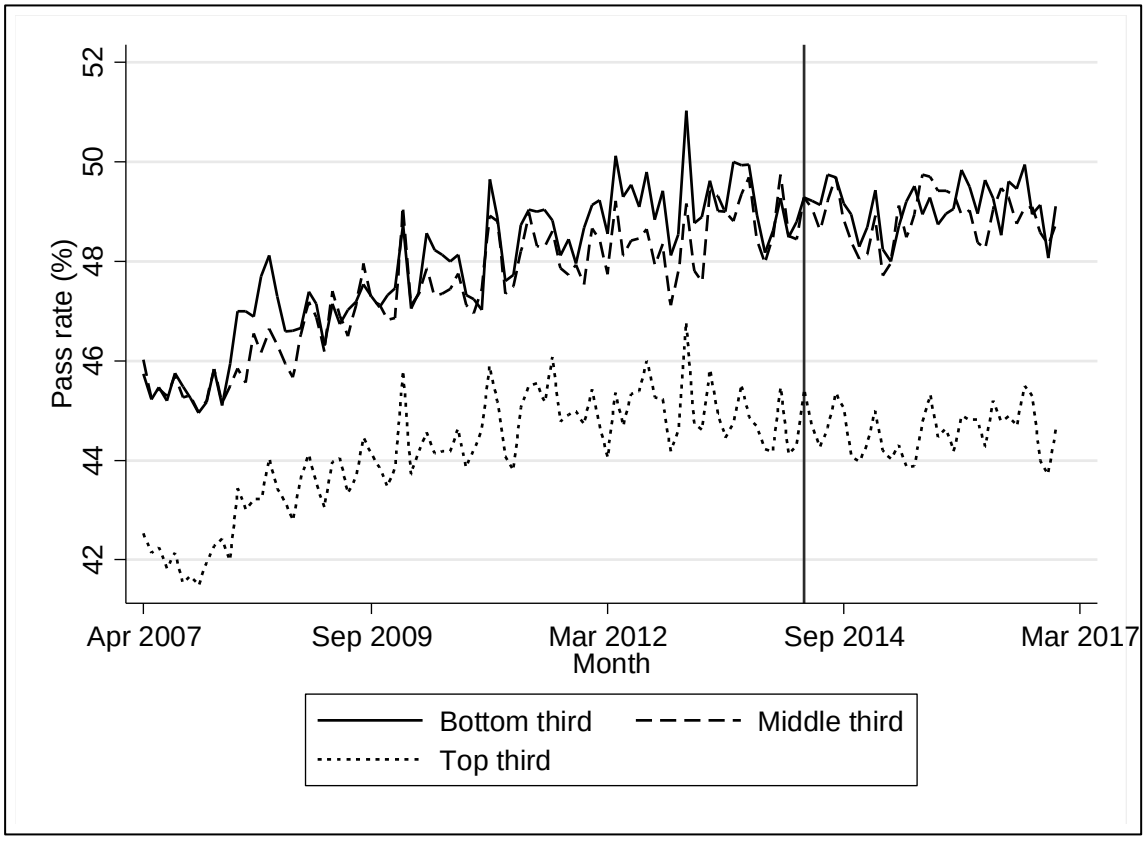


Figure 8. Learner practical test pass rates by pre-reform trainee instructor share tercile



Fixed Effects Estimates

We now switch to a multivariate approach estimating the two equations mentioned earlier for each of the measures of quality. An advantage of the fixed-effects approach is that the coefficients can be interpreted as differences in a given quality indicator before and after the reform within a given postcode area, stripping out all time-invariant factors across areas. In terms of exploring the effect of the reform on instructors, we examine four outcomes (which relate to our previous descriptive analysis). First, the number of trainee instructors. This is a proxy for the number of people wanting to enter the profession as an ADI. If this decreased after the reform, it could signify that new and possibly lower-skilled individuals were discouraged from entering the profession or decided to exit due to concerns about their ability to qualify (i.e. pass the Part 3 test). Second, the number of Part 3 tests taken. If this decreases, it could also signify discouraging new and low-quality entrants.[84] Third, the pass rate of instructors taking the practical test (defined as the percentage of passes over the number of tests taken). If the pass rate increases, it may signify an increase in quality of new instructors after the reform. Finally, we examine the number of qualified instructors before and after the reform proposals.

In terms of the effect on learner drivers we examine another four outcomes. First, the number of practical tests taken by learner drivers. If this falls after the reform proposals, it may indicate a fall in the access to instructors, and so fewer people are taking the test within a given time period. Second, the pass rate of the practical test (defined as the percentage of passes over the number of tests taken). This is a straightforward indicator of the quality of tuition. Third, pass rate at first attempt. Finally, the proportion of passes at the first attempt with zero faults (a perfect score). These latter two are also straightforward indicators of the quality of tuition. We take the natural logarithm of all dependent variables that are not measured in percentage terms. Descriptive statistics of all dependent variables are presented in Table 1.

Table 1. Descriptive statistics of dependent variables

	Mean pre-reform	Mean post-reform	Difference in means (t-test)
Number of trainees	2.418	1.492	0.926***
Number of instructor tests	4.679	3.873	0.806***
Instructor test pass rate	52.459	57.108	-4.648***
Number of qualified instructors	5.363	5.415	-0.053
Number of practical tests	6.768	6.754	0.014
Practical test pass rate	47.889	49.168	-1.279***
First attempt pass rate	47.193	48.930	-1.737***
Zero faults pass rate	0.646	1.185	-0.538***

Starting with the fixed-effects estimates for instructor outcomes (Table 2), we find that the number of trainee instructors fell in the period following the reform proposals relative to before it (Column 1), and that this is robust to the general time trend (Column 2). Exponentiating the latter model's coefficient translates to roughly 78 per cent fewer trainee instructors within postcodes on average in the period after the reforms. Nonetheless, the number of Part 3 tests being taken appears to be higher than the general trend after the reform proposals once the time trend is taken into account (Column 4), while the pass rate of trainees was no different before and after the reform proposals

[84] As before, we do not expect this to be affected by the cost of the Part 3 exam (i.e. £111 of a total of £750).

once general time trends are taken into account (Column 6). Finally, the number of qualified instructors was lower after the reforms than before it, but only marginally – by roughly 1.7 per cent (Column 8). Taken together, these results imply that more tests are being taken after the reform proposals than expected, but at the same time the number of instructors is falling even though the pass rate remains stable – perhaps due to exiting.

Table 2. Fixed-effects models of instructor outcomes pre- and post-reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of trainee instructors (log)		Number of Part 3 tests (log)		Part 3 test pass rate		Number of qualified instructors (log)	
Pre vs.	-1.017***	-1.531***	-0.811***	0.260***	4.243***	-1.442	-0.112***	-0.017***
post period	(0.047)	(0.071)	(0.054)	(0.066)	(1.122)	(1.272)	(0.014)	(0.004)
Time trend		0.015***		-0.019***		0.102***		-0.003***
		(0.002)		(0.001)		(0.017)		(0.000)
Constant	2.485***	1.379***	4.680***	5.517***	52.561***	48.121***	5.483***	5.687***
	(0.034)	(0.138)	(0.014)	(0.047)	(0.280)	(0.856)	(0.010)	(0.039)
R2	0.741	0.755	0.623	0.812	0.464	0.501	0.995	0.996
N obs	4752	4752	7128	7128	7116	7116	4752	4752
N panels	147	147	78	78	78	78	147	147

Notes: Robust standard errors clustered on postcode area in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Table 3. Fixed-effects estimates of learner outcomes pre- and post-reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of trainee instructors (log)		Number of Part 3 tests (log)		Part 3 test pass rate		Number of qualified instructors (log)	
Pre vs.	-0.009	0.127***	1.283***	-1.147***	1.752***	-1.533***	0.535***	-0.032
post period	(0.022)	(0.020)	(0.269)	(0.263)	(0.257)	(0.234)	(0.043)	(0.046)
Time trend		-0.002***		0.042***		0.061***		0.011***
		(0.000)		(0.005)		(0.005)		(0.001)
Constant	6.766***	6.866***	47.888***	46.107***	47.189***	44.426***	0.647***	0.170**
	(0.006)	(0.017)	(0.073)	(0.240)	(0.066)	(0.286)	(0.011)	(0.052)
R2	0.906	0.909	0.744	0.763	0.835	0.875	0.638	0.729
N obs	13412	13412	13400	13400	12396	12396	12396	12396
N panels	116	116	116	116	116	116	116	116

Notes: Robust standard errors clustered on postcode area in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Turning to the fixed-effects estimates for learner outcomes (Table 3), we begin by examining the number of practical tests taken by students and find that there is no difference before and after the reforms(Column 1), but when the general fall in the number of practical tests is taken into account through the time trend, the number of tests taken is about 13 per cent higher (Column 2). We also find higher pass rates after the reform than before it (Column 3), but once the general upward trend in pass rates is taken into account (as previously mentioned, perhaps this is due to the tightening up of the theory test procedures), we find the effect in fact becomes negative. A similar pattern of higher pass rates at first attempt and zero faults at first attempt being higher in the period following the reform is found but, yet again, once taking into account the general time trends in these two outcomes, the pass rates in the period after the reform become lower than before it.

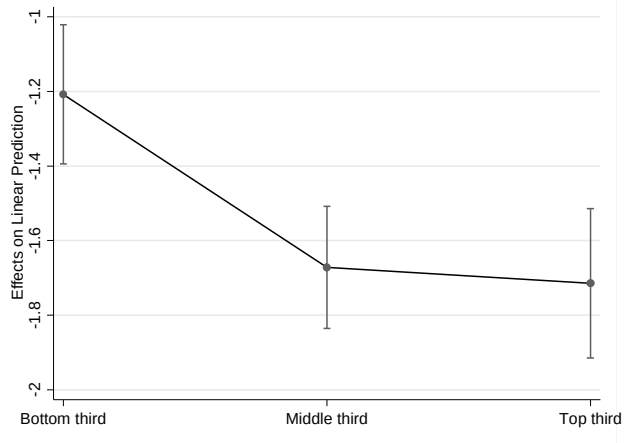
Variation Across Postcode Areas

In this section, we graph the results of the extension of the previous analysis. As mentioned, the fixed-effects models (with the time trend) are extended to include an interaction term between the reform dummy and whether a postcode area was in the top, middle or bottom tercile in terms of its share of trainee instructors prior to the reform. These are estimated marginal effects rather than direct regression estimates (so we can calculate an effect for the excluded postcode area type category). This is a particularly informative analysis as we expect the reform proposals to affect both instructor and learner outcomes through its effect on the ratio of trainees to fully qualified instructors.

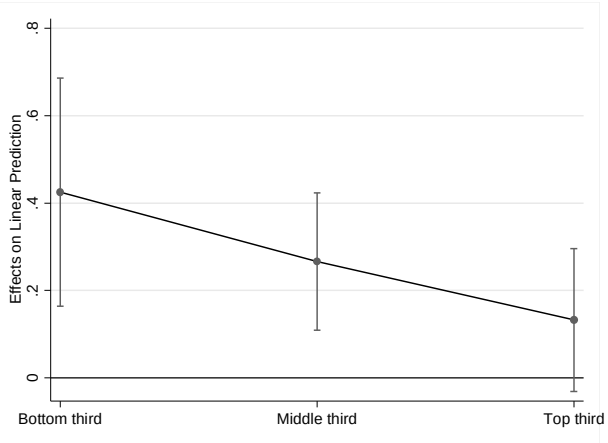
Beginning with instructor outcomes in Figure 9, we do not find any dramatic heterogeneity for most of the outcomes, although we do find that the drop in the number of trainees after the reform (perhaps unsurprisingly) was greater in magnitude in areas that had a higher share of instructors prior to it (Panel A). We also find the reform had no effect on the number of practical tests being taken by prospective instructors (a proxy for the number of potential entrants into the profession) in postcode areas that had a high share of trainees prior to the reform, while it did in the other two postcode area groups (Panel B).

Figure 9. Heterogeneity in instructor outcomes pre- and post-reform proposals by proportion of area who were trainees pre-reform

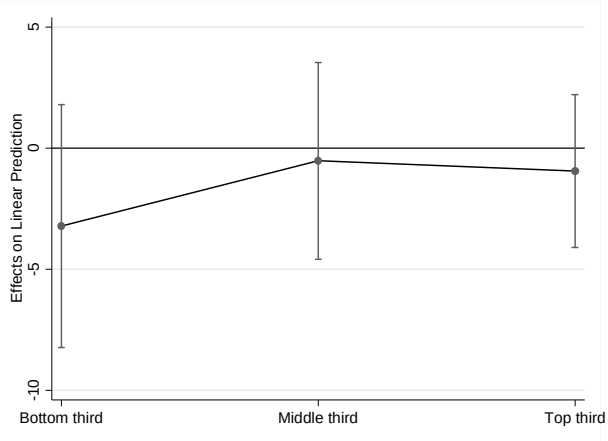
Panel A. Number of trainee instructors (log)



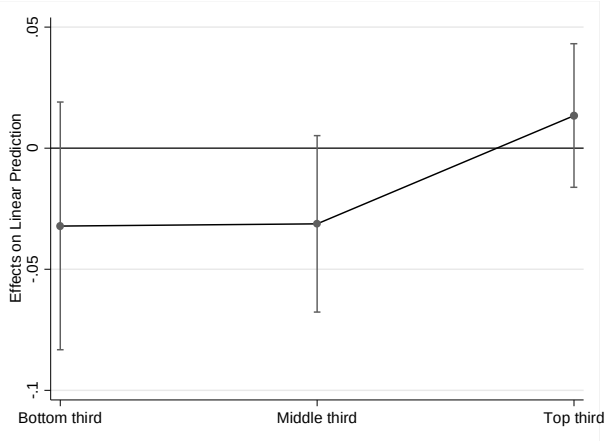
Panel B. Number of Part 3 tests taken (log)



Panel C. Part 3 test pass rate



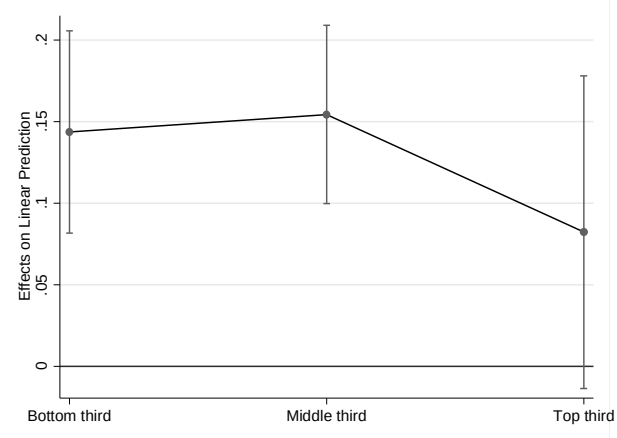
Panel D. Number of qualified instructors (log)



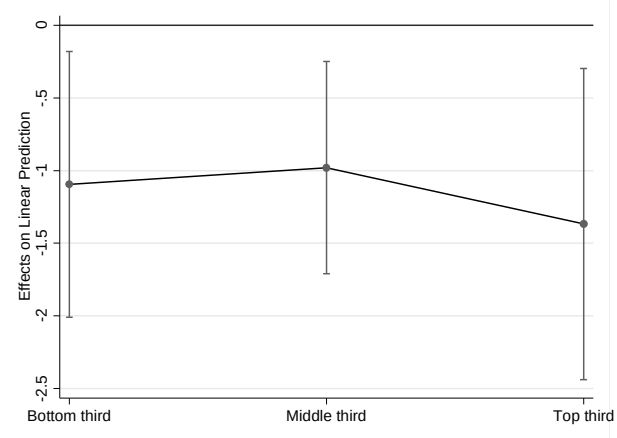
Notes: Vertical bars are 95% confidence intervals.

Figure 10. Heterogeneity in learner outcomes pre- and post-reform by proportion of area who were trainees pre-reform

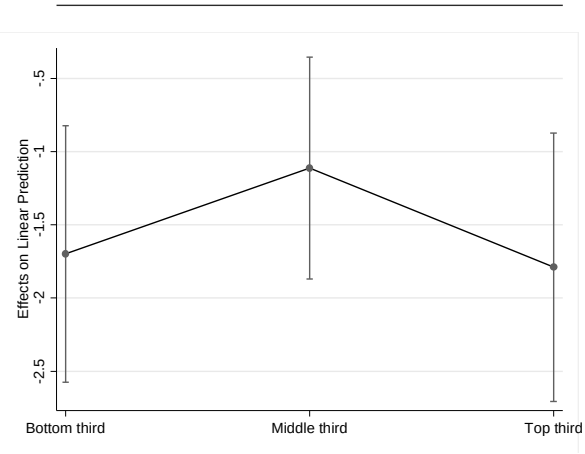
Panel A. Number of practical tests (log)



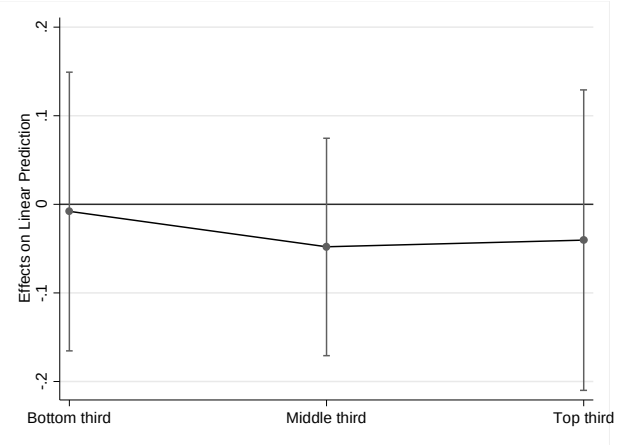
Panel B. Practical test pass rate



Panel C. First attempt pass rate



Panel D. First attempt zero faults pass rate



Notes: Vertical bars are 95% confidence intervals.

Turning to heterogeneity across postcode areas in the effect of the reform on learner outcomes in Figure 10, again we do not find much evidence of heterogeneity across postcode areas. Most notably, we find that the drop in practical test pass rates was fairly even across postcode areas (Panel B).

Overall, the fixed-effects part of the analysis shows that the reforms of the instruction by trainee instructors have had no discernible impact on the quality of instructors or the tuition. If anything, our results demonstrate that the quality declined as learner pass rates appeared to be lower before the reform proposals relative to after it.

Learner Driver vs. Learner Motorcyclist Practical Test Pass Rates

We now present the results of the DiD analysis in Table 4. The dependent variable here is the pass rate. The results compare the trends in pass rates for learner drivers compared to motorcyclists before and after the reform proposals in 2014 (see Figure 11). Column 1 presents the DiD estimate with no further controls, while Column 2 introduces year and quarter fixed-effects. Both models control for postcode area fixed-effects. We find that the DiD estimate in both columns is negative and of a similar magnitude – around 1.2 percentage points difference. This reinforces the earlier fixed-effects estimates that learner pass rates in practical tests were lower after the reform than before it. Overall, our results on this measure show that service quality (at least in terms of practical test pass rates) deteriorated.

Figure 11. Learner driver and motorcyclist practical pass rates

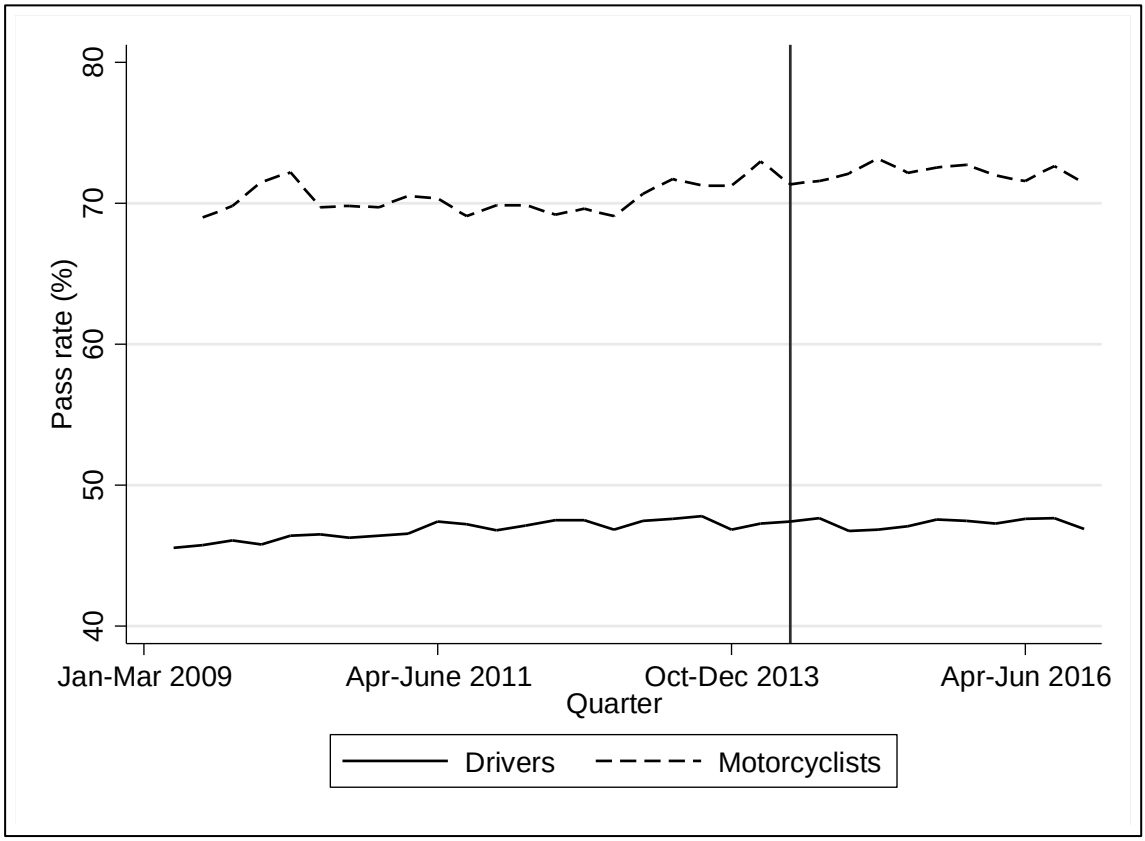


Table 4. Difference-in-difference estimates of learner driver vs. motorcyclist pass rates

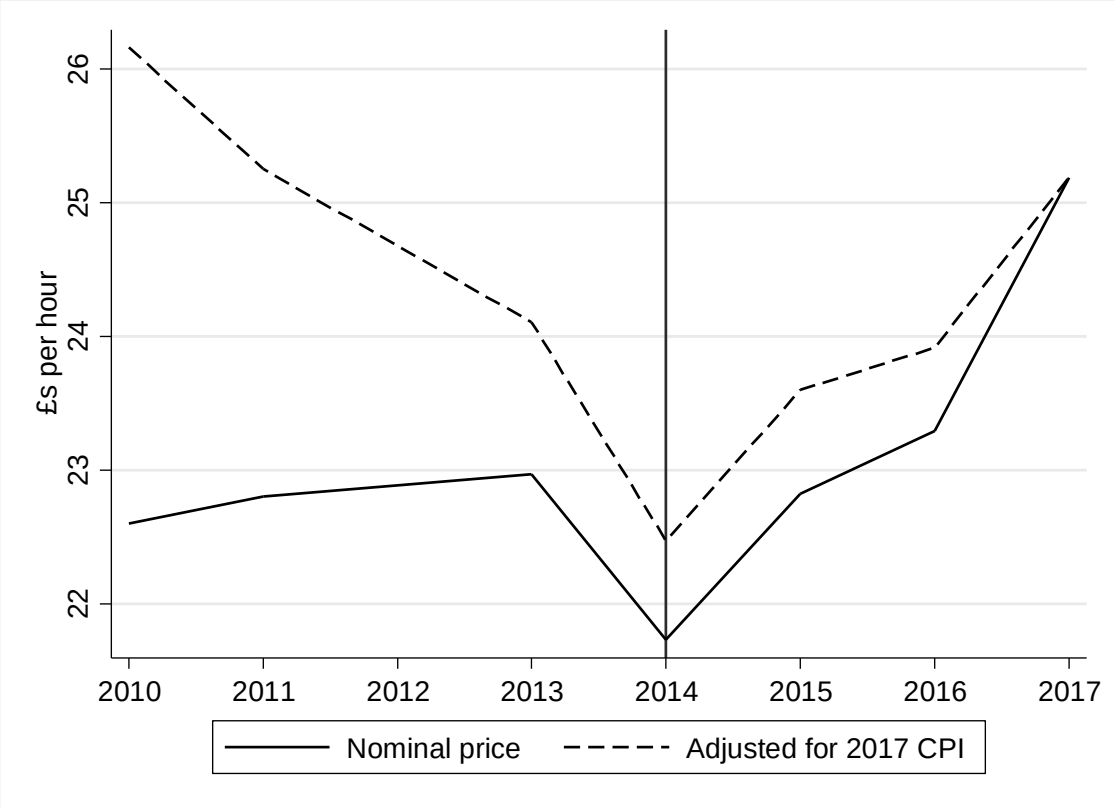
	(1)	(2)
	Practical test pass rate	
DiD	-1.234*	-1.258*
	(0.552)	(0.553)
Constant	70.864***	70.330***
	(0.424)	(0.498)
Postcode area FEs	✓	✓
Year FEs		✓
Quarter FEs		✓
R2	0.875	0.876
N obs	5968	5968
N panels	116	116

Notes: Robust standard errors clustered on postcode area in parentheses. Statistical signifiance * p<0.05, ** p<0.01, *** p<0.001.

Price of Driving Lessons

We use data from the UK Office for National Statistics to calculate the price of driving lessons (one hour) before and after the reform proposals. We present data on both the nominal price and the price adjusted for 2017 CPI. As can be seen in Figure 12, after a trend of falling real prices, there is a trend of increase in the price of driving lessons after the reform proposals. While at the cut-off date of the last available data the price of a driving lesson is still lower than it was before the reform proposals, there is a clear upward trajectory. One possible explanation for this trend relates to the lower availability of driving instructors in the market post the reform proposals, which has enabled incumbents to increase their fees.

Figure 12: Prices for one-hour driving lessons



Road Accidents

Finally, we explore trends before and after the reform proposals in road accidents. We linked our data set on the share of trainees by postcode area to data on serious road accidents. Accident data is recorded in geographical and time units (annual) such that the number of cells are small relative to our previous analysis, limiting the depth of our analysis. There are also further complicating factors in that the matching of geographical units with the accident data is not perfect. The accident data is recorded at a lower geographical unit (parliamentary constituency) than information on learner drivers and instructors (postcode area) and the overlap is not perfect (e.g. parliamentary constituencies may overlap with more than one postcode area). We therefore took the modal postcode areas mapping onto parliamentary constituencies and classified them accordingly. We then summed all the accident data across parliamentary constituencies by postcode and year to get annual postcode area-level totals. We are left with a data set of 585 postcode areas by year observations. We take the natural logarithm of total accidents to form the dependent variable in our analysis. We estimate a fixed-effects linear regression model with the reform dummy as a covariate. The coefficient can be interpreted as the average change in outcome averages across units before and after the reform. In this analysis, we control for region rather than postcode fixed-effects given the smaller N (12 regions), but still cluster standard errors on the postcode level. Our results show that serious road accidents were on average higher after the reform proposals than in the period before it (Table 5). As we are keen to stress, there are many confounding factors here that our data sources do not allow us to control for (e.g. weather, improvements in infrastructure and traffic management). We also have no way of distinguishing what proportion of these serious accidents were committed by newly qualified drivers (i.e. those that are likely to have been affected by the reform proposals). Taking this into account, however, it

is possible that the lower availability of instructors in the market (coupled with the trend of price increases we observed earlier) led learner drivers to take fewer lessons.

Table 5. Fixed-effects estimates of road accidents pre- and post-reform proposals

	(1)
	Serious accidents (log)
Pre vs. post	0.045**
period	(0.015)
Constant	4.872***
	(0.072)
R2	0.215
N obs	585
N panels	12

Notes: Robust standard errors clustered on postcode area in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

5.8 Conclusion

Entry into the profession of driving instructor has been subject to compulsory training requirements, although in practice there was a dual market of fully trained (ADIS) and trainee driving instructors that both could offer tuition for payment. The aim of this study was to explore the impact of the proposals to reform the driving instructor occupation in the UK. In particular, the state was concerned that the quality of tuition offered to prospective drivers was compromised due to the capacity of trainee driving instructors (who often were not able to finish their training) to provide paid tuition in the same manner as the fully licensed instructors did. The 2014 reform initiatives were aimed at increasing the stringency of regulations such that only ADIs or trainees accompanied by ADIs would be able to offer paid tuition. The adoption of these proposals would make the occupation subject to de facto licensing. In addition, the UK made the procedures used every four years to ensure instructors continued to provide tuition of high quality after they have qualified more stringent. In this study, we provide an assessment of the effect the proposed changes had on the quality of service. We do so by analysing government data on outcome and value-added measures of quality.

Starting with service availability, we looked at various outcomes for instructors. Our results show that the number of trainee instructors fell in the period following the reform proposals relative to before, with around 78 per cent fewer trainee instructors within postcodes on average in the period after. There was also a small drop in the number of trainee driving instructors qualifying as ADIs (1.7 per cent). We interpret this as a sign that the reform proposals discouraged instructors in the labour market from continuing to practise (e.g. those that had previously failed the part 3 test) or from entering the market in the first place. In that sense, the reform proposals meet their aims. However, the question is whether this reduced availability of driving instructors in the market (which can be adverse for consumers) translated to any benefits for the other quality indicators.

The reform proposals also seem to have put some pressure on trainee instructors to sit the relevant tests so they could qualify (as the reform proposals intended), but we do not observe an improvement in the Part 3 test pass rates before and after. As such, we find no evidence of improved quality of instructors. Turning to the outcomes for learners, we note that in addition to the proposed restrictions on the ability of trainee instructors to provide fully unsupervised tuition, we have the parallel improvements in the standards check (i.e. the review of instructional methods that happens once every four years), both of which were intended to affect the quality of instruction offered. Overall, we do not find any improvement in the three quality measures of learner performance. If anything, overall pass rates, pass rates at first attempt and zero faults all become negative once time trends are taken into account. These findings are further reinforced when we compare the trends in pass rates for learner drivers to those of motorcyclists (who were not subject to any changes in their tuition) before and after the reform proposals. Why these measures deteriorate is an open question, but our broad conclusion is that the reform proposals

are certainly not showing an improvement in the quality of learner drivers as would have been expected.

Our final measures of quality relate to prices and accidents. With regards to the former, we find that the price of a one-hour lesson has increased after the reform proposals. This is in line with expectations relating to how lower supply of practitioners in the market is likely to affect the cost of these services. Given that the state does not specify a minimum number of driving lessons learners have to do before taking the driving test, it is possible that the trend of increasing prices we observe is affecting demand for driving lessons. Turning to accidents, there appears to have been an increase in serious accidents, although due to data constraints, we cannot distinguish whether this is driven by newly qualified driving instructors. We treat this final part of our analysis with extreme caution given the difficulty in isolating the effects of regulation from the variety of factors that can affect accident rates in a given area (e.g. improvements in road quality and traffic management etc.).

To conclude, the reform proposals aimed to make access to the driving instructor occupation more stringent. From the evidence presented here, it is doubtful whether this has resulted in better quality outcomes. Our findings echo other studies that have on average failed to show that more intense driver education courses produce safer drivers. A recent review of the international evidence by the U.S. Department of Transport (2009), for example, concludes that there is little evidence that the crash records of those that underwent intensive formal training are different to those that did not, as beyond imparting basic car control and road law knowledge, their benefits cannot be scientifically established. Instead, according to the report, traffic management, attitudes, motivations, peer influences, cognitive and decision-making skills are more influential in shaping driving behaviours (see also Sümer 2003; French et al. 1993; Abdel-Aty and Radwan 2000 for further evidence). While a detailed analysis of this body of work is beyond the scope of this study, the findings have led many to question what should be expected of driving instructors and the stringency of the level of training that they should be subject to in order to meet such expectations (Waller 1997). We hope that the findings from this study make a small contribution to these wider debates about the optimal level of regulation of driving instructors.

References

Abdel-Aty, M. A., and Radwan, A. E. (2000). Modelling traffic accident occurrence and involvement. Accident Analysis & Prevention, 32(5), 633-642.

Angrist, J. D., and Guryan, J. (2004). Teacher testing, teacher education, and teacher characteristics. The American Economic Review, 94(2), 241-246.

Avrillier, P., Hivert, L., and Kramarz, F. (2010). Driven out of employment? The impact of the abolition of national service on driving schools and aspiring drivers. British Journal of Industrial Relations, 48(4), 784-807.

CARE (2014) EU Road Accidents Database: European Commission/Directorate General Energy and Transport, Available from: https://ec.europa.eu/transport/road_safety/specialist/statistics_en

Department for Transport (2016) National Travel Survey, available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/633077/national-travel-survey-2016.pdf

European Commission, DGGROW (2016) Mutual Evaluation of Regulated Professions: Overview of the Regulatory Framework in the Transport Sector by Using the Example of Driving Instructors, Available from: file:///C:/Users/bsw269/Downloads/150119-report-transport-sector-driving-instructors-final_en%20(2).pdf

French, D. J., West, R. J., Elander, J., and Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. Ergonomics, 36(6), 627-644.

Kane, T. J., and Staiger, D. O. (2005). Using imperfect information to identify effective teachers. Unpublished manuscript.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. Economics of Education review, 27(6), 615-631

Kleiner, M. M., and Petree, D. L. (1988). Unionism and licensing of public school teachers: Impact on wages and educational output. In When public sector workers unionize (pp. 305-322). University of Chicago Press.

Seim, K. and Vitorino, M.A. (2011). Efficiency Gains from Removing Entry and Price Controls: Evidence from a change in regulation, Wharton School Working Papers, Available from: http://maria-ana-vitorino.com/docs/dr_school_intro.pdf

Sümer, N. (2003). Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. Accident Analysis & Prevention, 35(6), 949-964.

U. S Department of Transport, National Highway Traffic Safety Administration (2009), Feasibility Study on Evaluating Driver Education Curriculum.

(This page intentionally left blank)

6 Ride-Hailing Drivers in Dublin and London

6.1 Introduction

Historically, the traditional taxi industry has been heavily regulated in most countries. Such regulations typically involve conferment of exclusive rights to ply for hire, quality controls (e.g. relating to vehicle age and appearance) the imposition of entry restrictions for drivers (e.g. topographical knowledge tests) and the control of fares (Beesley 1973). Public safety, consumer protection and congestion management have commonly been put forward to justify these policy decisions. From an occupational regulation perspective, the most common barrier to entry for individuals seeking to work in the sector is the requirement that drivers be 'licensed' by local governments. Such requirements are typically justified by regulators as ensuring a minimum level of safety and quality. Despite the worldwide prevalence of regulation for individual taxi drivers, there is a scarcity of empirical studies that focus on the entry requirements for these professionals (rather than the industry as a whole) and evaluate their effects on service quality. Ideally, a study seeking to address these questions would examine indicators of service quality before and after the removal of entry requirements to the occupation. However, since such deregulatory shocks are rare, an alternative approach is to focus on the differences in stringency of entry requirements for prospective drivers across jurisdictions. The relatively recent rise of ride-hailing services which have entered competition with traditional taxi services and to which regulators have responded in different manners, offers an appropriate research setting to examine how differences in regulation affect quality outcomes. Ride-hailing services traditionally refer to rides provided by drivers for remuneration by passengers who have ordered this service via a smartphone app.[85] There are various apps which attempt to connect users seeking a ride with providers that are prepared to provide one. The providers effectively operate in a similar manner to a taxi: the app enables them to locate the passengers, transport them to the destination of their choice and charge a fare based on time and distance of the trip set by the app provider.

This case study examines the influence of the variations in the regulations to become a ride-hailing driver on various quality outcomes using data from Uber, a firm that provides such ride-hailing services, for London, England and Dublin, Ireland. A key question we seek to address is whether more relaxed regulation requirements to become a ride-hailing driver are associated with inferior service quality outcomes. As the restrictiveness of the regulatory regime increases, does quality improve substantially or enough to justify costly regulatory intervention? More specifically, we examine the influence of regulations to provide taxi services via a ride-hailing app on consumer satisfaction and measures of road safety by exploiting the variation in the regulatory stringency to become a driver between these two jurisdictions. Moreover, we show how fares vary with the intensity of licensing requirements. We compare the quality of rides in London (where entry to the occupation is more relaxed relative to heavily regulated 'fully licensed' taxi drivers), and in Dublin (where ride-hailing drivers are subject to the same stringent requirements as those of 'fully licensed' taxi drivers). Our approach follows a long tradition in the study of occupational regulation that explores variations in the intensity of the regulatory regime (for a review, see Kleiner 2006), rather than a comparison between regulated and unregulated individuals and occupations[86].

[85] A definition of ride-hailing services and the following distinctive features are noted under https://www.bls.gov/ooh/transportation-and-material-moving/taxi-drivers-and-chauffeurs.htm?view_full#tab-2: The fare rate can fluctuate depending on demand relative to supply; however, passengers are notified if the current fare rate is higher than usual. Passengers pay for rides through a credit card linked to the app. Drivers use their own private vehicles and set their own hours.

[86] This is not a case study of Uber drivers or Uber as a company. Uber has agreed to provide us with the data it collects through its app. We viewed the data to be of good quality and suitable (a) to address the research questions of this study and (b) to the chosen research design (namely the study of variations in the intensity of the regulatory regimes). There were no incentives for Uber to provide data that might skew the data or results, since they were unaware of our approach or methods prior to providing us with the data.

6.2 Literature Review

Since providing ride-hailing services became a wide-spread business during the past 10 years when the use of smartphones reached a critical level, the research literature on these kinds of services is still sparse. However, since the main difference between ride-hailing and traditional taxi services is that drivers receive potential fares via smartphone rather than through a local taxi organization, the literature on traditional taxi services and its regulation are closely related. As such, we refer to this work to provide a theoretical and empirical context that informs our study.

The regulation of the taxi industry is a common phenomenon in many countries with such regulations setting entry restrictions for drivers, operators and vehicles (Schaller 2007). The rationale for taxi regulation (in general, and for taxi drivers in particular) stems from the experience good problem: goods for which quality cannot be determined by the consumer until after the good has been consumed thus giving rise to 'moral hazard' (e.g. the provider overcharging, operating a substandard vehicle or taking a longer route)[87]. Regulation in the form of minimum standards setting is assumed to solve this imperfect information problem (Shreiber 1975). Proponents argue that regulation is vital to ensure customers are not faced with longer journeys and higher fares through using unscrupulous providers or have their safety compromised by dangerous driving and criminal behaviour. Further, while they accept that price and supply controls might result in higher fares, this enables dense markets to cross-subsidise low-density ones while peak traffic cross-subsidizes off-peak service availability (Dempsey 1996). In the absence of regulation, service to low density and off-peak trips may decline or not be available at all. Critics of this approach argue that unregulated market entry and fares for taxi providers will shift the supply curve upwards, resulting in lower fares, shorter waiting times, improved availability (especially in marginal markets) and greater entrepreneurial opportunities (especially for minorities and immigrants) (Frankena and Pautler 1984; Cevero 1985).

From an empirical perspective, research has looked at the effect of regulatory changes on service availability and prices as they apply to the taxi service market in general and from a producer perspective. The removal of quantity controls on taxis in a number of US cities, for example, caused supply to increase by an average of between 19% and 23% (Dempsey 1996; Teal and Berglund 1987; Schaller 2007). Dempsey (1996) finds that in cities which abolished quantity and fares restrictions, prices in both cruising and dispatch markets rose by an average of 29% per year, but despite a slight surge in supply initially, service availability fell to pre-deregulation levels. Recent work by Rojeck and Masior (2016) looks at the deregulation of entry to the taxi driver occupation in Poland that took effect in 2013. The reform included authorities being given the right to control pricing practices as well as the removal of the requirement of a preparatory course and exam in cities with over 100,000 inhabitants. This provided an ideal context to compare the effects of the reform between cities, but also before and after reform comparisons of the regulations relative to prices. The study finds that in those cities where barriers to entry were relaxed, prices fell for the first time after the reform, while the supply of taxis rose as measured by the number of licenses issued. Using quality indicators such as licence withdrawals, complaints filed against taxi drivers and frequency of malpractices, the authors find mixed results on quality.

In this study we focus on the regulation of drivers and as such, a detailed review of the literature on providers, (i.e. companies providing taxi services) is beyond the scope of this analysis.[88] It should be emphasised that little is known about how regulating entry to the taxi and ride-hailing driving occupation (as opposed to the taxi and ride-hailing industry in general) affects the quality of services provided to consumers. It is also important to note that entry controls to the taxi and ride-hailing driver occupation and qualification requirements do not adhere to the binary picture (regulation/no regulation) painted by proponents and critics of regulation. In practice, entry to the occupations in different jurisdictions is covered by a wide spectrum of regulatory policies, and it is this variation in restrictiveness that is the focus of our study. Consequently, we are interested in analysing the levels of regulation rather than the existence of controls altogether.

[87] While these arguments were originally developed to explain the characteristics of the traditional taxi industry, they can also be applied to ride-hailing.

[88] The focus of our analysis is the occupation of taxi drivers, not the taxi industry and hence we focus our discussion on the requirements for drivers to enter the occupation, rather than requirements placed on companies. For reviews on the regulation of the taxi and ride-hailing industry more generally as well as the ride sharing (see Flores-Guri (2003) and Gallick and Sisk (1987); Harding et al. (2016).

6.3 Regulatory Context

As background, it should be noted that London's taxi service market has traditionally been served by two types of providers, namely the 'black cab/hackney' drivers and 'minicab' drivers. Black cab drivers have traditionally had the legal monopoly of being allowed to use taxi regulations to be the only service that that can be hailed on the street or from a taxi rank. In contrast, minicabs must be pre-booked and may not be fitted with a taximeter. The pre-booking takes place via taxi service companies who operate in the city and are only legally permitted to provide taxi services for pre-booked rides .Traditionally, these rides were booked by phone, but more recently, various ride-booking apps have been developed, Uber is one of the most popular ones in the city. London accounts for about 5 per cent of Uber's[89] global active user base of 65 million, and nearly a third of its active user base of 11 million in Europe (see Appendix B for a more detailed description of the firm and its evolution). There are about 40,000 licensed Uber drivers in London who service about 3.5 million users of the service. There is fierce opposition to Uber from the Licensed Taxi Drivers' Association, a union representing London's black cab drivers, whose drivers take several years to train for their job, in part by memorizing London's streets, an endeavour called 'The Knowledge'. In contrast, the governmental entry requirements that exist for Uber drivers in London are the same as those that apply to the minicab drivers and are substantially lower. These lower requirements have made entry into the ride-hailing business as a supplier of these services much easier than traditional black cabs. Uber drivers use technology through the app in the car to find locations in London. The relatively high cost of becoming qualified to offer taxi services for drivers using the 'fully licensed' black cab model has helped make Uber popular among prospective drivers in London. A four-mile trip in the middle of the week costs at least £16 in a black cab, compared with £8 on Uber (Wall Street Journal, September 23, 2017). Recently Uber regained its licence to operate in London in spite of opposition by traditional Black cab operators (Satariano, 2018).

Dublin, Ireland is also a capital city and the largest city in Ireland. It is a much smaller city than London, within a metropolitan area of almost 2 million. The taxi sector in Ireland has traditionally been highly regulated, involving severe restrictions on the number of new vehicle licenses (i.e. licenses linked to the taxi, not the driver) issued by the state (Barrett, 2003). As the industry regulator, the National Transport Authority also has the power to control the number of licensed taxi drivers by setting the pass rate, exam fees and other requirements. Ireland's rapid economic growth in the 1990s resulting in income prosperity and a booming tourist market increased demand for taxi services. However, due to the caps on the number of vehicle licenses that the state regulators were issuing, such demand could not be met resulting in increased customer discontent with the sector and its regulators (Barrett 2010). In 2000, the Irish High Court decided to somewhat liberalise entry to the taxi sector by gradually raising (but not abolishing) the cap on vehicle licenses.

Uber, as one knows it in London, is not available in Dublin. Irish law requires that anyone carrying passengers for money must have a taxi licence, meaning that Uber drivers in Dublin must go through the same city driver licensing procedures as apply to traditional taxi drivers. So Irish law requires Uber drivers in Dublin to meet the same regulatory requirements as a 'fully licensed' taxi driver. As a result, Uber drivers in Dublin are typically traditional taxi drivers that have also signed up to the Uber platform as an additional means to attract clientele.

One potential barrier to entry for ride-hailing companies is that drivers must be licensed by local authorities as is the case in both cities, and that the precise requirements vary across jurisdictions. As discussed above, these requirements are typically justified by regulators to ensure a minimum level of safety and quality.

To what extent do these requirements vary between Dublin and London?

As alluded to earlier, to operate as an Uber driver in Dublin, one needs to go through the same licensing process as any traditional "fully licensed" taxi driver. To become what is referred to by the state as a 'fully licensed'[90] taxi driver in Dublin, one must pass an SPSV Entry Test.[91] According to

[89] See Appendix for a summary of Uber, the company that provided us with the data used in our analysis.

[90] The term 'fully licensed' is used by local authorities in Dublin to refer to the fact that the driver has the right to practise in this particular jurisdiction as a regulated ride-hailing provider. Although this might coincide with 'licensing' (as understood in the occupational regulation literature), the two terms have different meanings in everyday use and are not interchangeable. Those who are not "fully licensed" cannot provide ride-hailing services.

the Irish Transport Authority, the test is designed to verify that new entrants to the industry are familiar with good practice and have a good working knowledge of the county in which they wish to operate.[92] This fairly detailed and in-depth test consists of 90 questions in two sections.

- The first section – the Industry Knowledge Module – consists of 54 questions relating to industry regulations, vehicle knowledge, map reading, fares and charges, customer service (including disability awareness and equality and diversity, business acumen and health and safety) as set out in the module's manual. The actual learning material that covers these areas of knowledge is very detailed and attempts to equip the candidates with an in-depth knowledge of all aspects of the job, from customer service to procedural to technical issues. It is therefore expected to ensure not only high standards of service but also high levels of safety (for sample questions asked in this test, see Appendix C).
- The second section – the Area Knowledge Module – consists of 36 questions relating to the administrative county in which one wishes to be licensed. This aspect of the taxi licensing process is aimed at ensuring driver familiarity with the area so that journeys are delivered in an efficient and safe manner. Table 1 summarises the key information needed for the Area Knowledge test as outlined in the 'Official Manual for Operating in the SPSV Industry' (2015) (for sample questions asked in this test, see Appendix C).

Overall, the process in Dublin is known to be more time-consuming due to the in-depth knowledge requirements. In addition to these tests, one must also produce a tax clearance certificate and undergo a criminal record and a medical check. The cost of the licensing process is approximately €370 and includes €90 towards the SPSV Driver Entry Test, €250 towards the SPSV Driver License Fee and the €30 criminal records check fee.[93] The licenses need to be renewed every five years. These requirements typically apply to all taxi drivers in the city, including those that offer ride-hailing services through the Uber platform. Practising without a licence is illegal for all taxi and ride-hailing drivers.

With regard to the impact of the cap on vehicle licenses in Dublin, the fact that there has been such a dramatic decline in the number of active taxi driver licenses (from more than 47,000 in 2009 to less than 27,000 in 2016) while there was only a small decline in the number of licensed vehicles (from about 27,000 in 2008 to 21,000 in 2016)[94], meaning that the ratio of licensed drivers to licensed vehicles fell from 1.75 in 2009 to 1.27 in 2016, shows that the effect of the cap became less important. Since the caps lost their effectiveness long before 2013 and the number of licensed vehicles remained roughly constant after 2013, these caps should not have been the cause of reduced entry into the profession. Also in London, since the number of licensed taxi vehicles remained roughly constant during this time[95], the results should not be confounded by vehicle licensing policy.

To drive for Uber in London, one is subject to the same regulations as the driver of a 'private hire' vehicle.[96] In particular, one must be at least 21 years of age, hold a full UK or EEA Member State driver's permit that is at least three years old and as of 2016 pass an English language test.[97] In addition, Uber drivers in London are required to undergo a criminal record check and a medical examination, as well as take a topographical knowledge test. The assessment structure that is related to this test differs substantially to the Dublin one. It consists of 5 modules assessing the

---

91 Details can be accessed here: https://www.nationaltransport.ie/taxi-and-bus-licensing/taxi/spsv-driver-licensing/apply-for-a-test/

92 These requirements are set out in the Taxi Regulation Act 2013 and the Taxi Regulation (Small Public Service Vehicle) Regulations 2015 (see www.irishstatutebook.ie for all legal texts).

93 Author's own calculations based on information from the National Transport Authority available here https://www.nationaltransport.ie/taxi-and-bus-licensing/taxi/spsv-driver-licensing/apply-for-an-spsv-driver-licence-2/

94Taxi Statistics for Ireland – Statistical Bulletin No. 1 2017. See https://www.nationaltransport.ie/taxi-and-bus-licensing/taxi/taxi-statistics/

95 See data tables link here: https://www.gov.uk/government/statistics/taxi-and-private-hire-vehicles-statistics-england-2015, Table 'Taxi0101'

96 Private hire vehicles are only permitted to pick up pre-arranged bookings and are not permitted to pick people up from the side of the road. Taxis (also known as London black cabs) are hackney carriages and are licensed to pick people up from the roadside, i.e. hailing a cab.

97 https://tfl.gov.uk/info-for/taxis-and-private-hire/licensing/private-hire-driver-licence

ability of prospective candidates to read maps, plan routes and their general understanding of London's topography (see Table 2).

Table 1: Key Information Needed for the Area Knowledge Test

Urban Centres: You will need detailed knowledge of:	
• One-way systems	• Districts
• Street names	• Adjoining streets
• Housing estates	• Routes to other towns or villages
National roads and motorways:	
The prospective drivers will need to know the numbers of national roads and motorways, for example, N3 or M1. Questions may reference the main routes out of the county being tested. For example, an applicant in Laois might be asked what road links Laois and Dublin.	
Places of interest:	
The prospective driver will need to be able to name the streets in your county on which places of interest in the following categories are located:	
• Hospitals and medical centres	• Sporting and leisure facilities (e.g. swimming pools, cinemas, theatres, libraries)
• Nearest airports	• Shopping centres and retail parks
• Local ferries and local boat services	• Public buildings and amenities
• Transport terminals (e.g. bus, train)	• Industrial estates and business parks
• Hotels and holiday centres	• Well-known pubs, restaurants and nightclubs
• Stadiums and prominent club grounds	• Schools, universities and other educational institutions
• Sports clubs (e.g. golf clubs, yacht clubs, tennis clubs)	• Garda stations
• Major landmarks and tourist attractions	
Towns and Villages:	
The prospective driver will need to know the location of all towns and villages within your operating area in relation to all regional and national routes. Questions in the test may reference towns and other locations in counties bordering the county being tested. For example, an applicant in Kilkenny might be asked about routes to Waterford City or New Ross.	

Source: The Official Manual for Operating in the SPSV Industry (2015).

Table 2. Assessment Structure for London Topography Test.

Module 1	Ability in using map indexes (Master Atlas of Greater London)
Module 2	Planning a short route (up to 5 miles) between two locations i.e. streets/roads/points of interest
Module 3	Planning a medium route (up to 15 miles) between two locations
Module 4	Planning a long route (up to 30 miles) between two locations
Module 5	General knowledge questions based on the Master Atlas of Greater London and compass based questions

During the assessment, candidates are provided with all the necessary equipment to answer the questions including an atlas and a magnifying glass; therefore, memorization is not necessary (as it is the case with the test administered in Dublin). As such it is deemed to be much less stringent than the Irish equivalent in that it tests one's ability to use maps and plan routes, rather than require detailed knowledge of urban centres, places of interest, street names in towns and villages (for sample questions asked in this test, see Appendix D). Contrary to the Irish case, there are also no requirements relating to industry knowledge, so the actual 'syllabus' that one has to study is

considerably less demanding. In total, the process costs approximately £724 (although some costs may vary slightly, for example, depending on the fees charged by doctors for medical certificates) (see Table 3 for a summary).[98]

As this review of entry to the occupation requirements has shown in Tables 1-3, there is considerable variation between London and Dublin in terms of the barriers faced by prospective Uber ride-hailing drivers. These differences provide an ideal setting to explore how variations in the stringency of regulation impact selected indicators of quality. It is important to stress that in both jurisdictions Uber ride-hailing drivers are subject to licensing since at least some requirements to enter the occupation are in operation, and these are legally binding. Consequently, Uber ride-hailing drivers in London and Dublin are subject to what is referred to as 'licensing' in the academic literature on regulation (see for example Kleiner 2006; Koumenta and Pagliero 2016).

Table 3: A Summary and Overview of Licensing requirements for Uber drivers in Dublin and London

	Dublin	London
Educational requirements	Area Knowledge Test	Topographical Knowledge Test (based on map reading)
	Industry Knowledge Test	
	Driving Permit	Driving Permit
Other entry requirements	Private Hire License (issued by the National Transport Authority)	Private Hire License (issued by Transport for London)
	Criminal Records Check	Criminal Records Check
	Medical Check	Medical Check
	Tax Clearance Certificate	
		Minimum driving experience of at least 3 years
		English language test (where applicable)
	Minimum age of 18 years	
		Minimum age of 21 years
Cost of driver licensing (approx.)	€370	£724
Renewal	Every 5 years	Every 3 years

The difference between the two contexts is the level at which these regulations are set, or else the ease with which one can enter the occupation. A prospective taxi driver might find it easier to become a licensed driver with more experience, because e.g. passing the exam might be easier for more experienced drivers, or for those with a higher proficiency in English, since the questions involve English terms. Therefore, we control for driver-specific differences to ensure that our analysis examines whether for persons who are identical along the characteristics that are available in the dataset, lower levels of regulation in London or higher standards in Dublin result in different outcomes in terms of correlates of safety or customer satisfaction.[99] Our analysis is on the magnitude of stringency rather than one based on a regulatory reform.

[98] Calculations based on Transport for London fees available here https://tfl.gov.uk/info-for/taxis-and-private-hire/licensing/private-hire-driver-licence.

[99] Experience of the driver and level of proficiency in English may serve as quality outcomes as well. We do not consider these variables as outcomes for the following reasons: First, a consumer of the taxi service might prefer a language other than English, e.g. because the costumer's native language might not be English or because she is a native speaker of another language, say Gaelic, Scots, Shelta, Urdu, French or German. In such a case the best quality would be a driver who speaks the preferred language of a consumer. Since we do not have this information, we cannot use it as a quality outcome. Second, in the simplest case, experience is a function of learning on the job and thus would be deterministic in time. Since the basic skill of driving a car is widespread, if at all, differences in stringency

6.4 Data

The data we use were obtained from Uber Technologies company files and are proprietary data. The unique aspect of the data is that, unlike survey data, which may be collected with less care since there are virtually no economic consequences to the firm of an inaccurate response, the information for our analysis is the data the company uses for decision-making and is collected and compiled with greater care. In all our analyses we use trip-level data from November 2013 (Uber expanded to London on July 2, 2012) through December 2016. The data was provided as part of the company's regular compilation of data without any information on how the data might be used for analysis or public policy purposes. Unfortunately, no data on tips are available at this time from the company, and since many times tips are given in cash, this may be an inaccurate measure of earnings and are often determined by culture rather than performance. To provide a balanced sample for analysis for the two cities, we used the universe of 162,386 rides from Dublin. For London, Uber provided a random sample of 260,081 or 0.5 percent of all rides.[100] In Appendix A we provide definitions of the variables in the data set.[101]

6.5 Quality Indicators

Based on the arguments made by proponents and critics of regulation of taxi drivers, in theory, various quality indicators can be used in this context. First, regulation is expected to be associated with higher levels of customer safety. As such, relevant indicators can include accidents and collision rates caused by taxi drivers. Correlates of safety (but also ride comfort) can also include sudden stopping and starting as well as speeding. Information asymmetry between drivers and customers can also compromise service quality. This can manifest itself through criminal and fraudulent behaviour by taxi drivers. The latter can involve indicators of cheating behaviour, for example, in the choice of routes (e.g. longer routes to maximize fare) or tariffs charged (e.g. manipulate base fare for individuals with no local knowledge and such as tourists). Balafoutas et al. (2011) for example using a field experiment methodology finds that passengers with inferior information about optimal routes are taken on longer detours while asymmetric information on the local tariff system leads to manipulated bills amongst a sample of Greek taxi drivers.

Regarding service availability, possible indicators include service provision in areas and at times where it did not previously exist (e.g. remote areas, non-peak times), reduced waiting times as well as an increase in the number of taxi trips taken per individual passengers. Prices as well as the emergence of differential pricing components (thus giving customers more choice to reduce cost by choosing time of travel and pick up points) may be affected too, although any such effects are more likely to come from changes in labour supply to the occupation (see for example Morrison 1997).

Quality can be assessed using indicators of service innovation such as expansion in the types of services offered by taxi drivers (e.g. executive services, taxi vans and large carriers) or addition of new payment methods and credit systems (e.g. electronic payments, business accounts) which attempt to make it easier for customers to use the service. Finally, customer satisfaction data such as customer ratings and complaints would also be suitable indicators of quality as perceived by service users.

In practice, it is very rare (if at all) to see all these quality indicators used in studies that attempt to assess the effects of regulation of the taxi driving profession. This is not only due to the scarcity of such deregulatory interventions (i.e. specifically targeting the dimensions of professional regulation such as driver skills and knowledge), but also due to the difficulty in simultaneously obtaining data on all these indicators and concerns about its quality (when such data can be obtained).

---

would only have minor effect on learning on the job. Since we cannot observe the investments of drivers in higher skills, we cannot measure the regulatory effect on learning on the job.

[100] From all Uber data, observations from London and Dublin were selected if they have rating and telematics information (some observations do not have both). The random sample for London was obtained using the simple random sampling method with replacement by repeating the following steps ten times: 1. random numbers were generated and assigned to each observation. 2. A subset of these observations was selected for the final sample according to the population ratio between London and Dublin.

[101] The specific data are available on the time and dates of pickups and drop-offs from Uber. The data we used had at least 1,000 pickups and drop-offs for every time and day and several thousand for peak periods.

For the purposes of this study we can collect data on the following indicators. First, we obtain data on the quality of ride-hailing rides through driver quality ratings for the most widely used Uber service, which is UberX. After completing a trip, riders rate the driver on a scale of one to five stars, with one star associated with the lowest quality and a five with the highest. We treat this as a 'measure of process' or satisfaction since it reflects the customer experience from the ride. Second, we take consumer safety as an important dimension of quality of service received. We treat correlates of the safety of the journey as an 'outcome indicator' of quality. We employ two measures of safety made available to us by Uber: (i) hard accelerations and (ii) hard braking on individual trips. We deem these to be representative indicators of safety as in the relevant literature, hard braking and accelerations have been shown to be some of the most predictive factors for future car crashes and accidents (Claims Journal, 2015). The values are obtained when drivers, who use the Uber app, have a passenger in the vehicle and have the app turned on. The information is gathered and stored and available for use by researchers. In our analysis, we include controls for journey duration and distance. Uber drivers in London could take generally longer for the same distance or generally go on longer trips than in Dublin. Including these variables allows us to measure the differences of quality of service that are not associated with differences in journey time or distance. For example, the quality of a ride could deteriorate due to congestion, but not differences in regulatory intensity. To account for this, we also use 37 dummy variables for year and month (as one variable), (2) six dummies indicating day of the week (Sunday-Saturday), and (3) 23 dummy variables for hour of travel. If congestion occurs regularly at a given time during a day, these time fixed effects would pick up this effect. In addition to the time fixed effects, we control for location-specific effects.[102] For instance, if a statistically identical driver in Dublin on a trip of identical duration and distance breaks more frequently using a dirt road than in London on a paved road, the behaviour results from the characteristics of the route chosen for the specific trip and not from differences in regulatory intensity. To account for this, we use a set of 1023 dummy variables clustering pickup and drop-off location by longitude and latitude (see Horton, 2016).[103]

6.6 Methodology

The analysis uses multivariate techniques to examine how the variation in licensing requirements affects the quality of the ride as perceived by Uber riders and the safety outcomes of rides provided by drivers operating on the Uber app for London and Dublin. Since the sample sizes of London and Dublin vary, we use boot-strapping and jackknife resampling (see Appendix E for a description) to account for the potential differences in the samples across both cities. We present estimates of our results using the boot-strapping approach to adjust the standard errors using this technique. We compare the ratings received by drivers who have gone through the full licensing process (Dublin) with the ratings received by drivers on the same general Uber app service with lesser requirements than full black cab licensing (London). We use ordered probit models on the pooled trip-level observations to estimate the effect of regulation and other factors on the customer rating of a ride. Our basic model relates (latent) quality outcomes and measures of safety with asterisks to observable variables for each individual trip t:

$$Q_t^* = \beta_0 + \beta_1 City_t + \beta_2 X_t' + \eta_t + \nu_t + \epsilon_t \quad (1)$$

$$\begin{Bmatrix} HB_t^* \\ HA_t^* \\ HB20_t^* \\ HA20_t^* \end{Bmatrix} = \beta_0 + \beta_1 City_t + \beta_2 Y' + \eta_t + \nu_t + \epsilon_t \quad (2)$$

The first set of outcome variables are in equation 1) Q denoting the quality rating given to the driver by the customer. In the second set shown in equation (2) HA -is a measure of the number

[102] The idea is to control for changes in quality due to the characteristics of a specific trip in the same way as the Least Squares Dummy Variable (LSDV) Estimator does. The LSDV estimator is numerically identical with the Fixed Effects estimator and consistent under the same assumptions. We do not report the LSDV estimators of the additional parameters for the driver-specific dummy variables, because they are inconsistent. In the non-linear model, we cannot include the dummy variables due to the incidental parameters problem. Instead, we include location clusters.

[103] The algorithm developed by Horton uses Python as its engine to generate a map with number of observations in a form of matrix, and then forms rectangular groups that have approximately the same total number of observation in the matrix using the TripMatchR' program (Horton, 2016).

of hard accelerations and $HA20_t^*$ *is* if the fraction of hard accelerations on a trip was over 20 percent of all accelerations, HB_t^* *is the number of hard brakes and* $HB20$ denotes if the fraction of hard brakes on a trip is greater than 20%. $City$ is an indicator variable denoting whether a ride is performed by a driver in a city which imposes full licensing (Dublin), or in London, where the driver is not required to have a full black cab licence. X' is a vector of covariates controlling for observable characteristics that may impact the dependent variable on a ride, which includes, trip distance (miles), trip duration (seconds), overall fare amount (USD), driver experience (previous number of Uber trips), rider experience (previous number of Uber trips), and gender, Y' is a vector of trip characteristics such as trip distance, trip duration, and driver experience η are time fixed effects, which include month between November 2013 and December 2016, day of the week, and hour fixed effects. Even though observations may vary by time of day and day of the week, there are well over one thousand observations for each time period, and over one hundred thousand for each day of the week[104]. Moreover, geographical clusters ν are included. In our preferred ordered probit models, additional cut-off parameters need to be estimated for each but one outcome of the dependent variable. The very large number of observations in each of the cities allows us to specify very flexible models including a large number of variables. We have 1023 location dummies that cover both Dublin and London. Even with time dummies, we still have over 160,000 observations per city. Consequently, issues of degrees of freedom would not be an econometric issue. The fuller listing of variables in our model are presented in Appendix A.

6.7 Results

In Table 4 we show the means and standard deviations for the key variables in our analysis of the two cities, Dublin and London. As can be seen by looking at the means and standard deviations for trip duration and trip distance, the two samples are similar so one can make meaningful comparisons between them. It is, however, interesting to see that London has a shorter estimated time of arrival at the client's pickup location, despite being a busier city with more traffic.[105] A likely reason for this is that since London has more drivers, they may be closer to the rider than drivers in Dublin. Both riders and drivers are also more experienced on the Uber app platform in London, indicating that it is an important control variable in the regression analysis as Londoners are more likely to use the service compared to those living in Dublin, but driver ratings are very similar and as we will discuss below, highly skewed.

Turning to indicators of service availability, we can estimate the population to Uber driver ratio for both cities. This information is summarized in Table 5. We pool together data on the city's total population in 2016 and divide it by the total number of Uber drivers in the period 2013–2016 (controlling for attrition) for each city. As can be seen from the table, the ratio of the population to Uber drivers in Dublin is more than twice the ratio of population to Uber drivers in London. This is in line with our expectations given the more stringent entry regime in Dublin.

Table 5: Estimates of Driver to Population

	London	Dublin
Total Number of Licensed drivers who use the Uber app (2013-2016)	61,892	1,609
Population (2016)	8.788 million	527,612
Proportion of Population to Driver	142	328

[104] Even though the observations vary by time of the day and day of the week there are well over one thousand observations for each day and hour time period, and over one-hundred thousand observations for each day of the week. Using "TripMatchR", we attempted to generate clusters per one geography variable (pickup and drop-off). When the distribution of the location (longitude and latitude) is ideal, then we can generate all 1024 clusters in the form of dummy variables. However, in most cases (based on academic research in the US (Hall, et. al.2018), we generally got less number of clusters than what we attempted (even though the numbers are still big enough). Since we obtained around 928,000 observations for both cities, the approximate number of observations per cluster is around 928,000/1,024=906.

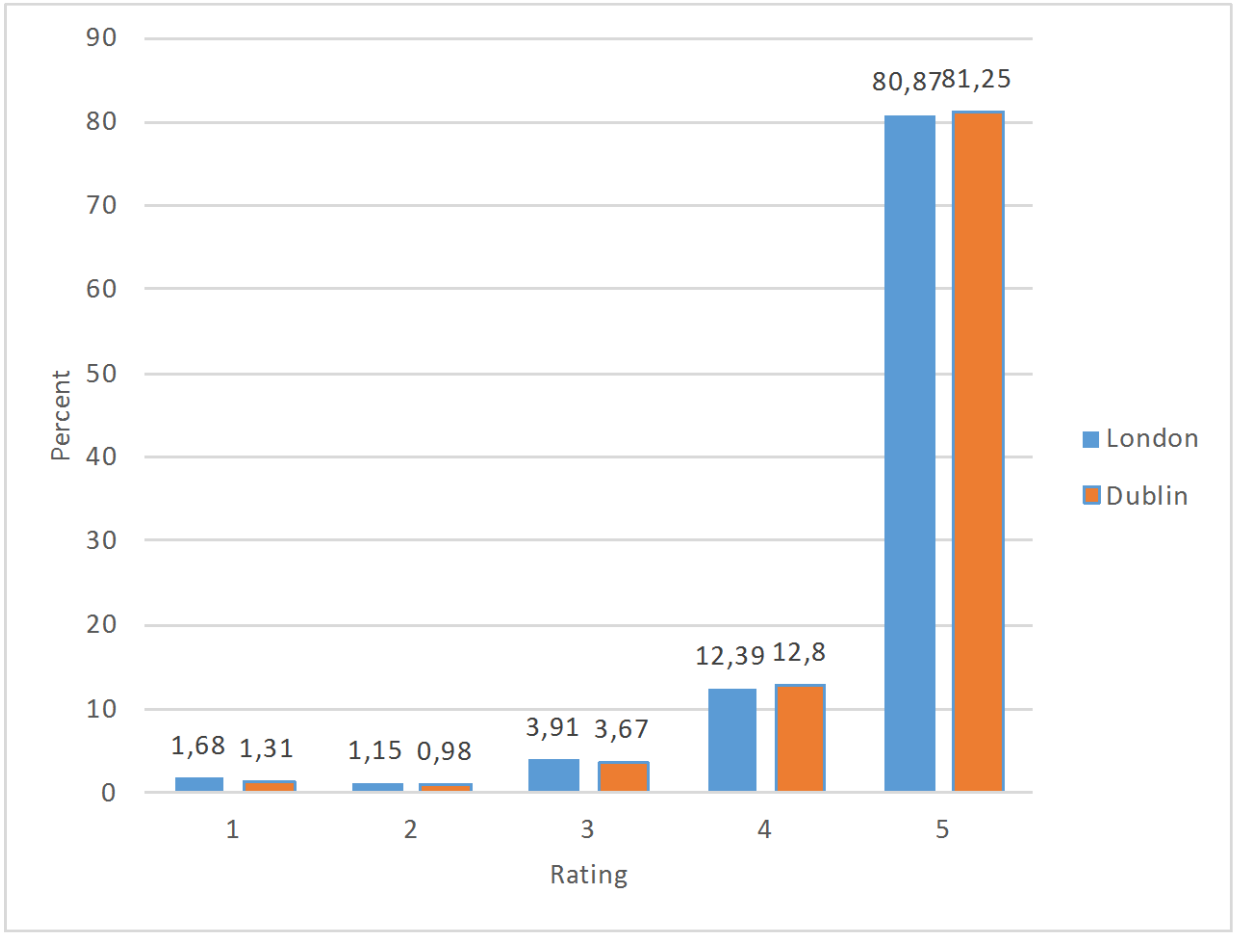
[105] There was one outlier in Dublin, likely due to a coding error, which increases the standard errors, and deleting it has little impact on the other variables.

Table 4: Means and Standard Deviations for the Key Variables in our Analysis for Dublin and London

Variable	City	N	Mean	SD
Trip Duration	Both	928,721	1,218.7450	756.8347
(in seconds)	London	471,877	1,219.9190	757.5157
	Dublin	456,834	959.8389	527.3873
Trip Distance	Both	928,721	4.6582	3.7284
(in miles)	London	471,877	4.6594	3.7300
	Dublin	456,834	4.3947	3.3273
Trip Fare	Both	928,721	17.7627	10.1188
(in US$)	London	471,877	17.8833	10.7795
	Dublin	456,834	17.6381	9.3861
Surge	Both	928,721	1.0705	0.2067
Multiplier	London	471,877	1.0755	0.2171
	Dublin	456,834	1.0001	0.0063
Estimated	Both	928,308	204.5293	143.5293
Time of Arrival	London	471,877	204.3132	143.6638
(in seconds)	Dublin	456,834	252.1905	135.6856
Driver	Both	928,721	1920.9100	1,766.2100
Experience	London	471,877	1927.8160	1,767.0310
(# of trips)	Dublin	456,834	397.9217	387.2624
Rider	Both	928,721	94.3996	148.0589
Experience	London	471,877	94.5536	148.1818
(# of trips)	Dublin	456,834	60.4299	112.8493
Rating	Both	617,573	4.6962	0.7485
(1-5 stars)	London	314,255	4.6961	0.7487
	Dublin	303,318	4.7169	0.7014
Fraction of Hard	Both	456,352	0.0824	0.0974
Brakes	London	240,890	0.0823	0.0974
	Dublin	215,462	0.0960	0.1135
Fraction of Hard	Both	347,694	0.0560	0.0818
Acceleration	London	183,678	0.0560	0.0818
	Dublin	164,016	0.0607	0.0902
Fraction of Hard	Both	456,352	0.1245	0.3302
Brake	London	240,890	0.1243	0.3300
over 0.2	Dublin	215,462	0.1727	0.3780
Fraction of Hard	Both	347,694	0.0677	0.2512
Acceleration	London	183,678	0.0676	0.2511
over 0.2	Dublin	164,016	0.0851	0.2791

Finally, we explore how customer satisfaction with the rides compares in both cities. Figure 1 includes data for all rated personal transportation trips completed in Dublin and London between November 2013 and December 2016. This depicts the overall skewness of the ratings system towards five-star ratings with the distribution of driver quality ratings being relatively consistent across the two locales. As can be seen from Figure 1, quality ratings are highly right skewed with nearly 81 per cent of trips receiving a five-star rating using both Dublin and London. In contrast, less than 2 per cent of trips receive one star. As can be seen, no substantial differences in how customers rate their experience in Dublin and London can be observed. This could indicate that the more intense training that ride-hailing drivers are subjected to in Dublin does not translate to better customer evaluations as one would expect.

Figure 1: Ratings Distribution on UberX Trips in Dublin and London (November 2013– December 2016)

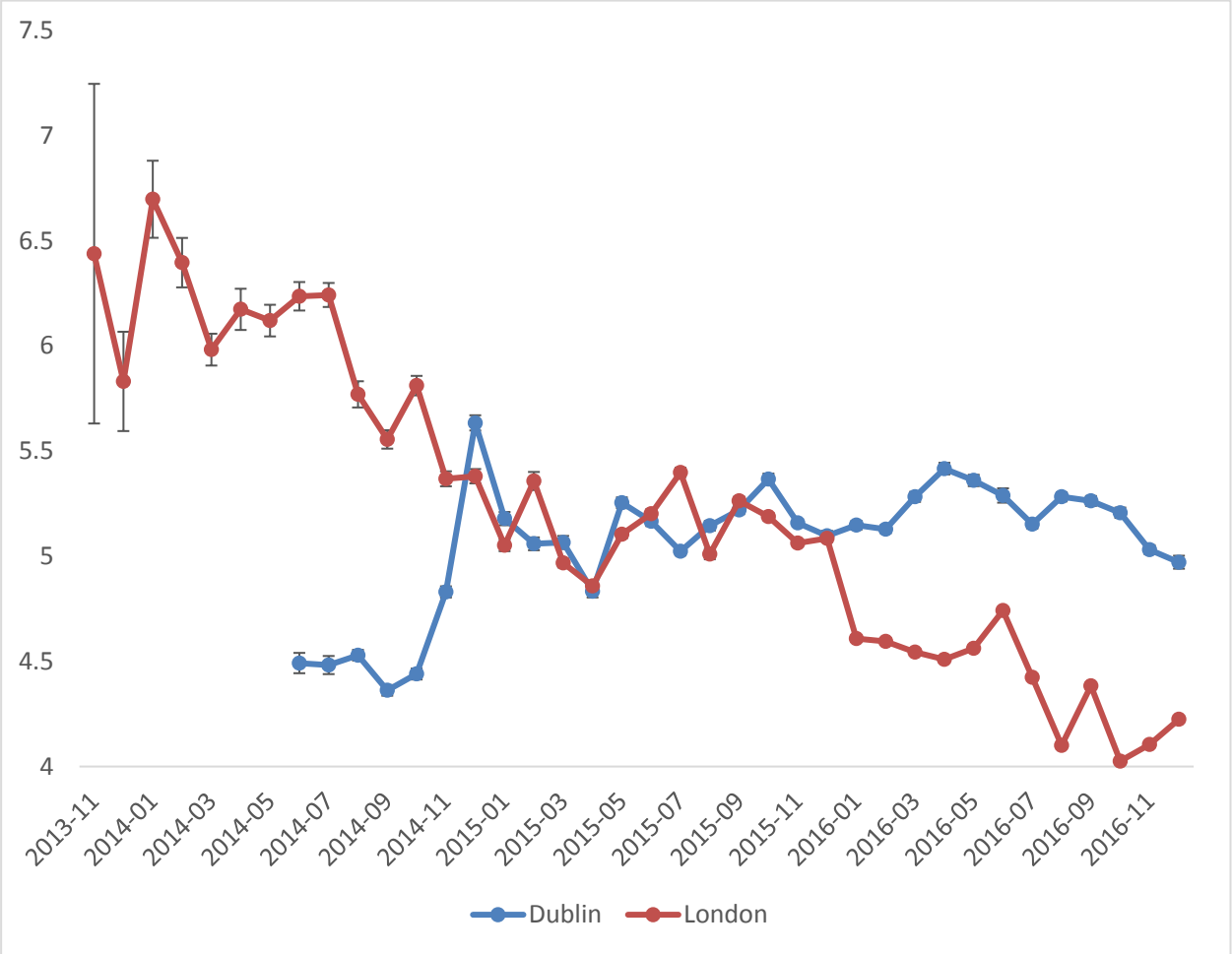


Finally, we examine the average per-mile fare in the two cities to gauge how the price of ride-hailing services by Uber compares between the two cities. Descriptively, we find that over the period from November 2013 through December 2016 the average base price in U.S. dollars based on the Uber administrative data was $5.58 in Dublin with a range from a low of $3 to a high of $10. In contrast, the average base price was $3.31 in London. The range of prices was $2.47 to $4 for the base price based on time and location. When surge pricing is considered, which is a measure of spot price based on supply and demand, the values in Dublin increase to a mean of $8.08, with a range from $4.16 to $25. The equivalent mean London prices are $5.81 with a range from $3.90 to $9.80. Using both methods of measuring base prices with and without the surge component used by Uber, the prices are higher in Dublin than in London.

Figure 2 depicts this analysis for the period 2013 through 2016. Initially, the London fare was higher, but they converged in 2015, and then the London fares were lower throughout 2016 and stayed lower throughout the year. Although in the latest period London prices were rising slightly, this seems to be in part due to the greater variations in London Uber services over the period examined. In contrast, Dublin fares have remained relatively stable (with some minor fluctuations)

during the same period. One explanation for the lower prices in London relates to the relatively higher supply of drivers compared to Dublin, which as discussed before is likely to be an outcome of the less stringent regulatory framework.

Figure 2: Average Uber Fare (in US$) per mile in Dublin and London (2013–2016)



Measures of quality take various forms in different industries. In this analysis, it is the satisfaction of the customer and the safety of the ride. If more relaxed occupational regulation has an influence on the quality of the ride, it is through these two outcomes in our models. The place where the ride takes place, however, is also an important determinant of both customer satisfaction and the safety of the ride. Higher levels of congestion could be associated with more customer frustration, and a greater chance of a collision. Therefore, where the ride takes place within the city is a key variable in our analysis of the role for regulation in ride-hailing. We use the method developed for Uber by Information, Operations and Management Sciences Professor John Horton at New York University to construct a grid based on the longitude and latitude of the place of the ride. Location dummies were generated by the R package called 'TripMatchR' (the package is not publicly available, but we were able to use it in our analysis by permission from Uber). The program generates a map with numbers of observations in a form of a matrix, and then forms rectangular groups that have approximately the same total number of observations to generate the location of the ride. More specifically the TripMatchR algorithmically implements a geography-based clustering approach that partitions trip locales into “regions” or “clusters” by partitioning the trips into iso-count. In Appendix F we provide more detailed information on the procedure used to develop the location variable. The approach has been used in other academic analysis of quality and demand for ride-sharing (Hall et. al., 2018 and Hall et.al. 2017). In our analysis we have the most confidence in the estimates that control for the location of the ride.

Regression estimates of the overall models for customer satisfaction and for the telematics data are provided in Tables 6, 7 and 8.[106] From left to right the columns start from a specification with only the city indicator, include control variables, time fixed effects, and finally location cluster indicators. Using ordered probit models for customer satisfaction,[107] the results show that Dublin drivers (where relative regulation is more stringent) have statistically significantly higher driver ratings in the first three columns, but insignificant estimates in the last one, when location of the pickup and drop off are accounted for. All the models are statistically significant using a standard chi-squared test.[108]

The results are consistent with satisfaction being about 3 percentage points higher in the relatively more highly regulated Dublin than in London in column one showing the results of the baseline specification that only includes the indicator variable for Dublin. However, since as previously discussed, the mean is skewed towards five, this is a very small improvement in the satisfaction of the rider in the more heavily regulated city of Dublin relative to London. Consider the probabilities to get a three- and a five-star rating, which are according to Figure 1 about four percent for three-star ratings and about 81 percent for five-star ratings. If, for example, higher regulation reduces three-star ratings by 3 percentage points and increases the five-star ratings by 3 percentage points, the overall distribution would not change very much, since the five-star ratings occur with very high probability in the first place. However, when trip location is added to the models (in column 4) the results suggest that London and Dublin have no significant difference on customer satisfaction. The location variable picks up both the effects specific to the location of the rides and partly the effects of the stringency of the regulation. Since there is no overlap between the two cities in observations, part of the result reflects stringency as well as any unique characteristics of the city not contained in our observable results. However, the location of the rides would likely be a serious omitted variable if left out of the analysis. The variable controls for factors such as high-density traffic and poor road conditions and other similar factors in estimating the quality of the ride. By leaving out this factor, regulation could pick up the variance in outcomes between the two cities, when the real difference were the conditions of the location of the ride.

As the demand for taxi services is highly elastic with respect to price, we include a full set of time fixed effects (Hall, Horton and Knoepel, 2017). The indicators for each month between November 2013 and December 2016 account for effects resulting from Uber's rollout, business cycle and seasonal effects. The day of the week fixed effects capture behavioural differences during a week. For instance, a person might be in an exceptionally good mood looking forward to the weekend on each Friday night or in an exceptionally bad mood on each Monday morning. The hour fixed effects capture fluctuations in quality that arise during a day. For example, rush hour related congestions might drive down the quality indicators at a given time each day. The time fixed effects account for the congestion issue, and the analysis now turns to further indicators of safety, namely hard accelerations and hard braking.

---

[106] An important aspect of regression outputs is the overall goodness of fit of a model. A standard measure of goodness of fit in the linear regression model is R^2. However, it is not an appropriate measure for the models with discrete dependent variables, such as the ones we estimate in this report. While there are many alternative ways to assess goodness of fit in non-linear models, McFadden (1973) pseudo-R^2 is one of the most widespread. This measure is defined as $R^2 = 1 - {\ln L_{fit}}/{\ln L_0}$, where $\ln L_{fit}$ is the log-likelihood function for the estimated model, and $\ln L_0$is the log-likelihood function in the model with only an intercept. If the covariates have no explanatory power, the pseudo-R^2 is zero, just as the usual R^2. Since for multinomial models, the maximum value of the log-likelihood theoretically is zero, McFadden's pseudo-R^2 can be interpreted as the fraction of the maximum potential gain in log-likelihood that is achieved by the fitted model. As we note a statistically significant Chi-squared result for all our models, as a result, we are able to reject the null hypothesis of no significance.

[107] We also estimate a linear probability model as a robustness check. Results remain the same.

[108] The number of observations differs between specifications due to missing values for some of the independent or dependent variables in each of the estimated models.

Table 6. Ordered Probit Estimates on Customer Satisfaction

	Rating			
Variable	(1)	(2)	(3)	(4)
City	0.0262***	0.0475***	0.0515***	-0.0600
(1=Dublin)	(0.00353)	(0.00437)	(0.00472)	(0.0512)
Trip Duration		-9.71e-05***	-8.94e-05***	-8.97e-05***
(in seconds)		(5.91e-06)	(6.40e-06)	(6.62e-06)
Trip Distance		0.0291***	0.0267***	0.0267***
(in miles)		(0.00149)	(0.00159)	(0.00167)
Client Fare		-0.00700***	-0.00670***	-0.00675***
		(0.000547)	(0.000604)	(0.000604)
Driver Experience		7.88e-06***	6.76e-06***	6.77e-06***
(# of Trips)		(1.48e-06)	(1.58e-06)	(1.58e-06)
Gender		-0.0886***	-0.0938***	-0.0936***
(1=Male)		(0.0270)	(0.0270)	(0.0270)
Rider Experience		0.000842***	0.000847***	0.000849***
(# of Trips)		(2.33e-05)	(2.38e-05)	(2.39e-05)
Constant Cut 1	-2.125***	-2.239***	-1.694***	-1.692***
	(0.00548)	(0.0279)	(0.260)	(0.271)
Constant Cut 2	-1.906***	-2.018***	-1.473***	-1.471***
	(0.00455)	(0.0277)	(0.260)	(0.271)
Constant Cut 3	-1.496***	-1.604***	-1.058***	-1.057***
	(0.00342)	(0.0276)	(0.260)	(0.271)
Constant Cut 4	-0.873***	-0.975***	-0.428*	-0.427
	(0.00257)	(0.0274)	(0.260)	(0.271)
Observations	617,573	617,573	617,573	617,573
Pseudo R^2	0.0000	0.0075	0.0081	0.0082
Time FE			✓	✓
Location				✓

Notes: Robust standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

To more directly assess safety, we also used telematics dependent variables in our analyses. These variables include the fraction of hard brakes on a trip ($\frac{\#\ of\ hard\ breaking\ events\ on\ a\ trip}{Total\ \#\ of\ breaking\ events\ on\ a\ trip}$), the fraction of hard accelerations on a trip,$\left(\frac{\#\ of\ hard\ acceleration\ events\ on\ a\ trip}{Total\ \#\ of\ acceleration\ events\ on\ a\ trip}\right)$, and indicators for whether a trip had greater than 20% hard brakes or greater than 20% hard accelerations. Any brake or acceleration on a trip with a force greater than 3.06 m/s2 is considered a hard brake or hard acceleration, which is consistent with transportation industry standards (Claims Journal 2015). Both sets of telematics metrics are used because the distribution of the percentage of hard brakes or hard accelerations on trips could influence the proportion of trips that are identified as “safe” trips using the 20% hard brakes or hard accelerations threshold. For example, the proportion of trips with greater than 20% hard brakes may differ significantly if hard braking events tend to occur in relatively high or relatively low proportions on individual trips versus if hard braking events tend to be relatively evenly distributed across trips. Importantly, hard braking has been shown to be one of the most predictive factors for predicting future crashes (Claims Journal 2015). Braking and acceleration events and the rates of braking and acceleration on individual trips are identified through the Uber app on a driver’s smartphone. Both sets of telematics metrics are used in panels A and B (Table 7). The estimates show that there is no difference in the fraction of hard

accelerations between Dublin and London when the location of the trip is controlled for in the analysis. Moreover, the fractions of all rides where 20 per cent or more of the accelerations are hard are not significantly lower in Dublin than in London when relevant factors are controlled for.

Table 7: Hard Acceleration, OLS and Probit

	Hard Acceleration			
Variable	(1)	(2)	(3)	(4)
City	0.00471***	-0.000742**	-0.000639*	0.00228
(1=Dublin)	(0.000293)	(0.000369)	(0.000380)	(0.00263)
Trip Duration		-1.01e-05***	-8.47e-06***	-8.72e-06***
(in seconds)		(3.56e-07)	(3.76e-07)	(3.99e-07)
Trip Distance		0.00261***	0.00225***	0.00234***
(in miles)		(7.15e-05)	(7.68e-05)	(8.55e-05)
Driver Experience		-1.86e-06***	-1.78e-06***	-1.78e-06***
(# of Trips)		(9.01e-08)	(9.05e-08)	(9.06e-08)
Gender		0.00439**	0.00416**	0.00416**
(1=Male)		(0.00172)	(0.00172)	(0.00172)
Constant	0.0560***	0.0563***	0.0570***	0.0441***
	(0.000191)	(0.00175)	(0.00218)	(0.00477)
Observations	347,694	347,694	347,694	347,694
R^2	0.000	0.008	0.012	0.012
Time FE			✓	✓
Location				✓
	Frac(Hard Acceleration)>0.2			
Variable	(1)	(2)	(3)	(4)
City	0.122***	-0.0216***	-0.0115	0.0466
(1=Dublin)	(0.00630)	(0.00787)	(0.00817)	(0.0513)
Trip Duration		-0.000628***	-0.000596***	-0.000592***
(in seconds)		(1.42e-05)	(1.46e-05)	(1.54e-05)
Trip Distance		0.0569***	0.0501***	0.0494***
(in miles)		(0.00216)	(0.00231)	(0.00253)
Driver Experience		-3.02e-05***	-3.06e-05***	-3.05e-05***
(# of Trips)		(2.40e-06)	(2.42e-06)	(2.42e-06)
Gender		0.0560	0.0539	0.0541
(1=Male)		(0.0453)	(0.0454)	(0.0454)
Constant	-1.494***	-1.055***	-1.070***	-1.451***
	(0.00448)	(0.0461)	(0.0554)	(0.123)
Observations	347,694	347,694	347,694	347,694
R^2	0.0000	0.0410	0.0428	0.0429
Time FE			✓	✓
Location				✓

Notes: Robust standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Table 8: Hard Brakes, OLS and Probit

	Hard Brakes			
Variable	(1)	(2)	(3)	(4)
City	0.0137***	0.00716***	0.00798***	0.00218
(1=Dublin)	(0.000315)	(0.000386)	(0.000396)	(0.00306)
Trip Duration		-2.16e-05***	-1.81e-05***	-1.68e-05***
(in seconds)		(3.73e-07)	(3.90e-07)	(4.11e-07)
Trip Distance		0.00523***	0.00445***	0.00420***
(in miles)		(7.44e-05)	(7.91e-05)	(8.79e-05)
Driver Experience		-1.39e-06***	-1.07e-06***	-1.03e-06***
(# of Trips)		(9.62e-08)	(9.66e-08)	(9.66e-08)
Gender		0.00622***	0.00513***	0.00513***
(1=Male)		(0.00186)	(0.00184)	(0.00184)
Constant	0.0823***	0.0809***	0.0829***	0.100***
	(0.000198)	(0.00189)	(0.00268)	(0.00596)
Observations	456,352	456,352	456,352	456,352
R^2	0.000	0.017	0.029	0.030
Time FE			✓	✓
Location				✓
	Frac(Hard Brakes)>0.2			
Variable	(1)	(2)	(3)	(4)
City	0.210***	0.0796***	0.0919***	0.0379
(1=Dublin)	(0.00457)	(0.00570)	(0.00598)	(0.0420)
Trip Duration		-0.000678***	-0.000636***	-0.000616***
(in seconds)		(9.76e-06)	(1.01e-05)	(1.05e-05)
Trip Distance		0.0795***	0.0702***	0.0663***
(in miles)		(0.00152)	(0.00161)	(0.00176)
Driver Experience		-1.48e-05***	-1.22e-05***	-1.17e-05***
(# of Trips)		(1.72e-06)	(1.75e-06)	(1.75e-06)
Gender		0.0257	0.0118	0.0113
(1=Male)		(0.0336)	(0.0337)	(0.0337)
Constant	-1.154***	-0.759***	-0.856***	-0.735***
	(0.00328)	(0.0341)	(0.0517)	(0.0901)
Observations	456,352	456,352	456,352	456,352
R^2	0.0001	0.0399	0.0469	0.0473
Time FE			✓	✓
Location				✓

Notes: Robust standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

The results show a higher probability of 'hard breaking' in Dublin (columns (1) through (3) of Table 8), even controlling for base fare, trip distance (miles), trip duration (seconds), gender, driver experience (trips a driver conducted in their lifetime for Uber), rider experience (number of Uber trips of the rider who issued the rating) and time fixed effects. However, once the location fixed

effects are included in the model, we do not find a statistically significant difference between the two regulatory regimes, i.e. safety is not higher in the more regulated Dublin than in London. Regarding the overall fit of the model, since it is practically never the case that the log-likelihood equals zero, even small values of this measure indicate a reasonable fit of the basic model given the statistical technique used. In other words, the results suggest that the rides are no smoother in Dublin where the intensity of regulation is higher than in London.

6.8 Conclusions

The case study examines how variations in the regulatory requirements affect quality outcomes. In particular, the analysis exploits how the differences in the stringency of becoming a licensed driver offering ride-hailing services in London and Dublin affect one 'process indicator' of quality (customer satisfaction) and two 'outcome indicators' (safety as measured by hard accelerations and hard braking) obtained from Uber using the drivers' app with controls for area of the city, time of the day, day of week, week of the year. Each of these locales has different levels of stringency for obtaining an occupational licence with barriers to entry being higher in Dublin compared to London. The analysis further describes and explores service availability and prices as 'value-added' measures of quality. The approach that was used allows for the examination of the role of regulation when one area provides a more relaxed form of occupational regulation. Data is collected directly from Uber administrative data on satisfaction, hard breaking, and rapid acceleration for the largest online ride-hailing platform in the world and a prominent example of the workings in the gig economy.

Our descriptive results suggest that Uber service availability as measured by the Uber driver to population ratio is higher in London and prices are lower. From a theoretical perspective, these findings are not surprising since one would expect higher supply of Uber ride-hailing services to be associated with lower prices. Further, customer ratings are only slightly higher in the more stringently regulated Dublin and, when location of the ride is controlled for, the difference is not statistically significant, which suggests that higher barriers do not seem to be correlated with customer satisfaction. Turning to indicators of safety, while we find a higher percentage of trips with hard breaking in Dublin, this relationship ceases to be statistically significant when we control for location of the trip. Regarding hard accelerations, the results also show that users have no smoother journeys in Dublin where the regulations are more stringent when controls for the location of the trip are accounted for, in the empirical analysis across different specifications and statistical models. As such, based on these two safety indicators, we find no justification for the considerably higher hurdles that taxi drivers must pass to legally operate in Dublin.

From a methodological perspective, although we do control for the area of the city, time, day of the week and month of the year, there may be other factors that may be important and different across the two cities that our analysis is not capturing. At the same time, the use of the location of the trip as an additional control variable may be masking some of the influence of the regulations. Future research with additional cities that are for example adjacent should be examined, where regulations differ, and location-specific data also are available. This approach would add to our ability to measure the influence of regulation on quality outcomes in the market. Further, the study only examines two cities (and by implication their respective regulatory arrangements) which limit the generalizability of the results in other member states that do not share the same regulatory provisions. Further, as our discussion of quality indicators has demonstrated, there are various other measures that are relevant to these debates for which data is not available. We acknowledge these as limitations of the current study. Nevertheless, these limitations open avenues for future research and continued discussions of these issues.

Overall, this research finds little evidence for greater regulatory constraints (at least of the type observed in these two jurisdictions) on ride-hailing drivers. Further, the technological context within which the taxi and ride-hailing driving professions operate is considerably different to what was the case when the regulatory provisions for taxi drivers were initially being drawn up. GPS technology, for example, provides drivers with detailed route information as well as information on traffic congestion, thus enabling them to conduct more efficient journeys. The same technology is available to customers, so information asymmetry is lower. As such, it is questionable whether regulations with detailed topographical knowledge are as relevant as they were in the past. Technologies allowing customers to rate their trips and the driver can mitigate problems associated with experience goods by ensuring that service standards are maintained. As such, our broad conclusion is that a reassessment of the form that regulating entry to the taxi and ride-hailing driving professions (and occupations that share similar characteristics, for example with respect to the role of technology in reducing information asymmetries) should take, and the level at which these are set is a fruitful path for policy makers to take.

References

Balafoutas, L., Beck, A., Kerschbamer, R., and Sutter, M. (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. Review of Economic Studies, 80(3), 876-891.

Barrett, S.D (2010). The Sustained Impacts of Taxi Deregulation. Economic Affairs, 5,2.

Barrett, S. D. (2003). Regulatory capture, property rights and taxi deregulation: a case study. *Economic Affairs*, *23*(4), 34-40.

Beesley, M. E. (1973). Regulation of taxis. The Economic Journal, 83(329), 150-172.

Cairns, R. D., and Liston-Heyes, C. (1996). Competition and regulation in the taxi industry. Journal of Public Economics, 59(1), 1-15.

Carpenter, D. M., Knepper, L., Erickson, A. C., and Ross, J. K. (2012). License to work: A national study of burdens from occupational licensing. Institute for Social Justice.

Cervero, R. (1985). Deregulating urban transportation. Cato Journal, 5, 219.

Claims Journal. (2015). Hard Braking Most Likely Predictor of Future Crashes: Progressive Data. Retrieved from: https://www.claimsjournal.com/news/national/2015/05/19/263424.htm

Dempsey, P. S. (1996). Taxi Industry Regulation, Deregulation & (and) Reregulation: The Paradox of Market Failure. Transport Law Journal, 24, 73.

DMR Statistics. (2017). Accessed June 28, 2017 http://expandedramblings.com/index.php/uber-statistics/

Douglas, G. W. (1972). Price regulation and optimal service standards: The taxicab industry. Journal of Transport Economics and Policy, 116-127.

Financial Times. (2017). “How Uber and London ended up in a taxi war”, September 27. Accessed 29 March 2018.

Flores-Guri, D. (2003). An economic analysis of regulated taxicab markets. Review of Industrial Organization, 23(3-4), 255-266.

Gallick, E. C., and Sisk, D. E. (1987). A reconsideration of taxi regulation. Journal of Law Economics. And Organization, 3, 117.

Hall, J., Hicks, J., Kleiner, M. and Solomon, R. (2018). “Occupational Licensing of Uber Drivers,” Paper presented at the American Economic Association Meetings, Philadelphia, Pa., January 6.

Hall, J. Horton, J. and D.T. Knoepel (2017). Labor Market Equilibration: Evidence from Uber, Working Paper.

Harding, S., Kandlikar, M., & Gulati, S. (2016). Taxi apps, regulation, and the market for taxi journeys. Transportation Research Part A: Policy and Practice, 88, 15-25.

Horton, John, J (2016) "TripMatchR" Version 1.0 http://john-joseph-horton.com/software/

The Official Manual for Operating in the SPSV Industry, 5th Edition – May 2015, Irish Department for Transport, available at: https://www.nationaltransport.ie/taxi-and-bus-licensing/taxi/spsv-driver-licensing/spsv-entry-test/

Kleiner, M. (2006). Licensing Occupations: Ensuring Quality or Restricting Competition? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Koumenta, M. and Pagliero, M. (2016). “Measuring Prevalence and Labour Market Impacts of Occupational Regulation in the EU”, European Commission Report.

Morrison, P. S. (1997). Restructuring effects of deregulation: the case of the New Zealand taxi industry. Environment and planning A, 29(5), 913-928.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior, in Frontiers in Econometrics, P. Zarembka (Ed.), New York, Academic Press.

Pautler, P. A., and Frankena, M. W. (1984). An Economic Analysis of Taxicab Regulation. Bureau of Economics Staff Report.

Rojeck, M. and Masior, M. (2016) The Effects of Reforms Liberalising Professional Requirements in Poland, European Commission Report. Available at: https://ec.europa.eu/growth/content/effects-reforms-regulatory-requirements-access-professions-country-based-case-studies-0_en

Satariano, Adam, (2018). " Uber Regains Its License to Operate in London, a Win for Its New C.E.O".. New York Times,, June 26. https://www.nytimes.com/2018/06/26/technology/uber-london.html

Schaller, B. (2007). Entry controls in taxi regulation: Implications of US and Canadian experience for taxi regulation and deregulation. Transport Policy, 14(6), 490-506.

Shreiber, C. (1975). The Economic Reasons for Price and Entry Regulation of Taxicabs. Journal of Transport Economics and Policy, 268-279.

Teal, R. F., & Berglund, M. (1987). The impacts of taxicab deregulation in the USA. Journal of transport economics and policy, 37-56.

Wall Street Journal (2017). Uber's Financials: An Inside Look, Available from http://www.wsj.com/graphics/uber-financials/

The Official Manual for Operating in the SPSV Industry, 5th Edition – May 2015

Appendix A

Table A.1: Variable Description

Variable	Description
City	Indicator for city, 0=London and 1=Dublin
Trip Duration	Time spent for the driver to provide service from the departure location to destination, in seconds
Trip Distance	Distance travelled for providing the service from the departure location to destination, in miles
Client Fare	Amount charged for the given service, in US$
Driver Experience	Number of trips that the driver provided on the Uber platform until the time he/she provided the current service
Gender	Gender of the driver, 0=Female and 1=Male
Rider Experience	Number of trips that the rider received on the Uber platform until the time he/she provided the current service
Month	37 dummy variables for each month between November 2013 and December 2016.
Day of Week	6 dummy variables for day of the week, i.e. for Monday through Sunday.
Hour	23 dummy variables for hour of the day.
Begin Trip Cluster	Dummy variables generated by R packaged (TripMatchR) developed by Dr. John Horton of New York University that generates 1023 Boolean variables that indicate geographical clusters for the pickup and drop-off locations. see Horton, John, J (2016) "TripMatchR" Version 1.0
End Trip Cluster	

Appendix B

Uber Technologies Inc. is a ride-hailing company headquartered in San Francisco, California, with operations in 633 cities worldwide. Its platforms can be accessed via its websites and mobile apps. Uber began its first rides in 2010 in San Francisco and in New York City in 2011, to match individuals who needed rides to work or recreation with those individuals who were willing to provide those rides for a price. Uber expanded outside the US in 2013, and London became one of its largest and most successful operations with about 40,000 drivers and more than 3.5 million customers. In the past the firm has had up to 80 per cent of the total ride-sharing (cab-substitute) market in London and the UK and US, but the Dublin estimates have been harder to obtain (DMR Statistics 2017; Financial Times 2017). However, those numbers have declined with more competitive substitutes.

The creation of an app and accompanying software allowed this process of matching to be done and valuable data to be collected that allows us to compare indicators of quality for different levels of stringency in entry requirements for taxi drivers. The cost to the driver of the matching process is that Uber takes a percentage of the ride price for the company as their fee for the matching process (Hall, et. al. 2018). Under this business model, ride-hailing drivers pay a proportion of their fares to the ride-hailing platform operator (in this case Uber), a commission-based compensation model used by many internet-mediated service providers. To Uber drivers, this commission is known as the Uber fee. By contrast, traditional taxi drivers in most cities make a fixed payment independent of their earnings, usually a weekly or daily medallion lease, but keep every fare dollar net of expenses.

In 2017 the company had more than 1,500,000 active drivers worldwide. For Uber, drivers (Uber refers to them as ‘driver-partners’) provide transportation services to customers requesting rides via Uber’s app on their smartphones or other devices.

Appendix C

Sample questions – Industry Knowledge in Dublin

1. Which of the following describes a taxi standing for hire?

a. The taxi is parked on a road

b. The taxi is stationary and available for hire

c. The taxi is in motion and available for hire

d. The taxi has a passenger and is on route

2. The Fixed Payment Offence 'Failure to display tamper-proof disc' carries a penalty of:

a. €40

b. €60

c. €80

d. €250

3. All SPSV drivers must provide the National Transport Authority with details of the SPSV they will be operating. How many weeks prior to the expiry date are drivers advised to renew their SPSV driver licences?

a. 6 weeks

b. 8 weeks

c. 10 weeks

d. 12 weeks

4. The charge per kilometre on Tariff A Premium Rate of the National Maximum Taxi Fare is:

a. €1.35

b. €1.45

c. €1.10

d. €1.75

5. The charge per kilometre on Tariff B Premium Rate of the National Maximum Taxi Fare is:

a. €1.35

b. €1.80

c. €1.50

d. €1.75

6. The charge per kilometre on Tariff A Special Premium Rate of the National Maximum Taxi Fare is:

a. €1.45

b. €1.80

c. €1.75

d. €1.95

8. Which of the following is found on a taxi's In-Vehicle Information card?

a. Vehicle registration number

b. Operator's name

c. Taxi licence number

d. Information on how to understand a taxi fare

9. On a map with a scale of 1:450000, a distance measured on the map is

a. 450 times smaller than it is on the ground

b. 4,500 times smaller than it is on the ground

c. 45,000 times smaller than it is on the ground

d. 450,000 times smaller than it is on the ground

Sample questions – Area Knowledge

URBAN CENTRES – ONE-WAY STREETS.

Thomas Davis Street runs one-way between which of the following streets?

a. John Dillon Street to Francis Street

b. Francis Street to Meath Street

c. Francis Street to John Dillon Street

d. Meath Street to Francis Street

URBAN CENTRES – DISTRICTS

Taking the most direct route from Blackrock to Merrion Square, which of the following districts would you pass?

a. Booterstown, Merrion and Ballsbridge

b. Booterstown, Mount Merrion and Donnybrook

c. Stillorgan, Mount Merrion and Donnybrook

d. Merrion, Sandymount and Irishtown

URBAN CENTRES – HOUSING ESTATES

Off which road in Arklow is The Maples housing estate located?

a. Dublin Road

b. Sea Road

c. Coolgreaney Road

d. Wexford Road

TOWNS AND VILLAGES/NATIONAL ROADS AND MOTORWAYS

Travelling on the N25 from Midleton towards Youghal, which of the following town or village centres would you pass through?

a. Castlemartyr, Killeagh

b. Ballynacora, Cloyne

c. Carrigtwohill (Carrigtohill), Glounthaune

d. Ladysbridge and Gortaroo

PLACES OF INTEREST – HOSPITALS

On which street is the Coombe Hospital located?

a. Dolphin Road

b. Dolphin's Barn Street

c. Cork Street

d. Parnell Road

PLACES OF INTEREST – PUBLIC BUILDINGS AND AMENITIES

Where is Nenagh Town Hall located?

a. MacDonagh Street

b. Banba Square

c. Sarsfield Street

d. Kenyon Street

PLACES OF INTEREST – MAJOR LANDMARKS AND TOURIST ATTRACTIONS

Which of the following is NOT a tourist attraction in Cork?

a. Derrynane House

b. Dursey Island Cable Car

c. Schull Planetarium

d. The Jameson Experience

URBAN CENTRES – STREET NAMES

In Clonmel, which street does NOT adjoin Queen Street?

a. St Oliver Plunkett Terrace

b. Bolton Street

c. William Street

d. Cashel Road

NATIONAL ROADS AND MOTORWAYS

Which national road connects Galway City and Headford?

a. N6

b. N17

c. N59

d. N84

Appendix D

Sample Questions: London Topographical Test (map and magnifying glass supplied)

MODULE 1: IDENTIFYING LOCATIONS

This section will assess your ability on how to use the indexes in an atlas.

Example Question: Identify the grid reference and page number of Parliament Street, London.

MODULE 2: ROUTE SELECTION (UP TO 5 MILES)

This section will assess your ability to plan and write a route between two locations and answer questions relating to your route.

Example Question: Write down the most direct route from London Zoo to Piccadilly Circus. Please include all road names and directions you take.

MODULE 3: ROUTE SELECTION (UP TO 15 MILES)

This section will assess your ability to plan and write a route between two locations and answer questions relating to your route.

Example Question: Write down the most direct route from Oxford Circus to Stratford New Town. Please include all road names and directions you take.

MODULE 4: ROUTE SELECTION (UP TO 30 MILES)

This section will assess your ability to plan and write a route between two locations.

Example Question: Plan a route from Buckingham Palace to Heathrow Airport Terminal 4 using the most direct route and/or major arterial roads.

MODULE 5: GENERAL TOPOGRAPHY

In this section you will be required to demonstrate your knowledge of the Greater London area, examples may include knowledge of major arterial roads and/or directions of major towns and cities from/to central London.

You may also be required to answer compass-based questions.

Note: Central London is defined as Trafalgar Square.

Example Question 1: What direction is London City Airport from central London?

Appendix E

We use two standard ways to estimate the standard errors of our estimates: The Jacknife method and bootstrapping. The notion of the jackknife method is to repeatedly calculate the estimate, each time omitting one observation or a cluster. Assume that $\hat{\vartheta}$ is the estimate based on the entire sample and $\widehat{\vartheta_j}$ is the estimate based on the subsample resulting from leaving out the jth observation or cluster, where $j = 1,2,\dots,N$. The jackknife procedure defines pseudovalues $\widehat{\vartheta_j^*} = \widehat{\vartheta_j} + N(\hat{\vartheta} - \widehat{\vartheta_j})$ giving the jackknife estimate

$$\overline{\vartheta^*} = 1/N \sum_{j=1}^{N} \widehat{\vartheta_j^*}.$$

From this the standard error is calculated as

$$\left(\frac{1}{N(N-1)} \sum_{j=1}^{N} (\widehat{\vartheta_j^*} - \overline{\vartheta^*})^2 \right)^{1/2}.$$

The bootstrap method is based on a very similar idea. However, instead of relying on pseudovalues, N observations are drawn with replacement from the sample with a total of N observations. In this random drawing, a new bootstrapped sample is generated with some of the original observations appearing once, some more than once, and some not at all. The process is repeated k times and the estimate recalculated from the ith bootstrapped sample $\widehat{\vartheta_i}$ is stored, giving k estimates. Using the average of the bootstrapped estimates $\bar{\bar{\vartheta}}$ the the standard error is calculated as

$$\left(\frac{1}{k} \sum_{i=1}^{k} (\widehat{\vartheta_i} - \bar{\bar{\vartheta}})^2 \right)^{1/2}.$$

The bootstrap method is regarded as more efficient and robust but can in contrast to the jackknife method not be used to detect outliers. In order to estimate bias-corrected and accelerated bootstrapped confidence intervals, the bootstrap technique is augmented with an additional parameter obtained with the jackknife method.

Estimates using this approach as a robustness check for the data in our sample are presented below and are consistent with our other estimated models.

Table A.2: Estimates of the measures of Driver Quality Using the Bootstrapping Model with the OLS Model

Variable	(1)	(2)	(3)	(4)
City	0.0208***	0.0303***	0.0307***	-0.000379
(1=Dublin)	(0.00184)	(0.00240)	(0.00243)	(0.00409)
Trip Duration		-5.31e-05***	-5.27e-05***	-5.41e-05***
(in seconds)		(2.44e-06)	(2.57e-06)	(2.53e-06)
Trip Distance		0.0161***	0.0159***	0.0163***
(in miles)		(0.000603)	(0.000629)	(0.000691)
Client Fare		-0.00422***	-0.00415***	-0.00418***
		(0.000257)	(0.000257)	(0.000255)
Driver Experience		4.67e-06***	4.66e-06***	4.61e-06***
(# of Trips)		(9.91e-07)	(8.71e-07)	(8.59e-07)
Gender		-0.0326***	-0.0350***	-0.0350***
(1=Male)		(0.00479)	(0.00416)	(0.00412)
Rider Experience		0.000335***	0.000336***	0.000336***
(# of trips)		(5.06e-06)	(4.27e-06)	(4.28e-06)
Constant	4.696***	4.749***	4.303***	4.245***
	(0.00140)	(0.00536)	(0.00756)	(0.0106)
Observations	617,573	617,573	617,573	617,573
R^2	0.000	0.008	0.009	0.009
Replications	1,000	1,000	1,000	1,000
Time FE			✓	✓
Location				✓

Notes: Bootstrap standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Table A.3: Fraction of Hard Acceleration, Bootstrap OLS

Variable	(1)	(2)	(3)	(4)
City	0.00471***	-0.000742**	-0.000639**	0.00228**
(1=Dublin)	(0.000273)	(0.000353)	(0.000311)	(0.00106)
Trip Duration		-1.01e-05***	-8.47e-06***	-8.72e-06***
(in seconds)		(3.54e-07)	(2.88e-07)	(2.93e-07)
Trip Distance		0.00261***	0.00225***	0.00234***
(in miles)		(6.66e-05)	(5.29e-05)	(5.79e-05)
Driver Experience		-1.86e-06***	-1.78e-06***	-1.78e-06***
(# of Trips)		(7.63e-08)	(6.64e-08)	(6.52e-08)
Gender		0.00439***	0.00416***	0.00416***
(1=Male)		(0.000196)	(0.000161)	(0.000157)
Constant	0.0560***	0.0563***	0.0570***	0.0441***
	(0.000178)	(0.000319)	(0.00103)	(0.00166)
Observations	347,694	347,694	347,694	347,694
R^2	0.000	0.008	0.012	0.012
Replications	1,000	1,000	1,000	1,000
Time FE			✓	✓
Location				✓

Notes: Bootstrap standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Table A.4: Fraction of Hard Acceleration, Bootstrap OLS

Variable	(1)	(2)	(3)	(4)
City	0.0137***	0.00716***	0.00798***	0.002184***
(1=Dublin)	(0.000306)	(0.000376)	(0.0003468)	(0.0011129)
Trip Duration		-2.16e-05***	-1.81e-.06***	-0.0000168***
(in seconds)		(3.55e-07)	(3.12e-07)	(3.45e-07)
Trip Distance		0.00523***	0.0044517***	0.0042032***
(in miles)		(6.09e-05)	(0.0000529)	(0.0000643)
Driver Experience		-1.39e-06***	-1.07e-06***	-1.03e-06***
(# of Trips)		(1.04e-07)	(8.46e-08)	(8.31e-08)
Gender		0.00622***	0.005129***	0.0051342***
(1=Male)		(0.000711)	(0.0007285)	(0.0007183)
Constant	0.0823***	0.0809***	0.082913***	0.1004078***
	(0.000194)	(0.000774)	(0.0015132)	(0.0018675)
Observations	456,352	456,352	456,352	456,352
R^2	0.000	0.017	0.029	0.030
Replications	1,000	1,000	1,000	1,000
Time FE			✓	✓
Location				✓

Notes: Bootstrap standard errors in parentheses. Statistical significance * p<0.05, ** p<0.01, *** p<0.001.

Appendix F

Location fixed effects were basically two variables - "begin trip location" and "end trip (or drop off) locations". The purpose of using these fixed effects were to control for the road conditions in some detail. Time fixed effects were used for controlling the macro-level change in the road conditions, such as different weather conditions (snow/rain/fog), and traffic (i.e. rush hour). However, time fixed effects could not control for the road and traffic condition at the street-level. For example, road condition in the London suburbs may not have 8-lanes or they may not have significant traffic from 8 a.m. until 6 p.m, but downtown London may have congested traffic conditions all day long. These sorts of road conditions cannot be controlled for unless we use location-specific variables that were generated using location-specific indicators such as longitude and latitude.

Both "begin trip" and "end trip" location variables were generated from the longitude and latitude at the beginning of the trip and the end of the trip locations using the TripMatchR algorithm. Uber saved these data, and there are very specific numbers that can be tracked down to the exact location in the TripMatchR. However, these numbers cannot be used directly since longitude and latitude are the indicators of two-dimensional space; that is the same latitude with different longitude may not have a perfect relationship. Therefore, we needed a way to group the locations into the parcels. For this step, we used the R package called "TripMatchR". It is a Python-based R package developed by Dr. Horton of New York University that is not publicly available (i.e. not downloadable through typical R package download procedure). How this package generates groups through the following:

(1) Plots all the locations into two-dimensional spaces using longitude and latitude

(2) The program uses the longitude and latitude into a scale then creates a matrix to indicate how many trips occurred in this area.

(3) It creates rectangular spaces that sum the values in each rectangle that is approximately the same

(4) Each number is the indicator of each space.

We generated the two location fixed effect variables and included it in each model. Although there is multi-collinearity between the variables and the city variables given the model there is no easy way to correct for this in the model.

Therefore, we can only control for the macro-change in the road and traffic conditions by using time fixed effect and controlling for the road and traffic conditions specifically since there are limitations for the approach that future research may be able to resolve.

(This page intentionally left blank)

II. Conclusions

Regulation of professions is an understudied area that affects product markets, labour markets and markets for education with the objective of improving quality of services. However, little is known about the empirical relation between regulation of professions and – broadly defined – quality of services. This study makes an exceptional attempt to contribute to the debate on the quality effects of regulation. We study six markets (lawyers, architects and engineers, pharmacists, tourist guides, driving instructors and ride hailing drivers) in six EU member states (Poland, Germany, Italy, Greece, UK and Ireland) and address the issue of measuring quality in each. Moreover, we exploit specific features in each regulated market to estimate the impact of regulations on quality.

Our work clearly shows that quality is not only difficult to measure, but that even the very definition of quality varies a lot across and within markets. The quality of some services can be verified before purchase, others during or after provision of the service, and for yet others quality can be verified only if additional costs are incurred, or not at all. Consumer satisfaction, complaints and peer judgments for lawyers; peer rankings for architects and engineers; product availability for pharmacists; consumer evaluations for tourist guides; service availability, pass rates and fatal accidents for driving instructors; and consumer satisfaction, hard accelerations and braking for taxi drivers are all examples of the different meanings that the idea of quality might take in different markets.

Moreover, in most markets quality is multidimensional, and even a longer list of measures might not fully capture the full experience of a customer. Still, we take on the challenge and try to measure quality using the available data. While admittedly imperfect, the long list of measures used throughout the study provides a surprising variety of sources of information on quality.

In order to estimate the impact of regulation on quality, one not only needs measures of quality, but also variability in the intensity of regulation. Our case studies focus on a number of reforms, but also exploit variability occurring across municipalities, regions and EU member states. An important lesson from our work is that there is more than one way to approach this empirical question. Taken together, our case studies make use of an extensive empirical methods toolkit: simple difference in means, difference-in-differences, synthetic control methods, panel data regressions, instrumental variable regressions and regression discontinuity designs. This variety in empirical methods reflects the variety of data used for the analysis, which came in the form of individual data from service satisfaction surveys (lawyers), peer ratings and census data (architects and engineers), confidential administrative data (pharmacists), labour force survey data and online booking website (tourist guides), administrative data (driving instructors) and even big data from a giant of the gig economy (ride-hailing drivers).

In the market for lawyers, we find little change in the overall quality of legal services following Polish relaxation of entry requirements, yet the number of complaints per active advocate did decrease and a decline in good manners before the court as a quality component was reported. Overall quality of services provided by architects and engineers decreased with higher market concentration in response to higher insurance costs and higher service prices. In the pharmacy market, the availability of pharmacies seems to be correlated with a decrease in the number of hospital admissions related to influenza, suggesting a possible link between the availability of the services offered by pharmacists and consumer health. Tourist guides' level of education increase post reform but the guides that entered the market via the new regime are more likely to receive lower consumer ratings. An increase in the stringency of regulations pushed many driving instructors out of the market with no evidence of improved quality of instructors or of learner drivers. Finally, more stringent licensing regulations for ride hailing drivers have no effect on customer satisfaction ratings or measures of hard breaking and accelerations (see Table II.1 for a summary).

Taken together, these case studies indicate that an increase in availability of service providers and/or competition does not have negative effects on the quality of the services provided or survey measures of consumer satisfaction and well-being. To the contrary, in a number of cases we find positive effects of increased availability and competition Still, our work does not provide final conclusions. We provide a series of examples on how to approach the issue of measuring quality and how to use existing data to investigate the impact of regulation on quality. As for any empirical work, it is difficult to extrapolate from a sample and more work in this area is certainly needed.

Table II.1. Summary of the six cases studies, the quality indicators, and the findings.

Occupation and Country	Nature of Regulatory Reform	Quality Indicator	Summary of Findings
Advocates and Legal Advisors (Poland)	Change in restrictions relating to educational requirements (more transparent bar exam, shorter bar training, access to bar without training for selected legal professionals)	Subjective assessment by clients and peers Complaints and disciplinary cases Enterprise creation and turnover Employment creation Price levels	Small effects
Architects and Civil Engineers (Germany)	Increase in price floors and ceilings Introduction of a minimum (professional indemnity) insurance coverage requirement	Score of an international peer-ranking of architectural firms Exit rates Continued education Professional-to-inhabitants-ratio Number of firm owners/employees	Negative effect
Pharmacists (Italy)	Relaxation of quantitative/structural restrictions	Availability of pharmacies at the municipality level	Positive effect
Tourist Guides (Greece)	Relaxation of the start-up process Modifications to educational requirements (broadening access to training while increasing required entry level qualifications (university degrees in a range of subjects))	Indicators of tourist guides' labour market performance Quality proxies based on customers' evaluations (index) and tourist guide profile	Mixed effects
Driving Instructors (UK)	Reform proposals related to restrictions on educational requirements	Service availability Indicators Indicators of student performance Road accidents Price levels	Mixed effects
Ride Hailing Drivers (London and Dublin)	Explore variations in the stringency of the regulatory regime	Customer satisfaction ratings Hard breaking and hard accelerations	No effects

Since the main obstacle to the study of the impact of regulation on quality is the lack of data, there could be significant benefits from more systematic data collection on this topic. At the EU level, specific surveys could provide evidence on consumer satisfaction in a variety of service markets. The survey approach has the advantage of providing comparable data across member states and over time, but has the disadvantage of capturing subjective evaluations. More objective quality

measures are generally occupation- and country-specific. These could be systematically collected with the collaboration of the professional associations or regulatory agencies in each member state. Regulatory agencies should always attempt to evaluate regulatory changes that could affect service quality, as well as prices and employment. Anticipating the need to evaluate policy changes, forward-looking regulatory agencies could then start collecting the appropriate data before the reform. More generally, they could implement it in such a way that evaluation ex post becomes easier. Finally, consumer associations could play an active role in monitoring the quality of services in many markets. They are in a unique and privileged position to collect information on quality of services, as they have a direct link with the final users.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);

- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries
  (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm)
  or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union
  (http://publications.europa.eu/others/agents/index_en.htm).

ET-04-18-894-EN-N

Publications Office

doi:10.2873/910094