

IMT School for Advanced Studies, Lucca
Lucca, Italy

Essays in Applied Economics

PhD Program in Institutions, Markets and Technologies
Track in Economics, Management and Data Science
XXXII Cycle

By
Duygu Buyukyazici
2022

The dissertation of Duygu Buyukyazici is approved.

PhD Program Coordinator: Prof. Rocco De Nicola, IMT School for Advanced Studies Lucca

Advisor: Prof. Francesco Serti, IMT School for Advanced Studies Lucca

The dissertation of Duygu Buyukyazici has been reviewed by:

Prof. Tiziano Razzolini, The University of Siena

Prof. Juan Alberto Sanchis Llopis, The University of Valencia

IMT School for Advanced Studies Lucca
2022

To my grandparents Gülbeyaz and Saadettin, my lovely sister Ebru and all the members of my beloved family.

Abstract

The present thesis consists of three independent chapters. The first chapter analyses the effect of religiosity on innovativeness. The empirical literature on the relation between religion and innovation hitherto applied regressions without considering endogeneity in the estimates, raising questions about spurious correlations. This chapter provides the first empirical study to build a causal link between religion and innovation by employing the instrumental variables method to untangle possible endogeneity. The results strongly suggest that higher religiosity has a somewhat negative effect on innovativeness. Three possible causality channels from religiosity to innovation are discussed: time allocation argument, the fear of uncertainty, and traditional roles empowered by religion. The second chapter explores the importance of regional capabilities, in the form of workplace skills, in the industrial diversification process of regions by exploiting two recently developed approaches: relatedness and economic complexity. Building on the network-based approach of evolutionary economic geography, the study shows that workplace skills form two highly polarised clusters into social-cognitive and technical-physical skills. The econometric analysis indicates that industries have a higher (lower) probability of developing a comparative advantage if their required skill set is (not) similar to those available in the region, regardless of the skill type. Nevertheless, similarity to technical-physical skills and higher complexity in social cognitive skills yield the highest regional competitive advantage probabilities. The third chapter analyses the Ramsey pricing of pharmaceuticals by using recently developed the debiased/double orthogonal machine learning method that allows for heterogeneous treatments in a dynamic panel setting. The study assesses the validity of the inverse elasticity rule by providing a cross-country analysis of pharmaceutical demand at the molecule level. The results show that pharmaceutical prices vary inversely with price elasticities, both in high-income and low-middle-income countries, signalling the existence of Ramsey pricing.

Contents

Abstract	vii
List of Figures	x
List of Tables	xi
Introduction	xii
1 Religiosity and Innovation Attitudes	1
1.1 Introduction	1
1.2 Literature	3
1.2.1 Religion, Economic Outcomes, and Innovation	3
1.2.2 Innovation as a Mindset	5
1.3 Estimation Strategy	7
1.3.1 Data	7
1.3.2 Measuring Innovation	8
1.3.3 Measuring and Instrumenting Religiosity	8
1.3.4 Dealing with Endogeneity	10
1.3.5 Model	12
1.4 Results	13
1.4.1 OLS Estimates	13
1.4.2 Instrumental Variables Estimates	16
1.4.3 Heterogeneity of Religiosity: Gender and Age	20
1.5 Sensitivity Analysis	23
1.5.1 Alternative Dependent Variable	23
1.5.2 Alternative Model Specifications	23
1.5.3 Reduced Sample Estimates	25
1.5.4 Possible Violations of Exogeneity	26
1.6 Possible Channels of Causality	29
1.6.1 Time Allocation	29
1.6.2 The Fear of Uncertainty	30
1.6.3 Roles Reinforced by Religion	32
1.7 Conclusion	34
2 Workplace Skills as Regional Capabilities: Relatedness, Complexity, and Industrial Diversification	36
2.1 Introduction	36
2.2 Literature	38
2.2.1 Regional Capabilities, Diversification, and Skill Relatedness	38
2.2.2 Regional Capabilities, Diversification, and Skill Complexity	40
2.3 Data	40
2.4 Construction of Skill Space	42
2.4.1 Skill Relatedness of Industries	42

2.4.2	Skill Complexity of Industries	47
2.4.3	Diversification Potential of Italian Regions	53
2.5	Econometric Analysis: Industrial Diversification	54
2.6	Robustness Checks	61
2.6.1	Average Marginal Effects: Linear Probability, Logistic, and Probit Models	61
2.6.2	Different Specifications for Dependent Variables	62
2.6.3	Alternative Model Specification: Industry–Time and Region–Time Fixed Effects	66
2.7	Concluding Remarks	66
3	Ramsey Pricing of Pharmaceuticals? A Debiased Machine Learning Analysis	69
3.1	Introduction	69
3.2	Literature Review	71
3.2.1	Second Best: Ramsey Pricing	71
3.2.2	Differential Pricing of Pharmaceuticals and Ramsey Pricing	72
3.2.3	Empirical Evidence of Price Elasticity of Demand	75
3.3	Data	76
3.4	Methodology	78
3.4.1	Double Machine Learning with Heterogeneous Treatments	78
3.4.2	Demand Model	80
3.4.3	Endogeneity and DML	82
3.5	Results	83
3.5.1	Overall Results	83
3.5.2	Ramsey Pricing? Price and Price Elasticity Relation	85
3.5.3	Ramsey Pricing? Income and Price Elasticity Relation	87
3.6	Sensitivity Analysis	88
3.6.1	Different Second Stage Models	88
3.6.2	Heterogeneity in Treatment Effect: Generic Pharmaceuticals	90
3.6.3	DML with Cross-Sectional Setting	92
3.6.4	Placebo test	95
3.7	Conclusion	96
	Conclusion	98
	Appendix: Religiosity and Innovation Attitudes	101
	Appendix: Workplace Skills as Regional Capabilities: Relatedness, Complexity, and Industrial Diversification	118
	Appendix: Ramsey Pricing of Pharmaceuticals? A Debiased Machine Learning Analysis	125

List of Figures

1	The Skill Space (2013-2019)	43
2	Example Industries and Effective Use of Skills	44
3	The Properties of the Skill Space	45
4	Skill Relatedness Density and Skill Complexity Index of Italian Regions	47
5	Differences in Relatedness Densities Across Industries, Regions, and Years	48
6	Skill Diversity of Industries and Ubiquity of Skills (2013-2019)	49
7	Interaction of Average Skill Diversity and Ubiquity of Effectively Used Skills (2013-2019)	51
8	Schematic Presentation of $K_{s,0} - K_{s,1}$ Space	51
9	Complexity Levels of Workplace Skills	52
10	Skill Complexity Scores of Industries	53
11	Diversification Potential of Italian Regions	54
12	Distributions of Independent Variables	56
13	Average Yearly Price of Standard Units by Country	77
14	Normalised Price Trends of Molecules	78
15	Price Elasticities by Country and Molecule	83
16	Price Elasticities by Molecule-Country Pairs	84
17	Relationship Between Price Elasticities and Price	85
18	Relationship Between Price Elasticities and Price for Low, Middle, and High Income Countries	86
19	Relationship Between Price Elasticities and Price of Molecules by Country	86
20	Relationship Between Price Elasticities and GDP Per Capita (PPP) Income	87
21	Relationship Between Price Elasticities and Income for Low, Middle, and High Income Countries	88
22	Different Second Stage Models	89
23	Price Elasticities by Country and Molecule for Generic Pharmaceuticals	90
24	Relationship Between Price Elasticities and Price for Generic Pharmaceuticals	91
25	Relationship Between Price Elasticities and GDP Per Capita (PPP) Income for Generic Pharmaceuticals	91
26	Price Elasticities by Molecule-Country Pairs for Generic Pharmaceuticals	92
27	DML Estimates: Relationship Between Price Elasticities and Price	93
28	DML Estimates: Relationship Between Price Elasticities and Price for Low, Middle, and High Income Countries	93
29	DML Estimates: Relationship Between Price Elasticities and GDP Per Capita (PPP) Income	94
30	DML Estimates: Relationship Between Price Elasticities and Income for Low, Middle, and High Income Countries	94
31	Heat Map of Placebo Price Elasticities Between the Molecules of Ecuador and Germany	95

List of Tables

1	OLS Estimates: Religiosity and Average Positive Innovation Attitudes	14
2	OLS Estimates: Religiosity and Average Negative Innovation Attitudes	15
3	OLS Estimates: Religious Denominations and Innovation Attitudes	16
4	IV Estimates: Religiosity and Average Positive Innovation Attitudes	17
5	IV Estimates: Religiosity and Average Negative Innovation Attitudes	18
6	Disaggregate IV Results for Age and Gender	21
7	Religiosity and Summary Innovation Attitudes	22
8	Alternative Model Specifications	24
9	Reduced Sample Estimates	25
10	IV Estimates: Fixed Effects for Religion Combinations	27
11	IV Estimates: Instruments of Non-Neighbouring Countries	28
12	IV Estimates: The Fear of Uncertainty	31
13	IV Estimates: Individualism-Collectivism	33
14	IV Estimates: Gender Roles	34
15	Entry and Exit Models: Full Skills	57
16	Entry and Exit Models: Social-Cognitive Skill Cluster	58
17	Entry and Exit Models: Technical-Physical Skill Cluster	59
18	Entry and Exit Models: Manufacturing and Service Industries	60
19	Average Marginal Effects for LPM, Logistic, and Probit Estimates	62
20	Entry and Exit Models: RCA in Number of Enterprises	64
21	Entry and Exit Models: Decrease of Time Span	65
22	Entry and Exit Models: Industry-Time, Region-Time Fixed Effects	66

Introduction

The present thesis consists of three independent chapters.

The first chapter analyses the effect of religion on innovation. Estimating the impact of religion is rather challenging due to potential inherent endogeneity that may stem from omitted variable bias, reverse causality, or measurement error. The empirical literature on the relation between religion and innovation hitherto applied regressions without considering endogeneity in the estimates, leading to spurious correlations. This chapter provides the first empirical study to build a causal link by untangling the endogeneity of the relation between religion and innovation. To this end, we employ several strategies. First, we adopt an individualistic approach and measure religion and innovation at the individual level to mitigate the impact of probable unobservables, including institutional level factors. Religion is addressed by personal religiosity, while innovation is measured by innovation attitudes, *i.e.*, personal traits that are favourable or unfavourable to innovation, presenting an individual's propensity to innovate. Second, we consider different types of religiosity and innovation attitudes to account for inherent heterogeneity in both concepts. Four variables are nominated to measure religiosity: self-reported degree of being religious, frequency of attending religious activities, frequency of praying, and an index of overall religiosity. Innovation attitudes are broken down into positive, *i.e.*, favourable to innovation—creativity, being free, being different, and being adventurous—and negative, *i.e.*, unfavourable to innovation—following traditions and following rules—innovation attitudes. Third, we employ the instrumental variable method. The degree of religiosity of individual i from country c who belongs to a given religion (does not belong to any religion) is instrumented by the average religiosity of people who belong to the same religion (do not belong to any religion), share the same age range and gender with i and live in countries that have the same dominant religious denomination of country c . Correspondingly, the instruments take into account both religious intensity and religious affiliation.

The two-stage least squares (2SLS) estimates strongly suggest that each measure of religiosity has a somewhat negative effect on innovation attitudes. Diagnostic tests and several sensitivity analyses support the main findings. We propose three causality channels from religiosity to innovation: time allocation argument, the fear of uncertainty, and traditional roles empowered by religion.

The second chapter explores the importance of regional capabilities, in the form of workplace skills, on the industrial diversification of regions by exploiting two recently developed approaches: skill relatedness and economic complexity.

The literature reaches a unanimous agreement that industrial diversification is path-dependent because new industries build on preexisting capabilities of regions that are partly embodied and reflected in the skills of regions' workforce. Although many scholars have underlined the significance of workplace skills in regional branching, empirical analysis rarely includes them. Moreover, little evidence is provided on the role of different skills. Building on the network-based approach of evolutionary economic geogra-

phy, we allow horizontal and vertical differentiation of workplace skills to unfold their inter-dependent and complex nature to enhance our understanding of their impact on industrial specialisation and regional diversification. We first construct skill relatedness and complexity measures to formalise them as a network, *i.e.*, the skill space, to make further structural analyses. Data-driven techniques we use reveal that workplace skills form two highly polarised clusters into social-cognitive and technical-physical skills. Afterwards, we perform econometric analyses to explore the effect of each detected skill cluster on the industrial diversification of regions. The entry and exit models we use indicate that industries have a higher (lower) probability of developing comparative advantage if their required skill set is (not) similar to those available in the region regardless of the skill type. We find systematic evidence that similarity to technical-physical skills and higher complexity in social cognitive skills yield the highest regional competitive advantage probabilities.

To the best of our knowledge, this chapter is the first empirical study that (1) explicitly analyses the relationship between skill relatedness based on workplace skills and industrial diversification, *i.e.* entry and exit probability, at the regional scale in a developed country; (2) examines the role of different skill types in industrial specialisation and regional diversification literature; (3) combines skill relatedness and complexity approaches; (4) provides skill complexity analyses for Italy.

The third chapter focuses on Ramsey pricing of pharmaceuticals by using the recently developed debiased/double orthogonal machine learning method that allows for heterogeneous treatments in a dynamic panel setting.

Differential pricing of pharmaceuticals has long been examined in the literature with three distinct theoretical approaches: price discriminating monopoly, Ramsey pricing, and value-based differential pricing. Among them, Ramsey pricing has gained theoretical attractiveness due to its ability to charge prices that vary inversely with the price elasticities of demand, known as the inverse elasticity rule. Despite theoretical foundations, no study provides empirical evidence on Ramsey pricing in pharmaceutical markets by testing the validity of the inverse elasticity rule. In other words, there is not any empirical evidence on whether pharmaceutical prices vary inversely with the price elasticity of demand. This chapter provides the first empirical study in this vein. It assesses the validity of the inverse elasticity rule by providing a cross-country analysis of pharmaceutical demand at the molecule level. We estimate the price elasticity of demand using a recently developed method of double/debiased orthogonal machine learning (DML) with heterogeneous treatment effects in a panel setting. The ability to use very large numbers of covariates decreases the probability of standard omitted variable bias, which was an essential drawback of conventional demand estimates. Moreover, the orthogonalisation step of the algorithm removes the influence of unobserved unit effects and other confounding factors from treatment and outcome, providing an effective solution to the long-standing problem of demand estimates: endogeneity of prices.

In order to test the validity of the inverse elasticity rule in pharmaceutical markets, we first analyse the relationship between pharmaceutical prices and the price elasticities of pharmaceutical demand at the country level. We find that prices vary inversely with price elasticities, both in high-income and middle-low income countries, meaning that the inverse elasticity rule holds. We then examine the relationship between country income and the price elasticities of pharmaceutical demand. The results indicate that price elasticity has a negative association with country income, signalling further evidence for Ramsey pricing in cross-country pharmaceutical markets. The main findings are robust to several sensitivity analyses.

To the best of our knowledge, this chapter is the first study that (1) estimates the cross-country price elasticities of pharmaceutical demand; (2) examines the validity of the inverse elasticity rule in pharmaceutical markets; (2) applies DML with heterogeneous treatment effects in a panel setting, excluding the applications for expository purposes.

Chapter 1

Religiosity and Innovation Attitudes¹

1.1 Introduction

There have been numerous studies on the economic consequences of religion since Max Weber's *The Protestant Ethic and the Spirit of Capitalism* (1905/2001), where he discusses that the emergence of modern capitalism in Europe was a product of Protestant Reformation that fostered certain traits of people including work ethic and thrift. Despite Weber's thesis being more of a verbal observation rather than an empirical argument, and many empirical studies prove Weber to be wrong (see Samuelsson, 1961; Becker and Woessmann, 2009), the idea that religion affects economic outcomes through shaping and transforming individual preferences and behaviours is far from being a myth. The most important connections between economics and religion happen through the effects of religion on economically important individual behaviours—including consumption patterns, saving patterns, time allocation decisions, marriage, fertility, gender roles in family and society—and traits—such as trust, honesty, thriftiness, tolerance to dissimilarity, willingness to work hard, openness to strangers, being prone to crime. The present chapter explores one dimension of this connection. It examines the effect of religion on innovation at the individual level by focusing on some economically important individual beliefs, behaviours, and traits that we call innovation attitudes.

To our knowledge, Benabou et al. (2013) is the first empirical study at the intersection of religion and innovation. Following their contribution, the research in this field has begun to flourish (Perret, 2014; Chen et al., 2014; Benabou et al., 2015; Huang et al., 2016; Cinnirella and Streb, 2017; Assouad and Parboteeah, 2018; Recio-Román et al., 2019). Previous studies have consistently demonstrated a correlation between religion and innovation yet recognise that observed associations probably do not identify whether religion has a causal impact on innovation due to the potential endogeneity of religion. Given the difficulty of defining and measuring all potentially non-ignorable factors, some unobservable components of the error term are likely to be correlated both with religion and innovation, leading to an omitted variable bias. Moreover, the process of causation is generally considered from religion to innovation. Nevertheless, also higher levels of innovation can affect religion through economic growth and development as conceptualised in the secularisation hypothesis (Iyer, 2016), raising the question of reverse causality.

Motivated by these premises, the present study introduces the first attempt to untangle the endogeneity of religion with respect to innovation, aiming to provide a causal interpretation of the effect of religion by focusing on individual religiosity and innova-

¹Joint work with Francesco Serti.

tion attitudes. We use eight waves —between 2002 and 2016— of the European Social Survey (ESS), an academically driven, cross-sectional, and an individual-level data set containing observations for 36 European countries. To our knowledge, the ESS has not been used by prior studies on innovation and religion. We define four different measures of religiosity —the degree of being religious, the frequency of attendance religious activities, the frequency of praying, and religiosity index, which is the principal component of the first three measures— since prior work has shown that different religiosity measures are likely to have different effects on economic outcomes (Barro and McCleary, 2006). Innovation is measured at the individual level with self-reported personal traits and beliefs related to the propensity to innovate of an individual. These innovation attitudes are broken down into positive and negative innovation attitudes. Positive innovation attitudes (PIA) are creativity, being free, being different, and being adventurous; negative innovation attitudes (NIA) are following traditions and following rules. Using attitudes, instead of an aggregate measure of innovation output such as patents per capita, allows the influence of institutional level confounding factors to be mitigated. Moreover, a large set of fixed effects is used to minimise the effect of (1) country-level economic and institutional confounding factors, (2) individual-level unobservables such as human capital and ability, and (3) global shocks. Finally, we instrument each measure of religiosity with the average religiosity of people of the same sex, age range, and religious affiliation who live in countries with the same dominant religious denomination. In other words, the religiosity of individual i from country c who belongs to a given religion (does not belong to any religion) is instrumented by the average religiosity of people who belong to the same religion (do not belong to any religion), share the same age range and gender with i , and live in countries that have the same dominant religious denomination with country c . By doing so, we aim to eliminate the effect of individual-level unobservables. Furthermore, we use religious affiliation along with religious intensity to mitigate reverse causality since religious affiliation is inherited and thus is not responsive to changes in other factors.

We start by estimating a linear model with OLS; then we apply our IV strategy. The first-stage results show that all the excluded instruments are strongly and significantly correlated with the religiosity measures. The OLS results show that each measure of religiosity is significantly and positively correlated with NIA, suggesting that religiosity has a negative effect on innovativeness. Regarding PIA, religiosity measures are positively correlated with creativity and being different, while negatively correlated with being free. In the IV results, the significant and positive association remains robust between the measures of religiosity and NIA. But the overall picture changes with PIA. The religiosity measures are no longer significantly related to creativity and being different. In contrast, the negative relation between religiosity and being free persists, meaning the OLS results suggesting that religiosity fosters some favourable attitudes to innovation may be driven by endogeneity. This main finding is robust to a series of sensitivity analyses.

Lastly, we discuss three possible causality channels from religion to innovation: time allocation, the fear of uncertainty, and roles reinforced by religion (individualism- collectivism and conventional gender roles). First, we consider the opportunity cost of time spent on religious activities, arguing that religious participation might negatively affect the human capital formation, which is essential to innovation. Since time is scarce, if an individual allocates a certain amount of time to religious activities, then there will be a decrease in the maximum amount of potential time devoted to human capital formation. Our empirical results support this argument. *Attendance in religious services* is the most

robust religiosity measure throughout the analysis. Second, we argue that for religious people, personal uncertainty is mitigated by their faith and by the support of religious institutions such as churches, mosques, or religious social organisations. Therefore, they have less experience dealing with risks and uncertainty inherent to the innovation process. Third, we assert that high religiosity may foster a collectivist culture that highly values established rules and might leave limited space for reformist and creative endeavours, negatively affecting innovation. We also consider the long-standing argument that religions foster gender discrimination by imposing stricter rules for women, influencing the access of women to education, liberty, labour force, and social and legal rights. We empirically test these hypotheses and find supporting evidence.

The remaining of the chapter is structured as follows. Section 1.2.1 briefly reviews the existing literature on religion and innovation. Section 1.2.2 looks at innovation from a behavioural point of view and describes innovation as a mindset, establishing the theoretical foundation for our innovation measures. Section 1.3 extensively describes the estimation strategy and discusses possible endogeneity channels together with our strategies to overcome them. Section 1.4 reports OLS and IV results and discusses the main findings. Section 1.5 introduces the sensitivity analyses. Section 1.6 discusses three possible channels from religion to innovation. Section 1.7 provides concluding remarks. In the Appendix, we present data summary tables and a detailed variable index (Section A), the first stage results for IV estimates (Section B), the OLS results for each innovation attitude (Section C), and the IV results for each innovation attitude (Section D).

1.2 Literature

The present chapter mainly relates to the expanding literature on the effects of religion on economic outcomes. The first subsection briefly overviews the literature on the relation between religion and economic outcomes, emphasizing innovation. The second subsection summarizes the literature on the individual and behavioural aspects of innovation, presenting a theoretical framework for our innovation measures.

1.2.1 Religion, Economic Outcomes, and Innovation

The interaction of economics and religion has long been a topic of sociological analyses. Until recently, economists have not focused on the topic, possibly due to a lack of reliable data on religion and potential methodological issues such as endogeneity. Nevertheless, recent decades have witnessed expanding literature on the economics of religion² in which religion has been integrated into economic research in three different ways: (i) applying the methodology of economics to religion (*e.g.* microeconomic theory to analyse religious behaviour of individuals, groups, or institutions); (ii) analyzing economic outcomes of religion (*e.g.* the effect of religiosity on economic growth); (iii) making use of holy books and theological norms to praise or criticize economic behaviours and economic policies (Iannaccone, 1998), (*e.g.* Islamic economics is critical about interest rates since Islamic law, *Sharia*, prohibits any interest paid on loans of money). The present study mainly contributes to the second category by examining the effect of religion on innovation.

The work of Benabou et al. (2013), to our knowledge, is the first empirical analysis of religion and innovation, while qualitative studies were present before (Kalliny and

²For an in-depth literature review, please refer to Iannaccone (1998) and Iyer (2016).

Hausman, 2007; Preble and Hoffman, 2012). Using the World Intellectual Property Organization (WIPO) data, they measure innovation with patents per capita. The two religiosity measures, namely, belief in God and being religious, are retrieved from the World Values Survey (WVS). Both cross-country and within-country (USA) analyses indicate a negative relationship between religiosity and innovativeness. In a follow-up paper, Benabou et al. (2015) enlarge the analysis by focusing on the innovation attitudes at the individual level instead of patents per capita which is an ex-post and macro-level measure. They define eleven different innovation attitudes and five alternative measures of religiosity that are retrieved from the WVS. They find an overall negative relation across 52 model specifications. Almost in every specification, greater religiosity is significantly associated with less favourable attitudes toward innovation. Though both papers do not directly try to deal with endogeneity, the latter study decreases the influence of institutional-level confounding factors by using individual-level data.

Perret (2014) examines the relationship between religious affiliation and innovativeness in Russia. Innovation is measured by the number of patents issued by the Russian Federal Service for Intellectual Property (Rospatent). He finds that only Hindu, Buddhist, and Jewish faiths exhibit significant and negative effects on innovativeness. Christians, Muslims, Jews, and Atheists do not display any significant impact at all.

Chen et al. (2014)³ analyse the relation between local gambling culture and corporate innovation. "Many innovative endeavours, *i.e.* attempts to come up with new products, services and methods, represent gambles because they promise relatively small probabilities of large success and large probabilities of failure" (Adhikari and Agrawal, 2016, pp. 229). They use religious beliefs as a measure of gambling preferences of firms' local communities. Taste for gambling is addressed by a higher Catholic-to-Protestant ratio assuming that Catholics are more likely to take risks. They find that the firms headquartered in areas with a taste for gambling tend to be more innovative.

Cinnirella and Streb (2017) analyse the effect of religious tolerance on innovation in Prussia by assuming that tolerance and diversity are conducive to technological creativity and innovation. Religious tolerance is proxied by the index of the religious diversity of the population across 1278 cities in Prussia. Innovation is measured by 1740 patents issued in Prussia between 1877 and 1890. They find that higher levels of religious tolerance have a strong positive impact on innovation during the second industrial revolution. They also show that the culture of tolerance did not stem from a particular denomination but rather from the presence of different denominations and churches.

The research at the intersection of religion and innovation has begun to flourish following the aforementioned studies (Huang et al., 2016; Assouad and Parboteeah, 2018; Recio-Román et al., 2019). However, no empirical study has considered the potential endogeneity of religion so far, despite reverse causality and identification problems have been recognized in the literature (Guiso et al., 2003).

The process of causation is generally considered as religion affects economic outcomes by playing an integral role in shaping culture, individual preferences, traits, and beliefs. However, the direction does not have to be from religion to economic outcomes. The secularization hypothesis suggests that economic outcomes, such as individual and country-level income and growth, may affect religious behaviour, presenting a possible reverse causality channel for our study. The core argument of secularization is that religiosity decreases as a nation gets richer. There has been a long debate on whether the secularization hypothesis holds. Many economists have analysed the main channels that

³This paper follows Kumar et al. (2011) and Kumar (2009) that find a positive relationship between the propensity to gamble and risky, lottery-like financial market investments.

might facilitate the secularization effect, yet it is unlikely to say there is a consensus in the literature. Ruiter and Tubergen (2009) argue that modernization reduces the need for religious reassurance because it enables the creation of more securities —whether it be financial, social, or political— for the population, thus decreasing the level of religiosity. Iannaccone (2008) provides valuable insights by analyzing the secularization with retrospective questions from 30 nations that stretch from the 1920s through the 1980s. She finds that secularization cannot explain the continued vitality of religion in the USA, one of the world's most modernized nations. On the contrary, she finds some favourable evidence for secularization in Britain, France, and Germany.

Apart from religion, there is a relatively vast literature on culture and innovation that has adopted the Hofstede model (1980/2001) (Shane, 1992, 1993, 1995; Gorodnichenko and Roland, 2010; Taylor and Wilson, 2012; Petrakis et al., 2015; Chen et al., 2017; Kostis et al., 2018). Hofstede analyses the differences in national cultures using the data of business employees from more than 50 countries. He empirically identifies and validates five independent dimensions affecting economic attitudes and outcomes —namely, power distance, uncertainty avoidance, individualism versus collectivism, masculinity versus femininity, long-term versus short-term orientation— on which national cultures differentiate ⁴. These five dimensions together form a model of differences among national cultures and have been pretty influential in the literature on culture and economic outcomes. Kirkman et al. (2006) review 180, and Sondergaard (1994) reviews 61 published studies that used Hofstede's dimensions, and they both confirmed their relevance to the differences in national cultures. In this chapter, we consider uncertainty avoidance and individualism-collectivism as potential causality channels from religion to innovation (see Section 1.6).

1.2.2 Innovation as a Mindset

Innovation can be defined in diverse forms. Wang and Ahmed (2004) review the literature on innovation and propose five different innovation types that jointly determine the overall innovativeness of an organisation: (1) product innovativeness (newness, novelty, originality, or uniqueness of products), (2) market innovativeness (newness of approaches that companies adopt to enter and exploit the targeted market), (3) process innovativeness (introduction of new production methods, new technology, and new management approaches that can be used to improve management and production processes), (4) strategic innovativeness (an organisation's ability to manage ambitious organisational objectives and to identify a mismatch of these ambitions and existing resources to use limited resources creatively) and (5) behavioural innovativeness (enables the formation of an innovative culture, the general internal capacity for new ideas and innovation and screens through teams, individuals, and management). Different types of innovation are expected to be related to different individual and organisational traits. Camps and Marques (2014) name these traits as innovation enablers, *i.e.* a set of general capabilities driven by social capital that contributes to favour innovation. They associate with innovation enablers and innovation types as follows: goal alignment and knowledge enhancement for product innovation; knowledge enhancement and cooperation for process innovation; cooperation and associability for strategic innovation; associability, risk-taking and creative environment for market innovation; risk-taking, creative environment, communication and information flow for behavioural innovation. Among

⁴Hofstede initially defines four dimensions. In the second edition of the book (2001), he adds *long-term versus short-term orientation* as the fifth dimension.

the innovation types, behavioural innovativeness is not only crucial for overall innovativeness but also for the other types of innovation, given that it enables the formation of an innovative culture by being present at the various levels for individuals, teams, and management (Wang and Ahmed, 2004). Cultivating an innovative culture through behavioural innovativeness induces other types of innovation mainly by creating an environment where new ideas are supported, failure is tolerated, creativity is promoted, and risk-taking is encouraged.

Many studies underline the importance of innovative culture and mindset by focusing on behavioural aspects of innovation. Goldstone (1987) argues that what separated the East from the West in the early modern world was not capitalism nor rationalisation of institutions; it was the willingness to innovate, which was fostered by revolting from orthodoxy and by cultivating tolerance to internal diversity, which enhanced openness to taking risks. He compares three historical crises in the 17th-century —the Stuart king crisis in England, Celali revolts in Ottoman Turkey, and revolts against the Ming dynasty in China—that have economic roots. He argues that England fostered tolerance to diversity and openness to take risks by adopting a relatively new, risk-taking path, which resulted in a higher propensity to innovate. On the contrary, Ottoman Turkey and China followed cultural orthodoxy and suppressed alternatives, resulting in an unfavourable environment to innovate. England, therefore, reached dynamism and growth, while Ottoman Turkey and China had stagnation. Even though the analysis of Goldstone is at the macro level, the cultural traits mentioned —tolerance to diversity, openness to taking risks, and revolting from tradition— are perfectly plausible at the individual level as well.

Dzallias and Blind (2019) review the extensive literature on organisational innovation indicators published between 1980 and 2015. They find that *innovation culture* has been one of the most used company-specific innovation indicators. Innovation culture is addressed by different indicator categories such as creativity, attitudes toward science and technology, social innovation climate and trust, support of new ideas, openness to new fields, the openness of the company towards change and innovation, resistance to change, willingness to exchange ideas, tolerance for innovation failures, et cetera. All these indicators are related to behavioural aspects of innovation.

Kahn (2018) underlines that innovation should be recognized as three different things: (1) *innovation is an outcome* (product innovation, process innovation, marketing innovation, business model innovation, supply chain innovation, organizational innovation), (2) *innovation is a process* (innovation process, product development process), and (3) *innovation is a mindset* (individual mindset, organization culture). “Mindset aligns employees and manifests the culture needed for innovation to happen. Encompassing a mindset that predisposes individuals and organisations to be risk-taking, cross-disciplinary, and open to varied ways of thinking helps establish the state necessary for innovation; state implies something habitual and lasting. It is about instilling and ingraining a mindset that prepares the individual and organisation for innovation so that there is proper engagement in the innovation process to achieve the desired innovation outcome.” (Kahn, 2018, pp. 459).

The literature mentioned above demonstrates that innovation is strongly associated with individual traits such as tolerance to diversity, risk-acceptance, creativity, and deviating from traditional paths and rules, which jointly form innovative cultures of teams, groups, and organisations. Nevertheless, previous research has generally measured innovation as an outcome, using patents, R&D expenditure and technology improvements’ expenditure at the country or firm level. There might be a couple of drawbacks in using

patents as a sole measure of innovation outcome. Moser (2012) argues that almost a thousand empirical studies have used patent counts to address innovation, mostly without controlling for variation across industries and over time. However, the percentage of patents varies significantly across industries. Furthermore, relying on the historical evidence, the majority of innovations were outside of the patent system, even in countries with a developed patent system, such as the mid 19th century USA (Moser, 2013). He uses the exhibition data of 8079 innovations between 1851 and 1915 and underlines that 89% of British innovations and 85% of American innovations were not patented. The probability of an innovation/invention being patented mainly depends on the country's patent system and law, among many other determinants. Furthermore, inventors do not have to use patents to protect their invention; they may use other property rights such as lead-time advantage and secrecy. Hence, patents as an indicator of innovation would be a biased measure in cross-country and cross-industry studies.

Another critical issue is that patents are ex-post measures of innovation that reflect country-specific institutional constraints. In other words, patents measure some portion of occurred innovation. However, they have little to say about the propensity to innovate, mainly determined by the behavioural and cultural traits mentioned above. One cannot address unhappened, *i.e.* blocked by unfavourable institutional constraints or cultural and religious traits, innovation with an ex-post measure. Many papers in this regard address culture and religion by individual measures —belief in god, degree of religiosity, willingness to take a risk, tolerance to diversity, et cetera— but address innovation by country-level patents, which is a macro and ex-post measure. Therefore some economists (Guiso et al., 2003, 2006; Esteban et al., 2015, 2018, 2019) suggest using individual attitudes as a measure of innovation rather than aggregate macro outcomes such as patents.

1.3 Estimation Strategy

1.3.1 Data

We use pooled cross-sectional data from the European Social Survey (ESS). The ESS is a biennial, academically-driven, and individual-level survey that includes large groups of observations about preferences, beliefs, and attitudes toward many different subjects, including immigration, religion, political choices, trust, and markets. The ESS is available for 36⁵ countries for nine waves from 2002 to 2018. We use the first eight waves that result in more than 380 000 observations. The survey does not homogeneously include all countries across all waves, meaning that some countries are not surveyed in certain waves and thus are represented in different proportions. Table A1 displays the surveyed countries across the waves.

The main variables we are interested in are presented in different groups below. Detailed descriptions of all variables are presented in Appendix A. Descriptive statistics of the variables are presented in Table A2.

⁵Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, and United Kingdom. We have dropped Israel from the sample since it is not located in mainland Europe and is the only dominantly Jewish country.

1.3.2 Measuring Innovation

Innovation is a complex phenomenon as religion and is challenging to measure due to its many dimensions. It is important to underline that innovation mainly occurs in two phases: initiation and implementation of innovation (Glynn, 1996; William and McGuire, 2010). The first phase is about creating ideas which is, in general, an individual task. On the other hand, the second phase might include teams, groups, and management activities, requiring an innovative organisational culture and mindset. Most papers on innovation literature focused on the second phase by using patents as the only indicator of innovation, as underlined in Section 1.2.2. On the contrary, our innovation measures are mainly related to the first phase. Nonetheless, individual innovation measures are likely to stay vital in the second phase as well.

Following Guiso et al. (2003), we address innovation by the individual propensity to innovate by introducing innovation attitudes that are designated under the guidance of the literature briefly reviewed in Section 1.2. In this regard, we exploit several variables in the ESS: *creative* (important to be creative), *different* (important to be different), *free* (important to be free), *adventurous* (important to be adventurous), *traditions* (important to follow traditions), *rules* (important to follow rules). We categorise these variables as *positive* or *negative innovation attitudes*. Positive innovation attitudes (PIA) are essential behavioural conditions and favourable personal traits to innovate. Negative innovation attitudes (NIA) are personal traits that contradict with an innovative mindset, thus are unfavourable to innovation. We nominate four favourable —*creative, different, free, adventurous*— and two unfavourable —*traditions, rules*— traits to innovation.

We create two average measures: *inposav* and *innegav*. *Inposav*, average positive innovation attitudes, is the mean of *iprcrtiv, different, free* and *rules*; while *innegav*, average negative innovation attitudes, is the mean of *traditions* and *rules*. *Innosum* is a summary measure of innovation attitudes computed as the sum of *creative, different, free, adventurous* minus *traditions* and *rules*.

Inposav and *innegav* are the primary dependent variables in our analysis, yet we also analyse each attitude that composes them. The corresponding tables of OLS and IV estimates of each attitude are presented in Appendix C and D for the sake of brevity. Detailed information on how we construct innovation variables is presented in Appendix A.

1.3.3 Measuring and Instrumenting Religiosity

The variables *belonging, pbelonging, denomination, pdenomination, degree, attendance, pray* are religion-related observations in the ESS. *Belonging, pbelonging* stand for respectively present and past belonging to a religion; *denomination, pdenomination* present respectively current and past denomination of an individual; *degree* is the degree of religiosity, *attendance* is the frequency of attendance to religious activities, *pray* is the frequency of praying. *Belonging, pbelonging, denomination, pdenomination* indicate religious affiliation —belonging to a particular religion or denomination— while *degree, attendance, pray* reveal religious intensity that we simply name religiosity.

Different measures of religiosity are likely to have different effects on socioeconomic variables, reflecting the multi-dimensional nature of religion (see Barro and McCleary, 2006). For instance, the religion market model (Finke and Stark, 1992; Finke and Iannaccone, 1993) suggests that if a state has a formal religion, imposing the state religion and limiting the other religions, participation in religious services is likely to decrease because the variety of services is subject to suffer (Barro and McCleary, 2006). Nevertheless, people might keep firm religious beliefs without attending state-regulated formal

religious services. Accordingly, low attendance may be caused by the supply side of the religion market and may not mean low religiosity. Motivated by this reasoning, we use three different measures of religious intensity (*degree, attendance, pray*) as explanatory variables. We also create an index of general religiosity, *religiosity*, that is the principal component of the three religious intensity variables.

Our IV strategy is inspired by Esteban et al. (2019) where they examine the role of religiosity, along with personal liberties, in influencing the decision of labour effort. They construct an instrumental variable for religious intensity by computing the average religious intensity of people of the same sex, age bracket, and religious denomination in neighbouring countries, assuming that religiosity is a cultural trait shared by people over national borders. By doing so, they could eliminate the possible omitted variable problem for individual unobservables, such as ability, because the instrument is other people's average religiosity, which is likely to be uncorrelated with individual i 's personal characteristics. We partly adopt this strategy to build instruments for religious intensity variables by computing the average religiosity of people of the same sex, age range, and religious affiliation who live in countries with the same dominant religious denomination. In other words, the religiosity of individual i from country x who belongs to a religion (does not belong to a religion) is instrumented by the average religiosity of people—who belong to a religion (do not belong to a religion), share the same age range and gender with i —who live in the countries that have the same dominant religious denomination with the country x has.

Briefly, our instruments are constructed by taking into account four elements: age range, gender, religious affiliation, and dominant religious denomination. We use 11 age ranges (15-20,..., 60-65, 65+)⁶. Gender is a dummy variable 0= male and 1=female. Religious affiliation is based on variable *belonging* which is a dummy, 0=does not belong to a religion and 1=belongs to a religion. We determine dominant religious denominations with the help of variables *denomination* and *country* after weighted by population size weights. We calculate the percentages of each religious denomination in each country in the sample and identify the religious denomination that has the highest percentage in a country as *the dominant religious denomination*. The dominant religious denomination of each country and corresponding percentages are presented in Table A1 in Appendix. There are five dominant religious denominations: Roman Catholics, Protestants, Eastern Orthodox, Muslims, and Atheists (do not belong to a religion).

The rationale behind our IV strategy emerges from the following points. The dominant religious denomination in a county is the leading actor that forms the characteristics of the religious environment in which individual preferences and beliefs are being shaped. Some individuals inherit more, some less, but the dominant religious denomination determines the rules. Thus, instrumenting individual i 's religiosity by the average religiosity of people who are subject to the same dominant religious denomination in their country (along with the same age range, gender, and affiliation) means that individual i 's average religiosity is instrumented by the average religiosity of people who live in a similar religious environment as that of i . Here we assume that, as Esteban et al., religiosity is not a national trait but a cultural trait that transcends national borders. People raised in a particular religious environment are likely to share a significant part of their culture with those raised in a similar religious environment regardless of national borders. However, they do not necessarily share the same institutional environment. We, therefore, use a large set of fixed effects to eliminate country-level institutional dif-

⁶We have also tried 6 age ranges (15-25, 25-35...55-65, 65+). The estimation results were pretty similar to those estimated with 11 age ranges.

ferences. Moreover, our sample consists of European countries that share many common values and customs.

We use religious affiliation variables, along with religious intensity variables when constructing instruments. Religious affiliation generally passes down from generation to generation and is thus inherited and subject to slow change. The ESS data justify this proposition. Only 9.8% of individuals in the sample have changed their religious denomination⁷, meaning that religious affiliation is pretty stable throughout one's lifetime and does not significantly change along with changes in other factors, including innovation. On the other hand, religious intensity is more likely to change over one's lifetime, making it more vulnerable to reverse causality⁸. Moreover, it is expected that the preferences and beliefs of an individual who belongs to a religion would be affected by the religious environment around her more than a person who does not belong to a religion.

Another crucial point is that the average religiosity of people who belong to a particular religious denomination may inherently be more intense than those who belong to another religious denomination. For instance, Muslims may be inherently more religious than Protestants⁹. We, therefore, use religious denomination fixed effects to be able to compare the religiosity of people from different religious denominations.

In our IV strategy, the exclusion restriction would be violated if belonging to a religion is not inherited from past generations but chosen during one's lifetime. In this case, belonging to a religion would be correlated with individual unobservables and thus cannot be considered exogenous. Nevertheless, as mentioned above, only 9.8% of our sample have changed their religious affiliation (once-believers). As a robustness check, we drop once-believers from the main sample and re-estimate the main specifications (Table 9, columns 7 and 8). The results are very similar to the full sample. Another scenario in which our instruments violate the exclusion restriction is that countries with the same dominant religious denomination are subject to correlated shocks. We use survey year fixed effects to capture global shocks along with country and country-survey year interaction fixed effects.

It is likely that within-group observations are correlated. In our case, those groups are the five elements we use to construct instruments: age, gender, religious affiliation, denominations, and country. For instance, the religiosity of a Catholic woman at age 36 from a dominantly Catholic country could be correlated to the religiosities of other middle-aged Catholic women from dominantly Catholic countries. It is known that ignoring within-group correlations leads to understated standard errors (Shore-Sheppard, 1996). We thus use clustered standard errors at the instruments' level to avoid within-group correlations in error terms.

1.3.4 Dealing with Endogeneity

Religious beliefs are embedded in social behaviour and culture, meaning that extracting the pure effect of religion can be challenging. One, therefore, must extensively consider endogeneity when working on religion. We ultimately address the possible inherent endogeneity of religion by employing the instrumental variable strategy. In addition, we consider the primary sources of endogeneity below and discuss how we overcome them.

⁷Calculated by using the observations of *denomination* and *pdenomination*.

⁸Botticini and Eckstein (2005) and Guiso et al. (2006) underline that not only religious affiliation but also religious practices are modified only over centuries; thus any aspect of religion can largely be assumed invariant over one's lifetime.

⁹The data support this proposition. For instance, Table A3 in Appendix A shows that Muslim countries exhibit significantly higher means for the religiosity measures.

Omitted variables. Religion is a complex phenomenon. Therefore, it is plausible to assume that possible confounding factors and unobservable components could be added to the error term when including religion-related variables into regression, causing omitted variable bias. To minimize the effect of confounding factors, we employ a multi-way fixed effects strategy along with controls of individual characteristics.¹⁰ We control for the following fixed effects: country, survey year, country and survey year interaction, religious denominations, occupation categories, and income level. Country and survey year fixed effects eliminate the impact of institutional features and country-level economic determinants, which are the same for the entire country. Occupation and income level fixed effects aim to minimize the differences of individual-level unobservables across respondents; religious denomination fixed effects help us compare the religiosity of people who belong to different religious denominations. Controlling for a large set of fixed effects conveys the possibility of underestimating the impact of religion, given that religion affects many determinants whose effects might be partially absorbed by fixed effects. Therefore, our estimates can be interpreted as a lower bound of the effect of religiosity on innovation attitudes.

In our model, one possible omitted variable is human capital. It is reasonable to assume that human capital has a feature somewhat unobservable, which gives rise to the omitted variable problem. Occupation and income level fixed effects can eliminate, to some degree, human capital differences among respondents, thereby reducing the possible omitted variable bias. Furthermore, we follow Barro and Lee (2011) and use educational attainment, namely completed years of education, as a measure of human capital. We also control for father's and mother's education levels, which are essential to human capital formation, especially at younger ages. Akcigit et al. (2017) examine a large data set on innovation and inventors collected for the period between 1880 and 1940 in the USA. They uncover many micro and macro stylized facts at the individual level. One of them is that father's education is a crucial determinant of being an inventor, especially through the channel of a child's education. Hence we control for the father's and the mother's education level since the mother's education is likely to be equally important, if not more, given that mothers spend more time with children in most cultures. Another uncovered fact by Akcigit et al. is that inventors tend to migrate from their birth state to more eligible states for innovation. We, therefore, control for the born-in country of respondent.

Reverse causality. The direction of causation may go both ways when working on religion. We aim to measure the effect of religiosity on innovation. However, changes in innovation may affect one's religiosity as well, mainly through economic growth and development, which is conceptualized in the secularization hypothesis mentioned in Section 1.2.1. To deal with reverse causality in the case of culture and religion-related research, Guiso et al. (2006) suggest focusing only on inherited dimensions of culture, such as religious affiliation and ethnicity, rather than accumulated ones such as social capital. We follow Guiso et al. and use religious affiliation along with religious intensity when measuring religiosity, as explained above.

Another strategy to mitigate reverse causality is to focus on economic attitudes rather than outcomes, given that economic outcomes are partly determined by the institutional and political environment, which is rather challenging to account for (Guiso et al., 2003, 2006). Economic attitudes reveal *individual propensity* to something; in other words, they give insights into the possibility of a particular economic outcome being real. As Guiso

¹⁰See Guimaraes and Portugal (2010) and Correia (2016) for further information on multi-way fixed effects in linear models.

et al. (2003) pointed out, asking somebody if she has ever cheated on taxes is different from asking somebody her opinion on cheating on taxes.

Measurement error. Data may be measured with errors, such as coding and reporting errors. When an independent variable is measured with error, endogeneity arises since it violates the zero conditional mean assumption. Our data source is the ESS, a comparative cross-national survey that collects measures of individuals' preferences and social and political attitudes across 36 European countries. These measures may contain errors due to differences in concepts measured across the participating countries. The ESS aims to minimize such measurement errors and improve data reliability, validity, and comparability; therefore, it undertakes a range of data quality assessment activities, including the Multitrait-Multimethod (MTMM) and Survey Quality Predictor (SQP). MTMM is an experimental project in which the same respondents are asked three survey questions twice in different concepts by using different response scales each time.¹¹ MTMM measures the measurement quality of individual questions, and it was implemented for all the waves of the ESS. SQP is an open-source database on which the complete set of the ESS questions is evaluated through MTMM design. Due to the rigorous approach of the ESS, we may say that it is unlikely that our independent variables suffer from measurement errors.

Furthermore, the ESS imposes a minimum target response rate of 70% in each country to minimize non-response bias which may lead over or under-representation of some individuals with certain characteristics. To adjust for non-response bias and others, the ESS data come with three weighting variables: design weights (*dweight*), post-stratification weights (*pspwght*) and population size weights (*pweight*). Design weights correct for sample selection bias, given that some countries use complex sampling designs and respondents have different probabilities of being a part of the sample. Post-stratification weights adjust for uneven representation of sub-groups and sub-populations with certain characteristics, correcting for sample and non-response biases. Population size weights are used when data from more than two countries are combined. Since each country has a different population size but a similar sample size, *pweight* corrects for over or under-representation of countries concerning their population. We analyse more than one country and total averages of countries in some cases, hence we must use a combination of either *dweight* and *pweight* or *pspwght* and *pweight*. Given that *pspwght* includes *dweight*, we use the combination of *pspwght* and *pweight* by generating a new variable $gweight = pswght * pweight$.¹²

1.3.5 Model

Based on the strategy above, we aim to estimate a causal relationship that goes from religion to innovation. We first estimate equation 1.1 with OLS. We then instrument religiosity measures, R , with Riv and estimate the baseline model with IV/2SLS¹³. The first stage model is specified in equation 1.2. All models include standard demographic controls and fixed effects. Control variables are gender (*gender*), age (*age*), completed years of education (*education*), parent's highest education level (*mothere*, *fathere*), born in country (*bornnc*), paid work status (*paidwork*), whether had a paid job before (*pwbefore*), whether have a partner or wife living in the house (*partner*), whether have a child living in the house (*child*), subjective health status (*health*). Fixed effects are country dummies

¹¹The detailed test data obtained from MTMM is available online:
https://www.europeansocialsurvey.org/data/download_mtm.html

¹²For detailed information on weighting the ESS data, please refer to the guide "Weighting European Social Survey Data".

¹³We use "ivreghdfe" command of Stata which allows IV/2SLS estimation with multi-way fixed effects.

(*country*), survey year (*essround*), country and survey year interactions (*cness*), religious denominations (*denomination*), occupation categories (*occupation*), and income level (*income1* & *income2*).

$$I_i = \alpha_0 + \beta R_i + \theta \mathbf{X}_i + \delta \mathbf{F} + \varepsilon_i \quad (1.1)$$

where I denotes innovation attitudes (*creative, different, free, adventurous, traditions, rules, inposav, innegav*) of individual i , R stands for religiosity variables (*degree, attendance, pray, religiosity*), β is the coefficient of interest, the vector X denotes control variables, the vector F represents the complete set of fixed effects absorbed into estimates, ε is the unobserved error term.

$$R_i = \alpha_1 + \zeta Riv_i + \vartheta \mathbf{X}_i + \eta \mathbf{F} + \epsilon_i \quad (1.2)$$

where Riv denotes instrumental variables that we use as an exogenous predictor of each religiosity measure.

1.4 Results

1.4.1 OLS Estimates

We begin by examining the relationship between religiosity and innovation attitudes estimated with OLS. All specifications include multi-level fixed effects that are fully absorbed into all specifications. We gradually control for socio-demographic indicators. Regression errors are likely correlated within groups since we expect endogenous religiosity variables. Moreover, the data set we use has a grouped structure, *i.e.* countries. We, therefore, report robust standard errors clustered at the level of instruments, namely clustered for the country, gender, age, religious affiliation and denomination groups. We also performed OLS estimates with robust standard errors and found very little difference so that the significance levels of the coefficients did not change. However, we prefer to report the clustered version due to the unity of the analyses throughout the chapter.

Table 1 presents the variations in *inposav* concerning different measures of religiosity. *Religiosity* is the independent variable in the first three specifications. The first specification, column 1, includes control variables for gender, age, age squared, education, and the complete set of fixed effects. The insignificant correlation coefficient in column 1 endures in columns 2 and 3 after adding other control variables. Other measures of religiosity display different patterns. *Degree* is statistically significant at the 99% level in column 4 and stays robust to the further controls in column 5. On the contrary to the other religiosity measures, *attendance* shows a negative relationship with *inposav* in column 6. It stays robust when we add further controls in column 7. *Pray* is positive and significant at the 99% level in column 8 and stays robust to the further controls, yet the coefficient and significance level decrease in column 9.

The detailed OLS estimates, broken down into each measure of innovation attitudes, are presented in Appendix C. Table C1 reports a stable, positive, and significant relation between *importance of creativity* and the religiosity measures, except for *attendance*, which has insignificant coefficients. In Table C2, all religiosity measures are positively and significantly associated with *importance of being different*, including *attendance*. Differently, *importance of being free* is negatively related to all religiosity measures in Table C3. As Table C4 displays, all religiosity measures have insignificant coefficients for *importance of being adventurous*, except *attendance*, which is negative and statistically significant at

the 90% level. When compared to Table 1, the positive coefficients for *degree* (0.008) and *pray* (0.006) seem to be driven by *importance of creativity* and *importance of being different* as evidenced in Tables C1 and C2. The negative relation between *attendance* and *inposav* (-0.012) is apparently fostered by the negative relation between *attendance* and *importance of being free* in Table C3.

Table 1. OLS Estimates: Religiosity and Average Positive Innovation Attitudes

	(1) <i>inposav</i>	(2) <i>inposav</i>	(3) <i>inposav</i>	(4) <i>inposav</i>	(5) <i>inposav</i>	(6) <i>inposav</i>	(7) <i>inposav</i>	(8) <i>inposav</i>	(9) <i>inposav</i>
<i>religiosity</i>	0.002 (0.004)	0.006 (0.004)	0.003 (0.004)						
<i>degree</i>				0.009*** (0.003)	0.008*** (0.003)				
<i>attendance</i>						-0.010*** (0.004)	-0.012*** (0.004)		
<i>pray</i>								0.008*** (0.003)	0.006** (0.003)
<i>gender</i>	-0.025*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.022*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.024*** (0.002)
<i>age</i>	-0.005*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)
<i>age2</i>	0.000*** (0.000)	0.000** (0.000)	0.000 (0.000)	0.000** (0.000)	0.000 (0.000)	0.000** (0.000)	0.000 (0.000)	0.000** (0.000)	0.000 (0.000)
<i>education</i>	0.004*** (0.000)	0.003*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
<i>paidwork</i>		-0.001 (0.002)	0.001 (0.002)	-0.000 (0.002)	0.001 (0.002)	-0.001 (0.002)	0.001 (0.002)	-0.000 (0.002)	0.001 (0.002)
<i>pwbefore</i>		-0.115** (0.051)	-0.069*** (0.013)	-0.115** (0.052)	-0.068*** (0.011)	-0.121** (0.049)	-0.077*** (0.013)	-0.115** (0.050)	-0.069*** (0.012)
<i>partner</i>		-0.019*** (0.002)	-0.018*** (0.002)	-0.020*** (0.002)	-0.018*** (0.002)	-0.019*** (0.002)	-0.018*** (0.002)	-0.019*** (0.002)	-0.018*** (0.002)
<i>health</i>		0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)
<i>child</i>		-0.019*** (0.002)	-0.020*** (0.002)	-0.019*** (0.002)	-0.020*** (0.002)	-0.019*** (0.002)	-0.020*** (0.002)	-0.019*** (0.002)	-0.020*** (0.002)
<i>bornnc</i>			-0.010*** (0.003)		-0.009*** (0.003)		-0.011*** (0.003)		-0.009*** (0.003)
<i>fathere</i>			0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)
<i>mothere</i>			0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)
<i>cons</i>	0.749*** (0.010)	0.767*** (0.051)	0.698*** (0.017)	0.761*** (0.053)	0.691*** (0.016)	0.773*** (0.050)	0.707*** (0.018)	0.764*** (0.051)	0.694*** (0.017)
<i>N</i>	234,528	228,327	202,300	231,306	204,788	231,733	205,176	229,606	203,384
<i>Adj. R²</i>	0.135	0.146	0.150	0.147	0.151	0.147	0.151	0.147	0.150

Notes: OLS estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Overall, the OLS results show that higher values of *degree* and *pray* have a somewhat positive relationship with innovation through *importance of creativity* and *importance of being different*, while *attendance* is negatively related to innovation. *Religiosity* shows no significant relationship with *inposav*, which is expected to some degree, given that *religiosity* is the principal component of other three measures (*degree*, *attendance*, *pray*) that exhibit both negative and positive relationships, yielding to cancel out each effect when the measures combined.

In contrast with Table 1, Table 2 displays a pretty stable relationship between different measures of religiosity and *innegav*. Each specification is statistically significant at the 99% level and positively related to *innegav*, regardless of the gradual inclusion of con-

trols. Another salient result is that all coefficients are substantially higher in absolute value than those in Table 1, suggesting that the religiosity measures are more strongly associated with NIA than with PIA. In other words, religiosity seems to foster individual traits that are unfavourable to innovation.

Table 2. OLS Estimates: Religiosity and Average Negative Innovation Attitudes

	(1) <i>innegav</i>	(2) <i>innegav</i>	(3) <i>innegav</i>	(4) <i>innegav</i>	(5) <i>innegav</i>	(6) <i>innegav</i>	(7) <i>innegav</i>	(8) <i>innegav</i>	(9) <i>innegav</i>
<i>religiosity</i>	0.180*** (0.005)	0.176*** (0.005)	0.175*** (0.005)						
<i>degree</i>				0.132*** (0.005)	0.130*** (0.005)				
<i>attendance</i>						0.129*** (0.005)	0.129*** (0.005)		
<i>pray</i>								0.077*** (0.003)	0.075*** (0.003)
<i>gender</i>	-0.009*** (0.003)	-0.007*** (0.003)	-0.007*** (0.003)	-0.004* (0.003)	-0.004* (0.003)	0.000 (0.003)	0.000 (0.003)	-0.005** (0.003)	-0.005** (0.003)
<i>age</i>	0.001*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001 (0.000)	-0.000 (0.000)	-0.001 (0.000)	-0.000 (0.000)	-0.001 (0.000)
<i>age2</i>	0.000 (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>		0.003 (0.002)	0.005** (0.002)	0.002 (0.002)	0.004* (0.002)	0.002 (0.002)	0.004* (0.002)	0.003 (0.002)	0.004* (0.002)
<i>pwbefore</i>		-0.244*** (0.034)	-0.209*** (0.031)	-0.256*** (0.037)	-0.218*** (0.034)	-0.253*** (0.032)	-0.221*** (0.032)	-0.276*** (0.027)	-0.257*** (0.034)
<i>partner</i>		0.025*** (0.002)	0.023*** (0.002)	0.026*** (0.002)	0.024*** (0.002)	0.025*** (0.002)	0.022*** (0.002)	0.027*** (0.002)	0.025*** (0.002)
<i>health</i>		0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
<i>child</i>		0.003 (0.002)	0.001 (0.002)	0.004** (0.002)	0.002 (0.002)	0.004** (0.002)	0.003 (0.002)	0.004** (0.002)	0.002 (0.002)
<i>bornc</i>			-0.032*** (0.004)		-0.036*** (0.003)		-0.037*** (0.003)		-0.035*** (0.004)
<i>fathere</i>			-0.001 (0.001)		-0.001 (0.001)		-0.001 (0.001)		-0.001 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>cons</i>	0.510*** (0.010)	0.756*** (0.036)	0.776*** (0.032)	0.765*** (0.038)	0.787*** (0.035)	0.793*** (0.033)	0.823*** (0.033)	0.818*** (0.028)	0.859*** (0.035)
<i>N</i>	234,387	228,192	202,191	231,156	204,668	231,582	205,055	229,462	203,268
<i>Adj. R²</i>	0.239	0.242	0.246	0.236	0.240	0.231	0.236	0.231	0.235

Notes: OLS estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Tables C5 and C6 display detailed results for the two NIA. All religiosity measures are positively and significantly (at the 99% level) associated with both *following traditions* and *following rules*, but the coefficients are almost four times higher for *following traditions*.

Table 3 shows the OLS estimates broken down into various religious denominations and innovation attitudes. The dependent variables are PIA for the first four columns, NIA for columns 5 and 6, and average innovation attitudes for the last two columns. The complete set of control variables and fixed effects are included in each specification. Briefly, Roman Catholics, Protestants and Other Christians exhibit a negative relation with *inposav*, while Jewish, Other Christian and Eastern religions are positively correlated. When it comes to *innegav*, there is a consistent relationship across denominations, each of them is positively and significantly correlated with *innegav*, meaning

that all kinds of religions are highly correlated to the unfavourable individual traits to innovation.

Table 3. OLS Estimates: Religious Denominations and Innovation Attitudes

	(1) <i>creative</i>	(2) <i>different</i>	(3) <i>free</i>	(4) <i>adventurous</i>	(5) <i>traditions</i>	(6) <i>rules</i>	(7) <i>inposav</i>	(8) <i>innegav</i>
Roman Catholic	-0.006* (0.004)	-0.007** (0.004)	-0.016*** (0.004)	-0.025*** (0.004)	0.171*** (0.004)	0.048*** (0.004)	-0.014*** (0.003)	0.110*** (0.003)
Protestant	-0.009** (0.004)	-0.018*** (0.004)	-0.017*** (0.004)	-0.027*** (0.003)	0.151*** (0.005)	0.059*** (0.005)	-0.018*** (0.003)	0.105*** (0.005)
Eastern Orthodox	0.008 (0.009)	0.005 (0.008)	0.003 (0.009)	-0.000 (0.008)	0.111*** (0.006)	0.034*** (0.007)	0.003 (0.006)	0.072*** (0.005)
Other Christian	0.007 (0.009)	-0.000 (0.010)	-0.027*** (0.009)	-0.026*** (0.010)	0.111*** (0.010)	0.075*** (0.011)	-0.011* (0.007)	0.093*** (0.008)
Jewish	0.051 (0.033)	0.074** (0.033)	0.011 (0.029)	0.073** (0.036)	0.205*** (0.033)	-0.044 (0.045)	0.052* (0.027)	0.079*** (0.021)
Muslim	-0.006 (0.009)	0.005 (0.009)	0.001 (0.010)	-0.013* (0.008)	0.224*** (0.009)	0.090*** (0.009)	-0.003 (0.007)	0.157*** (0.007)
Eastern Religions	0.015 (0.014)	0.022 (0.015)	0.015 (0.013)	0.045*** (0.015)	0.141*** (0.016)	0.000 (0.020)	0.024** (0.010)	0.070*** (0.015)
Other Non-Christian	0.055*** (0.016)	0.013 (0.018)	-0.009 (0.016)	0.030 (0.022)	0.127*** (0.018)	-0.025 (0.019)	0.021* (0.012)	0.051*** (0.013)
Not declared	0.005 (0.012)	0.009 (0.009)	-0.015 (0.009)	-0.006 (0.011)	0.143*** (0.013)	0.041*** (0.010)	-0.002 (0.007)	0.092*** (0.010)
<i>N</i>	204,634	204,801	204,940	204,732	204,960	204,087	205,773	205,647
<i>Adj. R</i> ²	0.082	0.081	0.074	0.174	0.194	0.134	0.151	0.222

Notes: OLS estimates for religious denominations are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include controls for age, age squared, gender, education, paid work status, children, health, mother's and father's, education, born-in country and the following fixed effects: country, survey year, country-survey year, religious denomination, occupation and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

1.4.2 Instrumental Variables Estimates

The OLS estimates reported above indicate both negative and positive associations between religiosity and innovation attitudes, though the negative relation is more substantial. Nevertheless, the potential endogeneity of religion makes it challenging to interpret the results as causal links. Therefore, we use instruments to cope with endogeneity fundamentally. Tables 4 and 5 report the main results estimated with IV. The first stage results of each specification can be found in the corresponding columns of Tables B1 and B2 in Appendix B. The OLS version of each specification is presented in the corresponding columns of Tables 1 and 2, meaning that, for instance, column 4 in Table 1 is the OLS version of the specification in column 4 in Table 4. All specifications include the complete set of fixed effects.

We report a couple of post-estimation diagnostic tests. Kleibergen-Paap rk LM statistic (*ldp*) is for under-identification. Kleibergen-Paap rk Wald F statistic (*widstat*), that is F test of excluded instruments valid in non-*i.i.d* case, and Cragg-Donald Wald F statistic (*cdf*) are diagnostics for weak identification. *Widstat* is equivalent to the first stage F statistic. Anderson-Rubin Wald F statistic (*AR test*), which is weak-instrument-robust, tests if the coefficients of endogenous variables are equal to zero in the reduced form estimation. We do not need to consider the over-identification problem because there is only one instrument for each endogenous religiosity variable in all estimates; thus, our model is just-identified. All the results and statistics are efficient only for homoscedasticity since

all are estimated with heteroscedasticity-robust options.

Table 4. IV Estimates: Religiosity and Average Positive Innovation Attitudes

	(1) <i>inposav</i>	(2) <i>inposav</i>	(3) <i>inposav</i>	(4) <i>inposav</i>	(5) <i>inposav</i>	(6) <i>inposav</i>	(7) <i>inposav</i>	(8) <i>inposav</i>	(9) <i>inposav</i>
<i>religiosity</i>	-0.155** (0.070)	-0.151** (0.065)	-0.198*** (0.062)						
<i>degree</i>				-0.099 (0.099)	-0.177* (0.096)				
<i>attendance</i>						-0.157*** (0.052)	-0.185*** (0.050)		
<i>pray</i>								-0.080** (0.040)	-0.119*** (0.039)
<i>gender</i>	-0.015*** (0.004)	-0.014*** (0.004)	-0.011*** (0.004)	-0.017*** (0.006)	-0.012** (0.005)	-0.019*** (0.003)	-0.019*** (0.003)	-0.013*** (0.005)	-0.009** (0.005)
<i>age</i>	-0.005*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.004*** (0.000)	0.003*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
<i>paidwork</i>		-0.001 (0.002)	-0.000 (0.002)	-0.001 (0.002)	0.001 (0.002)	-0.001 (0.002)	0.000 (0.002)	-0.001 (0.002)	-0.000 (0.002)
<i>prbefore</i>		-0.158*** (0.040)	-0.148*** (0.037)	-0.146*** (0.045)	-0.152*** (0.055)	-0.167*** (0.038)	-0.151*** (0.034)	-0.135*** (0.048)	-0.102*** (0.019)
<i>partner</i>		-0.017*** (0.002)	-0.015*** (0.003)	-0.018*** (0.003)	-0.016*** (0.003)	-0.016*** (0.003)	-0.014*** (0.003)	-0.019*** (0.002)	-0.017*** (0.002)
<i>health</i>		0.017*** (0.001)	0.017*** (0.001)	0.018*** (0.001)	0.017*** (0.001)	0.019*** (0.001)	0.018*** (0.001)	0.017*** (0.001)	0.016*** (0.001)
<i>child</i>		-0.016*** (0.002)	-0.017*** (0.002)	-0.017*** (0.002)	-0.017*** (0.002)	-0.017*** (0.002)	-0.018*** (0.002)	-0.017*** (0.002)	-0.017*** (0.002)
<i>bornnc</i>			-0.024*** (0.006)		-0.021*** (0.007)		-0.019*** (0.004)		-0.023*** (0.005)
<i>fathere</i>			0.005*** (0.001)		0.005*** (0.001)		0.006*** (0.001)		0.006*** (0.001)
<i>mothere</i>			0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)
<i>N</i>	234,528	228,327	202,300	231,306	204,788	231,733	205,176	229,606	203,384
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,338	2,380	2,161	846	771	2,732	2,533	2,012	1,730
<i>widstat</i>	187	194	160	84	79	141	131	192	160
<i>AR test</i>	0.022	0.016	0.001	0.297	0.052	0.002	0.000	0.046	0.002

Notes: IV estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), and Anderson-Rubin Wald F statistic (*AR test*) are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 4 displays the IV results for *inposav* with respect to different measures of religiosity. The first stage results in Table B1 show that all the excluded instruments are strongly and positively correlated with the corresponding religiosity measures, with a correlation coefficient between 0.581 and 0.784. P-values of Kleibergen-Paap rk LM test (*idp*) show that the null of under-identification is rejected, and the full rank condition is satisfied in all estimates. For all IV estimates, the first stage F statistics (*widstat*) are well above Stock-Yogo critical values of weak identification. The *AR test* results indicate that the instruments are relevant, except for *degree*, which we will discuss later.

The IV results for *inposav* exhibit a somewhat different pattern than the corresponding OLS results in Table 1. It is essential to underline that the OLS and IV results are not directly comparable since OLS estimates average treatment effect (ATE), while IV/2SLS estimates local average treatment effect (LATE) (Angrist et al., 1996). However, we emphasize the main differences to discuss the effects of vanishing endogeneity. First of all,

all religiosity measures display substantial increase in the coefficients, which is also the case in other studies using a more aggregated instrument than endogenous variable (Iyer et al., 2017; Cutler and Gruber, 1996). *Religiosity* becomes significant and negative in column 3; it is positive and insignificant with OLS. *Degree* gets negative and less significant in columns 5 and 6. *Attendance* keeps the sign of OLS estimates, which is negative. *Pray* in columns 8 and 9, changes the sign to be negative as well.

Table 5. IV Estimates: Religiosity and Average Negative Innovation Attitudes

	(1) <i>innegav</i>	(2) <i>innegav</i>	(3) <i>innegav</i>	(4) <i>innegav</i>	(5) <i>innegav</i>	(6) <i>innegav</i>	(7) <i>innegav</i>	(8) <i>innegav</i>	(9) <i>innegav</i>
<i>religiosity</i>	0.139*** (0.052)	0.177*** (0.050)	0.210*** (0.055)						
<i>degree</i>				0.182*** (0.067)	0.223*** (0.077)				
<i>attendance</i>						0.205*** (0.044)	0.216*** (0.047)		
<i>pray</i>								0.088** (0.036)	0.118*** (0.040)
<i>gender</i>	-0.006 (0.004)	-0.007* (0.004)	-0.009** (0.004)	-0.007 (0.005)	-0.010* (0.005)	-0.001 (0.003)	-0.002 (0.003)	-0.007 (0.005)	-0.010** (0.005)
<i>age</i>	0.001*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001 (0.000)
<i>age2</i>	0.000* (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>		0.003 (0.002)	0.005** (0.002)	0.003 (0.002)	0.004* (0.002)	0.003 (0.002)	0.004* (0.002)	0.003 (0.002)	0.005** (0.002)
<i>prbefore</i>		-0.243*** (0.037)	-0.195*** (0.037)	-0.242*** (0.049)	-0.176*** (0.050)	-0.230*** (0.039)	-0.184*** (0.036)	-0.273*** (0.028)	-0.246*** (0.034)
<i>partner</i>		0.025*** (0.002)	0.022*** (0.002)	0.025*** (0.002)	0.022*** (0.002)	0.023*** (0.002)	0.020*** (0.002)	0.027*** (0.002)	0.024*** (0.002)
<i>health</i>		0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.002** (0.001)	0.002** (0.001)	0.005*** (0.001)	0.005*** (0.001)
<i>child</i>		0.003 (0.002)	0.000 (0.002)	0.003 (0.002)	0.000 (0.002)	0.003* (0.002)	0.002 (0.002)	0.004* (0.002)	0.001 (0.002)
<i>bornnc</i>			-0.029*** (0.005)		-0.030*** (0.006)		-0.033*** (0.004)		-0.031*** (0.005)
<i>fathere</i>			-0.001 (0.001)		-0.001 (0.001)		-0.001 (0.001)		-0.001 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>N</i>	234,387	228,192	202,191	231,156	204,668	231,582	205,055	229,462	203,268
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,327	2,369	2,147	839	763	2,729	2,525	2,001	1,722
<i>widstat</i>	186	193	159	82	77	141	130	190	159
<i>AR test</i>	0.001	0.000	0.000	0.007	0.004	0.000	0.000	0.012	0.002

Notes: IV estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), and Anderson-Rubin Wald F statistic (*AR test*) are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Innegav reported in Table 5, in contrast to *inposav*, displays pretty stable and consistent patterns. All religiosity measures are positively and significantly correlated to *innegav*, as they are in the corresponding OLS estimates in Table 2, with the increase in the coefficients. The corresponding first stage results, reported in Table B2, show that all the excluded instruments are strongly correlated to the religiosity measures in each specification, with a correlation coefficient between 0.581 and 0.784.

Appendix D provides the IV estimates broken down into each measure of innovation attitudes. Tables D1 and D2 indicate that *importance of creativity* and *importance of*

being different are negatively affected by *attendance*, while other religiosity measures are insignificant. Table D3 shows that all religiosity measures are negatively related to *importance of being free*, yielding the highest coefficients among IV estimates. Religiosity does not affect *importance of being adventurous* as demonstrated in Table D4. These results signal that the estimated positive associations between religiosity and PIA in OLS Tables C1 (*importance of creativity*) and C2 (*importance of being different*) are driven by endogeneity. Tables D5 and D6 show that religiosity measures positively affect NIA, similar to the OLS estimates. Overall, it is plausible to argue that higher degrees of religiosity are not advantageous for innovation since it fosters some personal traits un conducive to an innovative mindset. Moreover, it falls away from conducive personal traits such as freedom.

Apart from the main coefficients of interest, the control variables also provide insightful results. The effect of religiosity on *importance of creativity*, *importance of being adventurous*, and *following rules* are stronger for men (*gender*); while it is stronger for women regarding *importance of being different* and *following traditions*. Age is negatively related to *importance of being different*, *importance of being adventurous*, and *following rules* and is insignificant for *importance of creativity*, *importance of being free*, and *following traditions*. Education seem to foster PIA (see Tables D1-D4) and mitigate NIA (see Tables D5 and D6). Having paid work in the last seven days (*paidwork*) is mostly insignificant. Having paid work before (*pwbefore*) yields the highest coefficients among control variables, implying that individuals with work experience are less affected by each measure of religiosity. Individuals living with a partner are less affected by religiosity with respect to PIA (except for *importance of creativity*), while they are more affected for NIA. *Health* and born-in country (*bornc*) do not introduce heterogeneous effects for PIA and NIA. *Health* is positively related to all innovation attitudes. *Bornc* is negative in all specifications, indicating the fact that being born in a different country is a factor that increases the effect of religiosity on innovation attitudes. Interestingly, individuals living with a child (*child*) are less affected by religiosity regarding PIA, while it does not matter for NIA. Both mother's (*mothere*) and father's education (*fathere*) are positively related to PIA, while only (*mothere*) significantly and negatively relates with NIA.

Overall, the negative effect of religiosity on innovativeness is in line with the findings of prior studies (Benabou et al., 2013, 2015). Using OLS estimates, Benabou et al. (2015) find that religiosity is positively related to creativity, which they describe as "puzzling". We contribute to this literature by uncovering that the positive relationship between religiosity and creativity may be driven by endogeneity. We find a negative relationship between religiosity and creativity once we rule out endogeneity by applying the IV method.

The question is that, compared to OLS estimates, why does *inposav* change its behaviour with respect to the religiosity measures while *innegav* stays stable? The answer is hidden in the relationship between religious belonging (*belonging*) and other religiosity measures, which are summarized in Tables C7 and C8 in Appendix C.

In the ESS, a respondent answers the questions regarding *degree*, *attendance* and *pray* regardless of being belong to a religion. In other words, even though the respondent does not believe in a religion, she still answers the questions such as "How religious are you?", "How frequently do you pray?", "How frequently do you attend religious services?" Interestingly, the answers are generally not zero. Table C7 displays the mean values of all religiosity measures, broken down into religious belonging. For instance, the mean value of *religiosity* is 0.51 for believers —people who belong to a religion at present (*belonging*=1 & *belongingp*=1)— , while it is 0.14 for never-believers —people who have never belonged to a religion (*belonging*=0 & *belongingp*=0)— , and is 0.18 for once-believers —people who do not belong to a religion at present but used to belong at past

(*belonging*=0 & *belongingp*=1).

Table C8 shows disaggregated OLS results for religious belonging. For each sub-sample, all religiosity measures are positively and significantly associated with *innegav*. This pattern disappears when we look at *inposav*. For the sub-sample of believers, only *attendance* is significantly and negatively related to *inposav*, meaning that other religiosity measures of believers are not associated with *inposav*. On the contrary, all religiosity measures of never-believers and once-believers are positively and significantly related to *inposav*. Under the light of this disaggregation, one can see that the OLS results in Table 1, namely, the positive correlations between *degree*, *pray*, and *inposav* are seemed to be driven by the sub-sample of non-believers, raising the concerns for a spurious correlation. We, therefore, consider religious belonging in our instrumental variables strategy and construct the instruments by taking it into account. We instrument the religiosity of a believer with the average religiosity of believers, along with other measures explained in Section 1.3. By doing so, we address the differences between the religiosities of believers and non-believers. Thus, the IV results for *inposav* in Table 4 display different signs than the overall OLS results.

1.4.3 Heterogeneity of Religiosity: Gender and Age

We re-estimate the baseline specifications to examine whether the main findings hold for different sub-samples of age and gender. Table 6 presents the results. Columns 1-12 show OLS and IV estimates disaggregated by six age brackets¹⁴. The last four columns (13-16) display the baseline OLS and IV estimates disaggregated by gender. The dependent variable is *inposav* on the upper part of the table and *innegav* on the lower part. Each coefficient comes from a different regression and includes the complete set of control variables and fixed effects which are not reported due to space constraints. The number of observations is reported below standard errors, followed by adjusted R^2 for OLS and AR test for IV estimates.

Considering age, OLS results show that the religiosity measures, when statistically significant, are negatively related to *inposav* in early and lower-middle ages, *i.e.* 15-25 and 35-45, but are positively related in middle and upper-middle ages, *i.e.* 45-55 and 55-65. On the other hand, IV results are always negatively related to *inposav*, regardless of age bracket. As mentioned above, this alteration probably stems from the vanishment of spurious correlations thanks to the indirect inclusion of religious belonging via the instruments.

The OLS and IV results in the lower part of Table 6 show that the religiosity measures are always positively associated with *innegav* for each age bracket except for the 15-25, which is the only negative association between religiosity measures and *innegav* throughout this chapter. One explanation would be that young adults are generally pretty open to questioning the values and beliefs they inherited from their parents as well as from the culture in which they live, assuming that in the early ages of life, people make up their minds, see the world from their own eyes, form their own beliefs and values. Moreover, fulfilling religious duties —such as going to church or mosque— to meet the expectations of parents and society is not uncommon in the early ages, yet, being exposed to religion by doing so might make them question traditions and rules more, which can explain the negative association between *attendance* and *innegav*.

Regarding gender, the OLS results for *inposav* show that *attendance* is negatively and

¹⁴We use eleven age brackets to construct the instruments. However, here we only consider six age brackets for brevity. We have estimated the model also with eleven age brackets, but the overall results do not change significantly.

Table 6. Disaggregate IV Results for Age and Gender

<i>inposav</i>	(15-25)		(25-35)		(35-45)		(45-55)		(55-65)		(65+)		(female)		(male)	
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
<i>religiosity</i>	-0.014 (0.013)	-0.087 (0.143)	-0.004 (0.009)	-0.178 (0.173)	-0.004 (0.008)	-0.191 (0.133)	0.027*** (0.008)	-0.241** (0.100)	0.012 (0.008)	-0.243** (0.097)	-0.004 (0.008)	-0.192 (0.119)	0.000 (0.005)	-0.154* (0.084)	0.010** (0.005)	-0.213 (0.149)
	16.883	16.898	33.172	33.200	38.457	38.493	38.165	38.199	35.236	35.269	40.387	40.404	104.447	104.447	98.016	98.016
	0.065	0.542	0.094	0.312	0.086	0.127	0.72	0.010	0.096	0.011	0.154	0.076	0.160	0.057	0.130	0.115
<i>degree</i>	-0.000 (0.010)	0.020 (0.122)	0.002 (0.007)	1.110 (3.310)	0.006 (0.007)	0.458 (0.338)	0.024*** (0.007)	-0.158 (0.191)	0.018** (0.007)	-0.282** (0.122)	-0.003 (0.007)	-0.160 (0.135)	0.006 (0.004)	-0.148 (0.111)	0.012*** (0.004)	-0.054 (0.158)
	17.031	17.047	33.607	33.636	38.944	38.982	38.641	38.675	35.695	35.729	40.870	40.887	105.725	105.725	99.231	99.231
	0.067	0.869	0.093	0.464	0.086	0.035	0.073	0.411	0.097	0.053	0.155	0.198	0.161	0.146	0.131	0.727
<i>attendance</i>	-0.025** (0.012)	-0.145 (0.113)	-0.011 (0.008)	-0.111 (0.125)	-0.026*** (0.008)	-0.329* (0.177)	-0.001 (0.009)	-0.250** (0.099)	-0.012 (0.008)	-0.181** (0.091)	-0.008 (0.007)	-0.266* (0.101)	-0.017*** (0.005)	-0.172** (0.080)	-0.005 (0.005)	-0.304** (0.147)
	17.063	17.079	33.621	33.651	39.005	39.043	38.721	38.756	35.799	35.833	40.967	40.984	105.912	105.912	99.434	99.434
	0.068	0.120	0.094	0.308	0.087	0.013	0.074	0.002	0.097	0.029	0.154	0.030	0.161	0.026	0.130	0.019
<i>pray</i>	-0.001 (0.008)	0.059 (0.182)	-0.001 (0.005)	-0.195** (0.097)	0.006 (0.005)	-0.107 (0.089)	0.024*** (0.005)	-0.144** (0.059)	0.011** (0.005)	-0.229** (0.149)	0.000 (0.005)	-0.108 (0.113)	0.007** (0.003)	-0.011 (0.076)	0.009*** (0.003)	-0.176 (0.109)
	16.970	16.985	33.335	33.365	38.643	38.679	38.369	38.404	35.448	35.481	40.619	40.636	104.992	104.992	98.558	98.558
	0.066	0.741	0.093	0.065	0.085	0.219	0.073	0.009	0.095	0.007	0.155	0.313	0.161	0.887	0.130	237
<i>innegav</i>	(15-25)	(25-35)	(35-45)	(45-55)	(55-65)	(65+)	(female)	(male)	(15-25)	(25-35)	(35-45)	(45-55)	(55-65)	(65+)	(female)	(male)
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
<i>religiosity</i>	0.220*** (0.015)	-0.244* (0.144)	0.216*** (0.011)	0.323** (0.163)	0.184*** (0.009)	0.647*** (0.161)	0.167*** (0.009)	0.354*** (0.128)	0.153*** (0.009)	0.378** (0.154)	0.151*** (0.008)	0.609** (0.286)	0.170*** (0.006)	0.292*** (0.113)	0.182*** (0.006)	0.189 (0.138)
	16.879	16.894	33.154	33.182	38.443	38.479	38.149	38.183	35.208	35.241	40.358	40.375	104.396	104.396	97.958	97.958
	0.240	0.073	0.236	0.029	0.240	0.000	0.242	0.001	0.220	0.001	0.167	0.003	0.248	0.007	0.248	0.174
<i>degree</i>	0.168*** (0.012)	0.033 (0.155)	0.141*** (0.009)	2.274 (3.875)	0.134*** (0.008)	0.860* (0.462)	0.118*** (0.008)	0.352 (0.223)	0.116*** (0.008)	0.477* (0.248)	0.123*** (0.007)	0.546*** (0.192)	0.122*** (0.005)	0.142 (0.113)	0.136*** (0.005)	0.305** (0.154)
	17.027	17.043	33.587	33.616	38.926	38.964	38.623	38.657	35.665	35.699	40.840	40.857	105.671	105.671	99.165	99.165
	0.231	0.830	0.229	0.115	0.232	0.001	0.236	0.110	0.213	0.011	0.163	0.002	0.241	0.197	0.243	0.058
<i>attendance</i>	0.143*** (0.015)	-0.236* (0.125)	0.168*** (0.010)	0.221** (0.108)	0.140*** (0.009)	0.649*** (0.232)	0.136*** (0.009)	0.323** (0.126)	0.125*** (0.009)	0.332** (0.139)	0.093*** (0.007)	0.622* (0.317)	0.133*** (0.005)	0.332*** (0.091)	0.132*** (0.006)	0.198 (0.130)
	17.058	17.074	33.601	33.631	38.987	39.025	38.704	38.739	35.770	35.804	40.935	40.952	105.858	105.858	99.367	99.367
	0.218	0.021	0.227	0.006	0.227	0.000	0.236	0.000	0.212	0.000	0.155	0.001	0.240	0.000	0.236	0.123
<i>pray</i>	0.103*** (0.010)	-0.371* (0.217)	0.101*** (0.007)	0.081 (0.111)	0.079*** (0.006)	0.317*** (0.094)	0.066*** (0.005)	0.175** (0.074)	0.057*** (0.006)	0.264* (0.141)	0.071*** (0.005)	0.435 (0.347)	0.068*** (0.003)	0.268** (0.121)	0.084*** (0.004)	0.119 (0.113)
	16.965	16.980	33.317	33.347	38.627	38.663	38.352	38.387	35.420	35.453	40.587	40.604	104.939	104.939	98.495	98.495
	0.222	0.028	0.226	0.455	0.228	0.000	0.231	0.009	0.209	0.021	0.159	0.051	0.237	0.011	0.240	0.279

Notes: Disaggregated OLS and IV estimates for age and gender are reported. Robust standard errors clustered at the level of instruments are in parentheses. The observation numbers are below standard errors, followed by adjusted R squared for OLS and Anderson-Rubin Wald F statistic (*AR test*) for IV estimates. All regressions include controls for age, gender, education, paid work status, partner, children, health, born in country, mother's and father's, education and the following fixed effects: country, survey year, country-survey year, religious denomination, occupation and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

pray is positively correlated for women, while higher religiosity favour *innosav* for men. When it comes to IV results, only *attendance* is significant for both genders, introducing a higher negative effect for men.

The overall picture changes when we consider *innegav* as the dependent variable. Considering OLS estimates, each measure of religiosity is positively and significantly correlated to *innegav* with relatively high coefficients both for men and women. IV results suggest that this pattern stays the same for women except for *degree*, while *degree* is the only significant effect for men. These findings are sensible and consistent with the existing literature on gender. Religions, generally, have different approaches and prescribed behavioural rules for men and women, imposing various restrictions on women's liberty and rights. In contrast, men face very little of them, if not none. Therefore, the results are reasonable, suggesting that higher values of *attendance* and *pray* causes women to follow traditions and established rules more.

Table 7. Religiosity and Summary Innovation Attitudes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>	<i>innosum</i>
<i>religiosity</i>	-0.056*** (0.003)				-0.193*** (0.042)			
<i>degree</i>		-0.038*** (0.003)				-0.176*** (0.065)		
<i>attendance</i>			-0.051*** (0.003)				-0.185*** (0.038)	
<i>pray</i>				-0.021*** (0.002)				-0.112*** (0.027)
<i>gender</i>	-0.012*** (0.002)	-0.014*** (0.002)	-0.015*** (0.002)	-0.013*** (0.002)	-0.005 (0.003)	-0.006 (0.004)	-0.012*** (0.002)	-0.003 (0.003)
<i>age</i>	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
<i>age2</i>	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<i>education</i>	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
<i>paidwork</i>	-0.000 (0.002)	0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	-0.001 (0.002)	-0.000 (0.002)	-0.001 (0.002)	-0.001 (0.002)
<i>pwbefore</i>	0.022* (0.012)	0.026** (0.013)	0.021* (0.012)	0.039*** (0.011)	-0.031 (0.025)	-0.036 (0.039)	-0.036 (0.024)	0.015 (0.012)
<i>partner</i>	-0.020*** (0.002)	-0.021*** (0.001)	-0.020*** (0.001)	-0.021*** (0.002)	-0.018*** (0.002)	-0.019*** (0.002)	-0.017*** (0.002)	-0.020*** (0.002)
<i>health</i>	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.009*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.011*** (0.001)	0.009*** (0.001)
<i>child</i>	-0.014*** (0.001)	-0.014*** (0.001)	-0.014*** (0.001)	-0.014*** (0.001)	-0.011*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)
<i>bornc</i>	0.004* (0.002)	0.006*** (0.002)	0.005** (0.002)	0.006** (0.002)	-0.005 (0.004)	-0.003 (0.005)	-0.001 (0.003)	-0.004 (0.004)
<i>fathere</i>	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>mothere</i>	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>cons</i>	0.546*** (0.014)	0.538*** (0.015)	0.536*** (0.014)	0.516*** (0.013)				
<i>N</i>	198,145	200,453	200,808	199,128	198,145	200,453	200,808	199,128
<i>Adj. R²</i>	0.242	0.241	0.242	0.239				
<i>idp</i>					0.000	0.000	0.000	0.000
<i>cdf</i>					2,080	716	2,426	1,711
<i>widstat</i>					158	68	130	167
<i>AR test</i>					0.000	0.003	0.000	0.000

Notes: OLS and IV estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), and Anderson-Rubin Wald F statistic (*AR test*) are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

1.5 Sensitivity Analysis

1.5.1 Alternative Dependent Variable

So far, we have analysed innovation attitudes separately, concluding that higher religiosity is negatively related to PIA and positively correlated to NIA, which means that being religious is not a catalyser for innovation-related traits and is possibly an obstacle. In order to assess the sensitivity of this proposition, we introduce an overall, summary measure of innovation attitudes: *innosum* = *creative* + *free* + *different* + *adventurous* - *traditions* - *rules*, which is inspired by Tabellini (2005)¹⁵. *Innosum* is rescaled to be between 0 (low propensity to innovate) and 1 (high propensity to innovate); thus, higher values favour innovation.

We re-estimate the baseline specifications with OLS and IV by identifying *innosum* as the dependent variable. The results are presented in Table 7. Columns 1-4 are estimated with OLS, while columns 5-8 are estimated with IV. All specifications include the complete set of control variables and fixed effects. The OLS results favour the negative association between religiosity and innovation attitudes through each measure of religiosity. The IV results are also in line with the previous results.

1.5.2 Alternative Model Specifications

Religiosity is defined as the principal component of three measures: *degree*, *attendance*, and *pray*. However, *degree* differs from the other two in one important dimension. *Pray* and *attendance* are measures of an activity, an action, while *degree* is self-evaluation of a belief, a value. It is plausible to think that one cannot easily overestimate or underestimate the frequency of an activity but can do so to evaluate a belief. Therefore an interpersonal comparison of *degree* might be biased since some individuals may overvalue or undervalue their religiosity. This point is especially crucial to our instrumental variables, given that we instrument individual *i*'s *degree* by the average *degree* of other people. If many respondents tend to misvalue their *degree*, then this self-report bias will lead to weak instruments. Indeed, the diagnostic tests' results for *degree* in Tables 4 and 5 support this evaluation. Furthermore, *degree* is a component of *religiosity*; it thus endangers the instrument for *religiosity* as well. In order to see if the results are sensitive to the possible self-report bias of *degree*, we modify the IV strategy. Table 8 displays the results.

The first column is estimated with 2SLS and *religiosity* is instrumented with *attendance* and *pray*. Columns 2 and 3 follow the specification of column 1 but are estimated with *k*-class estimators. Prior studies suggest that Limited Information Maximum Likelihood Method (LIML) and Fuller edited LIML have a better finite-sample performance than 2SLS in the presence of weak instruments (Baum et al., 2007). Accordingly, in case of strong instruments, the estimates of 2SLS, LIML and Fuller should yield very similar results. Column 2 is estimated with LIML, and column 3 is with Fuller edited LIML with Fuller parameter $k = 1$. Column 4 is estimated with 2SLS, but the dependent variable is *innosum* and the instruments are *attendance* and *pray*. In specification 5, we add *iv_degree* as the third instrument for *religiosity*. The results of 2SLS, LIML, and Fuller estimates are the same, suggesting that our instruments are strong. The coefficients of *religiosity* in columns 1, 2, and 3 are qualitatively and quantitatively similar to the corresponding coefficient of *religiosity* (-0.198) in column 3 in Table 4, indicating that the initial instrument

¹⁵Tabellini (2005) introduces a summary variable of cultural traits which is the sum of the three positive beliefs (control, respect, trust) minus the negative belief (obedience).

of *religiosity* (*iv_religiosity*) is not significantly affected by the potential self-report bias of degree.

Table 8. Alternative Model Specifications

	(1)	(2)	(3)	(4)	(5)
	<i>inposav</i>	<i>inposav</i>	<i>inposav</i>	<i>immosum</i>	<i>inposav</i>
<i>religiosity</i>	-0.204*** (0.058)	-0.204*** (0.058)	-0.204*** (0.058)	-0.201*** (0.041)	-0.201*** (0.059)
<i>gender</i>	-0.011*** (0.004)	-0.011*** (0.004)	-0.011*** (0.004)	-0.004 (0.003)	-0.011*** (0.004)
<i>age</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000 (0.000)	0.000*** (0.000)
<i>education</i>	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.002*** (0.000)
<i>paidwork</i>	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	-0.001 (0.002)	-0.000 (0.002)
<i>prwbefore</i>	-0.151*** (0.036)	-0.151*** (0.036)	-0.151*** (0.036)	-0.035 (0.025)	-0.149*** (0.036)
<i>partner</i>	-0.015*** (0.002)	-0.015*** (0.002)	-0.015*** (0.002)	-0.018*** (0.002)	-0.015*** (0.002)
<i>health</i>	0.017*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.010*** (0.001)	0.017*** (0.001)
<i>child</i>	-0.017*** (0.002)	-0.017*** (0.002)	-0.017*** (0.002)	-0.011*** (0.002)	-0.017*** (0.002)
<i>bornnc</i>	-0.024*** (0.005)	-0.024*** (0.005)	-0.024*** (0.005)	-0.006 (0.004)	-0.024*** (0.005)
<i>fathere</i>	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	0.004*** (0.001)	0.005*** (0.001)
<i>mothere</i>	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>N</i>	202,300	202,300	202,300	198,145	202,300
<i>idp</i>	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	1,100	1,100	1,100	1,064	738
<i>widstat</i>	83	83	83	82	58
<i>AR test</i>	0.000	0.000	0.000	0.000	0.000
<i>Hansen J</i>	0.480	0.480	0.480	0.195	0.612

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), Anderson-Rubin Wald F statistic (*AR test*), and Sargan-Hansen test *Hansen J* are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

1.5.3 Reduced Sample Estimates

Table 9 displays the estimates from four different sub-samples. Each coefficient comes from a different IV estimate. The number of observations is below standard errors, which is followed by Cragg-Donald Wald F statistic (*cdf*), Kleibergen-Paap rk Wald F statistic (*widstat*), and Anderson-Rubin Wald F statistic (*AR test*) results.

First, we drop Muslim countries. Turkey, Albania, and Kosova are the only dominantly Muslim counties in the ESS. The rest of the sample consists of dominantly Christian countries, if not dominantly Atheist¹⁶. Despite controlling for country and denomination fixed effects, combining Muslim and Christian countries may complicate the interpretation since we consider inter-religion religiosity regardless of denomination. Moreover, Muslim countries exhibit outlier values for religiosity, especially for *degree*, such as 0.7-0.73 on average (see Table A3), which is pretty high given that the overall cross-country average of *degree* is 0.46. Therefore, we exclude Muslim countries from the sample and re-estimate the baseline specification as a robustness check. Columns 1 and 2 report the results. The overall negative effect of *attendance* increases compared to the full sample results; from -0.185 to -0.256 on *inposav* and from 0.216 to 0.281 on *innegav*. On the other hand, the effect of *degree* on *innegav* becomes less significant, which is not surprising since the Muslim countries exhibit outlier values of *degree*. All in all, the overall picture stays robust.

Table 9. Reduced Sample Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	<i>inposav</i>	<i>innegav</i>	<i>inposav</i>	<i>innegav</i>	<i>inposav</i>	<i>innegav</i>	<i>inposav</i>	<i>innegav</i>	<i>inposav</i>	<i>innegav</i>
<i>religiosity</i>	-0.214*** (0.065) 199,429 2,041 151 0.000	0.199*** (0.058) 199,324 2,028 150 0.000	-0.235*** (0.056) 184,551 1,632 105 0.000	0.212*** (0.062) 184,458 1,624 105 0.000	-0.259*** (0.059) 181,680 1,530 99 0.000	0.192*** (0.065) 181,591 1,523 98 0.003	-0.162*** (0.061) 181,332 1,999 155 0.004	0.200*** (0.055) 184,230 1,987 156 0.001	-0.153** (0.063) 174,626 1,916 133 0.012	0.206*** (0.057) 174,533 1,903 133 0.000
<i>degree</i>	-0.178* (0.093) 201,906 877 101 0.043	0.164** (0.074) 201,790 868 99 0.031	-0.251*** (0.091) 186,578 1,632 105 0.004	0.332*** (0.107) 186,474 1,624 105 0.001	-0.259*** (0.056) 183,696 1,530 99 0.002	0.220** (0.093) 183,596 1,523 98 0.019	-0.157* (0.088) 183,690 1,999 155 0.061	0.228*** (0.074) 183,577 1,987 156 0.002	-0.125 (0.106) 176,822 1,916 133 0.228	0.236*** (0.081) 176,718 1,903 133 0.006
<i>attendance</i>	-0.256*** (0.066) 202,287 1,653 113 0.000	0.281*** (0.060) 202,170 1,646 113 0.000	-0.196*** (0.048) 187,023 1,632 105 0.000	0.192*** (0.048) 186,920 1,624 105 0.000	-0.287*** (0.062) 184,134 1,530 99 0.000	0.256*** (0.064) 184,035 1,322 81 0.000	-0.140*** (0.051) 184,024 2,248 120 0.003	0.194*** (0.047) 183,910 2,243 120 0.000	-0.161*** (0.052) 177,171 2,262 108 0.001	0.192*** (0.049) 177,067 2,253 108 0.000
<i>pray</i>	-0.117*** (0.039) 200,499 1,776 162 0.002	0.104*** (0.040) 200,387 1,769 161 0.008	-0.148*** (0.039) 185,498 1,411 113 0.000	0.105** (0.044) 185,399 1,406 113 0.014	-0.147*** (0.038) 182,613 1,457 115 0.000	0.084* (0.044) 182,518 1,452 115 0.049	-0.099** (0.039) 182,319 1,586 152 0.008	0.121*** (0.041) 182,210 1,580 152 0.002	-0.072* (0.040) 175,562 1,645 156 0.067	0.123*** (0.042) 175,462 1,638 155 0.002

Notes: Robust standard errors clustered at the level of instruments are in parentheses and followed by the number of observations, Cragg-Donald Wald F statistic (*cdf*), Kleibergen-Paap rk Wald F statistic (*widstat*), and Anderson-Rubin Wald F statistic (*AR test*). All specifications include the full set of control variables and fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Second, we exclude Russia and Norway from the main sample because the percentages of the first and the second dominant denominations among respondents are pretty

¹⁶The second dominant religion is Christianity in each dominantly Atheist countries.

close, less than 1%, for those countries. The majority of respondents from Norway are Atheists with 47.02%; Protestants come second with 46.36%. For Russia, the majority is Eastern Orthodox with 46.03%; Atheists come second with 45.83%. Since the dominant denomination is one of the elements we construct instruments, any misidentification would cause bias in the instruments of individuals who share the misidentified dominant denomination. We thus drop Russia and Norway and re-estimate the baseline specifications to assess the difference. Columns 3 and 4 in Table 9 report the results. Column 3 shows that the coefficient of each religiosity measure increases for *inposav* and *degree* becomes significant at the 99% level compared to the full sample results in Table 4.

Third, we exclude Muslim countries along with Russia and Norway to obtain a relatively homogeneous sample. Columns 5 and 6 display the results. Coefficients of all religiosity measures substantially increase for *inposav* and *degree* becomes more significant compared to Table 4. For *innegav*, *degree* and *pray* become less significant compared to the full sample.

Fourth, we exclude once-believers. Non-believers are split into two categories among themselves: never-believers and once-believers. Table C7 shows that once-believers exhibit slightly higher religiosity means. The reason behind is that religiosity tends to be an absorbed value. Even though a person has stopped believing in a religion, there might be persistent effects, let alone the fact that personal values and traits might have been mostly grown when the person was a believer. Thus, to examine whether the higher religiosity of once-believers substantially affects the overall results, we drop once-believers from the full sample and leave the respondents who have been either an always-believer or never-believer. The corresponding results are presented in columns 7 and 8 in Table 9. Column 8 shows that the overall results for *innegav* do not change substantially compared to Table 5. However, the coefficient of each religiosity measure somewhat decreases in column 7 compared to Table 4.

Lastly, we exclude second-generation immigrants. Immigrants¹⁷ and their descendants might possess peculiar traits due to different cultural and institutional environments of their origin country. Including second-generation immigrants in the sample might introduce some unobservable effects regarding innovation attitudes. Columns 9 and 10 report the results. The coefficients for *inposav* slightly decrease and *degree* turns to be insignificant compared to Table 4. On the contrary, the results of *innegav* are pretty stable and similar to Table 5.

1.5.4 Possible Violations of Exogeneity

The dominant denomination is one of the factors by which we construct the instruments. As Table A1 indicates, a very high percentage of the population belongs to the dominant religion in many countries, while some countries have relatively low percentages of the dominant religion. For instance, the prevailing denomination is atheism in Switzerland (34%) and Germany (38%), yet they comprise roughly one-third of the population. Apparently, there are other religions in those countries with a significant presence. However, the significant presence of religions that are different from the dominant religion of a country, and their combination, might introduce some unobservable traits related to innovation attitudes. Suppose the instruments for religiosity are computed on a group of individuals that exhibit some peculiar values of the unobservable traits (that stem from the presence of different religion combinations in the country) related to innovation attitudes. In that case, the exogeneity condition may not be met. Since the instrument is the

¹⁷Note that immigrants are controlled for in the baseline specifications with variable *bornc*: born-in country.

average religiosity of a group of countries sharing the same dominant denomination, it is unlikely that such unobservable traits prevail in every country in the group.

Nevertheless, as a robustness check, we include fixed effects for each pair of religion-the dominant religion to account for diverse religious environments resulting from different combinations of religions in a country. The results are presented in Table 10. Despite moderate decreases in the coefficients, the results are pretty similar to the baseline specifications reported in Tables 4 and 5, suggesting that the instruments are relevant and representative of various religious environments.

Table 10. IV Estimates with Fixed Effects for Religion Combinations

	(1) <i>inposav</i>	(2) <i>inposav</i>	(3) <i>inposav</i>	(4) <i>inposav</i>	(5) <i>innegav</i>	(6) <i>innegav</i>	(7) <i>innegav</i>	(8) <i>innegav</i>
<i>religiosity</i>	-0.193*** (0.074)				0.180*** (0.066)			
<i>degree</i>		-0.227* (0.123)				0.205** (0.093)		
<i>attendance</i>			-0.140** (0.055)				0.161*** (0.054)	
<i>pray</i>				-0.118*** (0.043)				0.110** (0.045)
<i>gender</i>	-0.011** (0.004)	-0.009 (0.007)	-0.019*** (0.002)	-0.009* (0.005)	-0.008* (0.005)	-0.009 (0.006)	-0.001 (0.003)	-0.009* (0.006)
<i>age</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001 (0.000)	-0.001 (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.002*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>	-0.000 (0.002)	0.001 (0.002)	0.000 (0.002)	-0.000 (0.002)	0.005** (0.002)	0.004* (0.002)	0.004* (0.002)	0.005** (0.002)
<i>pwbefore</i>	-0.139*** (0.042)	-0.171** (0.070)	-0.127*** (0.035)	-0.093*** (0.022)	-0.224*** (0.046)	-0.198*** (0.060)	-0.223*** (0.044)	-0.267*** (0.040)
<i>partner</i>	-0.015*** (0.003)	-0.015*** (0.003)	-0.015*** (0.003)	-0.017*** (0.002)	0.022*** (0.002)	0.022*** (0.002)	0.021*** (0.002)	0.024*** (0.002)
<i>health</i>	0.017*** (0.001)	0.017*** (0.001)	0.018*** (0.001)	0.016*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.003** (0.001)	0.005*** (0.001)
<i>child</i>	-0.017*** (0.002)	-0.017*** (0.003)	-0.018*** (0.002)	-0.017*** (0.002)	0.001 (0.002)	0.001 (0.003)	0.002 (0.002)	0.001 (0.002)
<i>bornc</i>	-0.023*** (0.006)	-0.023*** (0.007)	-0.017*** (0.004)	-0.023*** (0.005)	-0.029*** (0.005)	-0.029*** (0.006)	-0.033*** (0.004)	-0.029*** (0.005)
<i>fathere</i>	0.006*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
<i>mothere</i>	0.004*** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
<i>N</i>	202,298	204,787	205,175	203,383	202,189	204,667	205,054	203,267
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	1,768	584	2,356	1,536	1,755	575	2,348	1,528
<i>widstat</i>	126	51	114	139	125	50	113	138
<i>AR test</i>	0.003	0.027	0.003	0.004	0.008	0.101	0.008	0.010

Notes: IV estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), and Anderson-Rubin Wald F statistic (*AR test*) are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Another possible violation of exogeneity might stem from using neighbouring countries when computing the instruments. We argue that religiosity is a transnational trait; therefore, one's religiosity can be instrumented by other people's religiosity who live in a similar religious environment. However, suppose there are substantial similarities in the institutional environments of countries that share the same dominant denomination. In

that case, individual innovation attitudes could be affected by these similarities in a particular way, violating the exogeneity assumption. We, therefore, use a large set of fixed effects to eliminate the institutional level factors. However, using neighbouring countries when computing instruments might introduce another channel of institutional similarity, given that sharing a border and spatial proximity can reinforce a similar institutional environment. Motivated by this reasoning, we change the computation of instruments as a robustness check. We build instruments for religious intensity variables by computing the average religiosity of people of the same sex, age range, and religious affiliation who live in countries with the same dominant religious denomination that are not neighbours, *i.e.* do not share a border. Table 11 presents the results. All coefficients for *inposav* and *innegav* are increased compared to the main results in Tables 4 and 5. The negative effect of religion on innovation attitudes persists.

Table 11. IV Estimates with Instruments of Non-Neighbouring Countries

	(1) <i>inposav</i>	(2) <i>inposav</i>	(3) <i>inposav</i>	(4) <i>inposav</i>	(5) <i>innegav</i>	(6) <i>innegav</i>	(7) <i>innegav</i>	(8) <i>innegav</i>
<i>religiosity</i>	-0.253*** (0.071)				0.270*** (0.062)			
<i>degree</i>		-0.271** (0.133)				0.324*** (0.098)		
<i>attendance</i>			-0.251*** (0.065)				0.259*** (0.057)	
<i>pray</i>				-0.127*** (0.042)				0.160*** (0.043)
<i>gender</i>	-0.008* (0.004)	-0.007 (0.007)	-0.017*** (0.003)	-0.008* (0.005)	-0.013*** (0.004)	-0.016** (0.006)	-0.003 (0.003)	-0.015*** (0.006)
<i>age</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.002*** (0.000)	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>	-0.001 (0.002)	0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	0.005** (0.002)	0.004* (0.002)	0.005** (0.002)	0.005** (0.002)
<i>pwbefore</i>	-0.170*** (0.043)	-0.194*** (0.075)	-0.180*** (0.042)	-0.104*** (0.019)	-0.171*** (0.038)	-0.131** (0.061)	-0.165*** (0.039)	-0.235*** (0.033)
<i>partner</i>	-0.014*** (0.003)	-0.014*** (0.003)	-0.013*** (0.003)	-0.017*** (0.002)	0.021*** (0.002)	0.021*** (0.003)	0.020*** (0.002)	0.024*** (0.002)
<i>health</i>	0.017*** (0.001)	0.017*** (0.001)	0.019*** (0.001)	0.016*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.002 (0.001)	0.005*** (0.001)
<i>child</i>	-0.016*** (0.002)	-0.016*** (0.003)	-0.017*** (0.002)	-0.017*** (0.002)	-0.001 (0.002)	-0.001 (0.003)	0.001 (0.002)	0.000 (0.002)
<i>bornc</i>	-0.027*** (0.006)	-0.026*** (0.009)	-0.022*** (0.005)	-0.023*** (0.006)	-0.025*** (0.005)	-0.024*** (0.007)	-0.031*** (0.004)	-0.026*** (0.006)
<i>fathere</i>	0.006*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	-0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)
<i>mothere</i>	0.004*** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)
<i>N</i>	202,300	204,788	205,176	203,384	202,191	204,668	205,055	203,268
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	1,621	469	1,823	1,408	1,610	463	1,815	1,401
<i>widstat</i>	121	42	95	144	120	42	94	144
<i>AR test</i>	0.000	0.024	0.000	0.002	0.000	0.001	0.000	0.000

Notes: IV estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. Kleibergen-Paap rk LM statistic (*idp*), Kleibergen-Paap rk Wald F statistic (*widstat*), Cragg-Donald Wald F statistic (*cdf*), and Anderson-Rubin Wald F statistic (*AR test*) are reported. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

1.6 Possible Channels of Causality

1.6.1 Time Allocation

Starting from Gary Becker's pioneering work "*A Theory of the Allocation of Time*" (Becker, 1965), numerous papers have analysed agents' time allocation among various activities. The work of Azzi and Ehrenberg (1975) is the first paper that considers religious participation in the context of household members' time allocation, presenting the first model of consumer choice in religious markets. They adopt a microeconomic approach to examine the demand of religion, introducing a multi-period utility-maximising model of household behaviour. In the model, religious participation is a part of household members' utility function and is a time-consuming activity as well as an investment in after-life consumption. Household members are men and women who have different opportunity costs of time due to different wages they are subject to in the labour market. They allocate their time by considering not only their wage but also the marginal utilities of after-life and in-life consumption. Suppose household members allocate more time for religious participation, favouring after-life consumption. In that case, they will have less time for productive activities for in-life consumption, resulting in lower total household production.

The point we are interested in the model of Azzi and Ehrenberg is the opportunity cost of time spent on religious activities, given that religious participation is a time-consuming activity and time is scarce. The concept of *opportunity cost of religious participation* can also be considered for other relevant time-consuming activities such as human capital formation, which requires many years of education, a lifelong habit of reading, and attending scientific and intellectual activities. All these activities demand time, and time is limited. If an individual is religious and allocates a certain amount of time to religious activities, then there will be a decrease in the maximum amount of potential time that could be devoted to human capital formation. It is well-known that human capital is one of the most critical drivers of innovation. Thus, any unfavourable activity to human capital formation can be plausibly assumed unfavourable to innovation.

On the contrary, the social capital approach is a perspective that contrasts with the negative effect of time spent on religious activities. The main idea is that participating in religious services is a form of networking and helps the attendant build social capital. Barro and McCleary (2006) underline that social capital formation through religious participation is favourable for the economy only if it fosters economically relevant individual traits such as work ethic, thrifty, honesty, and trust.

"The social capital and cultural aspects of religion (communal services, rituals, religious schools) are significant only to the extent that they influence beliefs and, hence, behaviour. In fact, for given beliefs, more time spent on communal activities would tend to be an economic drag, at least as measured by market output (GDP)." (Barro and McCleary, 2006, pp. 51).

Consequently, the net effect of religious participation, positive or negative, on economic outcomes would be determined by taking into account the comparative consequences of the two approaches mentioned: the opportunity cost of time spent on religious activities and the social capital formation through religious participation. In our case, it is somewhat arguable to claim that social capital formation fosters favourable individual traits to innovation. As mentioned in Section 1.2.2, taking risks, being different, adventurous, and creative are the main individual traits that are considered favourable to innovation. However, as shown by our analyses, religious participation does not seem to foster any of them. Moreover, it is likely to impose traditional values and rules. The

results throughout the present chapter support this argument. *Attendance* has been the most robust religiosity measure across different specifications, and it negatively affects innovation attitudes.

1.6.2 The Fear of Uncertainty

As mentioned in Section 1.2.1, Hofstede (1980/2001) identifies uncertainty avoidance as one of the five dimensions on which national cultures differ. Innovation requires newness and risk tolerance, thus is related to uncertainty which hints at unpredictability and a lack of structure and information (Rogers, 1983). Individuals or the decision-making unit involved in the first stages of the innovation process would be naturally unsure of the new idea's results and face the inherent uncertainty of newness. Therefore, an individual's propensity to innovate is expected to be positively related to the ability to deal with uncertainty¹⁸. Low levels of the fear of uncertainty refer to more openness toward new ideas and change, more willingness to take risks, and less fear of novelty.

Shane (1993) shows that uncertainty avoidance is one of the values that have a negative impact on a country's overall innovativeness, and it explains, to some degree, the variation in national rates of innovativeness across countries. Shane (1995) uses the data of 4405 individuals in 43 organizations from 68 different countries and examines the relationship between uncertainty avoidance and individual preferences for four innovation championing roles: the organizational maverick, the network facilitator, the transformational leader, and the organizational buffer. He shows that individuals from uncertainty-avoiding cultures are less likely to prefer championing roles. "One might argue that uncertainty-accepting societies are more innovative (Shane, 1993) because championing roles which overcome organizational inertia to innovation are more likely to be accepted in those societies." (Shane, 1995, pp. 64).

Chen et al. (2014) find that higher levels of uncertainty avoidance have negative effects on corporate innovation. Firms located in countries with higher levels of uncertainty avoidance generate fewer and less critical patents, and their R&D expenditures are less efficient. William and McGuire (2010) show that uncertainty avoidance has a negative effect on economic creativity which facilitates innovation implementation.

Religion is one of the ways to cope with uncertainty (Hofstede 2001). It can be argued that religious people tend to fear uncertainty more than irreligious people. The first empirical study that used risk analysis in the context of religiosity is that of Miller and Hoffmann (1995). They consider religious acceptance a risk-averse behaviour and the rejection of religious beliefs as risk-taking behaviour. One salient feature of religion is to provide a sort of protection, both materially and mentally. Religion mitigates the uncertainties and risks of daily life by providing spiritually rewarding networks of welfare activities for the community, such as charity for the poor, assistance for individuals who experience personal disasters, elder care, medical assistance, orphanages, and education (Gill and Lundsgaarde, 2004; Barro and McCleary, 2006; Scheve and Stasavage, 2006). Those activities are provided by religious organizations (e.g. churches, temples, mosques, or synagogues) and may be crucial in encouraging individuals to attend religious activities and be a part of religious organizations. Religion also helps one deal with uncertainty and fear mentally by introducing the salvation motive and after-life

¹⁸Uncertainty and risk are initially different concepts. Risk is the perceived probability that a particular event will happen. Uncertainty is about ambiguity. Any event may happen, and there is no probability attached to it (Hofstede et al., 2010). However, for simplicity, we use uncertainty as inclusive of risk since any degree of uncertainty contains risk, whether with a known or unknown probability. Hence, we use the term "the fear of uncertainty" instead of "uncertainty avoidance" to distinguish it from Hofstede's concept.

rewards. Therefore, religious people are used to being protected by divine power and religious organizations. They have preset answers for the unknown, meaning they have less experience dealing with uncertainty than irreligious people. In order to test this hypothesis, we examine the relationship between religiosity and the fear of uncertainty as follows.

We nominate five indicators from the ESS that are related to the fear of uncertainty; then, we examine their relation with religiosity measures. Each indicator corresponds to one or more differences between weak and strong uncertainty avoidance societies in Hofstede's model. For instance, societies with strong uncertainty avoidance exhibit the following behaviours: "*The uncertainty inherent in life is felt like a continuous threat that must be fought.*", "*Higher stress, emotionality, anxiety, neuroticism.*", "*Staying in jobs even if disliked.*" (Hofstede, 2011, pp.10). More broadly, Hofstede defines three components of uncertainty avoidance: rule orientation, employment stability, and stress. The variables *unemployed* (ever been unemployed and seeking job more than 3 months), *securejob* (important when choosing job: secure job) refer to employment stability; *trust* (most people can be trusted), *safe* (important to live in secure surroundings), *government* (important: government ensures safety) refer to stress. Apart from these measures, *following rules*, one of the NIA, also refers to rule orientation.

Table 12. IV Estimates: The Fear of Uncertainty

	(1) <i>unemployed</i>	(2) <i>securejob</i>	(3) <i>trust</i>	(4) <i>safe</i>	(5) <i>government</i>
<i>religiosity</i>	-0.128*** (0.024) 206,589 0.000	0.028** (0.012) 43,307 0.021	0.027*** (0.009) 206,639 0.002	0.109*** (0.011) 201,571 0.000	0.098*** (0.010) 200,322 0.000
<i>degree</i>	-0.100*** (0.023) 209,106 0.000	0.022** (0.011) 43,773 0.045	0.027*** (0.008) 209,164 0.001	0.102*** (0.010) 204,032 0.000	0.090*** (0.009) 202,752 0.000
<i>attendance</i>	-0.189*** (0.030) 209,522 0.000	0.030** (0.015) 43,901 0.041	0.036*** (0.011) 209,566 0.004	0.135*** (0.013) 204,416 0.000	0.125*** (0.012) 203,132 0.000
<i>pray</i>	-0.100*** (0.020) 207,701 0.000	0.022** (0.010) 43,547 0.030	0.021*** (0.007) 207,743 0.000	0.088*** (0.009) 202,632 0.000	0.080*** (0.008) 201,377 0.000

Notes: Robust standard errors clustered at the level of instruments are in parentheses, followed by sample size and Anderson-Rubin Wald F statistic (*AR test*). All specifications include the full set of control variables and fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The results are presented in Table 12. *Unemployed* is negatively, *securejob* is positively related to the religiosity measures. It can be said that religiosity and employment stability are associated. The variables referring to stress are positively and significantly correlated with religiosity measures, suggesting that religious people do not necessarily feel stressed. The IV results for *following rules* (Table D6) are positively and significantly correlated with all religiosity measures. Thus, rule orientation relates positively to religiosity. Overall, there are signals that religiosity and the fear of uncertainty are related,

but further evidence is needed.

1.6.3 Roles Reinforced by Religion

Roles reinforced by religion can be read through various concepts. Yet, we consider two of them: individualism-collectivism and conventional gender roles¹⁹.

In Hofstede's work, individualism-collectivism is one of the five dimensions on which national cultures differ and has been widely used in the empirical literature related to culture (Sondergaard, 1994). In individualist societies, the interests of individuals come before the interests of the group. On the other hand, in collectivist societies, the group's interests come first. Individuals work for the group, not for themselves. Therefore, complying with the group's rules and staying loyal to the group are crucial behaviours in collectivist societies that require one to follow the established rules and traditions, leaving limited space for reformist and creative endeavours. Nevertheless, creativity, which is one of the main ingredients of innovation, is said to be the act of an individual, at times in contradiction with the norms and values of the group (Amabile, 1996; William and McGuire, 2010).

Individualism may facilitate innovation by fostering a tendency to accept novelty (Steenkamp et al., 1999), giving courage to individuals to defend new ideas in the face of resistance, enabling the emergence of champion roles (Shane, 1995). Collectivism, on the other hand, may damage innovation by fostering the ideas that are acceptable to all interested parties only (William and McGuire, 2010).

William and McGuire (2010) show that individualism positively influences economic creativity, which facilitates innovation implementation. Chen et al. (2017) show that higher levels of individualism foster corporate innovation by generating more and higher impact patents and by being more efficient in converting R&D into innovative output. Taylor and Wilson (2012) find that most measures of individualism have a strong and positive effect on national innovation rates. They also find that certain types of collectivism, such as patriotism and nationalism, may also foster innovation at the country level, while other types of collectivism, such as familism and localism, harm national innovation rates.

Gorodnichenko and Roland (2010) examine the effect of individualism and collectivism on the long-run economic growth. They find that individualism has a strong and positive effect on economic growth, mainly because higher degrees of individualism lead to more innovation, given that innovation is associated with higher personal and social rewards in an individualistic culture. Individualist countries, in general, are more prosperous than collectivist societies (Hofstede et al., 2010).

Religion is likely to favour collectivist culture since following rules and staying loyal to the group, church, and god are inherent to religion. One cannot modify religion to best self-interest, as an individualistic approach would require. Some denominations are said to be more individualistic than others, such as Protestantism, yet we do not compare denominations and only discriminate between being religious and irreligious.

We examine the relationship between individualism-collectivism and religiosity based on the approaches above. We define four indicators from the ESS data that are somewhat related to individualism and collectivism. *Devote* (important to devote himself to people close to him), *family* (family should be priority in life), *success* (important: being success-

¹⁹Masculinity-femininity is one of the five dimensions of Hofstede's model, and it is a combination of conventional gender roles and the degree of orientation to material achievement (William and McGuire, 2010). However, there is no consensus in the literature that masculinity (or femininity) favours innovation. We, therefore, simplify this dimension and take only conventional gender roles into account.

ful and recognised achievements), *lookafter* (everyone should look after himself). *Lookafter* is related to personal freedom and self-care, *success* is related to personal achievement, which both are stronger in individualistic cultures than they are in collectivist cultures (Gorodnichenko and Roland, 2010). *Devote* and *family* refer to *we consciousness* and superiority of group over individual, thus expected to be stronger in collectivist cultures.

Table 13. IV Estimates: Individualism-Collectivism

	(1) <i>devote</i>	(2) <i>success</i>	(3) <i>lookafter</i>	(4) <i>family</i>
<i>religiosity</i>	0.035*** (0.007) 201,446 0.000	0.043*** (0.012) 201,068 0.000	-0.035 (0.021) 21,621 0.101	0.120*** (0.016) 21,699 0.000
<i>degree</i>	0.032*** (0.007) 203,904 0.000	0.037*** (0.011) 203,497 0.001	-0.039* (0.020) 21,798 0.054	0.112*** (0.015) 21,874 0.000
<i>attendance</i>	0.045*** (0.008) 204,293 0.000	0.058*** (0.014) 203,871 0.001	-0.040 (0.027) 21,858 0.135	0.147*** (0.019) 21,937 0.000
<i>pray</i>	0.028*** (0.006) 202,510 0.000	0.033*** (0.010) 202,119 0.001	-0.030* (0.017) 21,715 0.085	0.097*** (0.014) 21,793 0.000

Notes: Robust standard errors clustered at the level of instruments are in parentheses, followed by sample size and Anderson-Rubin Wald F statistic (*AR test*). All specifications include the full set of control variables and fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 13 presents the results. *Devote* and *family* are positively and significantly related to the religiosity measures, suggesting that higher religiosity favours collectivist culture. On the other hand, *success* and *lookafter* show contrasting results, leaving the question open whether religiosity and individualism negatively relate.

When it comes to gender differences and traditional roles, there is a vast literature and a lot to say, yet we only discuss the matter non-exhaustively for the sake of clarity.

Higher degrees of religiosity may cause individuals to be more submissive, obedient, and passive regardless of gender. Women are said to be more inclined toward those characteristics than men, mainly because of two reasons. Firstly, religions can foster gender differences since many major denominations impose strict rules on women, such as wifely submission, veiling, et cetera. Therefore, women are taught to be homemakers, fulfil childcare duties, and be committed to men and family, rather than being encouraged to work outside the home. Secondly, women are generally more religious than men because women have lower labour force participation rates and greater responsibility for housework and childcare, leading them to greater (time-wise) involvement in religion²⁰ (Miller and Hoffmann, 1995).

Being taught to be submissive and obedient is expected to restrain novelty, creativity, and innovative endeavours, thus a negative channel from religion to innovation for both genders, yet expected to be more influential for women. Conventional gender roles

²⁰See Miller and Stark (2002), Vaus and McAlliste (1987) for gender differences and religiosity.

empowered by religion are a significant drawback for innovation and the economy, allegedly a major reason for the lower female labour supply. If nothing, half of a given population is being treated differently regarding access to education, liberty, and social and legal rights means that up to 50% of the capacity to create and produce cannot be properly used.

Table 14. IV Estimates: Gender Roles

	(1) <i>lgbt</i>	(2) <i>womenwork</i>	(3) <i>menwork</i>
<i>religiosity</i>	-0.196*** (0.011) 200,719 0.000	0.187*** (0.018) 79,002 0.000	0.136*** (0.020) 106,126 0.000
<i>degree</i>	-0.176*** (0.011) 202,984 0.000	0.175*** (0.016) 79,994 0.000	0.128*** (0.019) 107,436 0.000
<i>attendance</i>	-0.249*** (0.014) 203,336 0.000	0.236*** (0.021) 80,200 0.000	0.166*** (0.025) 107,666 0.000
<i>pray</i>	-0.158*** (0.009) 201,706 0.000	0.151*** (0.015) 79,491 0.000	0.113*** (0.017) 106,709 0.000

Notes: Robust standard errors clustered at the level of instruments are in parentheses, followed by sample size and Anderson-Rubin Wald F statistic (*AR test*). All specifications include the full set of control variables and fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

We define three variables from the ESS that could reflect the strong connection between traditional gender roles and religiosity. *lgbt* (gay men and lesbians should be free), *womenwork* (a woman should be prepared to cut down on her paid work for the sake of her family), *menwork* (when jobs are scarce, men should have more right to a job than women). Table 14 displays the results. *lgbt* is negatively and significantly correlated, while *womenwork* and *menwork* are positively correlated with all measures of religiosity, supporting the hypothesis that religiosity favours conventional gender roles.

1.7 Conclusion

Religion is a multi-dimensional and complex phenomenon whose effect is embedded in social and economic behaviour. Hence, using religion in regression without considering its multi-dimensionality and embeddedness would yield spurious correlations. More formally, the potential sources of the endogeneity of religion, including omitted variables, reverse causality, and measurement error should be considered to reach an accountable and causal effect of religion on economic outcomes; otherwise, estimates will be unreliable.

The present study provides the first attempt to tackle the inherent endogeneity of religion with respect to innovation. We use different measures of religiosity, adopt an

individual-level approach to innovation, use multi-way fixed effects, and employ the instrumental variables method. The results show that higher religiosity negatively affects positive innovation attitudes and positively affects negative innovation attitudes, implying that being religious is not a catalyser for innovation-related traits and is possibly an obstacle. This finding is in line with the literature (Benabou et al., 2013, 2015). Using OLS estimates, Benabou et al. (2015) find that religiosity is positively related to creativity, which they describe as “puzzling”. We contribute to this literature by uncovering that the positive relationship between religiosity and creativity is driven by endogeneity. We find a negative relationship between religiosity and creativity once we rule out endogeneity by applying the IV method.

The negative effect of religiosity is robust to a series of sensitivity analyses. First, we introduce a different dependent variable, an overall summary measure of innovation attitudes: *innosum*. OLS and IV results of *innosum* support the negative effect of religiosity on innovativeness. Second, we reconsider the religiosity index (*religiosity*) and drop the degree of being religious (*degree*) from its components, thinking that *degree* might be prone to self-report bias since it is a self-evaluation of a belief, a value. In contrast, the other components of *religiosity* (frequency of pray and attendance to religious services) are measures of an activity, an action that is difficult to over or underestimate. The estimated coefficients are pretty close to the initial one, meaning that the initial instrument of *religiosity* is not significantly affected by the potential self-report bias of *degree*. Third, we re-estimate the baseline model by reducing the main sample in four ways to analyse if minorities and/or outliers affect the main findings. Fourth, we consider two possible violations of the exogeneity condition. Sensitivity analyses, overall, demonstrate that coefficient values might change depending on the specification choices, but the main results stay robust.

As possible channels from religion to innovation, we empirically examine and discuss three approaches: time allocation, the fear of uncertainty, and roles reinforced by religion, such as traditional roles and gender roles.

Although the instrumental variables method is seen as effective in dealing with endogeneity and the diagnostic tests suggest non-weak instruments, the results should be approached with caution since we use observational survey data. As Iannaccone (1998) points out, nothing less than a genuine experiment that is probably unattainable will demonstrate the true causal effect of religion. In this study, we aim to approximate the causal effect of religion on innovation, yet further research is needed. First, the negative effect of religion on innovation should be tested with different instruments and sophisticated estimation methods. Second, different data sources, especially not self-reported attendance data, can be used to rule out the potential self-report bias fundamentally. Third, the causal channels from religion to innovation and their underlying mechanisms should more deeply be examined since we discuss them non-exhaustively.

Chapter 2

Workplace Skills as Regional Capabilities: Relatedness, Complexity, and Industrial Diversification¹

2.1 Introduction

A large strand of the evolutionary economic geography (EEG) literature has shown that regions build on existing capabilities and diversify into related activities (Boschma, 2017; Hidalgo et al., 2018; Balland et al., 2019; Balland and Boschma, 2019). *Existing capabilities* is an extensively used, broad term for productive inputs that range from tangible assets, *i.e.*, physical capital, labour force, to intangible assets, *i.e.*, norms, institutions, knowledge and skills. What the literature has been somewhat reluctant to do is to explicitly address capabilities, despite diversification is seen as a process in which new activities emerge from new combinations of existing capabilities in regions (Antonietti and Boschma, 2021). In this context, the meaning and measurement of relatedness have been questioned (Tanner, 2014; Boschma, 2017) in the sense that relatedness stands for the similarity of entities that share common capabilities, yet the measurement of relatedness seldom include capabilities.

The mainstream practice, co-occurrence based relatedness, uses co-location numbers of entities as input data to compute location quotients to further use them to calculate the minimum of conditional probabilities that two entities co-occur together. Then, these two entities are assumed to share similar capabilities if they have a high relatedness score, yielding an outcome-based assessment and interpretation of regional capabilities. This approach has generated criticism in the literature. Kogler (2017, pp. 2) underlines that *"co-location of activities is frequently an a priori assumption of connectedness."* Whittle and Kogler (2019, pp. 8) ask *"Related to what?"* and as a solution, they emphasise the potential of skill relatedness to embrace capabilities at micro level: *"... this approach has the benefit of capturing relatedness at a micro level via occupational data and labour mobility, something that the previous relatedness measures have either failed to do or done so inadequately."*

Skill relatedness concept is introduced by Neffke and Henning (2013) who developed the revealed skill relatedness method (RSR) that assumes two industries to be skill-related if they exhibit intensive labour mobility. Skill similarity is, therefore, an a priori assumption and relatedness is indirectly observed (Neffke and Henning, 2008). In this chapter, we argue that regional capabilities are partly embodied and reflected in the knowledge, abilities, and skills of a region's workforce employed by the region's industry

¹Joint work with Leonardo Mazzoni, Massimo Riccaboni and Francesco Serti.

mix. Based on this argument, we aim to further elaborate the existing practice by explicitly accounting for regional capabilities as workforce skills in the measurement process of relatedness. We also consider the complexity approach, along with relatedness, to adequately account for different aspects of capabilities.

Relying on a unique data set on workplace skills, the Italian Sample Survey on Professions (ICP), we construct skill relatedness and complexity matrices for 534 industries for each of the 107 Italian NUTS 3 regions² over the period 2013-2019, based on the 161 workplace skills' intensities. First, we provide an overview of these measures, formalised as *the skill space* of the industrial profile of Italian regions. By employing data-driven methods, we investigate the similarity between skill types in terms of their effective use by industries and how their degree of relatedness and complexity differ. The results indicate that workplace skills are clustered into two main communities: social-cognitive skills and technical-physical skills. We find that social-cognitive skills have the highest complexity scores while technical-physical skills are generally below average. We then deploy econometric models to estimate the impact of relatedness and complexity of different skill types on the process of industrial specialisation of regions. The results indicate that, regardless of the skill type, higher skill relatedness and complexity enhance the probability of building a new comparative advantage, whilst lower skill relatedness and complexity increase the probability of losing an established comparative advantage. However, similarity to technical-physical skills yields a higher (lower) probability of entry (exit) than social-cognitive skills. Conversely, higher complexity in social-cognitive skills is associated with a slightly higher probability of entry. We also find that social-cognitive skill relatedness does not enhance or mitigate specialisation process of service industries, while technical-physical skill relatedness does. On the contrary, specialisation process of manufacturing industries is affected by the relatedness of both types of skills.

The present work contributes to the literature in four ways. First, to the best of our knowledge, it is the first empirical study that explicitly analyses the relationship between skill relatedness based on workplace skills and industrial diversification, *i.e.* entry and exit probability, at the regional scale in a developed country.

Second, this is the first study that considers the role of different skill types in industrial and regional diversification literature. So far, skill relatedness research for regional diversification has employed the RSR method, forming *implicitly* skill-related industries (Elekes et al., 2021). The presence of detailed and reliable data on workplace skills enables us to use skill scores in input matrices to *explicitly* define skill relatedness measures for different skill types.

Third, this chapter provides the first empirical attempt to consider the skill complexity of industries and to jointly analyse skill relatedness and skill complexity in explaining path-dependencies of industrial diversification of regions. The present study shows that the skill complexity of industries, along with skill relatedness, plays a role in diversification process.

Fourth, we contribute to the relatedness literature by providing more geographical wisdom to relatedness as the literature tends to treat relatedness as a global phenomenon (Boschma, 2017). Nevertheless, region specific capabilities might cause the degree of relatedness between activities to differ across regions. The high granular analysis we conduct shows that skill relatedness of a particular industry dramatically changes across regions, indicating that relatedness is not independent of its spatial context. Differences in regional workforce skills of the same industry might conceptualise a convincing answer to the question of why some industries are related in one spatial context and not

²NUTS 3 regions known as *provinces* in Italy.

related in another.

The structure of the chapter is as follows. Section 2.2 briefly overviews the related literature. Section 2.3 describes data sources. Section 2.4 introduces skill relatedness and skill complexity measures. Section 2.5 presents econometric analyses. Section 2.6 provides sensitivity analyses. Section 2.7 overviews the main findings and concludes.

2.2 Literature

2.2.1 Regional Capabilities, Diversification, and Skill Relatedness

The main argument behind regional diversification is that new activities emerging in a region are prone to be related to the region's preexisting activities in terms of capabilities, knowledge, and skills, which is called *relatedness* (Hidalgo et al., 2007; Neffke et al., 2011; Hidalgo et al., 2018). As Neffke and Henning (2013) pointed out, there are three different approaches to measure relatedness: (1) hierarchical measure based on standard industry classification systems such as NACE and SIC (Chang, 1996; Farjoun, 1998; Lee and Lieberman, 2010), (2) outcome-based co-occurrence methods (Neffke and Henning, 2008; Hidalgo et al., 2007), (3) resource-based measures such as technological resources (Breschi et al., 2003), human-capital resources (Farjoun, 1994)³.

Among the three measures the literature has identified thus far, co-occurrence based relatedness, only *relatedness* hereafter, has gained prominence. Notably, after the pioneering work of Hidalgo et al. (2007), scholars have developed relational networks using the relatedness concept to shed light on the diversification capabilities of a variety of activities such as trade (Hidalgo et al., 2007; Boschma et al., 2013; Guo et al., 2020), technologies (Kogler et al., 2013; Boschma et al., 2015; Tanner, 2016; Montesor and Quatraaro, 2017; Balland et al., 2019), jobs (Muneepeerakul et al., 2013; Farinha et al., 2019), skills (Neffke and Henning, 2013; Neffke et al., 2017) and industries (Neffke et al., 2011; Xiao et al., 2018) at multiple spatial scales, including countries, regions or cities (Whittle, 2019). In this literature, capabilities have generally been captured by co-location patterns of different entities. If two entities co-exist, their spatial units are assumed to have similar capabilities. For instance, Hidalgo et al. (2007) measure relatedness between trade products based on their co-location patterns, arguing that two trade products demand similar capabilities if product pairs happen to be frequently present in the same location. Neffke et al. (2011) capture regional capabilities by product relatedness between manufacturing plants. In Rigby (2015), capabilities are addressed by technological relatedness that is derived from the co-occurrence of patent classes on patent documents. Accordingly, capabilities are seen as an enabling but implied source of regional diversification whose exact nature is not directly observed (Boschma, 2017). This partly blind approach has generated criticism in the literature (Kogler, 2017), while some relatedness types are seen as more capable of providing a more explicit measure, such as skill relatedness (Whittle and Kogler, 2019).

Studies on skill relatedness, still in its infancy, can be divided into two major approaches with respect to their methodology. The first approach employs *the revealed skill relatedness* (RSR) method developed by Neffke and Henning (2013) and applied by Timmermans and Boschma (2014); Boschma et al. (2014); Diodato and Weterings (2015); Fitjar and Timmermans (2017); Neffke et al. (2017); Neffke et al. (2018); Elekes et al. (2021). Neffke and Henning (2013) underline that different industries in the economy seem to

³See Neffke and Henning (2013) and Whittle (2019) for brief reviews.

be interconnected by linkages of skill similarity, and firms are much more likely to diversify into the activities related to their core competencies. They take a micro perspective and assume that, when switching jobs, individuals tend to remain in industries related to their previous job in terms of skill content. Based on this assumption, they develop a method that employs co-occurrence techniques to assess skill-relatedness between industries by using cross-industry labour flows in Sweden. They predict expected flows between industries to compare them to actual labour flows. Industries that exhibit mobility more than the expected flows are assumed to be skill-related. The results show that the RSR index is predictive of firm diversification. Fitjar and Timmermans (2017) propose an improvement to the RSR method with a measure at the industry and region level.

Inter-industry labour flows might unravel exciting insights, yet such data are not available for many countries, making the RSR method challenging to employ. Moreover, the RSR method provides an implied measure of skill similarity between industries as it does not use workplace skills data, deducing skill relatedness from labour mobility. Excessive labour mobility between industries does not necessarily indicate high skill similarity. Neffke and Henning (2013) emphasise that the RSR method is an indirect measure of skill relatedness, and intense labour flows might also stem from similar corporate culture and social network effects. Hence the RSR index may also capture other effects than skill similarity introducing a substantial risk of overvaluation of skill relatedness.

Furthermore, the RSR method does not allow horizontal differentiation of skills and employs an agnostic attitude to different skill types due to the lack of workplace skills data. Previous studies, therefore, have been limited to *industry space*, an industry-to-industry formalisation of labour-flow based skill relatedness, not being able to provide a skill-to-skill analysis of industries. Consequently, the impact of different skill types on the industrial diversification of regions has not been considered so far. In addition, these drawbacks have prevented previous studies from incorporating skill complexity into the analyses, leading to an either/or approach towards skill relatedness and skill complexity. However, sophisticatedness of skills may unravel valuable insights into the regional diversification process.

The second methodology for skill-relatedness uses data on workplace skills, employs co-occurrence methods à la Hidalgo, and combines them with network analyses to construct a skill space, a network formalisation of skill-to-skill relatedness matrix. This approach can differentiate between skill types and does not require labour-flow data, unlike the RSR method. Nevertheless, only a couple of studies hitherto applied this method to focus on the relationship between different skill types. Anderson (2017) uses online freelance job market data with network methods to categorise worker skills based on their relationship with other skills in the market. She argues that workers with diverse and synergistic skills earn higher wages than others. Alabdulkareem et al. (2018) construct the skill space of occupations using skill relatedness based on co-occurrence methods combined with network analyses. They apply a community detection algorithm to the skill space and find two highly polarised skill clusters associated with the polarisation of wages and the hollowing out of the middle-wage occupations. However, these studies apply skill relatedness to occupations globally; none considers skill relatedness at the industry and region level in the context of regional diversification.

The present study combines and elaborates on these two approaches. We use workplace skills to compute relatedness as in the second approach and consider skill relatedness in the industrial diversification process of regions as in the first approach. To our knowledge, no empirical study has analysed the relationship between skill relatedness,

based on workplace skills, and industrial diversification, *i.e.*, entry and exit probability, at the regional scale in a developed country.

2.2.2 Regional Capabilities, Diversification, and Skill Complexity

In their pioneering work, Hidalgo and Hausmann (2009) introduce a measure of economic complexity that draws valuable information from the structure of a bipartite country-product network, *i.e.*, product space. Their measure is based on the method of reflections (MOR), which iteratively combines two variables: products' ubiquity and countries' diversity. Consequently, countries with non-ubiquitous, untradable capabilities enjoy an exclusive source of comparative advantage, thereby producing more complex and privileged goods that only a small fraction of countries can produce. On the contrary, countries experiencing less exclusive, ordinary capabilities produce ubiquitous goods that many countries can produce; thus, they tend to have low complexity scores. The economic complexity approach based on MOR has been replicated by many scholars⁴ using either trade data and product space or different relational networks, including knowledge space (Balland and Rigby, 2017), digital proximity (Rahmati et al., 2021), technological complexity (Whittle, 2019; Mewes and Broekel, 2020) and skill complexity (Caines et al., 2017; LoTurco and Maggioni, 2020; Antonietti et al., 2021).

The empirical work related to skill complexity is still scarce. LoTurco and Maggioni (2020) construct an occupational complexity measure by exploiting the knowledge and skill requirements of occupations to investigate the labour content of complex products. They find that the new measure of occupational complexity is more predictive of GDP growth in the USA than other measures. Antonietti et al. (2021) build three complexity measures, occupation, task, and skill, to explore the impact of key enabling technologies (KEY) on workforce demand. They show that a larger share of KEY increases the demand for more complex occupations, tasks, and skills.

To our knowledge, no study has yet investigated skill complexity in industrial diversification literature, nor with skill relatedness. In this chapter, we aim to fill this gap by arguing that skill complexity is highly relevant in the diversification process.

Regional research has come to a consensus that regional resources which underpin the diversification process are often localised, non-tradable and non-ubiquitous (Neffke et al., 2018), implying that a region's workplace skills are mostly region-specific and spatially dependent. The complexity approach introduces an effective method to quantify these aspects of regional skills in terms of their diversity, ubiquity, and sophisticatedness in a comparable way to other regions' skills. Therefore, complexity is complementary to relatedness rather than being rival. In this chapter, we consider relatedness and complexity as two different mechanisms that are beneficial to unravelling the path-dependent nature of the diversification process.

2.3 Data

The data for the present study derives from a couple of sources. The primary occupational and industrial data source is the Italian Labour Force Survey (ILFS) provided by the National Institute of Statistics of Italy (Istat). The data on workplace skills, the Italian Sample Survey on Professions (ICP)⁵, is obtained from the National Institute for Public Policies Analysis (INAPP). ICP collects detailed information on the characteristics of all

⁴See Balland et al. (2022) for an overview.

⁵See <https://www.inapp.org/it/dati/ICP> for more details.

professions existing in the Italian labour market, particularly on the content of the work, knowledge and skills it requires, and the organisational structure where the work takes place. The ICP data provide extremely granular and valuable information on workplace skills and reflect the Italian labour market structure. The concept and survey questions are borrowed from the Occupational Information Network (O*net)⁶ that is run by the Bureau of Labour Statistics in the USA. Up to date, there are two ICP waves: 2007 and 2013. In each wave, almost 16.000 workers are interviewed, representing approximately 800 occupational units at the five-digit level in the context of the Classificazione delle Professioni (CP), which is the Italian version of ISCO classification⁷. We use only one wave, the ICP 2013, to match the occupational classification scheme of ILFS.

Each wave consists of seven sections, each of which captures one aspect of occupations:

- Knowledge (33 questions, both importance and complexity level);
- Skills (35 questions, both importance and complexity level);
- Attitudes (52 questions, both importance and complexity level);
- Generalised working activities (41 questions, both importance and complexity level);
- Values (21 questions);
- Working styles (16 questions);
- Working conditions (57 questions).

The major ICP sections can further be broken down into more specific, homogeneous, and task-related sub-categories, which are presented in Table A1. The sub-categories, or the ICP descriptors, are adapted from the O*net data descriptors. Each competence in sub-categories is cross-checked against the ICP questionnaire. The codes in parentheses stand for the ICP question number that we use to identify workplace skills throughout this chapter.

The four major sections of ICP data appear to be well suited for this study: knowledge, skills, attitudes, and generalised working activities, which sum up to 161 skill variables. Mainly because they use the same question design that accounts for both the importance and level of the competence in question⁸. The remaining sections have various question designs and scales, including only importance, frequency, time, agreement, et cetera. Hereafter, we use the term *skills* in a broader sense to refer to these four ICP sections.

ICP data refer only to occupational categories at the five-digit level. Therefore, we use ILFS data to connect workplace skills to spatial and industrial information. We constructed the main data set as follows. First, we transform ICP data from the five-digit occupational level (796) to be at the four-digit level (507), given that ILFS is available at the four-digit level. Then we generate skill intensity variables for each workplace skill by multiplying importance scores with level scores in a similar fashion as it was done for O*net variables in the works of Feser (2003), Gabe and Abel (2011), and Krenz (2014). This multiplicative approach increases the skill variation across occupations (Feser, 2003).

⁶See <https://www.onetcenter.org> for details.

⁷The ICP 2007 uses CP 2001, while the ICP 2013 uses CP 2011.

⁸Importance question: *How important is competence in carrying out your current profession?* Level question: *Among those indicated below, at what level is this competence necessary for the development of your current profession?* Importance questions rated on a scale from 1 (not important) to 5 (extremely important), while complexity level questions rated on a scale from 1 (least complex) to 7 (most complex). Then they are re-scaled to be between 0 and 100.

Second, we merge ICP and ILFS data sets on the four-digit occupational level. By doing so, we end up with 497⁹ occupational categories and 161 workplace skills. Occupations match on a one-to-one basis, given that both data sets are classified according to CP 2011.

Lastly, we compute average skill intensity scores for each industry. Recall that the merged ICP/ILFS data set provides skill distributions of occupations (from ICP), occupational distributions of industries (from ILFS) and industrial portfolios of regions in each year (from ILFS). We use these distributions to generate average skill intensity variables for each industry in each region and year. After excluding part-time workers, and individuals out of the 15-64 age range, the resulting sample consists of 161 average skill intensity variables for 534 industries and 107 regions for the period 2013-2019.

For the econometric analyses, we use employment (*Local Units and Persons Employed: Size Class of Persons Employed, Economic Activities, Geographical Areas*) and business register (*ASIA Business Register Database*) data from Istat to construct dependent variables. The data for control variables are extracted from both Eurostat (GDP per capita, business growth, churn) and Istat (population density, education, industrial ubiquity, regional diversity).

2.4 Construction of Skill Space

There are four main entities in this study: workplace skills ($N_s = 161$), regions ($N_p = 107$), industries ($N_i = 534$), and time ($N_t = 7$). We first construct industry-skill input matrices, $\mathbf{C}_{is}$, for each region in each year, summing up to 749 matrices. Each input matrix consists of 161 skills and 534 industries whose each cell $x_{i,s}$ contains the skill intensity of industry i for skill s ($i = 1, \dots, n; s = 1, \dots, m$) in region p in year t . We then compute skill relatedness matrices and skill complexity vectors by using these input matrices as defined below.

2.4.1 Skill Relatedness of Industries

We estimate skill relatedness between skill pairs based on the framework proposed by Hidalgo et al. (2007). The first step is to define the effective use of skills. Skill $s \in S$ is effectively used by industry $i \in I$ if its relative skill advantage (RSA) is greater than 1. RSA is the share of the relative importance of skill s to industry i (the numerator in equation 2.1) to the relative importance of skill s to all industries I (the denominator). RSA is a measure based on the Balassa index, also known as location quotient (LQ) and revealed comparative advantage (RCA). A higher value of RSA indicates a higher level of importance of skill s for industry i compared to the overall importance of skill s for all other industries. We apply the RSA formula to the industry-skill matrix $\mathbf{C}_{is}$. The result is $(N \times M)$ two-mode adjacency matrix $\mathbf{M} = (M_{i,s})$ where $M_{i,s} = 1$ if RSA is greater than 1, *i.e.* industry i effectively uses skill s , and $M_{i,s} = 0$ otherwise. Figure 2 presents several realisations of matrix $\mathbf{M}$.

$$RSA(i, s) = \frac{icp(i, s) \setminus \sum_{s' \in S} icp(i, s')}{\sum_{i' \in I} icp(i', s) \setminus \sum_{i' \in I, s' \in S} icp(i', s')} \quad (2.1)$$

⁹The ICP sample does not contain Armed Forces, we thus excluded these occupational categories. Legislators and Senior Officials are recoded due to aggregation differences between ICP and ILFS.

After we identify effectively used skills by each industry, we compute the skill relatedness of industries between each pair of effectively used skills based on the minimum conditional probability of their co-occurrences in industry classes as formulated in equation 2.2¹⁰.

$$R(s, s') = \frac{\sum_{i \in I} e(i, s) \cdot e(i, s')}{\max(\sum_{i \in I} e(i, s), \sum_{i \in I} e(i, s'))} \quad (2.2)$$

where effective use of skills denoted as $e(i, s) = 1$ if $RSA > 1$, and $e(i, s) = 0$ otherwise. The resulting matrix is the skill relatedness index ($M \times M$) of N industries which contains proximities between two skill types s and s' . Each cell (s, s') represents the probability that a random industry effectively uses skill $s(s')$ effectively uses skill $s'(s)$ as well.

The skill relatedness index is an adjacency matrix of 161 skills based on the co-occurrence analysis of 534 industries in region p in year t . Applying equations 2.1 and 2.2 to all input matrices leaves us with 749 skill-to-skill relatedness indexes. We first analyse these indexes at the national level, averaged for regions and years.

Figure 1 displays the skill relatedness matrix of Italy's industrial portfolio formalised as a one-mode network that we coin as *the skill space*¹¹ for the period 2013-2019. The individual nodes represent 161 skills and are coloured by the skill sub-categories presented in Table A1. The edges between skills indicate their degree of relatedness. The visualised network is fully connected, *i.e.*, we do not allow for any isolates. All skills are presented in the graph, although the network is thresholded (relatedness ≥ 0.45) for visualisation purposes.

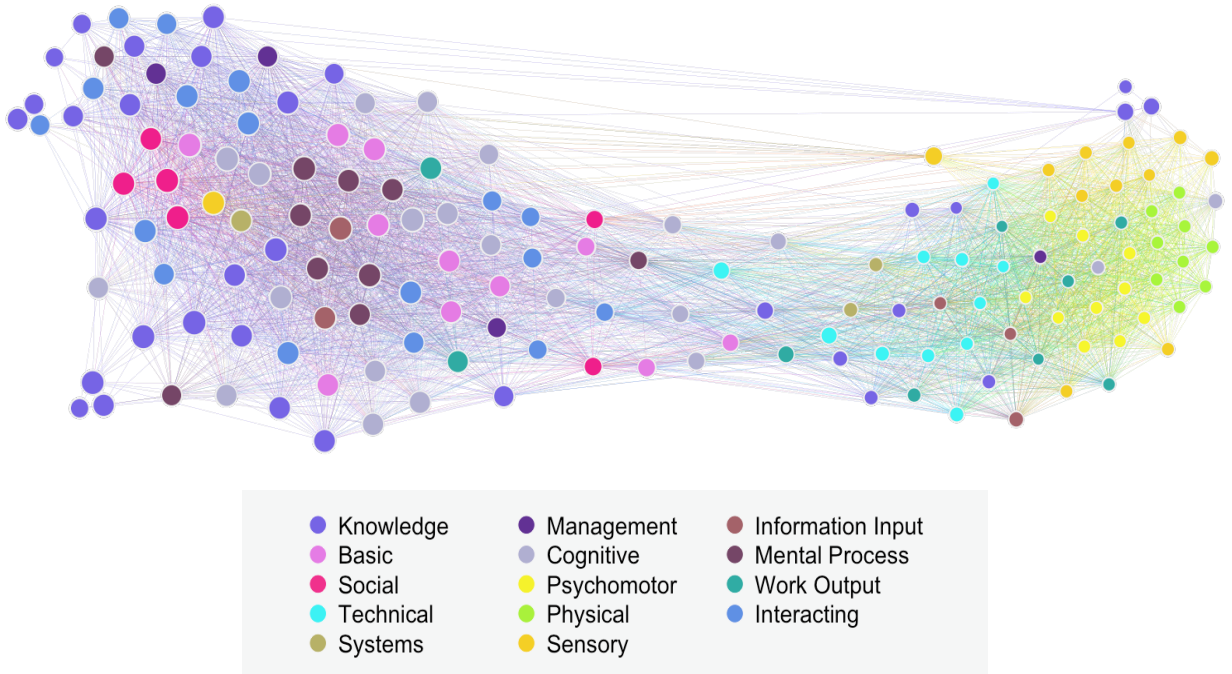
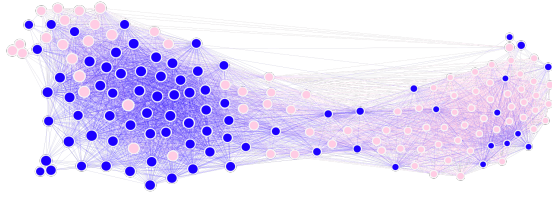


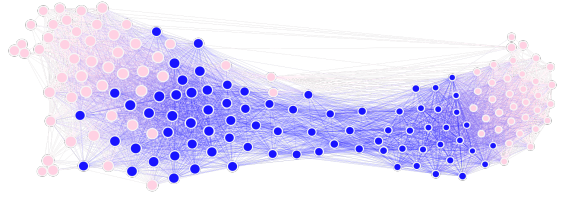
Figure 1. The Skill Space (2013-2019). Nodes represent skills. The size of each node is proportional to the complexity level of the skill the node represents. Nodes are coloured to the subcategories of skills. Edge lengths show the degree of relatedness between skill pairs.

¹⁰The standardisation method we used is 'cosine', and implemented with Econ Geo R package by Balland (2016). See VanEck and Waltman (2009) for the standardisation methods of co-occurrence data.

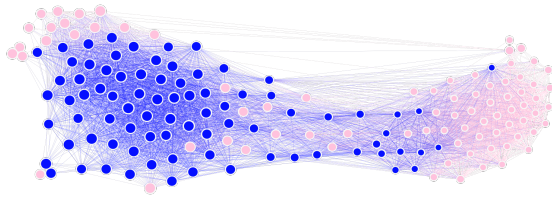
¹¹The network is visualised in Gephi software by using Multi-Scale Force Atlas, which is a force-based algorithm. The two highly polarised skill clusters are robust to different layout algorithm choices.



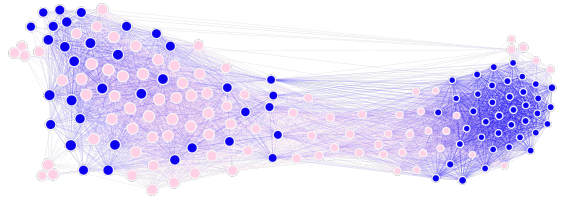
(a) Artistic creation



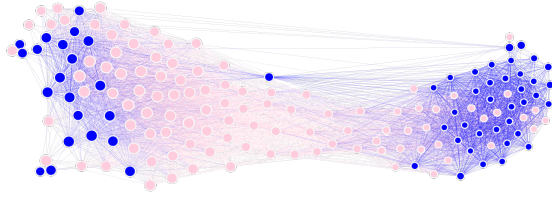
(b) Manufacture of electronic components



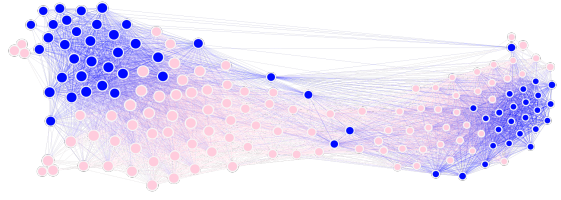
(c) Satellite telecommunications activities



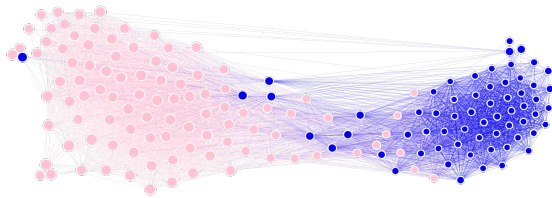
(d) Restaurants and mobile food service activities



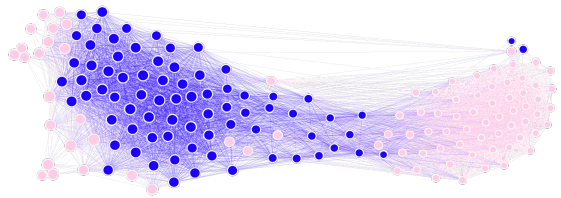
(e) Taxi operation



(f) Retail sale of textiles in specialised stores



(g) Construction of roads and motorways



(h) Activities of holding companies

Figure 2. Example Industries and Effective Use of Skills. Skills in which the industries have $RSA > 1$ are projected onto the skill space using navy nodes.

A striking feature at first glance is that the skill space explicitly forms two highly polarised skill clusters. Physical, psychomotor, sensory, systems, and technical skills are located on the right. Basic, social, management, interacting, and higher concentration of knowledge and cognitive skills are located on the left. More related skills tend to cluster together; therefore, the skill space gives a first impression of which skills are mostly used together by industries.

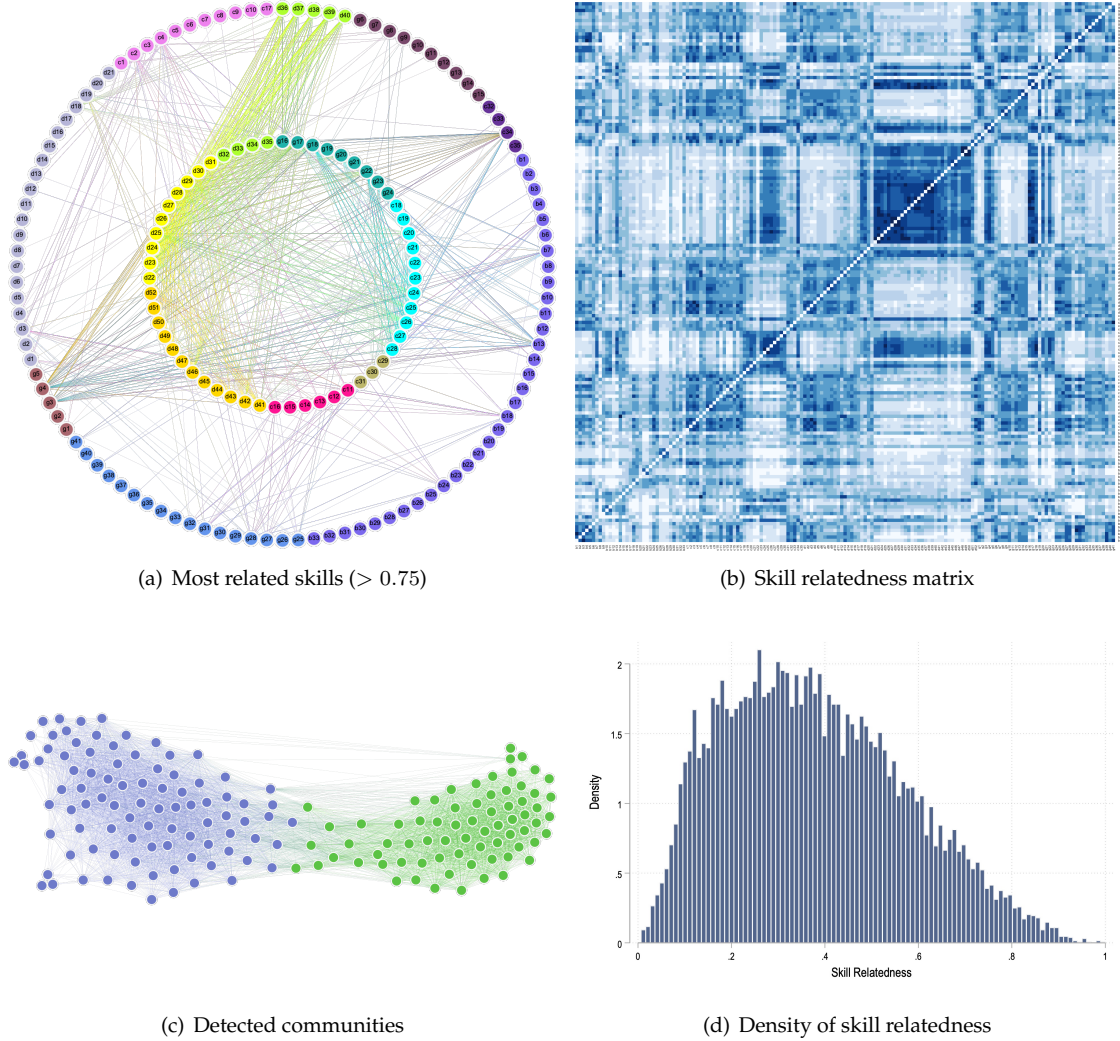


Figure 3. The Properties of the Skill Space

We further analyse the skill polarisation by employing a data-driven community detection method. Figure 3(c) presents the detected communities with the multi-level modularity optimisation (Louvain) algorithm (Blondel et al., 2008) and affirms the polarisation of skills. The resulting modularity is 0.42, showing that partition is solid and community structure is significant. The two communities of skills are unfolded in Table A2.

The polarisation of workplace skills is in line with the findings of Alabdulkareem et al. (2018). They analyse the skill relatedness of occupations in the USA labour market and find two polarised skill clusters, which they coin as *social-cognitive* and *sensory-physical*. We embrace a similar approach and term the two skill clusters as *social-cognitive* and *technical-physical* skills.

Figure 3 gives further insights into the structure of the skill space. Figure 3(a) shows the most related skills, *i.e.* relatedness score above 0.75, using a dual-circular layout. Figure 3(b) displays the skill space as a similarity matrix. Figure 3(d) presents the density of pairwise skill relatedness scores.

The illustrative and quantitative evidence provided thus far is representative at the national level. It is also possible to analyse the skill space with additional dimensions such as time and space. Recall that the skill relatedness formula in equation 2.2 yields a skill-to-skill matrix; therefore, the dimensional extension requires an additional indicator of relatedness. To this end, we employ a measure of relatedness density (equation 2.3) that is developed by Hausmann and Klinger (2007) and re-formulated by Balland et al. (2015) as *technological flexibility measure*. We use RSA (equation 2.1) and the skill relatedness index (equation 2.2) to construct the density measure.

$$AverageRelatednessDensity_{i,t}^p = \frac{\sum_{s \in i} \left(\frac{\sum_s \phi_{s,j,t} RSA_{s,i,t}}{\sum_s \phi_{s,j,t}} \times 100 \right)_{s,i,t}}{\sum_{s \in i} s} \quad (2.3)$$

where $\phi_{s,j,t}$ refers to relatedness between skill s and j at time t ; $RSA_{s,i,t}$ is a binary variable that takes the value of 1 if industry i effectively uses skill s at time t , and takes the value of 0 otherwise. Accordingly, skill relatedness density around industry i at time t in region p is defined as the sum of relatedness values between all the skill pairs that industry i effectively uses, divided by the sum of relatedness between all the skill pairs available in region p at time t . As a result, the skill relatedness density formula, located inside the parentheses in the numerator of equation 2.3, gives a matrix ($M \times N$) whose each cell indicates relatedness density between skill s and industry i . We use this skill relatedness density matrix to calculate the average skill relatedness density (ASRD) of industry i in region p as the sum of relatedness densities of skills that industry i effectively uses to the sum of all skill relatedness densities available in region p at time t . ASRD measure allows us to compare the required skill portfolios of different industries across regions and years by combining information on the use of 161 different workplace skills. ASRD will be high for a new industry if its skill space is similar to other industries in which the region has an RSA. In other words, ASRD measures how close a potential new industry is to the region's existing industry mix in terms of human capital.

Figure 4(a) demonstrates ASRD scores averaged for Italian regions for the period 2013-2019. The higher the ASRD of region, the closer its existing set of effectively used skills to the missing skills in the region. This is to say that the region is closer to forming the necessary capabilities to specialise in new industries that require different sets of workplace skills than the existing ones. The figure displays considerable differences in the branching potentials of regions. The northern and upper-central regions have higher potential than the lower-central and southern regions.

Figure 5 displays the dynamic nature of ASRD, exemplified by three industries for 2013 and 2019. The subfigures indicate that the same industry might use different skills and be skill-related to different industries in different locations, underlining the importance of regional scale and inadequateness of global relatedness measures.

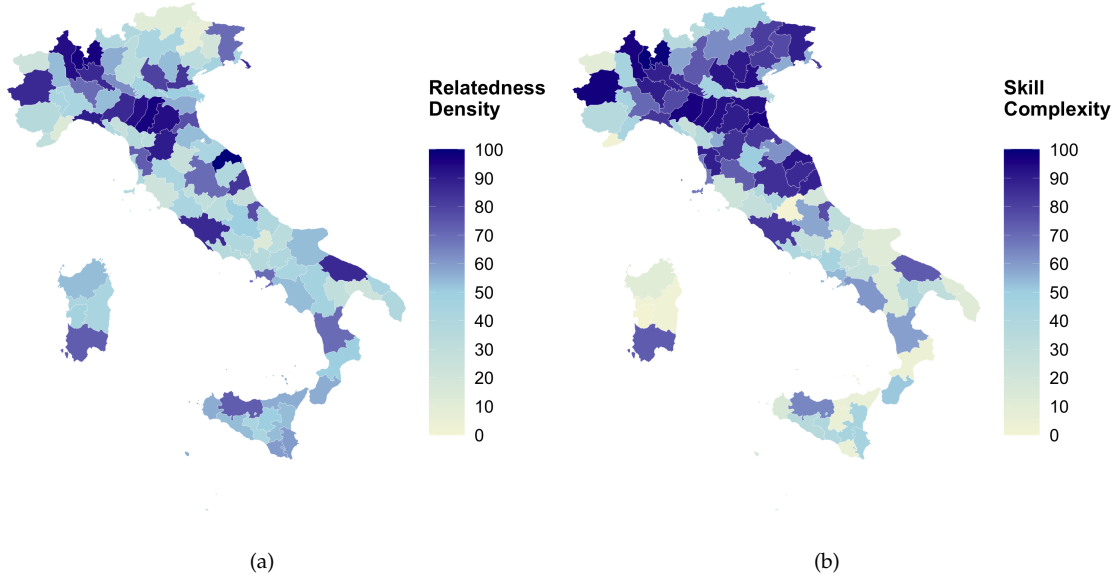


Figure 4. Skill Relatedness Density and Skill Complexity Index of Italian Regions

2.4.2 Skill Complexity of Industries

The complexity measure used in the present study is *the method of reflections* (MOR), introduced in the pioneering work of Hidalgo and Hausmann (2009). MOR sequentially combines two measures: diversity and ubiquity. Diversity ($K_{i,0}$) is the number of skills effectively used ($RSA > 1$) by industry i . Ubiquity ($K_{s,0}$) is the number of industries that effectively use a particular skill s .

$$Diversity = k_{i,0} = \sum_s M_{i,s} \quad (2.4)$$

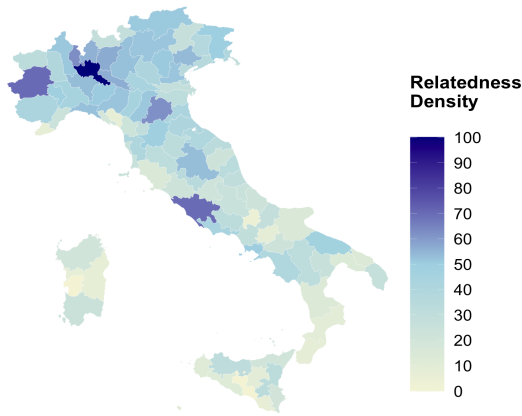
$$Ubiquity = k_{s,0} = \sum_i M_{i,s} \quad (2.5)$$

where $M_{i,s}$ is an adjacency matrix of industries and skills, resulting from equation 2.1. $M_{i,s} = 1$ if industry i effectively uses ($RSA > 1$) skill s , and $M_{i,s} = 0$ otherwise. These measures indeed summarise the structural properties of a bipartite industry-skill network. Diversity measures the number of edges between industry i and skills, while ubiquity measures the number of edges each particular skill has. After sequentially combining diversity and ubiquity measures for $N \geq 1$ steps, MOR is defined as iterative linear equations that are theoretically infinite. In other words, MOR iteratively calculates the average value of the previous-level properties of a node's neighbours as defined in the following equations.

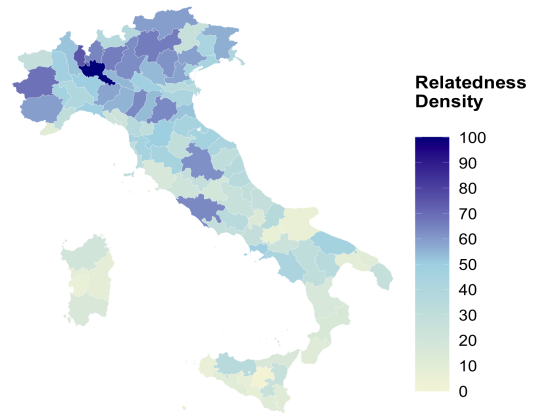
$$K_{i,N} = \frac{1}{k_{i,0}} \sum_s M_{i,s} k_{s,N-1} \quad (2.6)$$

$$K_{s,N} = \frac{1}{k_{s,0}} \sum_i M_{i,s} k_{i,N-1} \quad (2.7)$$

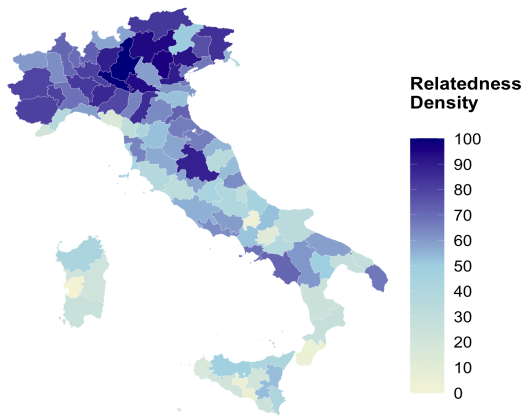
For industries ($K_{i,N}$), even variables ($K_{i,0}, K_{i,2}, K_{i,4}...$) are general measures of skill diversification; odd variables ($K_{i,1}, K_{i,3}, K_{i,5}...$) are measures of ubiquity of their effectively



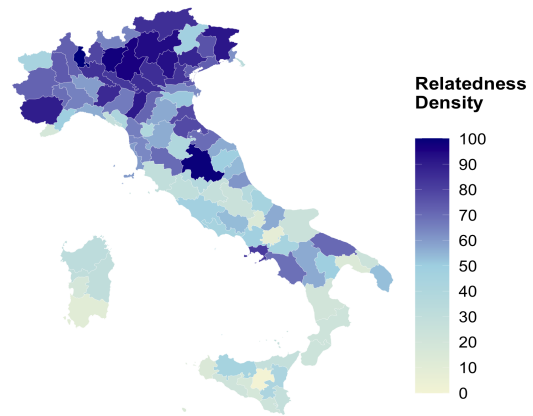
(a) Manufacture of electronic components, 2013



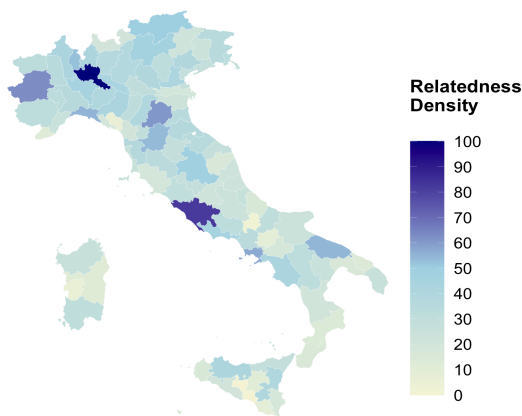
(b) Manufacture of electronic components, 2019



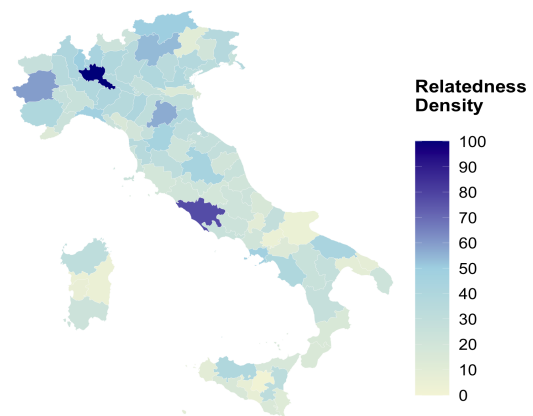
(c) Construction of roads and motorways, 2013



(d) Construction of roads and motorways, 2019

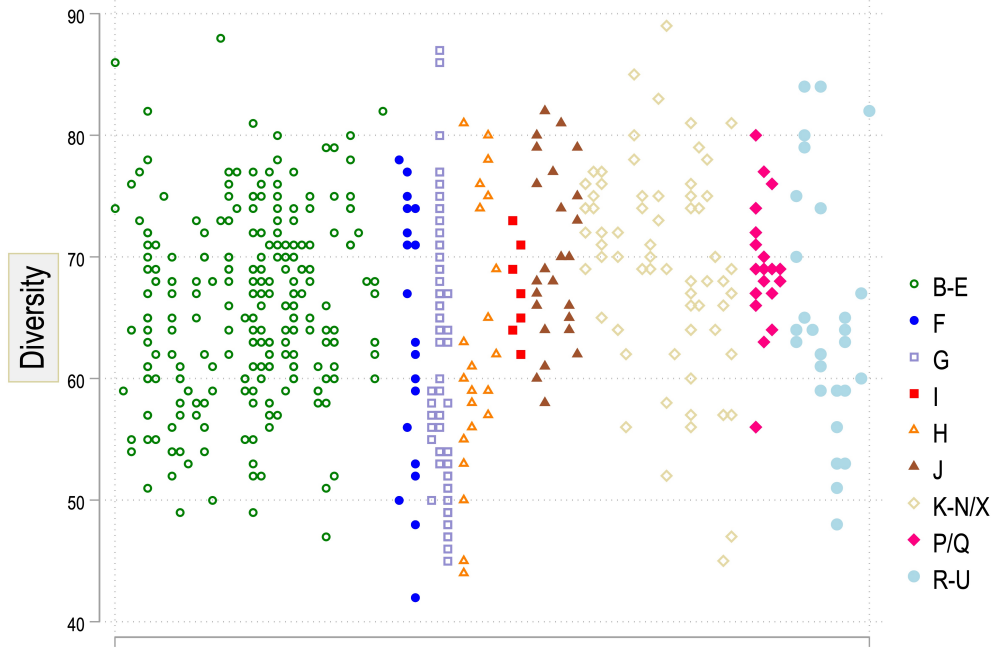


(e) Activities of holding companies, 2013

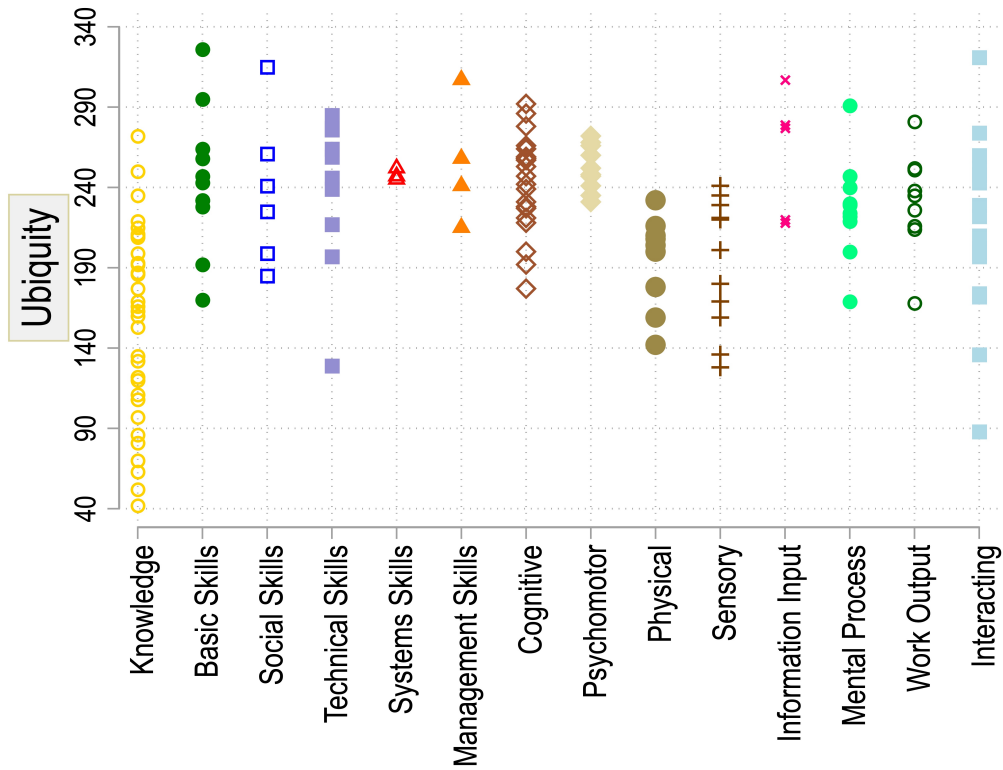


(f) Activities of holding companies, 2019

Figure 5. Differences in Relatedness Densities Across Industries, Regions and Years



(a)



(b)

Figure 6. Skill Diversity of Industries and Ubiquity of Skills (2013-2019). **Figure 6a. Skill Diversity of Industries.** Shows how many skills are effectively used by a particular industry. Each symbol corresponds to an industry at the four-digit NACE level and is coloured by aggregate industry classifications. **Figure 6b. Ubiquity of Skills.** Shows how many times a particular skill is effectively used by industries. Different colours correspond to the sub-categories (ICP descriptors) of skills.

used skills. For skills ($K_{s,N}$), even variables are related to their ubiquity, while odd variables are related to the diversification of industries that effectively use those skills.

Higher order iterations can be interpreted as a linear combination of the properties of all the nodes in the network that we briefly name complexity score. Complexity score converges to a certain value as the number of iterations increases.

Based on the equations above, the skill complexity of industry i is defined by the composition of the skill set necessary to perform tasks required by that industry and the relative composition of skill sets of all other industries. In other words, a higher skill complexity score for industry i would reflect a relatively higher number of effectively used skills (diversity component) that are relatively effectively used by a small number of industries (ubiquity and scarcity component).

We begin by illustrating diversity ($K_{i,0}$) and ubiquity ($K_{s,0}$). Figure 6(a) shows the skill diversity of industries (equation 2.4) averaged for the period 2013-2019. A particular industry¹² effectively uses 40 skills at the minimum and 90 skills at the maximum. Manufacture, construction, and wholesale trade exhibit the widest range of skill diversity values.

Figure 6(b) presents the ubiquity of effectively used skills by industries (equation 2.5) averaged for the period 2013-2019. A particular skill is effectively used by a minimum of 40 industries and a maximum of 340 industries. (C22) programming, (B21) medicine and dentistry, and (B22) therapy and counselling are the most scarce skills. If we look at sub-categories of skills, knowledge is highly non-ubiquitous as expected, while different knowledge types range between 40 and 270. One should note that ubiquity is an absolute measure; it does not assess the relative scarcity of skills. For instance, physical and sensory skills are more ubiquitous than some technical and cognitive skills, but it does not necessarily mean physical skills are scarcer. Figure 7 gives more insights into the relative scarcity of skills.

By following the approach of Hidalgo and Hausmann (2009), we examine the interaction of the first two sets of variables, namely, the ubiquity of skills ($K_{s,0}$) and the average skill diversity of industries ($K_{s,1}$) which are presented in Figure 7. A higher value of $K_{s,1}$ means that the necessary skill sets for the industries that effectively use skill s are highly diversified; a lower value of $K_{s,0}$ means that a relatively low number of industries effectively use skill s . Figure 7 shows a strong negative correlation between these measures meaning that more diversified industries effectively use less ubiquitous skills. The negative relationship between average diversity and ubiquity is in line with the findings of Hidalgo and Hausmann.

The graph is divided into four quadrants by plotting reference lines for the mean of ubiquity (vertical line) and average diversity (horizontal line). In order to interpret these four quadrants, we refer to Figure 8, which is adopted from Hidalgo and Hausmann (2009)¹³. Figure 8 is the schematic representation of Figure 7, displaying the properties of each quadrant. The top-left quadrant shows that relatively scarce skills, such as scientific and academic knowledge, programming, and technology design, are required by industries with high skill diversity which means more complexity in necessary tasks to perform a job. The bottom-left quadrant shows that physical, sensory, and technical skills related to physical activities, including repairing and installation, are effectively used by relatively few industries that are non-complex in terms of skill diversity. The top-right

¹²(B-E) Industry (except construction), (F) Construction, (G) Wholesale and retail trade; repair of motor vehicles and motorcycles, (I) Accommodation and food service activities, (H) Transportation and storage, (J) Information and communication, (K-N/X) Financial, real estate, scientific and technical, administrative and support service activities, (P/Q) Education; human health and social work activities, (R-U) Arts, entertainment and recreation; other service activities.

¹³Figure S7 for the product space.

quadrant accommodates the skills required by many complex industries, such as mental process skills. In the bottom-right quadrant, we find the most used skills such as sensory, psychomotor, work output, and a high fraction of technical skills. Cognitive skills are required by many industries with various skill complexity levels, given that they are located in both the top-right and bottom-right quadrant. Interacting and social skills are distributed all over the graph.

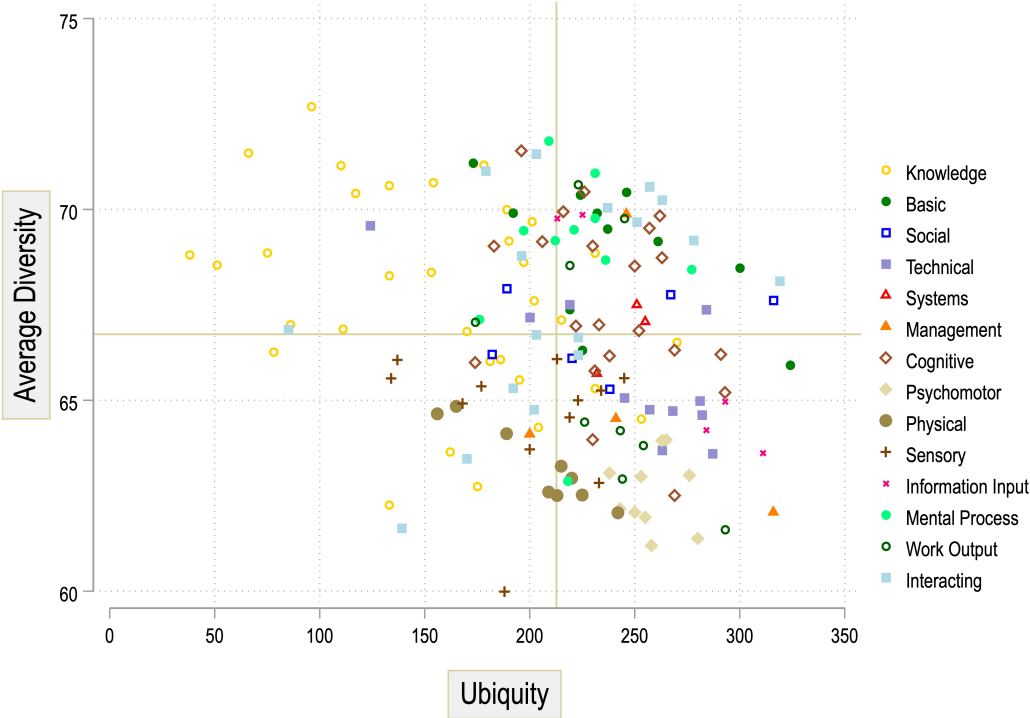


Figure 7. Interaction of Average Skill Diversity and Ubiquity of Effectively Used Skills (2013-2019). Each symbol corresponds to a workplace skill and is coloured by ICP descriptors presented in Table A1. The reference lines indicate the mean scores.

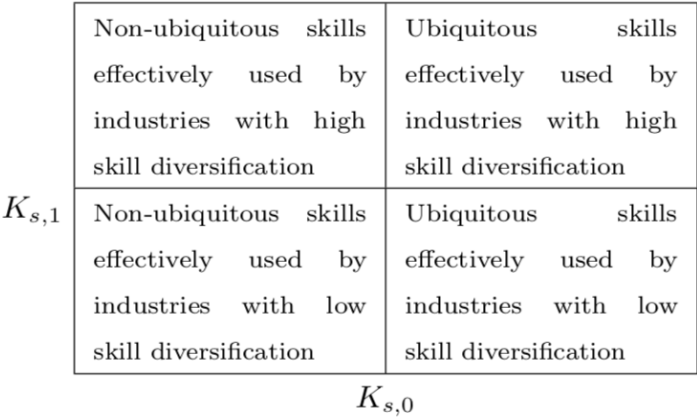


Figure 8. Schematic Presentation of $K_{s,0} - K_{s,1}$ Space

Figures 6 and 7 give interesting preliminary information on the effective use of various skills by different industries. Nevertheless, they are based on only the first two sets of MOR. A higher number of iterations would unravel further insights. Figure 9 dis-

plays skill complexity scores of workplace skills (obtained after 17 iterations of MOR) on the y-axis. The sharp divide between social-cognitive and technical-physical skills is striking, posing further evidence for the skill polarisation of industries. Almost all technical-physical skills have below-average complexity scores. Transportation (B33), science (C6), and programming (C22) are the most complex technical-physical skills. Conversely, social-cognitive skills exhibit much higher complexity scores. Communication and media (B32), persuasion (C13), and negotiation (C14) are the most complex social-cognitive skills.

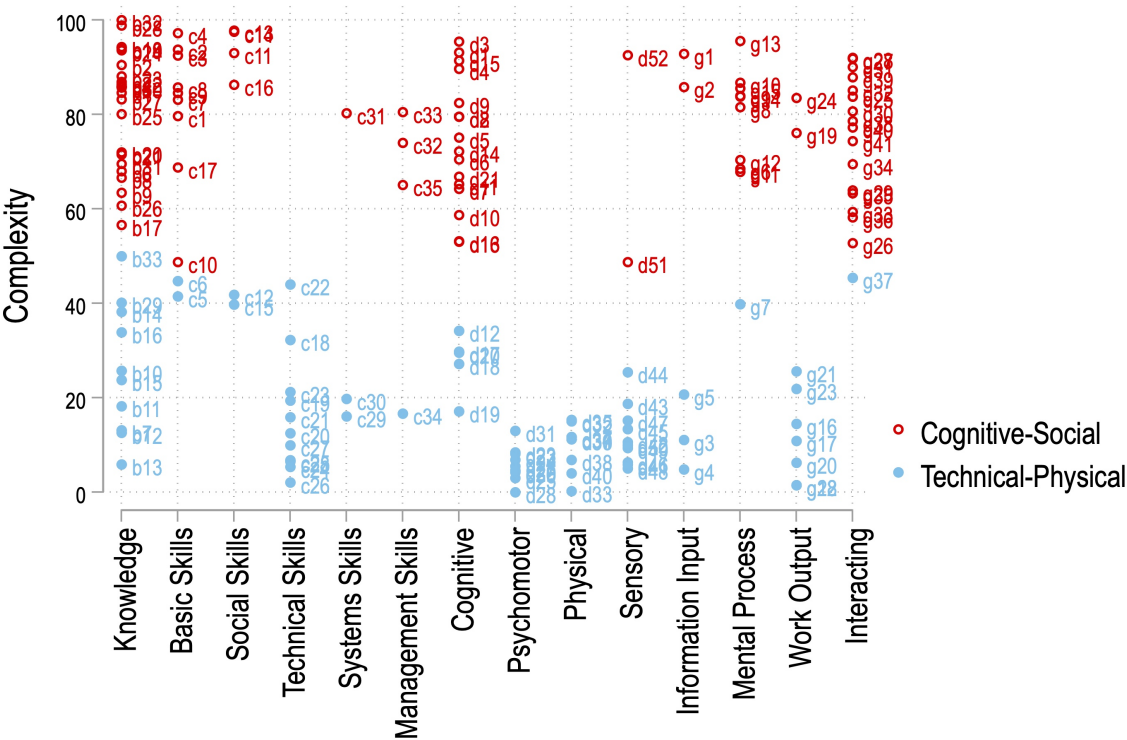


Figure 9. Complexity Levels of Workplace Skills. Each symbol corresponds to a workplace skill that is labelled following Table A1. Each symbol is coloured by the detected skill clusters.

Figure 10 exhibits skill complexity scores of industries at the four-digit NACE level. Construction, a high share of manufacturing, transportation, and storage industries have relatively lower skill complexity. In contrast, education, human health, social work activities, financial and scientific activities, and information and communication industries have the highest skill complexity scores.

Figure 4(b) presents the skill complexity index (based on 17 iterations of MOR) of Italian regions, averaged for the period 2013-2019. Higher skill complexity score for a region indicates diverse and exclusive workplace skills in the population. Apparently, the northern and central regions have higher complexity scores than the southern regions.

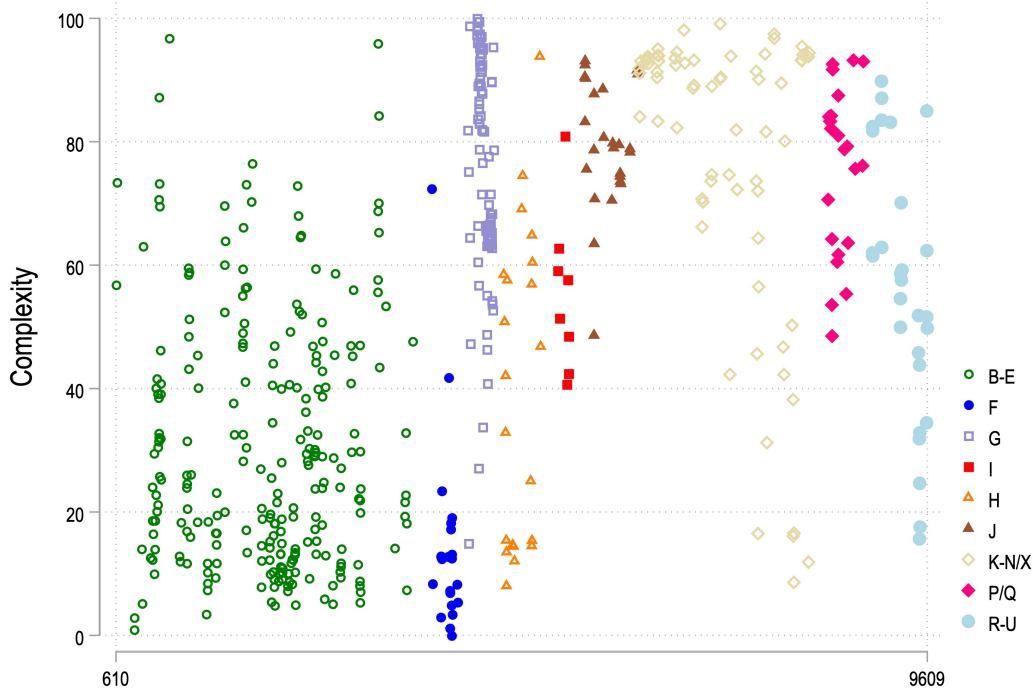


Figure 10. Skill Complexity Scores of Industries. Each symbol corresponds to an industry at the four-digit NACE level and is coloured by aggregate industry classifications displayed in the legend. See Footnote 11 for the detailed industry classifications.

2.4.3 Diversification Potential of Italian Regions

The concepts of relatedness density and complexity can be used together to identify regional specialisation opportunities. Balland et al. (2019) formalise relatedness density and knowledge complexity of technologies into a smart specialisation framework by placing the relatedness density (the complexity) scores of technologies that the region does not have RCA along the y-axis (x-axis). According to this scheme, the higher the relatedness density score, the lower the risk of a new specialisation; the higher the complexity score, the higher the benefits of a new specialisation.

Inspired by Balland et al. (2019), we analyse the potentials of Italian regions by combining skill relatedness density and skill complexity scores as presented in Figure 11. A positive relationship between skill complexity and skill relatedness density is evident. The figure is divided into quadrants by the reference lines of mean skill relatedness density and skill complexity scores. The regions that occupy the top-right quadrant, such as Torino, Roma, and Verona, have high skill complexity and skill relatedness density scores, meaning that their workforces are highly skilled and can fulfil tasks requiring unique and diverse skills. Therefore, new industries may quickly acquire a comparative advantage in these regions and can produce unique and valuable goods and services, given that the workforce they employ effectively uses almost all skills available in any other place in the country. Missing skills can be obtained easily since they are pretty similar to those the workforce of the region already held. The top-left quadrant accommodates regions with above-average skill complexity but below-average skill relatedness density, indicating that those regions specialized in a subsample of skills. If a new industry's skill set is similar to skills the region already held, then high-quality goods and services can be produced in that industry, and comparative advantage can be established easily. However, if the new industry's skill set is different, establishing a comparative

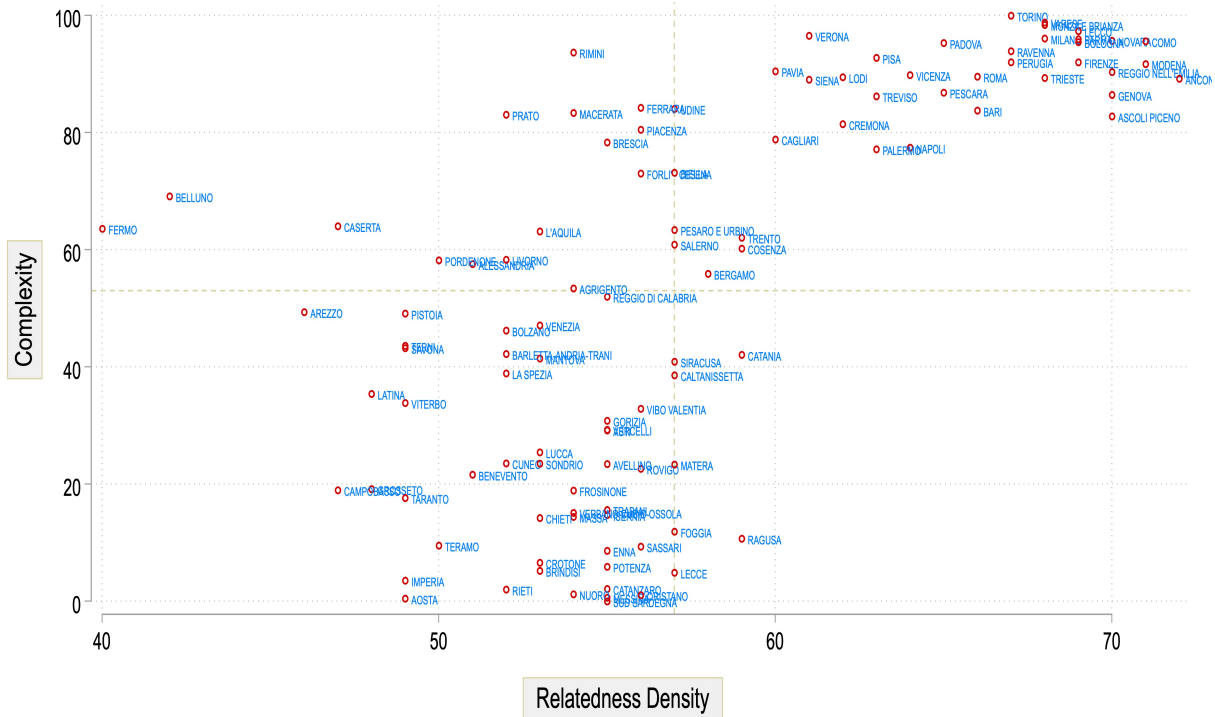


Figure 11. Diversification Potential of Italian Regions

advantage is risky because the region does not have the necessary human capital to specialize in this industry. Regions like Venezia, Livorno, Macerata fall into this category. The regions in the bottom-left quadrant, such as Aosta, Lecce, and Grosseto, have below-average skill complexity and relatedness density, meaning that their workforce held a subsample of skills that are ubiquitous and less diverse. A new industry might establish a comparative advantage in one of these regions if its skill set is very similar to those held in the region; if not, the industry should invest in training and education of the workforce to develop a comparative advantage. However, benefits tend to be low because goods and services produced in these regions can be produced in many other regions, given that regional capabilities, in the form of workplace skills, are ubiquitous and limited. The regions in the bottom-right quadrant, including Catania, Ragusa, and Sassari, have above average skill relatedness density. Therefore, compared to the bottom-left quadrant, a new industry's skill set is more likely to be similar to those already held in these regions, yet, benefits will be low due to low complexity.

2.5 Econometric Analysis: Industrial Diversification

This section conducts econometric analyses to assess the impact of skill relatedness and complexity on the probability that a region specialises in a new industry (entry) or loses it (exit). Entry and exit models are commonly used in the EEG literature. Farinha et al. (2019) study entry and exit of occupation classes in the Metropolitan Statistical Areas of the USA. Bahar et al. (2018) assess the effects of three Marshallian linkages—input-output relations, pooling of workers, and the sharing of technology and knowledge—on the take-off of export industries. Balland et al. (2019) analyse the entry and exit of technological classes into European regions.

Xiao et al. (2018) investigate the effect of industry-relatedness on industrial diversifi-

cation of regions between only *tradable* industries based on the assumption that tradable sectors are more likely to reflect the genuine capabilities of a region. They exclude one-third of the existing industries to get a smaller subsample. However, we include all the available industries in the data set because we do not wish to imply any ad-hoc assumption on the industry mix of regions, or their capabilities, given that we assess the *differences* of regional capabilities through skill-relatedness. As a result, our data consist of 161 workplace skills, 534 industries, and 107 Italian regions for the period 2013-2019, forming seven unbalanced panels¹⁴.

We construct two binary dependent variables, *Entry* and *Exit*, to account for the industrial specialisation of regions. *Entry* equals to one if industry i develops a specialisation in region p , *i.e.*, if industry i has revealed comparative advantage (RCA) in region p in period t while it did not have RCA in period $t - 3$, and *Entry* equals to zero otherwise. *Exit* equals to one if industry i had RCA in region p in period $t - 3$ and loses it in period t . More formally:

$$Entry_{i,p,t} = 1, \text{ if } RCA_{i,p,t} > 1 \text{ and } RCA_{i,p,t-3} \leq 1 \quad (2.8)$$

$$Exit_{i,p,t} = 1, \text{ if } RCA_{i,p,t} \leq 1 \text{ and } RCA_{i,p,t-3} > 1 \quad (2.9)$$

$$RCA_{i,p,t} = \frac{E_{i,p,t}/E_{i',p,t}}{E_{i,p',t}/E_{i',p',t}} \quad (2.10)$$

where $E_{i,p,t}$ is employment in industry i in region p in period t . A higher RCA means that industry i has higher employment in region p in period t , than its average employment in all regions p' in period t . In other words, higher RCA indicates over-presence of an industry in a given region. Tables A3 and A4 provide summary statistics for the dependent variables.

The dependent variables are binary; therefore, logistic, probit, and linear probability (LPM) models can be used. The Hausman test procedure¹⁵ suggests fixed effects with our data set. It is known that logistic and probit models can be inconsistent when estimated with large fixed effects (Greene, 2012). Accordingly, by following the literature (Bahar et al., 2018; Xiao et al., 2018; Farinha et al., 2019; Balland et al., 2019), we use LPM to estimate the following specification at the industry-region level for four overlapping periods of three-years¹⁶.

$$\begin{aligned} Y_{i,p,t} &= [Entry_{i,p,t}, Exit_{i,p,t}] \\ Y_{i,p,t} &= \beta_1 ASRD_{i,p,t-1} + \beta_2 SkillComplexity_{i,p,t-1} + \\ &\quad \beta_3 RegionControls_{i,p,t-1} + \beta_4 IndustryControls_{i,p,t-1} + \\ &\quad \beta_5 GeoDiversification + \rho_p + \iota_i + \gamma_t + \varepsilon_{i,p,t} \end{aligned}$$

The principal coefficients of interest are β_1 and β_2 . β_1 captures the impact of ASRD of industries on the probability of developing a comparative advantage in a region or losing an existing one. We expect a positive sign in the entry models. If an industry has a more similar skill set to the existing skill portfolios of industries located in a particular

¹⁴Some industries have zero employment in some region-year pairs; therefore, the resulting sample is unbalanced.

¹⁵The Hausman test has been performed with different models, including logistic, probit, and LPM. The null "difference in coefficients not systematic" was rejected in each model specification.

¹⁶2013-2016, 2014-2017, 2015-2018, 2016-2019.

region, then that industry would have a relatively higher probability of being specialised in that region. The region already possesses the necessary capabilities that the new industry can combine. Conversely, we expect a negative sign in the exit models, mainly because if an industry's required skill set is not similar to the skill sets of preexisting industries in a particular region, then that industry would have a higher probability of losing specialisation in that region since the region does not have the required capabilities thus the industry must nurture them. β_2 measures the impact of skill complexity on the probability of new industrial specialisations. We anticipate a positive coefficient in the entry models and a negative one in the exit models. If an industry has a complex skill set with respect to other industries in the region, it is more likely to produce more valuable and unique goods and services, thus more (less) likely to develop (lose) comparative advantage.

In the previous section, we showed that workplace skills are divided into two clusters with respect to their effective use by industries. In order to investigate how these skill clusters reflect themselves in the industrial diversification process, ASRD and skill complexity are computed for three different skill aggregation levels: full skills (161 skill types that are presented in Table A1), social-cognitive skills (one of the detected skill clusters, displayed in the first section of Table A2), technical-physical skills (another skill cluster, presented in the second section of Table A2). Figure 12 presents their distributions. The inner-quartile ranges of ASRD variables are pretty narrow compared to skill complexity variables. Regarding the social-cognitive skill cluster, 75% of scores are below 50 and below 45 for the technical-physical skill cluster. The skill complexity variables exhibit a more balanced distribution across quartiles. The average social-cognitive skill complexity of industries is around 62, while it is 46 for technical-physical skill complexity. The majority of industries exhibit higher ASRD and complexity scores for the social-cognitive skill cluster.

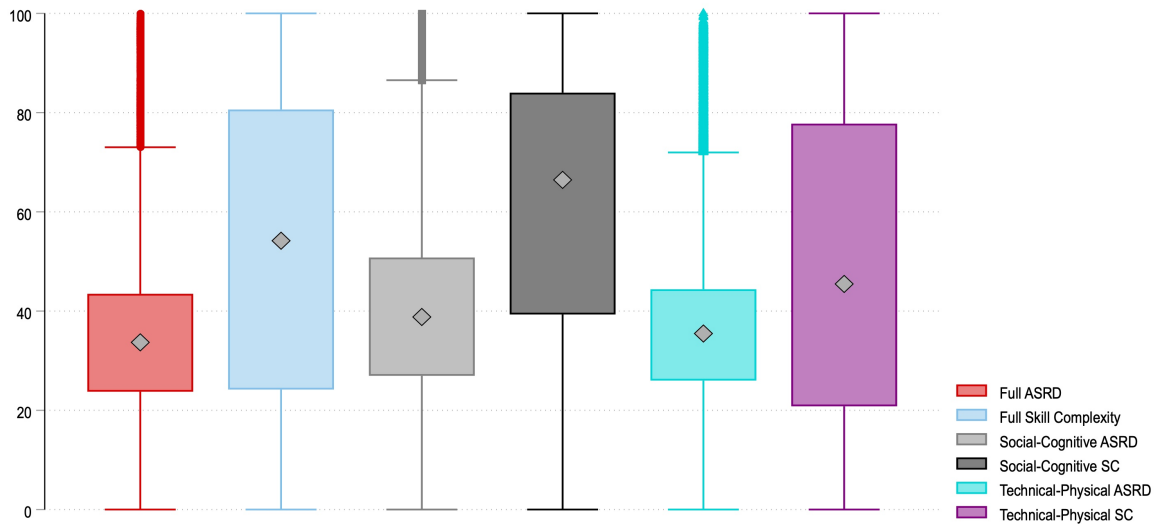


Figure 12. Distributions of Independent Variables

We include three-way fixed effects to all estimates, namely, fixed effects for industries ν_i , regions ρ_p , and time γ_t , assuming that there might be some time-independent features that correlate with the variables. In addition, we include three sets of control variables. The first set stands for regional economic and demographic differences. Population den-

sity (log), the number of inhabitants per square kilometre per region, is included to control for urbanisation. GDP per capita (log) controls for the level of economic development. Education is the share of tertiary education in the population and enters the specification as a standard control of human capital. The second set of control variables is at the industrial-regional level. Business growth (percentage) is the net population growth of industries. Churn (percentage) is the sum of growth and death rates of industries. This set controls for the differences in turnovers of industries. Some industries may be inherently more (less) mobile, which might increase (decrease) their entry or exit probability. Lastly, we add two more variables named geographical diversification controls. Industrial ubiquity (number of regions that a particular industry already specialised in, *i.e.* $RCA > 1$) controls for the rarity of industries since rare industries are expected to form relatively fewer specialisations. Regional diversity (number of specialised industries, *i.e.* $RCA > 1$, in a given region) accounts for the diversity of industry mixes of regions, given that more diversified regions might be more prone to attract new specialisations. Geographical diversification controls allow us to account for the geographical distribution patterns of industries and regions, helping to isolate the effect of skill-related co-occurrence patterns.

Table 15. Entry and Exit Models: Full Skills

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
<i>Full ASRD</i>	0.00207*** (0.00016)		0.00181*** (0.00016)	-0.00292*** (0.00036)		-0.00237*** (0.00036)
<i>Full Complexity</i>		0.00035*** (0.00003)	0.00030*** (0.00003)		-0.00057*** (0.00005)	-0.00051*** (0.00005)
<i>GDP (log)</i>	-0.06612** (0.03261)	-0.04227 (0.03258)	-0.06117* (0.03260)	0.07519 (0.07535)	0.03401 (0.07517)	0.06221 (0.07536)
<i>Pop. Density</i>	-0.07006 (0.06147)	-0.01994 (0.06132)	-0.06715 (0.06148)	0.08037 (0.13001)	0.00111 (0.12967)	0.05532 (0.13012)
<i>Education</i>	-0.00159* (0.00088)	-0.00113 (0.00088)	-0.00166* (0.00088)	-0.00300 (0.00197)	-0.00331* (0.00196)	-0.00285 (0.00196)
<i>Churn</i>	-0.00076* (0.00040)	-0.00095** (0.00041)	-0.00074* (0.00040)	0.00137 (0.00095)	0.00159* (0.00095)	0.00122 (0.00095)
<i>Bus. Growth</i>	0.00105*** (0.00029)	0.00119*** (0.00029)	0.00107*** (0.00029)	-0.00152** (0.00074)	-0.00181** (0.00074)	-0.00150** (0.00074)
<i>Reg. Diversity</i>	-0.00118*** (0.00014)	-0.00109*** (0.00014)	-0.00117*** (0.00014)	0.00293*** (0.00032)	0.00281*** (0.00032)	0.00289*** (0.00032)
<i>Ind. Ubiquity</i>	-0.00561*** (0.00037)	-0.00564*** (0.00037)	-0.00561*** (0.00037)	0.00932*** (0.00067)	0.00927*** (0.00067)	0.00924*** (0.00067)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes
<i>N</i>	157,602	157,602	157,602	65,814	65,814	65,814
<i>Adj.R²</i>	0.051	0.051	0.052	0.077	0.078	0.079

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

An examination of the correlation matrix reported in Table A5 reveals low correlations, below 0.49, between independent variables. The only exception is GDP per capita,

which is relatively highly correlated, around 0.66, with ASRD variables. However, we included it in the models since it is a standard control variable and is almost always insignificant. Variance inflation factors (VIF) are computed after each baseline specification, as displayed in Table A6. All VIFs are in the acceptable range, *i.e.* below 5, including GDP per capita.

Since errors are likely to be correlated within regions and industries, all estimates are presented with robust standard errors clustered at the industry-region level. To mitigate potential endogeneity, all independent variables are lagged by one period, denoted as $t - 1$ where t is the first year of each three-year period.

Table 16. Entry and Exit Models: Social-Cognitive Skill Cluster

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
<i>SC ASRD</i>	0.00234*** (0.00017)		0.00220*** (0.00017)	-0.00288*** (0.00032)		-0.00251*** (0.00032)
<i>SC Complexity</i>		0.00043*** (0.00003)	0.00041*** (0.00003)		-0.00084*** (0.00005)	-0.00081*** (0.00005)
<i>GDP (log)</i>	-0.07196** (0.03262)	-0.04730 (0.03258)	-0.07252** (0.03261)	0.08062 (0.07542)	0.04377 (0.07514)	0.07790 (0.07539)
<i>Pop. Density</i>	-0.08564 (0.06159)	-0.01682 (0.06136)	-0.08283 (0.06163)	0.10038 (0.13012)	0.01007 (0.12928)	0.08438 (0.12985)
<i>Education</i>	-0.00192** (0.00089)	-0.00107 (0.00088)	-0.00198** (0.00088)	-0.00287 (0.00197)	-0.00333* (0.00196)	-0.00270 (0.00196)
<i>Churn</i>	-0.00073* (0.00040)	-0.00095** (0.00041)	-0.00070* (0.00040)	0.00150 (0.00095)	0.00161* (0.00095)	0.00131 (0.00095)
<i>Bus. Growth</i>	0.00104*** (0.00029)	0.00115*** (0.00029)	0.00101*** (0.00029)	-0.00150** (0.00074)	-0.00161** (0.00074)	-0.00125* (0.00074)
<i>Reg. Diversity</i>	-0.00121*** (0.00014)	-0.00110*** (0.00014)	-0.00121*** (0.00014)	0.00294*** (0.00032)	0.00282*** (0.00032)	0.00291*** (0.00032)
<i>Ind. Ubiquity</i>	-0.00560*** (0.00037)	-0.00562*** (0.00037)	-0.00558*** (0.00037)	0.00934*** (0.00067)	0.00916*** (0.00067)	0.00915*** (0.00067)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes
<i>N</i>	157,602	157,602	157,602	65,814	65,814	65,814
<i>Adj. R²</i>	0.051	0.052	0.053	0.077	0.081	0.082

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

We first estimate the baseline specification for the full skill set, meaning that ASRD and skill complexity variables are computed with the full set of 161 workplace skills. Table 15 displays the results for both entry and exit models. In entry model 1, *Full ASRD* has a positive and highly significant effect on the probability that an industry specialises in a new region, as expected. In entry model 2, skill complexity *Full Complexity* also has a positive and significant effect on the probability of new industrial specialisation. In entry model 3, ASRD and skill complexity are present together, and significant effects prevail. The dependent variables are binary; therefore, coefficients must be interpreted as probabilities. Regarding entry, when ASRD increases by ten percentage points, the probability of a new industry specialisation in the region increases by about 1.8%. Skill

complexity shows relatively little relevance, a ten per cent increase is associated with a 0.3% increase in the probability of a new industry entry.

Exit models in Table 15 indicate that ASRD and skill complexity are negatively and significantly affect the probability that an industry loses comparative advantage in a particular region. When ASRD increases by ten percentage points, the probability of an industry losing comparative advantage in the region decreases by about 2.3%. Likewise, when skill complexity increases by ten percentage points, the probability of exit decreases by about 0.5%. In general, the effects of ASRD and skill complexity are more potent for exit models.

Table 17. Entry and Exit Models: Technical-Physical Skill Cluster

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
<i>TP ASRD</i>	0.00231*** (0.00017)		0.00218*** (0.00017)	-0.00274*** (0.00037)		-0.00252*** (0.00037)
<i>TP Complexity</i>		0.00038*** (0.00003)	0.00035*** (0.00003)		-0.00070*** (0.00006)	-0.00067*** (0.00006)
<i>GDP (log)</i>	-0.06515** (0.03260)	-0.04513 (0.03257)	-0.06426** (0.03258)	0.06728 (0.07526)	0.03891 (0.07517)	0.06285 (0.07527)
<i>Pop. Density</i>	-0.05940 (0.06140)	-0.01976 (0.06135)	-0.06128 (0.06143)	0.05413 (0.12971)	0.01454 (0.12953)	0.05025 (0.12968)
<i>Education</i>	-0.00151* (0.00088)	-0.00101 (0.00088)	-0.00154* (0.00088)	-0.00329 (0.00197)	-0.00348* (0.00196)	-0.00319 (0.00196)
<i>Churn</i>	-0.00058 (0.00040)	-0.00097** (0.00041)	-0.00057 (0.00040)	0.00126 (0.00095)	0.00173* (0.00095)	0.00118 (0.00095)
<i>Bus. Growth</i>	0.00102*** (0.00029)	0.00119*** (0.00029)	0.00102*** (0.00029)	-0.00157** (0.00074)	-0.00188** (0.00074)	-0.00156** (0.00074)
<i>Reg. Diversity</i>	-0.00118*** (0.00014)	-0.00110*** (0.00014)	-0.00118*** (0.00014)	0.00294*** (0.00032)	0.00281*** (0.00032)	0.00290*** (0.00032)
<i>Ind. Ubiquity</i>	-0.00562*** (0.00037)	-0.00562*** (0.00037)	-0.00559*** (0.00037)	0.00934*** (0.00067)	0.00924*** (0.00067)	0.00923*** (0.00067)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes
<i>N</i>	157,602	157,602	157,602	65,814	65,814	65,814
<i>Adj. R²</i>	0.051	0.051	0.052	0.077	0.078	0.079

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

In the following step, we turn to the detected skill communities in Section 2.4 and separately compute the variables ASRD and skill complexity for the two polarised skill clusters: social-cognitive skills and technical-physical skills. We then run the baseline specification with new independent variables. Table 16 shows the results for social-cognitive skills. Regarding entry models, a ten per cent increase in social-cognitive ASRD (*SC ASRD*) positively affects the probability of a new industry specialisation by 2.2%. Negative association endures for exit models; when social-cognitive ASRD increases by ten percentage points, the probability of exit decreases by 2.5%. Social-cognitive skill complexity *SC Complexity* exhibits a similar look to full skill complexity (*Full Complexity*), the coefficient almost doubles in exit models.

Table 18. Entry and Exit Models: Manufacturing and Service Industries

	Manufacturing						Service					
	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit	(7) Entry	(8) Entry	(9) Entry	(10) Exit	(11) Exit	(12) Exit
<i>Full ASRD</i>	0.00220*** (0.00019)			-0.00212*** (0.00036)			0.00040* (0.00023)			-0.00117** (0.00050)		
<i>Full Complexity</i>	0.00041*** (0.00004)			-0.00060*** (0.00005)			0.00020*** (0.00005)			-0.00051*** (0.00012)		
<i>SC ASRD</i>		0.00196*** (0.00018)			-0.00213*** (0.00031)			0.00012 (0.00022)			-0.00072 (0.00051)	
<i>SC Complexity</i>		0.00048*** (0.00003)			-0.00077*** (0.00006)			0.00023*** (0.00005)			-0.00059*** (0.00011)	
<i>TP ASRD</i>			0.00211*** (0.00016)			-0.00159*** (0.00034)			0.00071*** (0.00024)			-0.00140*** (0.00046)
<i>TP Complexity</i>			0.00045*** (0.00004)			-0.00063*** (0.00006)			0.00021*** (0.00005)			-0.00051*** (0.00013)
<i>GDP (log)</i>	-0.03444 (0.03725)	-0.03128 (0.03723)	-0.03942 (0.03722)	0.03595 (0.08652)	0.03529 (0.08646)	0.04114 (0.08646)	-0.06872 (0.06407)	-0.06452 (0.06408)	-0.07072 (0.06409)	0.08131 (0.14723)	0.06692 (0.14710)	0.08432 (0.14714)
<i>Pop. Density</i>	0.02474 (0.07304)	0.03260 (0.07307)	0.03286 (0.07312)	0.16489 (0.14710)	0.19492 (0.14690)	0.13185 (0.14723)	-0.18969* (0.11199)	-0.19140* (0.11203)	-0.19104* (0.11201)	-0.28059 (0.26465)	-0.27046 (0.26480)	-0.26816 (0.26447)
<i>Education</i>	-0.00347*** (0.00103)	-0.00338*** (0.00103)	-0.00340*** (0.00103)	-0.00273 (0.00225)	-0.00289 (0.00225)	-0.00283 (0.00225)	0.00180 (0.00169)	0.00188 (0.00169)	0.00182 (0.00169)	-0.00187 (0.00388)	-0.00159 (0.00388)	-0.00223 (0.00388)
<i>Churn</i>	-0.00112 (0.00072)	-0.00115 (0.00072)	-0.00105 (0.00072)	0.00213 (0.00137)	0.00215 (0.00136)	0.00230* (0.00136)	-0.00005 (0.00054)	-0.00005 (0.00054)	0.00000 (0.00054)	0.00282* (0.00146)	0.00287** (0.00146)	0.00272* (0.00147)
<i>Bus. Growth</i>	0.00156*** (0.00052)	0.00140*** (0.00052)	0.00167*** (0.00052)	-0.00147 (0.00104)	-0.00103 (0.00104)	-0.00184* (0.00104)	0.00051 (0.00036)	0.00050 (0.00036)	0.00050 (0.00036)	-0.00183* (0.00108)	-0.00175 (0.00108)	-0.00182* (0.00108)
<i>Reg. Diversity</i>	-0.00107*** (0.00016)	-0.00108*** (0.00016)	-0.00107*** (0.00016)	0.00252*** (0.00036)	0.00250*** (0.00036)	0.00252*** (0.00036)	-0.00119*** (0.00027)	-0.00119*** (0.00027)	-0.00118*** (0.00027)	0.00346*** (0.00062)	0.00346*** (0.00062)	0.00346*** (0.00062)
<i>Ind. Ubiquity</i>	-0.00624*** (0.00049)	-0.00618*** (0.00049)	-0.00622*** (0.00049)	0.00859*** (0.00080)	0.00855*** (0.00080)	0.00866*** (0.00080)	-0.00456*** (0.00055)	-0.00455*** (0.00055)	-0.00456*** (0.00055)	0.01035*** (0.00122)	0.01039*** (0.00122)	0.01036*** (0.00122)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
<i>N</i>	106,662	106,662	106,662	46,562	46,562	46,562	50,940	50,940	50,940	19,252	19,252	19,252
<i>Adj. R²</i>	0.058	0.059	0.058	0.060	0.063	0.060	0.048	0.048	0.048	0.112	0.112	0.112

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The results for technical-physical skills, presented in Table 17, are economically and statistically similar to those of full skills and social-cognitive skills. Overall, similarity to technical-physical skills and complexity in social-cognitive skills pose the most substantial impact on the industrial specialisation process, even though the size effects are pretty similar.

The overall picture changes when considering manufacturing (columns 1-6, Table 18) and service (columns 7-12, Table 18) industries separately. The results show that ASRD and skill complexity affect the specialisation process of manufacturing industries with pretty similar coefficients to the overall results. Regarding service industries, ASRD of social-cognitive skills is insignificant, suggesting that similarity to those skills does not enhance or mitigate the specialisation process. On the other hand, ASRD of technical-physical skills is significant at the 99% level with relatively lower coefficients than the overall results. The skill complexities of both skill types are relevant and significant.

Even though the effect of skill relatedness and complexity could be expected to be somewhat more sizeable, low relatedness and complexity coefficients with respect to entry and exit models are generally seen in the literature. Farinha et al. (2019) report coefficients for complementary and similarity relatedness densities, respectively, 0.003 and 0.004 concerning job entries in the USA, while job complexity coefficients range between 0.0001 and 0.001. Balland et al. (2019) report 0.004 for relatedness density in a model for technological class entries in European regions, while the knowledge complexity coefficient is around 0.00004.

2.6 Robustness Checks

2.6.1 Average Marginal Effects: Linear Probability, Logistic, and Probit Models

LPM has been extensively used in the applied microeconometrics and economic geography literature, partly due to its usefulness since the estimated coefficients can be interpreted on the probability scale. In a logistic model, the estimated coefficients are changes in the log-odds scale that is challenging to interpret. In a probit model, the estimated parameters are interpreted as shifts in the cumulative normal distribution, which is also intuitively challenging. Nevertheless, LPM is criticized for having a couple of pitfalls, including inherent heteroscedasticity, inappropriate significance tests due to non-normality of errors, and estimates out of the probability range (0 – 1) (Greene, 2012). The heteroscedasticity problem can be resolved by using robust options of statistical software. Hellevik (2009) shows that the significance probabilities from linear and logistic regressions are identical; therefore, the argument of inappropriate significance tests does not necessarily hold. On the other hand, estimations below zero and above one are possible. Therefore, we estimate the baseline specifications with logistic and probit models to mitigate potential concerns for LPM. Table A7 displays the results. The first six columns are entry models where columns 1-3 are estimated with logistic regression and columns 4-6 are estimated with probit regression. The second half of Table A7 displays exit models. Columns 7-9 are estimated with logistic, while columns 10-12 are estimated with probit models. The signs of the coefficients and the qualitative results of the estimates are in line with the LPM estimates reported in Tables 15, 16, and 17.

When it comes to size effects, as already mentioned, the estimated parameters of linear probability, logistic, and probit models should not be directly compared since they are at different scales. Hence, we calculate average marginal effects (AME) to be able to compare the estimated effects of these models. Table 19 exhibits AME computed after

the baseline specifications reported in the previous tables. Columns 1 and 4 for LPM correspond to Table 15 (column 3 for entry, column 6 for exit) for full range of skills (*Full ASRD*, *Full Complexity*), Table 16 for social-cognitive skill cluster (*SC ASRD*, *SC Complexity*), and Table 17 for technical-physical skill cluster (*TP ASRD*, *TP Complexity*). Columns 2 and 3 of Table 19 correspond to the first six columns of Table A7, while columns 5 and 6 correspond to the last six columns of Table A7. AME estimated with logistic and probit models suggest that a ten per cent increase in *Full ASRD* is associated with, respectively, a 1.82% and 1.89% increase in the entry probability, which is almost the same size effect estimated with LPM. AME of other variables estimated with different models are also pretty similar, suggesting that the estimated probabilities in previous specifications are not severely affected by model selection.

Table 19. Average Marginal Effects for LPM, Logistic, and Probit Estimates

	Entry			Exit		
	(1) LPM	(2) Logistic	(3) Probit	(4) LPM	(5) Logistic	(6) Probit
<i>Full ASRD</i>	0.00180 (0.00016)	0.00182 (0.00016)	0.00189 (0.00016)	-0.00237 (0.00036)	-0.00246 (0.00035)	-0.00247 (0.00034)
<i>SC ASRD</i>	0.00219 (0.00016)	0.00187 (0.00014)	0.00195 (0.00014)	-0.00251 (0.00032)	-0.00272 (0.00032)	-0.00272 (0.00031)
<i>TP ASRD</i>	0.00218 (0.00017)	0.00216 (0.00017)	0.00225 (0.00016)	-0.00252 (0.00036)	-0.00252 (0.00035)	-0.00255 (0.00035)
<i>Full Complexity</i>	0.00030 (0.00003)	0.00026 (0.00002)	0.00028 (0.00002)	-0.00051 (0.00005)	-0.00054 (0.00005)	-0.00054 (0.00005)
<i>SC Complexity</i>	0.00041 (0.00003)	0.00038 (0.00002)	0.00039 (0.00002)	-0.00081 (0.00005)	-0.00083 (0.00005)	-0.00084 (0.00005)
<i>TP Complexity</i>	0.00035 (0.00003)	0.00032 (0.00003)	0.00034 (0.00003)	-0.00066 (0.00006)	-0.00064 (0.00006)	-0.00065 (0.00006)

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

2.6.2 Different Specifications for Dependent Variables

The dependent variables, entry and exit, account for changes in comparative advantages of industries that are based on the number of employed people, as denoted in equations 2.8, 2.9, and 2.10. Nevertheless, industrial specialisations might also be defined by other types of comparative advantages. We thus aim to assess if the main findings are robust to a different specification of comparative advantage.

Furthermore, Boschma (2017) underline that the effect of relatedness can be overestimated when using the same data set for both dependent and independent variables. Even though we use different data sets (see data section) for dependent and independent variables, both are based on the same survey, ILFS. Motivated by this reasoning, we use data from another survey¹⁷ to construct alternative dependent variables with a different comparative advantage criterion. We change the computation of dependent variables by changing the RCA formula as follows.

¹⁷Statistical register of active enterprises (ASIA - Enterprises)

$$RCA_{i,p,t}^1 = \frac{ENT_{i,p,t}/ENT_{i',p,t}}{ENT_{i,p',t}/ENT_{i',p',t}} \quad (2.11)$$

where $ENT_{i,p,t}$ is the number of enterprises that are active in industry i , located in region p in period t . $RCA > 1$ indicates that the number of active enterprises in industry i , region p , and period t is greater than the average national level. Afterwards, we define entry and exit variables as in equations 2.8 and 2.9 to run the baseline econometric model with new dependent variables. Table 20 displays the results. Technical-physical ASRD has the highest impact, for both entry and exit models, compared to social-cognitive and full skills, even though the differences are negligible. When there is a ten per cent increase in the ASRD of technical-physical skills, the probability of entry (exit) increases (decreases) by 2.7% (3.2%). As in the previous models, social-cognitive skill complexity yields the highest impact. A ten per cent increase in social-cognitive skill complexity is associated with a 0.46% (0.81%) increase (decrease) in the probability of entry (exit).

Overall, the results indicate that the main findings are robust to a different comparative advantage criterion. Moreover, the relatedness effect is not overestimated in the baseline specification; indeed, it is underestimated given the increased coefficients in Table 20 compared to Tables 15, 16, and 17.

As another robustness check, we analyse if the main results are sensitive to a change in the time span of RCA. Recall that we construct variables *Entry* and *Exit* based on the changes in RCA (equation 2.10) from time $t - 3$ to time t , stands for three years. Here, we decrease the time span to one year, from time $t - 1$ to time t , and re-define the dependent variables with the two criteria of comparative advantage, RCA and RCA^1 , as follows.

$$Entry_{i,p,t} = 1, \text{ if } RCA_{i,p,t} > 1 \text{ and } RCA_{i,p,t-1} \leq 1 \quad (2.12)$$

$$Exit_{i,p,t} = 1, \text{ if } RCA_{i,p,t} \leq 1 \text{ and } RCA_{i,p,t-1} > 1 \quad (2.13)$$

$$Entry_{i,p,t} = 1, \text{ if } RCA_{i,p,t}^1 > 1 \text{ and } RCA_{i,p,t-1}^1 \leq 1 \quad (2.14)$$

$$Exit_{i,p,t} = 1, \text{ if } RCA_{i,p,t}^1 \leq 1 \text{ and } RCA_{i,p,t-1}^1 > 1 \quad (2.15)$$

Table 21 shows the results. The first panel presents results for the dependent variables constructed with RCA as denoted in equations 2.12 and 2.13 (*i.e.* comparative advantage in terms of the number of employed people). The second panel exhibits results for dependent variables constructed with RCA^1 as denoted in equations 2.14 and 2.15 (*i.e.* comparative advantage in terms of the number of enterprises). The results confirm the main findings in previous models; ASRD and skill complexity variables are significantly and positively related to entry while significantly and negatively related to exit. The coefficients of interest exhibit some degree of decrease, regardless of the specification. For the dependent variables constructed with RCA , a ten per cent increase in the ASRD for, respectively, full, social-cognitive and technical-physical skills increases the probability of developing a new specialisation by 1%, 1.2%, 1.3%; whilst decreases the probability of losing comparative advantage by 1.3%, 1.4%, 1.4%. For the dependent variables constructed with RCA^1 , the coefficients are slightly higher. A ten per cent increase in the ASRD for, respectively, full, social-cognitive and technical-physical skills increases the probability of entry by 1.3%, 1.5%, 1.7%; whilst decreases the probability of losing comparative advantage by 1.6%, 1.5%, 1.9%.

Table 20. Entry and Exit Models: RCA in Number of Enterprises

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
<i>Full ASRD</i>	0.00216*** (0.00020)			-0.00278*** (0.00031)		
<i>Full Complexity</i>	0.00039*** (0.00003)			-0.00058*** (0.00005)		
<i>SC ASRD</i>		0.00246*** (0.00020)			-0.00281*** (0.00029)	
<i>SC Complexity</i>		0.00046*** (0.00003)			-0.00081*** (0.00005)	
<i>TP ASRD</i>			0.00268*** (0.00020)			-0.00322*** (0.00032)
<i>TP Complexity</i>			0.00043*** (0.00004)			-0.00072*** (0.00006)
<i>GDP (log)</i>	-0.00851 (0.03949)	-0.01869 (0.03948)	-0.01271 (0.03948)	-0.00043 (0.06750)	0.01619 (0.06748)	0.00262 (0.06749)
<i>Pop. Density</i>	-0.06265 (0.07264)	-0.07509 (0.07266)	-0.05832 (0.07260)	0.09983 (0.12240)	0.12441 (0.12256)	0.09435 (0.12225)
<i>Education</i>	-0.00081 (0.00104)	-0.00103 (0.00104)	-0.00063 (0.00104)	0.00124 (0.00188)	0.00154 (0.00188)	0.00096 (0.00188)
<i>Churn</i>	-0.00100** (0.00047)	-0.00100** (0.00047)	-0.00083* (0.00047)	0.00281*** (0.00087)	0.00289*** (0.00086)	0.00267*** (0.00087)
<i>Bus. Growth</i>	0.00126*** (0.00033)	0.00119*** (0.00033)	0.00120*** (0.00033)	-0.00288*** (0.00068)	-0.00273*** (0.00068)	-0.00289*** (0.00068)
<i>Reg. Diversity</i>	-0.00132*** (0.00016)	-0.00138*** (0.00016)	-0.00131*** (0.00016)	0.00213*** (0.00027)	0.00220*** (0.00027)	0.00213*** (0.00027)
<i>Ind. Ubiquity</i>	-0.00618*** (0.00038)	-0.00615*** (0.00038)	-0.00618*** (0.00038)	0.00883*** (0.00063)	0.00875*** (0.00063)	0.00881*** (0.00063)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes
<i>N</i>	144,087	144,087	144,087	79,329	79,329	79,329
<i>Adj. R²</i>	0.050	0.051	0.050	0.073	0.075	0.074

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 21. Entry and Exit Models: Decrease of Time Span

	Dependent variable specification: number of employees				Dependent variable specification: number of enterprises							
	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit	(7) Entry	(8) Entry	(9) Entry	(10) Exit	(11) Exit	(12) Exit
<i>Full ASRD</i>	0.00104*** (0.00009)			-0.00130*** (0.00020)			0.00130*** (0.00011)			-0.00164*** (0.00019)		
<i>Full Complexity</i>	0.00021*** (0.00002)			-0.00039*** (0.00003)			0.00026*** (0.00002)			-0.00040*** (0.00003)		
<i>SC ASRD</i>		0.00117*** (0.00009)			-0.00137*** (0.00019)			0.00148*** (0.00011)			-0.00154*** (0.00018)	
<i>SC Complexity</i>		0.00024*** (0.00002)			-0.00052*** (0.00003)			0.00034*** (0.00002)			-0.00051*** (0.00003)	
<i>TP ASRD</i>			0.00130*** (0.00009)			-0.00144*** (0.00021)			0.00166*** (0.00012)			-0.00191*** (0.00019)
<i>TP Complexity</i>			0.00024*** (0.00002)			-0.00046*** (0.00004)			0.00033*** (0.00002)			-0.00047*** (0.00004)
<i>GDP (log)</i>	-0.02936 (0.02217)	-0.02996 (0.02217)	-0.03006 (0.02217)	0.00406 (0.05472)	0.00718 (0.05473)	0.00674 (0.05473)	-0.00838 (0.02760)	-0.00789 (0.02762)	-0.00893 (0.02760)	-0.00282 (0.04968)	0.00008 (0.04973)	0.00016 (0.04971)
<i>Pop. Density</i>	0.00432 (0.00402)	0.00520 (0.00402)	0.00475 (0.00401)	-0.00672 (0.00987)	-0.00750 (0.00986)	-0.00603 (0.00987)	0.01025** (0.00476)	0.01148** (0.00476)	0.01095** (0.00476)	-0.01167 (0.01011)	-0.01190 (0.01012)	-0.01127 (0.01012)
<i>Education</i>	-0.00168*** (0.00048)	-0.00179*** (0.00048)	-0.00164*** (0.00048)	-0.00232** (0.00112)	-0.00215* (0.00112)	-0.00238** (0.00112)	0.00091 (0.00059)	0.00077 (0.00059)	0.00093 (0.00059)	0.00093 (0.00108)	0.00107 (0.00108)	0.00087 (0.00108)
<i>Churn</i>	-0.00048*** (0.00022)	-0.00048** (0.00022)	-0.00038* (0.00022)	0.00151*** (0.00057)	0.00163*** (0.00057)	0.00153*** (0.00057)	-0.00076*** (0.00027)	-0.00077*** (0.00027)	-0.00067** (0.00027)	0.00173*** (0.00054)	0.00185*** (0.00054)	0.00166*** (0.00054)
<i>Bus. Growth</i>	0.00028 (0.00018)	0.00027 (0.00018)	0.00024 (0.00018)	-0.00051 (0.00049)	-0.00049 (0.00049)	-0.00054 (0.00049)	0.00014 (0.00022)	0.00013 (0.00022)	0.00010 (0.00022)	-0.00067 (0.00046)	-0.00071 (0.00046)	-0.00068 (0.00046)
<i>Reg. Diversity</i>	-0.00064*** (0.00009)	-0.00065*** (0.00009)	-0.00064*** (0.00009)	0.00199*** (0.00021)	0.00199*** (0.00021)	0.00198*** (0.00021)	-0.00080*** (0.00010)	-0.00084*** (0.00010)	-0.00080*** (0.00010)	0.00172*** (0.00017)	0.00175*** (0.00017)	0.00173*** (0.00017)
<i>Ind. Ubiquity</i>	-0.00364*** (0.00024)	-0.00362*** (0.00024)	-0.00363*** (0.00024)	0.00638*** (0.00042)	0.00636*** (0.00042)	0.00639*** (0.00043)	-0.00361*** (0.00025)	-0.00359*** (0.00025)	-0.00360*** (0.00025)	0.00720*** (0.00041)	0.00717*** (0.00041)	0.00720*** (0.00041)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
<i>N</i>	234,494	234,494	234,494	98,020	98,020	98,020	214,232	214,232	214,232	118,282	118,282	118,282
<i>Adj. R²</i>	0.030	0.031	0.031	0.047	0.048	0.046	0.032	0.033	0.032	0.045	0.046	0.045

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

2.6.3 Alternative Model Specification: Industry–Time and Region–Time Fixed Effects

Here, we present an alternative specification of the fixed effects that are already added to the previous estimates. We intend to eliminate the effects of time-varying heterogeneity across regions or industries. The aim is to test whether the effects of ASRD and skill complexity variables are driven by some unobservables that affect both dependent variables. We thus include industry-time and region-time fixed effects instead of their separate inclusion. The control variables are not included in the models, given that most of them become invariant in the present specification. According to the results in Table 22, ASRD and skill complexity indicators for all skill types exhibit positive (negative) and significant effects with increased coefficients on acquiring (losing) new industrial specialisations.

Table 22. Entry and Exit Models: Industry-Time, Region-Time Fixed Effects

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
<i>Full ASRD</i>	0.00264*** (0.00022)			-0.00237*** (0.00036)		
<i>Full Complexity</i>	0.00029*** (0.00003)			-0.00052*** (0.00005)		
<i>SC ASRD</i>		0.00385*** (0.00025)			-0.00410*** (0.00047)	
<i>SC Complexity</i>		0.00041*** (0.00003)			-0.00083*** (0.00005)	
<i>TP ASRD</i>			0.00300*** (0.00021)			-0.00344*** (0.00047)
<i>TP Complexity</i>			0.00035*** (0.00003)			-0.00069*** (0.00006)
Industry-Time FE	yes	yes	yes	yes	yes	yes
Region-Time FE	yes	yes	yes	yes	yes	yes
<i>N</i>	157,602	157,602	157,602	65,814	65,812	65,812
<i>Adj. R²</i>	0.049	0.051	0.050	0.075	0.077	0.073

Notes: Robust standard errors clustered at the region and industry level are in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

2.7 Concluding Remarks

Region-specific capabilities are increasingly accepted as a critical source of regional diversification, yet they are seldom included in the measurement process. In this chapter, we include regional capabilities into the measurement process of relatedness and complexity by using industry-skill input matrices per region and year, enabling fine-grained analyses of skill relatedness and complexity in the industrial diversification process of regions.

Building on the network-based approach of EEG, we first construct a skill-to-skill relatedness matrix and skill complexity vector, *i.e.*, skill space, of each region in each year. We then combine these measures at the national level and formalise them as a one-mode

network, *i.e.*, the skill space of Italy, that consists of 161 workplace skills connected with edges that represent their degree of relatedness. Each node is proportional to its complexity score. After applying a force-based algorithm, the skill space forms two highly polarised skill clusters. The first one consists of basic, social, cognitive, management, and interacting skills, therefore named social-cognitive skills; whilst another one consists of physical, psychomotor, sensory, systems, and technical skills, thus named technical-physical skills.

We then employ linear probability models with three-way fixed effects to quantify the effect of skill relatedness and skill complexity on the probability of (1) forming a new regional comparative advantage in terms of the number of employed people in a given industry (entry), or (2) losing an established comparative advantage (exit). The independent variables are constructed for three different aggregation levels: (1) full skills, (2) social-cognitive skills, and (3) technical-physical skills. The results show a highly significant and positive relationship between higher skill relatedness density and the probability of entry, regardless of the skill type. The more similar the skill set of the new industry to the skill sets of industries that are already established a comparative advantage in that region, the higher the probability of building a new comparative advantage for the new industry in that region. Even though the correlation coefficients are pretty close among the skill types, similarity to technical-physical skills yields a slightly higher probability of entry. When it comes to exit, all estimates show a highly significant and negative association between skill relatedness density and the probability of losing an established comparative advantage, regardless of the skill type. The skill complexity of an industry positively (negatively) affects its entry (exit) probability. We also show that skill relatedness density and skill complexity affect the entry and exit probability of manufacturing industries, regardless of the skill type. On the other hand, only technical-physical skill relatedness density enhances (decreases) the entry (exit) probability of service industries, while skill complexity affects both entry and exit probabilities. By controlling for industrial ubiquity and regional diversity, we show that our results are not just a reflection of the geographical distribution patterns of industries.

The findings mentioned above are robust to a couple of sensitivity analyses. First, we re-estimate the baseline specifications with logistic and probit models to assess the resulting average marginal effects. Second, we re-identified the dependent variables, entry and exit, with a different data set and comparative advantage criterion. Instead of RCA in the number of employed people, we used RCA in the number of active enterprises. The main findings endured with very similar coefficients. Third, we decreased the time span of the comparative advantage to be one year and estimated the baseline specification with both RCA criteria. The results were identical, with a slight decrease in the coefficients. Lastly, we employed region-time and industry-time fixed effects. The results did not change.

Although the present work is far from revealing all aspects of skill relatedness and complexity, the findings clearly underline the role of these measures in the industrial diversification of regions yet also call for further investigation. First, the analysis covers only one country due to data availability. Apparently, further research is needed to understand how skill relatedness and complexity affect the industrial diversification process of regions in a cross-country setting. Second, this study has analysed skill relatedness and complexity in isolation. However, the literature has demonstrated that other relatedness types, such as technological relatedness, are highly relevant in the diversification process. Therefore, further exploration of different relatedness and complexity types would unravel valuable insights regarding underlying mechanisms that cause regions to

diversify into different industries.

Regional skill relatedness and complexity analyses might yield rewarding insights for policymakers. They provide a recent panorama of regional human capital obtained with highly granular data and sophisticated methods, unlike previous measures such as completed years of education and work experience. Accordingly, these measures are highly beneficial for human capital formation because skill shortages can be analysed at the industry-region level to define necessary measures that local and national authorities can take. By doing so, regional innovation and smart specialisation policies can be more strategically designed to stimulate new industrial specialisations.

Chapter 3

Ramsey Pricing of Pharmaceuticals? A Debiased Machine Learning Analysis¹

3.1 Introduction

Cross-country pharmaceutical pricing has long been examined in the literature (Maleug and Schwartz, 1994; Danzon, 1997; Jack and Lanjouw, 2005; Towse et al., 2015; Towse et al., 2018; Danzon, 2019). Some argue that high pharmaceutical prices are necessary because manufacturers need funds to finance pharmaceutical research and development (R&D) to produce new breakthrough products (Danzon, 1997). Others draw attention to ethical aspects and assert that drugs cannot be treated as normal market goods since they are essential for public health; hence they should be accessible to low-income populations (Outterson, 2005). In theory, differential pricing can balance these two tensions: high prices charged in high-income countries support R&D activities, while low prices in low-income countries increase access to drugs (Sherer and Watal, 2002). Consequently, differential pricing has become a widely accepted policy with a long history in pharmaceutical markets.

Three distinct theoretical approaches provide analytical support for differential pricing in pharmaceutical markets: price discriminating monopoly, Ramsey pricing, and value-based differential pricing (Danzon, 2018). Among them, Ramsey pricing has been widely discussed in the literature partly due to its theoretical attractiveness. Ramsey pricing is based on the seminal contribution of Ramsey (1927), who developed a theory to solve the problem of raising a specific level of tax revenue with an optimal set of tax rates that maximises social welfare. Ramsey's solution was to charge prices that vary inversely with the price elasticities of demand, known as the inverse elasticity rule. In the '70s, Ramsey's theory had a renaissance after the influential work of Baumol and Bradford (1970) that uses the inverse elasticity rule to find an optimal set of prices for a monopoly firm to cover joint/sunk costs that cannot be causally allocated to particular groups of costumers. Thereafter, Ramsey pricing was applied mainly to utility firms in industries where joint costs are high and marginal costs are lower, such as telecommunications, airlines, railroads, and health care.

Danzon (1997) argues that pharmaceutical markets form conditions for Ramsey pricing because R&D is an essential part of drug production, and it cannot be allocated to a particular group of consumers or countries, given that such costs arise before product launch and thus cannot be captured by marginal cost. She demonstrates that Ramsey pricing is welfare superior to marginal-cost pricing in the presence of inherently high

¹Joint work with Federico Nuttonarelli and Massimo Riccaboni.

R&D costs. Further theoretical studies support the welfare superiority of Ramsey pricing to marginal pricing and uniform pricing (Danzon, 1998; Danzon and Towse, 2003; Barros and Martinez-Giralt, 2008), while empirical evidence is scarce. Empirical literature so far indirectly assessed Ramsey prices' existence by looking at the relationship between pharmaceutical prices and income. The rationale is that one should observe a high positive correlation between pharmaceutical prices and per capita income (GDP) in the presence of Ramsey pricing, assuming that GDP is a good proxy of the price elasticity of demand (Yadav, 2010). However, empirical evidence on the relationship between pharmaceutical prices and income does not necessarily predict high positive correlations (Maskus, 2001; Sherer and Watal, 2002; Lichtenberg, 2011); while some report no correlations at all (Waning et al., 2009b; Morel et al., 2011; Petrou and Vondoros, 2016).

To the best of our knowledge, no study provides empirical evidence on Ramsey pricing in pharmaceutical markets by testing the validity of the inverse elasticity rule. In other words, there is not any empirical evidence on whether pharmaceutical prices vary inversely with the price elasticity of demand. This chapter provides the first empirical study in this vein.

It is essential to underline that the price elasticity of demand for pharmaceuticals has been estimated by a stream of studies at the patient level, using within-country, individual level data of natural experiments stemming from changes in co-payment shares of insurance (health) plans (Manning et al., 1987; Goldman et al., 2004; Landsman et al., 2005; Chandra et al., 2014). Such studies analyse price responsiveness (own-price elasticities) of different groups of drugs—or the number of prescriptions—with respect to changes in insurance co-payments. Accordingly, those estimates include only the insured population within a country, generally in the USA, and none of them addresses Ramsey pricing.

This chapter aims to fill these gaps and assesses the validity of the inverse elasticity rule by providing a cross-country analysis of pharmaceutical demand at the molecule level. We use the IQVIA database that compiles data on 70 national pharmaceutical markets and provides high granular information concerning molecules, therapy classes, launch date, sales quantity, sales price, brands of products, manufacturers, et cetera. Our sample consists of 34 countries—including 10² middle and low income (MLIC) and 24³ high income (HIC) countries⁴—and 33⁵ molecules for the period 2008-2020.

We estimate the price elasticity of demand using a recently developed method of double/debiased orthogonal machine learning (DML) with heterogeneous treatment effects (Chernozhukov et al., 2018; Semenova et al., 2021). DML combines sample splitting and Neyman-orthogonality to deliver an asymptotically normal, debiased estimator of treatment effect parameter in a low-dimensional and cross-sectional setting, assuming no unobserved unit effects. Semenova et al. (2021) extend this framework by proposing an estimator that performs better in a panel setting with the presence of a large number of heterogeneous treatment effects with even a larger number of potential covariates and unobserved unit heterogeneity. The possibility of using very large numbers of covariates decreases the probability of standard omitted variable bias, which was an essential drawback of conventional demand estimates. The method incorporates previous realisations

²Algeria, Bulgaria, China, Colombia, Egypt, India, Romania, Russia, Serbia, Turkey.

³Australia, Austria, Belgium, Canada, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, UK, USA.

⁴According to World Bank country classification by income.

⁵Acetylcysteine, Acetylsalicylic Acid, Aciclovir, Allopurinol, Amlodipine, Betamethasone, Bisoprolol, Budesonide, Carbamazepine, Ciprofloxacin, Clindamycin, Dexamethasone, Diclofenac, Finasteride, Fluconazole, Furosemide, Ibuprofen, Ketoconazole, Lactulose, Levothyroxine Sodium, Loratadine, Mesalazine, Methylprednisolone, Metoprolol, Metronidazole, Omeprazole, Paracetamol, Ranitidine, Risperidone, Salbutamol, Simvastatin, Testosterone, Valproic Acid.

of the demand system, such as price lags and sales lags, accounting for unobserved unit heterogeneity and serial correlation. Moreover, we control for cross-price effects. These aspects of the demand model allow us to approximate to true own-price elasticities of pharmaceuticals.

The overall results show that all countries have average price elasticities less than zero, ranging from the highest (0.88) for India to the lowest (-0.03) for the USA, while molecules have a higher variation from -1.23 to 0.02 . In order to test the validity of the inverse elasticity rule in pharmaceutical markets, we first analyse the relationship between pharmaceutical prices and the price elasticities of pharmaceutical demand at the country level. We find that prices vary inversely with price elasticities at the country level, both in HIC and MLIC, indicating the existence of Ramsey pricing in cross-country pharmaceutical markets.

The results exhibit some degree of variation when we look at within-country pharmaceutical markets at the molecule level. In Austria, France, and Portugal, pharmaceutical prices at the molecule level do not inversely vary with price elasticities. Algeria, Belgium, Germany, and Hungary show a neutral relationship between price elasticities and prices. Australia, Bulgaria, Canada, Croatia, Denmark, Egypt, and Slovakia have a somewhat strong inverse relation between price elasticities and price. At the same time, the rest of the countries tend to have an inverse relation, yet not very strong. All in all, these results demonstrate moderate evidence on the existence of within-country Ramsey pricing in many national pharmaceutical markets.

As a second investigation of the inverse elasticity rule in pharmaceutical markets, we analyse the relationship between GDP per capita income based on purchasing power parity (PPP) and the price elasticities of pharmaceutical demand at the country level. We find evidence of the inverse relation between price elasticities and GDP per capita income in HIC and MLIC, supporting Ramsey pricing.

Our work contributes to pharmaceutical literature in two ways. First, this chapter is the first empirical work that analyses Ramsey pricing in pharmaceutical markets by testing the validity of the inverse elasticity rule. Second, we provide the first cross-country estimates of the price elasticity of pharmaceuticals at the molecule level. Previous studies estimated price elasticity at the patient level using data only on the insured population within a country.

The structure of the chapter is as follows. Section 3.2 overviews the related literature. Section 3.3 describes the data sources. Section 3.4 introduces the methodology. Section 3.5 presents the results. Section 3.6 provides sensitivity analysis. Section 3.7 overviews the main findings and concludes.

3.2 Literature Review

3.2.1 Second Best: Ramsey Pricing

In the presence of a perfect competition market structure, the first best pricing strategy from a welfare-maximising perspective is to set prices equal to marginal cost. However, perfect competition conditions may not hold in some cases, and market failures arise. For instance, in the case of a natural monopoly firm, a marginal-cost pricing strategy may not be enough to be break-even, *i.e.* zero profit, and may result in negative profit. If the firm is not subsidised, prices must be set adequately above marginal cost to continue to be solvent and break even. Deviation from marginal-cost pricing introduces various sets of output and price combinations that may result in zero profit. Then, the question

is that what is the best combination of price-output that would be desirable from a social welfare perspective⁶ if marginal-cost pricing is unattainable?

The mathematician Frank Plumpton Ramsey (1903–1930) was the first to conceptualise this question with analogous reasoning in the context of optimal taxation in his 1927 published article "*Contribution to the Theory of Taxation*". He developed a theory to analyse how to most efficiently collect a given level of tax revenue from different goods. Ramsey's theory had a renaissance after the seminal work of Baumol and Bradford (1970). They built on Ramsey's theory and transferred it to a utility pricing model, which they called *Ramsey pricing* or *the second-best pricing*. In the model, a multi-product monopoly firm divides costumers into groups based on their demand elasticities (market segmentation), assuming that costumers cannot re-sell among different groups. Then, the firm charges higher prices to the groups with inelastic demand (lower demand elasticity); lower prices to the groups with elastic demand (higher demand elasticity). More formally, the critical idea behind Ramsey pricing is the inverse elasticity rule which "requires that each price be set so that its percentage deviation from marginal cost is inversely proportionate to the item's price elasticity of demand" (Baumol and Bradford, 1970, pp. 267). Accordingly, Ramsey pricing is a version of price discrimination with a difference. The classic Ramsey pricing formulation assumes normal profit: the firm recovers fixed and marginal costs to break even but cannot earn excess profits. By contrast, the general aim of other types of price discrimination is to maximise profits (Shepherd, 1992).

Ramsey pricing yields static efficiency by maximising consumer surplus and total output consumed when marginal costs are lower than average costs, implying a fixed cost. Hence, it has been applied mainly to utility firms in industries where joint costs are high and marginal costs are lower, including telecommunications (Ford et al., 1993), airlines (Martin-Cejas, 1997), railroads (Sanchez-Borras et al., 2010), electrics (Bigerna and Bollino, 2016), health (Melnick et al., 1992) and other deregulated industries. Nevertheless, it is essential to underline that the efficiency of Ramsey pricing heavily depends on the presence of conditions that are mentioned above, such as natural monopoly, segmented markets, and inability to earn excess profits.

3.2.2 Differential Pricing of Pharmaceuticals and Ramsey Pricing

Specific features of pharmaceutical markets, *e.g.* monopoly rents stem from patented innovation, imperfect information (Kalo et al., 2013), sunk costs of research and development (R&D) (Danzon, 1997), violate several of the criteria for perfect competition; hence prices can be well above marginal cost. In this context, cross-national pharmaceutical pricing has long been examined in the literature (Maleug and Schwartz, 1994; Danzon, 1997; Jack and Lanjouw, 2005; Towse et al., 2015; Towse et al., 2018; Danzon, 2019). Scholars have mainly focused on the welfare effects of different pricing strategies, including marginal cost pricing, differential pricing⁷, and uniform prices. Danzon (1997) argues that marginal cost pricing may not be optimal for pharmaceuticals because R&D is an essential part of the production process and accounts for 30% of total costs. As there is no doubt development of new drugs serves all customers around the world, R&D costs should be seen as a global joint cost (Danzon and Towse, 2003). A critical aspect is that pharmaceutical R&D costs are fixed. In other words, R&D costs are invariant to the number of countries or people that purchase the drug, therefore cannot be easily allocated as

⁶Social welfare is the sum of consumer and producer surplus.

⁷Differential pricing (price discrimination, tiered pricing) is charging different prices for the same drug across different markets. It may occur across low-income and high-income countries as well as in a single country between different payers (Danzon, 2018).

marginal costs (Danzon, 1997), making marginal cost pricing fail as the first-best pricing strategy. Motivated by this reasoning, Danzon (1997) proposes Ramsey pricing as the most efficient option. Under the constraint of a specific profit level that covers all costs, including R&D, Ramsey pricing principles yield the most efficient set of differential prices⁸ (Danzon, 1997; Danzon, 2000; Danzon and Towse, 2003).

Differential pricing based on Ramsey principles suggests that pharmaceutical prices should vary inversely with demand elasticities (Danzon, 1997; Danzon and Towse, 2003), assuming that absolute price elasticity of demand is negatively correlated with net income. Put differently, the mark-up above marginal cost increases as the price elasticity of demand decreases and vice versa. This is to say, high-income countries pay higher prices sufficient enough to cover pharmaceutical R&D, while low-income countries pay lower prices to increase public access to medicines. Furthermore, high-income countries have a more inelastic demand for price shifts than low-income countries. Both customers and firms are better off with Ramsey prices than they would be with uniform prices (Danzon, 1998). If a uniform price is charged to everyone irrespective of demand elasticity, price-sensitive customers will reduce their consumption more than price-insensitive customers. Moreover, highly price-sensitive customers might drop out of the market, reducing total output consumed, thus reducing social welfare. Uniform prices might also hamper pharmaceutical R&D in the long run. If differential prices based on Ramsey principles are charged, firms can exploit the willingness to pay of different income groups. Since differential prices vary based on demand elasticities, firms can raise more funds for R&D by keeping customers in the market whose mark-up is below uniform price but above marginal cost.

The literature theoretically approves welfare superiority of Ramsey prices, while empirical evidence is scarce. Danzon (1997) conducts a theoretical analysis with multiple firms and products to compare the welfare effects of Ramsey pricing and marginal cost pricing in the USA and the EU. She demonstrates that price differentials in the USA may be explained by the price elasticity of demand that is consistent with Ramsey pricing. Mainly because the pharmaceutical market in the USA is relatively competitive thanks to different health plans that can reflect patients' preferences. However, the market structure is different in the EU, where individual governments are single buyers that exploit monopsony power. Thus, prices may deviate from optimal Ramsey prices. Danzon and Towse (2003) underline that marginal cost pricing may cover short-term production costs but cannot generate sufficient funds to finance pharmaceutical R&D; therefore, patent protection guarantees above-marginal-cost pricing, ensuring sustained incentives for R&D activity. They argue that patents do not necessarily mean higher prices if manufacturers apply differential prices based on Ramsey principles. Barros and Martinez-Giralt (2008) incorporate insurance costs to Ramsey pricing to examine how different insurance coverage rates alter the optimal Ramsey prices. They theoretically demonstrate that higher insurance coverage leads to higher Ramsey prices.

Empirical literature so far indirectly assessed Ramsey prices' existence by looking at the relationship between pharmaceutical prices and income. The rationale is that one should observe a high positive correlation between pharmaceutical prices and per capita income (GDP) in the presence of Ramsey prices, assuming that GDP is a good index of price elasticity of demand (Yadav, 2010). However, empirical evidence hitherto does not necessarily report strong correlations. Sherer and Watal (2002) examine whether pharmaceutical companies applied Ramsey principles to prices of HIV/AIDS drugs in the

⁸The literature provides other rules for differential pricing of pharmaceuticals such as value-based differential pricing (Danzon et al., 2015), indication-based differential pricing (Towse et al., 2018).

90s. They employ multiple regressions to see if there is a strong association between prices and gross national product (GNP) per capita. The results indicate a weak relation: an extra 1000\$ in income per capita increases relative prices by 0.018. Similarly, Maskus (2001) uses international data on 20 major molecules for the period 1994-1998 to analyse correlations between molecule prices and per capita income. According to the results, 17 out of 20 correlations are positive, and six are above 0.5, suggesting a positive but not very strong relationship between drug prices and income. Lichtenberg (2011) performs regressions to analyse price variations across countries by using data on 6500 drugs in 38 countries in the third quarter of 2018. He finds a positive correlation between prices and per capita income. However, he emphasises wide variability within low-income countries so that lower-income citizens pay higher prices than higher-income citizens in some countries.

Furthermore, several studies found that pharmaceutical prices do not necessarily inversely vary with per capita income, implying that some low-income countries may pay higher prices (Waning et al., 2009a). Using data from fourteen middle-income and three high-income countries from 1999 to 2008, Morel et al. (2011) calculate Laspeyres and Paasche indexes to examine if pharmaceutical prices systematically vary with income. They find no such variation; instead, they report high variability of prices across countries. Petrou and Vondoros (2016) calculate the Laspeyres index in an analysis of pharmaceutical price differences across eleven European countries. They report no evidence of an association between prices and income levels.

Despite the theoretical foundations, there is little consensus on whether Ramsey prices exist in pharmaceutical markets. The literature shows that differential pricing has been limited to some categories of drugs such as HIV/AIDS drugs, malaria drugs, and vaccines due to political pressures. In contrast, essential drugs generally remain out of differential pricing (Yadav, 2010). A primary opposition is that observed price differentials within and/or between countries do not approximate Ramsey pricing levels (Danzon and Towse, 2003) because of two main obstacles. First, the prices of pharmaceuticals are highly influenced by third-party public buyers who have monopsony power against a monopolist seller in price negotiations (Kalo et al., 2013). In such a case, public buyers often resorted to international reference pricing to control and compare pharmaceutical prices. Reference pricing is viewed as one of the critical practices that severely impairs differential pricing (Schmidt et al., 2001; Danzon, 2018). Second, differences in prices across countries give an incentive to cross-border pharmaceutical arbitrage, or parallel trade (Schmidt et al., 2001), which may reduce the financial benefits of certain countries and companies who finance pharmaceutical R&D and innovation, therefore can undermine price-makers' incentives to apply differential pricing strategies (Danzon and Towse, 2003; Jack and Lanjouw, 2005; Reisinger et al., 2019).

Moreover, Ramsey principles require a substantial qualification that may not be achieved in real markets due to several barriers. First, pharmaceutical markets are heavily regulated, especially in the EU, where individual governments can enjoy monopsony power. In such an environment, demand elasticities may not be accurate (Maskus, 2001). Second, Ramsey pricing assumes well-segmented markets where price strategies can be implemented independently (Schmidt et al., 2001). For instance, the price for a product in Italy should not be affected by the price for the same product in Germany. It is doubtful that this condition is met in the EU because of parallel trade and reference pricing as mentioned above (Towse et al., 2015). Third, the classical Ramsey problem applied to utilities when a sunk cost needs to be covered by a regulator that earns normal profit (Danzon et al., 2015). However, pharmaceutical companies can earn more than normal profit thanks

to patented products that provide monopolistic rents to manufacturers (Yadav, 2010).

Based on these premises, the debate on the existence of Ramsey prices in pharmaceutical markets remains inconclusive.

3.2.3 Empirical Evidence of Price Elasticity of Demand

To the best of our knowledge, no empirical study has tested the inverse elasticity rule by analysing the relation between prices and the price elasticities of pharmaceuticals. Previous studies used income as a proxy for price elasticity; thus, they indirectly addressed Ramsey pricing, as mentioned above. Therefore, the empirical evidence on demand elasticities reviewed in this section does not relate to Ramsey pricing but is provided to be a benchmark for our analysis.

Estimating demand elasticity for pharmaceuticals is challenging given the complexity of the market and the variety of factors that affect customers' preferences. Accordingly, estimates exhibit some degree of conflict due to differences in calculation methods, demand definitions, measurement of prices, data sources, and study periods (Li and Chollet, 2006).

A stream of literature has estimated the price elasticity of demand for pharmaceuticals at the patient level, using within-country individual-level data of natural experiments stemming from changes in co-payment shares of insurance (health) plans. Such studies analyse price responsiveness (own-price elasticities) of different groups of drugs—or the number of prescriptions—with respect to changes in insurance co-payments. Accordingly, those estimates include only the insured population. Thus, these studies perhaps should be named as *coinsurance elasticity* as in Manning et al. (1987).

Rand Health Insurance Experiment that ran in the '70s in the USA is the most widely known study in this field due to its randomised sample. The price elasticity of demand is estimated as -0.2 with respect to overall healthcare, including prescription drugs (Manning et al., 1987). Other studies estimated the price elasticity of demand on total drug costs ranging from -0.33 to -0.12 (Gilman and Kautter, 2008; Chandra et al., 2014; Brot-Goldberg et al., 2017). Some studies measured demand elasticities by medication class. For instance, Goldman et al. (2004) estimate elasticities with respect to drug days and report results as follows: NSAIDs (-0.45), antihistamines (-0.44), anti-hyperlipidemics (-0.34), anti-asthmatics (-0.32), anti-hypertensives (-0.26), antidepressants (-0.26) and antidiabetics (-0.25). Landsman et al. (2005) measure elasticities concerning the number of prescriptions and find lower elasticities for asymptomatic conditions' drugs such as ACE inhibitors, statins (from -0.10 to -0.16), and higher elasticities for symptomatic conditions' drugs such as Cox-2 inhibitors, NSAIDs, SSRIs (from -0.24 to -0.60).

However, Yeung et al. (2018) criticise this stream of studies because in a natural experiment with changes in co-payments, one must take cross price elasticities into account as well, given that pharmaceuticals are likely to have complements and substitutes whose prices are also affected by the same co-payment change. Consequently, own-price elasticities cannot be accurately estimated since *ceteris paribus* is violated. Yeung et al. (2018) propose a value-based formulary—that incorporates own drug prices as well as prices of its complements and substitutes—to overcome such an endogeneity problem when using natural experiment data. They report an overall price elasticity for pharmaceuticals to be -0.16. Further studies apply the instrumental variable method to generate exogenous variation and estimate price elasticities ranging from -0.14 to -0.4 (Ellis et al., 2017; Einav et al., 2018).

The only country-level empirical study on demand elasticities, to the best of our knowl-

edge, is that of Danzon et al. (2011). They estimate cross-national income elasticity for pharmaceuticals to treat HIV/AIDS, Malaria, TB, both for originator and generic drugs. They find a positive but small overall income elasticity (0.2) across the full range of countries—0.4 for originators and 0.6 for generics. For low and middle-income countries, the average income elasticity is 0.15 and insignificant or negative for originators.

3.3 Data

We use pharmaceutical data from IQVIA⁹, previously IMS Health and Quintiles. The database has a molecule-wise panel structure and comprises 48 quarters for each molecule in each country in a comparable way. Molecules are provided with an Anatomical Therapeutic code (ATC) at the fourth-level of aggregation. In the ATC classification system, drugs are divided into various groups with respect to the organ or system on which they act and their chemical, therapeutic and pharmacological properties. Products (molecules) in the same ATC have the same mechanism of action by definition. Other variables the database provides are quarterly sales, the launch year of the molecule, quarterly standard units expressed in thousands, the number of products referred to each molecule, the name of the producer firm, and the number of firms trading each molecule.

Our sample consists of 34 countries—including ten middle and low income (MLIC) and 24 high income (HIC) countries—and 33 molecules for the period 2008-2020 (48 quarter¹⁰), summing up to 53,856 observations at the molecule-country-quarter level. We retain a subset of countries and molecules available in the database due to methodological constraints. The DML algorithm is better off with balanced panel data, and some molecules did not have information in some countries. Therefore we formed a sample with equal distribution of molecules to form a balanced panel, resulting in fewer molecules and countries. The choice is not to impute zero values mainly due to the necessity of maintaining the data as close as possible to the original database.

The dependent variable is the total demand for molecule i , in country c , in quarter j as measured by log standard units. The treatment variable is the weighted log price of molecule i , in country c , in quarter j . The weights provide a measure of importance for drugs in terms of quantities sold. The covariates are the number of firms that produce a product inclusive of molecule i in country c in quarter j , percentage of generic products among all the products inclusive of molecule i in country c in quarter j , yearly log GDP per capita based on purchasing power parity (PPP), Gini index, fixed effects for quarter, country and molecule. In addition, previous realisations of the demand system in the form of 48 lagged variables of quarterly molecule prices and sales are included in covariates.

The number of firms controls for the level of competition in the market. If there are relatively more firms producing drugs inclusive of molecule i , some degree of price competition is expected, which may lead to lower prices and vice versa. The percentage of generic products also controls for the level of competition, but also for the degree of novelty of molecule i . Originator drugs are patented after their invention. Hence firms cannot produce the same drug during patent life, leading to a monopolistic market structure in which originator drugs can be sold at pretty high prices. After the originator drug becomes off-patent, other firms can produce the same drug (generators), and consequently, the price tends to decrease. The percentage of generator drugs accounts for this process. GDP per capita controls for the level of economic development. Gini index is included in

⁹<https://www.iqvia.com>

¹⁰The observations start at the third quarter of 2008 and end at the third quarter of 2020.

the model to account for income inequalities of countries. The literature on pharmaceutical pricing underlines that prices might be lower in low-income countries. Moreover, some studies argue that pharmaceutical firms tend to apply different prices for different income groups to benefit the income inequalities in low and middle-income countries. We thus control for GDP per capita and Gini index to partly address these pricing strategies. GDP PPP and Gini index data are obtained from the World Development Indicators database of World Bank.

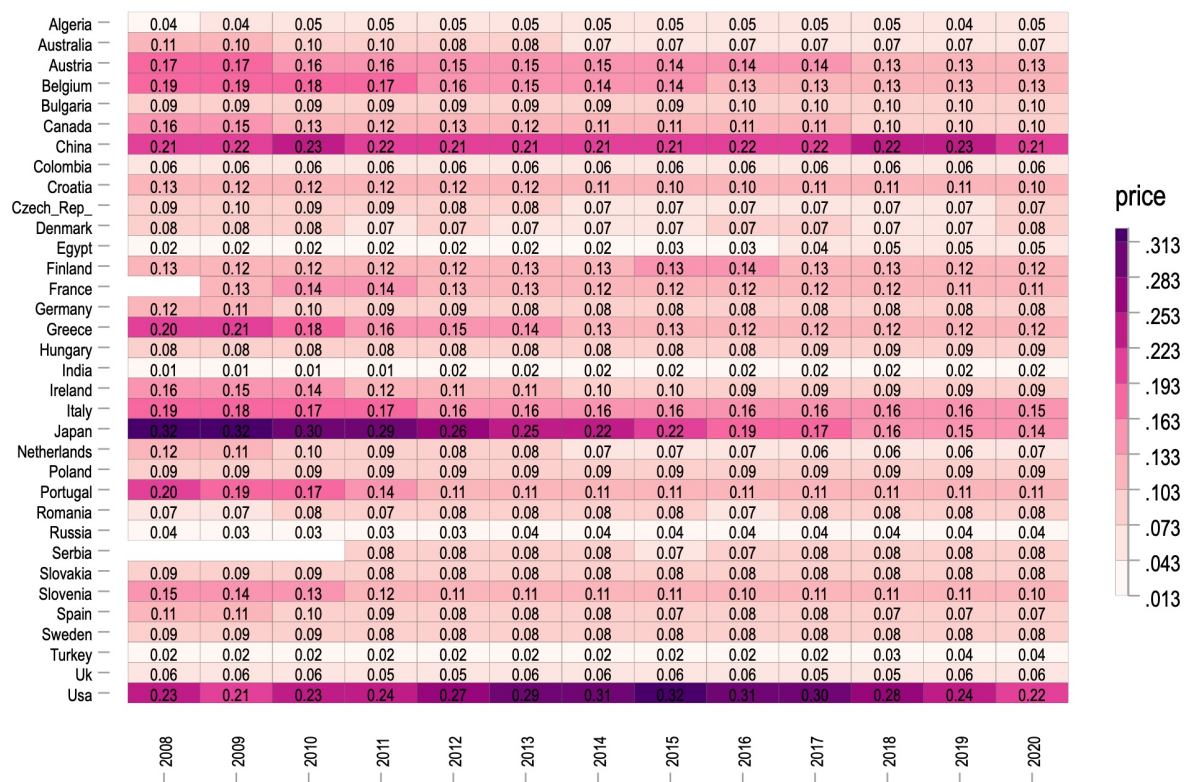


Figure 13. Average Yearly Price of Standard Units by Country

Figure 13 presents the average pharmaceutical prices per standard unit across years and markets. The price data is measured in US dollars from 2008 to 2020, averaged for 33 molecules and weighted by standard units. USA, China, Japan, and Italy have the highest average pharmaceutical prices, while India, Turkey and Russia have the lowest average prices. The figure clearly shows that average prices in some HIC are lower than the prices in some MLIC. For instance, UK and Netherlands have lower average prices than Serbia and Romania.

Figure 14 displays the price shifts of the molecules in the next 5, 10, and 15 years from the launch year to give a hint about price trends in the countries in the data set. The launch year is denoted as 0, and the normalised initial price is denoted as 100. The launch year is considered as the year in which the first product within molecule m has been sold in the national market of country c . The normalised initial price represents the average price of the products belonging to molecule m in country c in year t .

Figure 14 insightfully shows that the price of molecules tends to change linearly, *i.e.* either increase or decrease smoothly, in the majority of countries. This revealed pattern signals the need to control for the price history when computing the price elasticities of demand. Indeed, if no control were present, one might argue that the quantities sold of

a molecule in a country are prone to increase (decrease) since prices tend naturally to decrease (increase). Thus, the estimates of the price elasticity of demand turn out to be biased in absolute terms due to the price trends of molecules in the national markets. We avoid these concerns by considering the price lags in the demand model. A detailed explanation of where and how the price lags enter the analysis is provided in the next section.

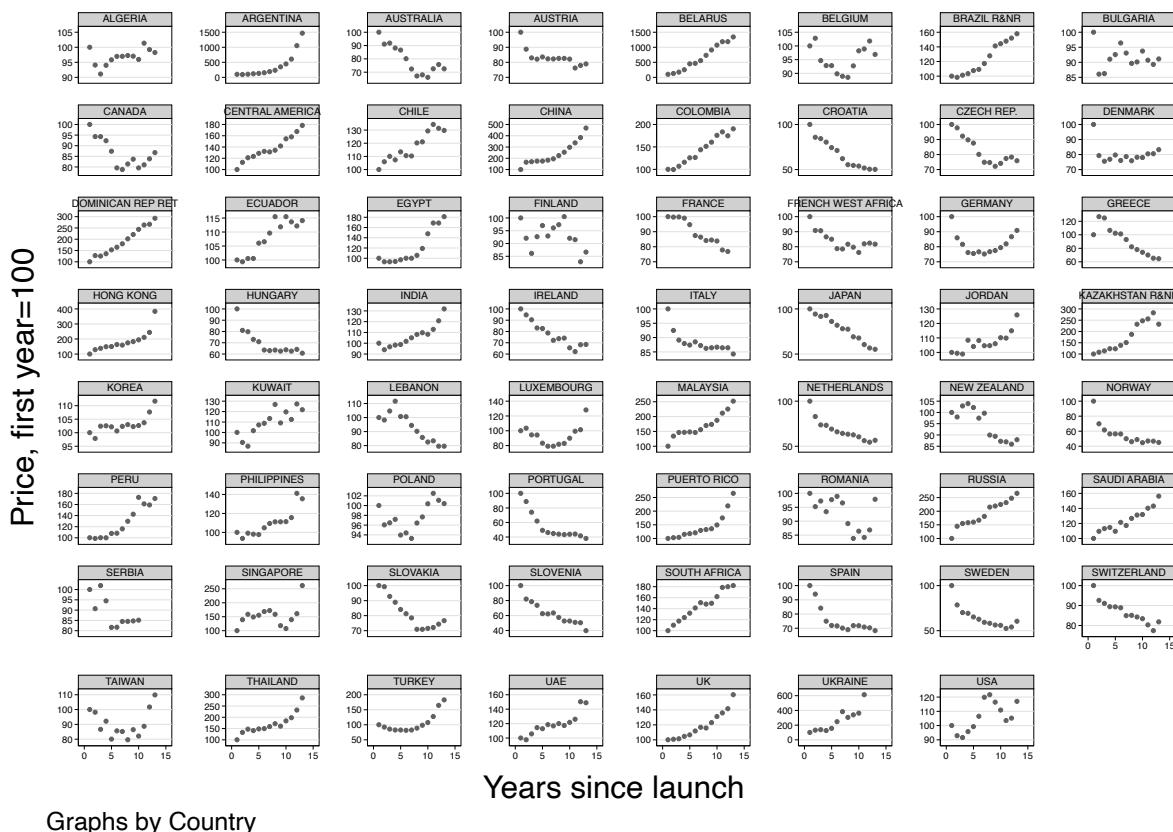


Figure 14. Normalised Price Trends of Molecules. Each subfigure displays the price trends of molecules from their launch year in a given country.

3.4 Methodology

The methodology of the present work —debiased/double machine learning with heterogeneous treatment effects in high dimensional panels— is based on Chernozhukov et al. (2018) and Semenova et al. (2021). Section 3.4.1 summarises the method. Section 3.4.2 postulates the demand model we estimate. Section 3.4.3 briefly comments on the potential endogeneity and how DML overcomes it.

3.4.1 Double Machine Learning with Heterogeneous Treatments

The availability of large data sets and sophisticated tools has raised interest in modelling consumer demand over the past decade (Bajari et al., 2015). Previously, estimating demand with large number of observations and covariates was challenging. Thanks to the recent advancements in machine learning methods, the ability to use a high-dimensional set of covariates decreases the probability of standard omitted variable bias, which was

an essential drawback of conventional demand estimates. The double (debiased) machine learning (DML) framework of Chernozhukov et al. (2018) automates model selection and allows large number of controls and interactions. The DML method combines sample splitting and Neyman-orthogonality to deliver an asymptotically normal, debiased estimator of treatment effect parameter in a low-dimensional and cross-sectional setting, assuming no unobserved unit effects (classical DML). Semenova et al. (2021) extend this framework to a panel setting by allowing (1) high-dimensional and sparse treatment effect parameters and (2) unobserved unit effects (extended DML). We briefly summarise the method in this section by following Semenova et al. (2021).

In particular, below is reported a scenario in which the price elasticity of demand is estimated by exploiting a partially linear panel model. The linear part is heterogeneous treatment effect that consists of baseline treatment effect and treatment modification effect. The nonlinear part is the effect of confounding factors on outcome.

Consider a demand model of N heterogeneous products ($C : \{1, 2, \dots, N\}$) that cluster into G independent clusters.

$$Y_{it} = D'_{it}\theta_0 + g_o(X_{it}) + \xi_i^E + U_{it}, \quad E[U_{it}|D_{it}, X_{it}, \phi_{it}] = 0 \quad (3.1)$$

$$D_{it} = [P_{it}(1, K_{it}), P_{-it}(1, K_{it})] \quad (3.2)$$

$$P_{it} = m_o(X_{it}) + \xi_i^P + V_{it}, \quad E[V_{it}|X_{it}, \phi_{it}] = 0 \quad (3.3)$$

$$P_{-it} = \frac{\sum_{j \in C(i), j \neq i} P_{jt}}{|C(i) - 1|}. \quad (3.4)$$

where Y_{it} is log sales and P_{it} is log price. X_{it} contains controls—including lags of Y_{it} and P_{it} , product characteristics, and exogenous controls—and affects the outcome variable via function $g_o(X_{it})$ and the price variable via $m_o(X_{it})$. Members of X_{it} can affect the outcome, the treatment or both. Functions g_o and m_o are unrestricted, unit specific, highly complex, and nonlinear, known as the nuisance parameters. ξ_i^E and ξ_i^P are unobserved unit effects. U_{it} and V_{it} are disturbances. ϕ_{it} is the full set of information per unit prior to period t , spanned by lagged realizations of the demand system and consists of observables and unobserved unit effects ξ_i^E, ξ_i^P , more formally $\phi_{it} = \{(X_{is}, D_{is}, Y_{is})_{s=1}^{t-1}; \xi_i^E, \xi_i^P\}$. It is assumed that U_{it} satisfies the conditional sequential exogeneity condition, denoted in equation 3.1, to enable the causal interpretation of θ_0 . D_{it} is the technical treatment vector obtained by interacting the base treatment variable P_{it} and P_{-it} with the observable product characteristics K_{it} , which is a subset of X_{it} . P_{-it} (equation 3.4) is the average price of the other products in the same cluster with product i (non-self price).

The main equation of interest is 3.1. Equations 3.2 and 3.3 are auxiliary and contain information on the dependence of treatment variable on confounding factors X_{it} . The parameter of interest is θ_o —the vector of heterogeneous treatment effects—that we interpret as the price elasticity of demand. In other words, θ_o is the causal effect of price on demand under the assumption that U_{it} is mean independent of the information set ϕ_{it} .

$$\theta_o = (\theta_o^{own}, \theta_o^{cross}) \quad (3.5)$$

where θ_o^{own} is the vector of own-price elasticity and θ_o^{cross} is the vector of cross-price elasticity.

$$\Delta D'_{it}\theta_o = \Delta P_{it}(1, K_{it})'\theta_o^{own} = \theta_{0,1}^{own} + K'_{it}\theta_{0,-1}^{own} \quad (3.6)$$

$$\Delta D'_{it}\theta_o = \Delta P_{it}(1, K_{-it})'\theta_o^{cross} = \theta_{0,1}^{cross} + K'_{it}\theta_{0,-1}^{cross} \quad (3.7)$$

where $\theta_{0,1}^{own}$ and $\theta_{0,1}^{cross}$ are baseline treatment effects (BTE), while $K'_{it}\theta_{0,-1}^{own}$ and $K'_{it}\theta_{0,-1}^{cross}$ are treatment modification effects (TME). Demand is affected by a change in the own price (ΔP_{it}) via equation 3.6 and is affected by a change in the average cluster price (ΔP_{-it}) via equation 3.7. Cross-price elasticities across different units are assumed to be zero.

The estimator proposed by Semenova et al. (2021) —termed Orthogonal Lasso (ℓ_1 - penalised least squares)— to estimate the model postulated above has two stages that are performed on different samples, using sample splitting criteria of DML.

In the first stage, the treatment variable is interacted with a subset of covariates to obtain the vector of heterogeneous treatments. Then, the unit-specific expectation functions of each treatment interaction (equation 3.8) and outcome (equation 3.9) conditional on product characteristics, lagged prices, and sales are estimated using machine learning methods and sample A^{11} —the auxiliary sample— then the residuals are obtained on the sample M^{12} —the main sample— (equation 3.10), known as the orthogonalisation step. By doing so, the effects of confounders, X_{it} , has been removed from outcome and treatment; therefore, any remained relation between them can be estimated and interpreted as the causal effect of treatment on outcome, which is performed in the second stage of the method.

$$p_{i0} = m_0(x) + \xi_i^P \quad (3.8)$$

$$l_{i0} = d_{i0}(x)\theta_0 + g_0(x) + \xi_i^E \quad (3.9)$$

$$\hat{V}_{it} = P_{it} - \hat{p}_{i0}(X_{it}), \quad \hat{Y}_{it} = Y_{it} - \hat{l}_{i0}(X_{it}), \quad it \in M \quad (3.10)$$

In the second stage, the residuals of outcome are regressed on the residuals of the vector of heterogeneous treatments by using Lasso in a high-dimensional sparse regime¹³ and OLS in a low-dimensional regime¹⁴ to estimate θ_o , the heterogeneous treatment effect parameter.

$$\hat{\theta}_L = \arg \min \frac{1}{NT} \sum_{it \in M} (\hat{Y}_{it} - \hat{V}'_{it}\theta)^2 + \lambda \|\theta\|_1 \quad (3.11)$$

where the estimator $\hat{\theta}_L$ of θ_0 is the minimiser of Lasso lost function and $\lambda > 0$ is a penalty parameter.

The bias introduced by l_{-1} shrinkage of Lasso estimator is corrected with Gaussian bootstrap to allow simultaneous inference on θ_o .

3.4.2 Demand Model

Following Semenova et al. (2021), we estimate a partially linear sequentially exogenous dynamic panel model for quarterly demand of pharmaceuticals:

$$\log(P)_{it} = \sum_{l=1}^{48} \sum_{h=1}^{33} 1_{\{i \in G_h\}} \log(P)_{it-l} \alpha_{lh}^P + \bar{M}'_i \gamma_0 + \rho_q^P + a_i^P + V_{it}^P \quad (3.12)$$

¹¹Since we estimate the unit specific functions, the main sample is split by time dimension. $A = (Y_{it}, D_{it}, P_{it}, X_{it})_{i=1}^N$ $1 \leq t \leq T$.

¹² $M = (Y_{it}, D_{it}, P_{it}, X_{it})_{i=1}^N$ $t+1 \leq t \leq 2T$.

¹³The dimension of treatment vector D_{it} is larger than sample size N , but the number of non-zero treatments is relatively small.

¹⁴The dimension of treatment vector D_{it} is lower than sample size N .

where $G = \{G_1, G_2, \dots, G_{33}\}$ is the partition of all pharmaceutical products into 33 exclusive and exhaustive molecule categories, $(P)_{it}$, the base treatment, is log price for a pharmaceutical product in molecule G , in country c ¹⁵ at time t , l is lag, α_{lh}^P is molecule-specific effects of price lags, $\bar{M}_i'$ is time-invariant covariate vector including binary covariates such as country and molecule fixed effects¹⁶, ρ_q^P is quarter fixed effect, a_i^P is unit-specific remainders (*i.e.* unobserved unit effect) vector $a_i^P = (a_1^P, a_2^P, \dots, a_N^P)$ ¹⁷.

$$\log(S)_{it} = \log(P)_{it} \cdot (1, K_{it})' \beta_0 + \sum_{l=1}^{48} \sum_{h=1}^{33} 1_{\{i \in G_h\}} (\log(P)_{it-l}, \log(S)_{it-l})' \alpha_h^E + \bar{M}_i' \kappa_0 + \rho_m^E + a_i^E + U_{it} \quad (3.13)$$

where K_{it} is a vector of covariates. Other variables, α_h^E , $\bar{M}_i'$, κ_0 , ρ_m^E , a_i^E are defined similar to equation 3.12. The parameter of interest is the average heterogeneous effect, in other words, the average price elasticity of demand:

$$\epsilon_{it} = (1, K_{it})' \beta_0 \quad (3.14)$$

We use DML with heterogeneous treatment effects to estimate the demand model specified above. As mentioned in the previous subsection, the method is performed in two stages. In the first stage, we define the reduced forms of equations 3.12 and 3.13. Equation 3.15, *the unit-specific treatment reduced form*, is the reduced form of equation 3.12.

$$\hat{p}_{i0}(X_{it}) = \sum_{l=1}^{48} \sum_{h=1}^{33} 1_{\{i \in G_h\}} \log(P)_{it-l} \hat{\alpha}_{lh}^P + \bar{M}_i' \hat{\gamma}_0 + \hat{\rho}_m^P + \hat{a}_i^P \quad (3.15)$$

where $X_{it} = (M_{it}, \bar{M}_i, P_{it-1}, \dots, P_{it-l})$ is a vector of pre-determined control variables. Similarly, equation 3.16, *the unit-specific outcome reduced form*, is the reduced form of equation 3.13.

$$\hat{l}_{i0}(X_{it}) = \hat{p}_{i0}(X_{it}) \cdot (1, K_{it})' \check{\beta}_0 + \sum_{l=1}^{48} \sum_{h=1}^{33} 1_{\{i \in G_h\}} (\log(P)_{it-l}, \log(S)_{it-l})' \cdot \hat{\alpha}_h^E + \bar{M}_i' \hat{\kappa}_0 + \hat{\rho}_m^E + \hat{a}_i^E \quad (3.16)$$

We estimate equations 3.15 and 3.16 with Lasso¹⁸ and obtain residuals for treatment (equation 3.17) and outcome (equation 3.18).

$$\tilde{P}_{it} = P_{it} - \hat{p}_{i0}(X_{it}) \quad (3.17)$$

$$\tilde{S}_{it} = S_{it} - \hat{l}_{i0}(X_{it}) \quad (3.18)$$

In the second stage of the estimation, we regress the residuals of outcome on the residuals of the vector of heterogeneous treatments to capture the remained causal relation between them. The covariates (X_{it}) in both stages include the number of firms that produce a product inclusive of molecule i in country c at time t , percentage of generic products among all the products inclusive of molecule i in country c at time t , yearly log GDP per capita based on purchasing power parity, Gini index, fixed effects for quarter, country and molecule. In addition, the model includes the previous realisations of log price

¹⁵Each pair of molecule G and country c forms one panel unit i .

¹⁶ $\bar{M}_i = 1/T \sum_{t=1}^T M_{it}$.

¹⁷The remainders vector obtained from the correlated random effects approach that projects the unobserved unit heterogeneity, ξ_i^P , onto the vector of average exogenous covariates: $\xi_i^P = \bar{M}_i' \gamma + a_i^P$, accordingly the remainder vector is $a_i^P := (a_1^P, a_2^P, \dots, a_N^P)$.

¹⁸Belloni and Chernozhukov (2013) suggest that using Lasso at the first stage increases the fit of the model at the second stage.

and log standard units (48 lags for 48 quarters) to account for the effects of price and sales history. Average non-self unit price takes into account the prices of substitutes and complements.

Based on the method and demand model above, we estimate the price responsiveness of pharmaceutical demand at the country-molecule level. We performed three different second-stage estimation models: OLS, cross-validated Lasso, and Debiased Lasso. In addition, we estimate the demand model separately for generic pharmaceuticals only to account for product heterogeneity that potentially contributes to the variation of price elasticities.

3.4.3 Endogeneity and DML

Endogeneity is an essential problem for demand estimates. Notably, two sources of endogeneity —omitted variable bias and simultaneity— are of major concern. As already mentioned, DML allows for a large number of covariates, including product fixed effects, lags of prices and sales. In a standard regression model, using a large set of covariates is an actual problem leading to poor parameter estimates and multicollinearity (Bajari et al. (2015)). Accordingly, variable selection becomes an integral part of the estimation strategy. DML automates variable selection in the sense that large numbers of covariates included in the model are regularised, *i.e.* covariates with zero coefficients are removed from the model. This process decreases the possibility of standard omitted variable bias.

Another major source of endogeneity, simultaneity, occurs if producers and consumers know the unobserved unit characteristics —such as product quality, fewer side-effects, and higher effectiveness— and producers set prices based on them. In this case, unobserved unit characteristics are correlated with prices and quantities (Berry et al., 1995). Simultaneity problem might lead to upward-biased elasticity estimates if unobserved characteristics are not accounted for. Introducing unobserved unit effects (ξ_i^E and ξ_i^P) in the extended DML algorithm alleviates the overfitting problem since they account for molecule-country-quarter specific characteristics that are difficult to quantify. They also capture quantifiable molecule-country-quarter characteristics not present in the data set we use. Notice that unobserved unit effects are unit-specific disturbances (see equations 3.8 and 3.9) estimated with unit-specific expectation functions. Afterwards, their effects have been removed from sales and prices during the orthogonalisation step (equation 3.10). Consequently, unobserved characteristics of pharmaceutical products cannot be correlated with prices and sales in the second stage of estimation. In other words, the orthogonalisation step of DML spirits away from the potential endogeneity.

However, we instrument quarterly pharmaceutical prices with the average pharmaceutical prices in other countries as a robustness check. More precisely, the instrument for the price of a drug inclusive of molecule G in country C at time t (the baseline treatment) is the average price of drugs inclusive of the same molecule G in other countries at time t . We then re-estimate the demand model postulated above with the instrument. The results, presented in Appendix, are pretty similar to the main findings, indicating the absence of endogeneity.

3.5 Results

3.5.1 Overall Results

Figure 15 shows the overall results estimated with Debiased Lasso. Given that we only estimate own-price elasticities at this step, the dimension of treatment is low, hence Lasso, Debiased Lasso, and OLS produce very similar results (see Figure 22).

Figure 15(a) displays the average price elasticities for countries, which are the average price elasticities of molecules available in a country for the period 2008-2020. Figure 15(b) exhibits the average price elasticities for molecules. According to the results, all countries have average price elasticities less than zero, ranging from the highest (-0.88) for India to the lowest (-0.03) for the USA, while molecules have a higher variation from the highest (-1.23) to the lowest (0.02).

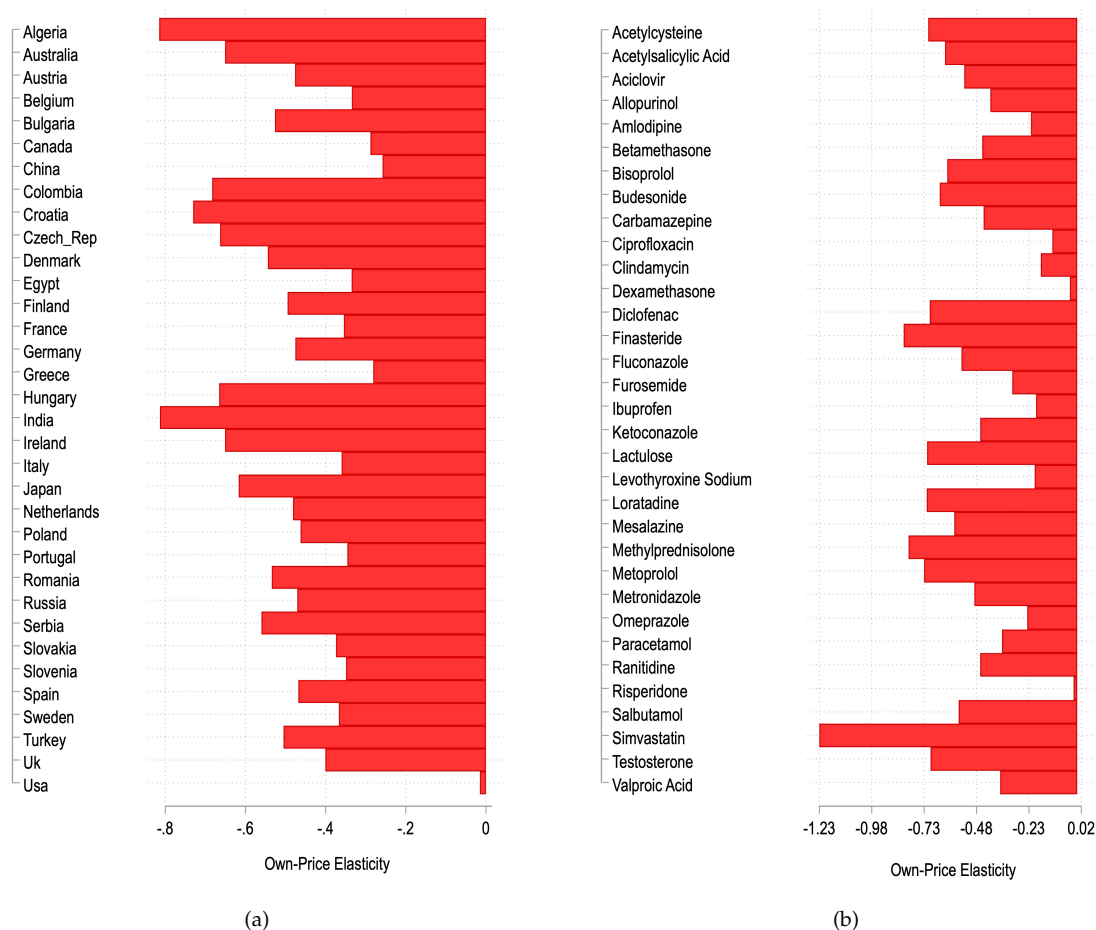


Figure 15. Price Elasticities by Country and Molecule

Figure 16 indicates how price elasticities vary by country - molecule pairs. Some molecules have significantly elastic demand —ranging from -1 to -1.61 — across all countries, represented by darker coloured columns; while others have relatively inelastic demand —ranging from -0.01 to -0.5 — represented by lighter-coloured columns. Interestingly, several molecule-country pairs have positive elasticities, which could be a signal

of the Giffen Paradox in pharmaceuticals. Nevertheless, almost all molecules show some degree of variation across countries. For instance, the price elasticity of dexamethasone is 0.43 in the USA, while it is -0.42 in India. Similarly, each country has varying elasticities across molecules despite some have a more elastic average, such as Algeria, Colombia and India, represented by darker coloured rows. In contrast, USA, Canada and China have a rather inelastic average with substantial within-variation.

Differences between Figures 15 and 16 underline the importance of high-granular level estimates. Figure 15 reports the highest (lowest) elasticity estimate to be -0.88 (-0.03) for countries and -1.23 (-0.02) for molecules. Figure 16 shows that elasticities can be more elastic (-1.81) than average levels, even positive (0.42).

Our results support the criticism of Yeung et al. (2018) who argue that the demand elasticities of pharmaceuticals are biased to be inelastic, *i.e.* close to zero, when prices of substitutes and complements are not taken into account, which was the case in previous elasticity estimates as mentioned in Section 3.2. The methodology and demand model we use allow us to incorporate non-self prices (equation 3.4) into the estimates; thus, we can rule out the cross-price effects and approximate true own-price elasticities. Correspondingly, our estimates are more elastic than the previous studies that range from -0.12 to -0.6 , even though they cannot be easily compared to our case in terms of data type, model selection and aggregation level as discussed above.

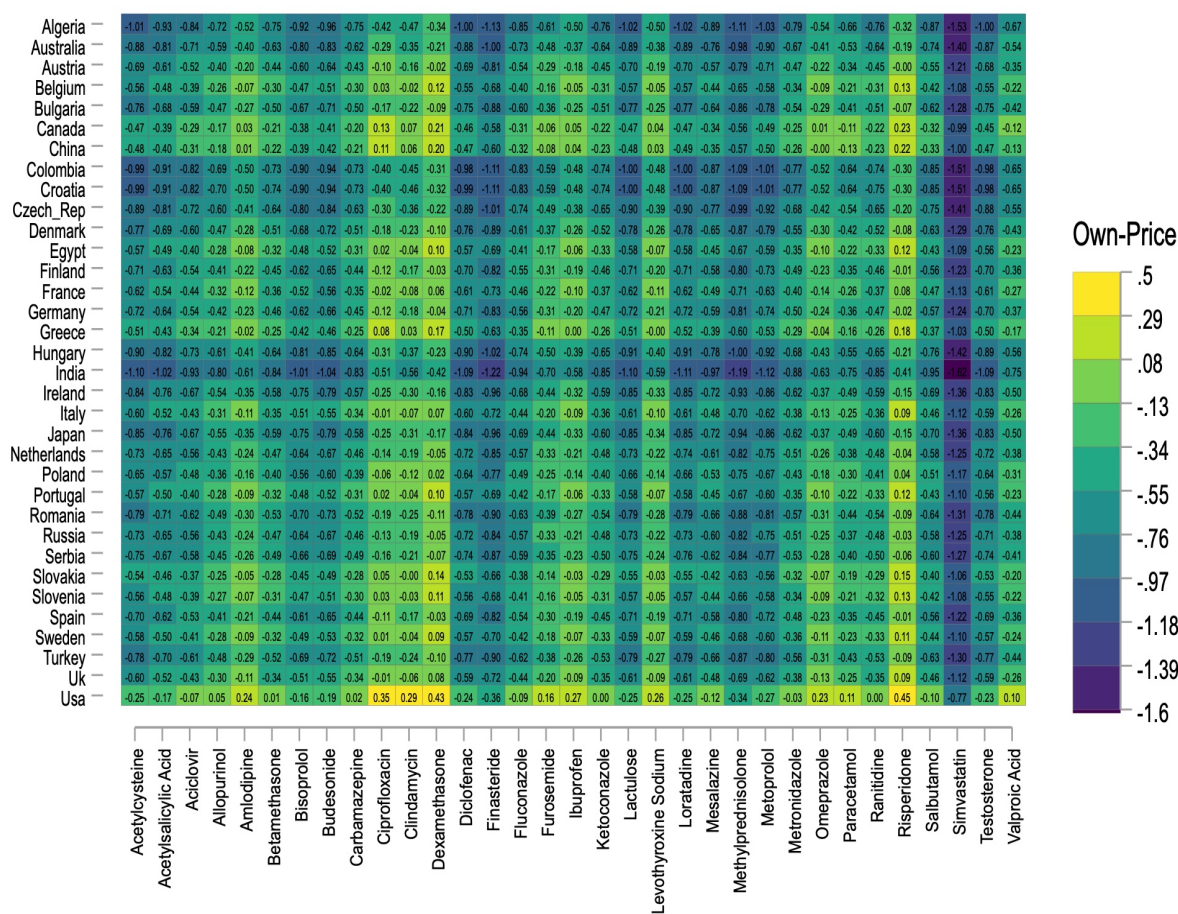


Figure 16. Price Elasticities by Molecule-Country Pairs

3.5.2 Ramsey Pricing? Price and Price Elasticity Relation

Recall that Ramsey pricing is built on the inverse elasticity rule, which states that prices should vary inversely with the price elasticity of demand. Accordingly, the existence of Ramsey prices can be tested by looking at the relationship between prices and the price elasticities of demand.

Consider two markets with different prices and price elasticities. Let price be 25, and elasticity is -0.8 in market one; while price is 22 and elasticity is -1.3 in market two. If Ramsey pricing exists in these two markets, price should be higher in the market with the lower —smaller in magnitude, indicating less price response— price elasticity of demand. Indeed, market one has a lower price elasticity and higher price than market two, meaning that the inverse elasticity rule holds and Ramsey pricing exists in these markets.

Motivated by this reasoning, we analyse the relationship between pharmaceutical prices and the price elasticities of pharmaceutical demand to test whether the inverse elasticity rule holds in national and international pharmaceutical markets. Figure 17 shows the prediction for price elasticity from a linear regression of price elasticity on average prices (log) weighted by standard units. The resulting line has a strong positive slope, indicating a negative association between price elasticity and price at the country level. Figure 18 presents the same practice applied separately for HIC and MLIC. The negative relation between price elasticity and price persists despite a slight decrease in the degree of slope for MLIC. These results validate that the inverse elasticity rule holds, thus, supporting the hypothesis that differential pricing based on Ramsey principles exists in cross-national pharmaceutical markets.

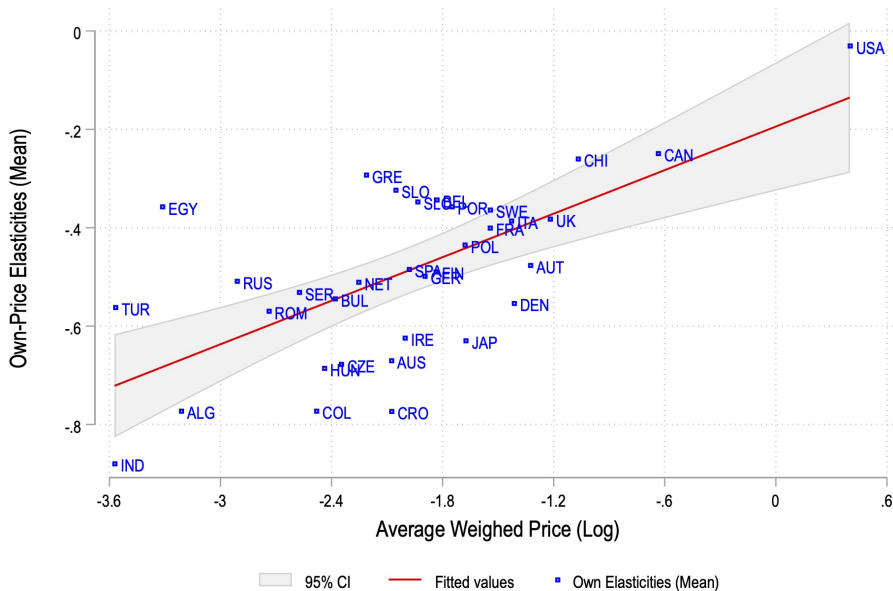


Figure 17. Relationship Between Price Elasticities and Price

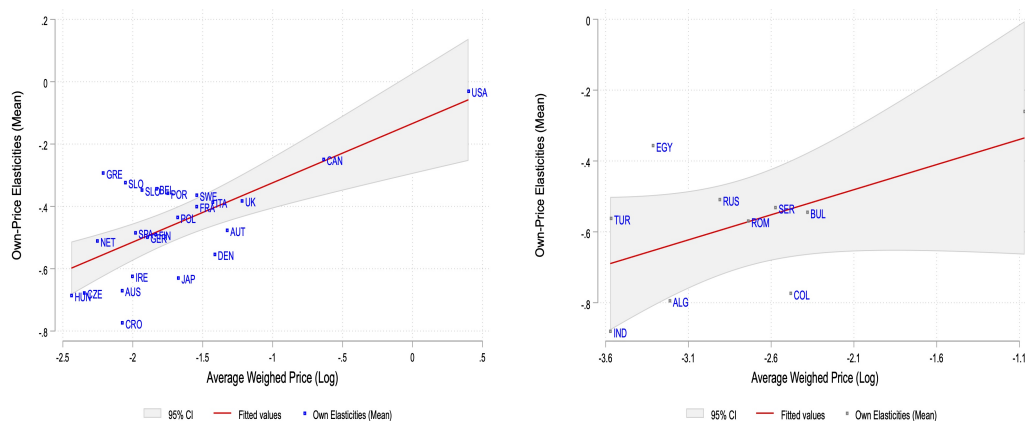


Figure 18. Relationship Between Price Elasticities and Price for (a) High Income Countries (HIC); (b) Middle and Low Income Countries (MLIC)

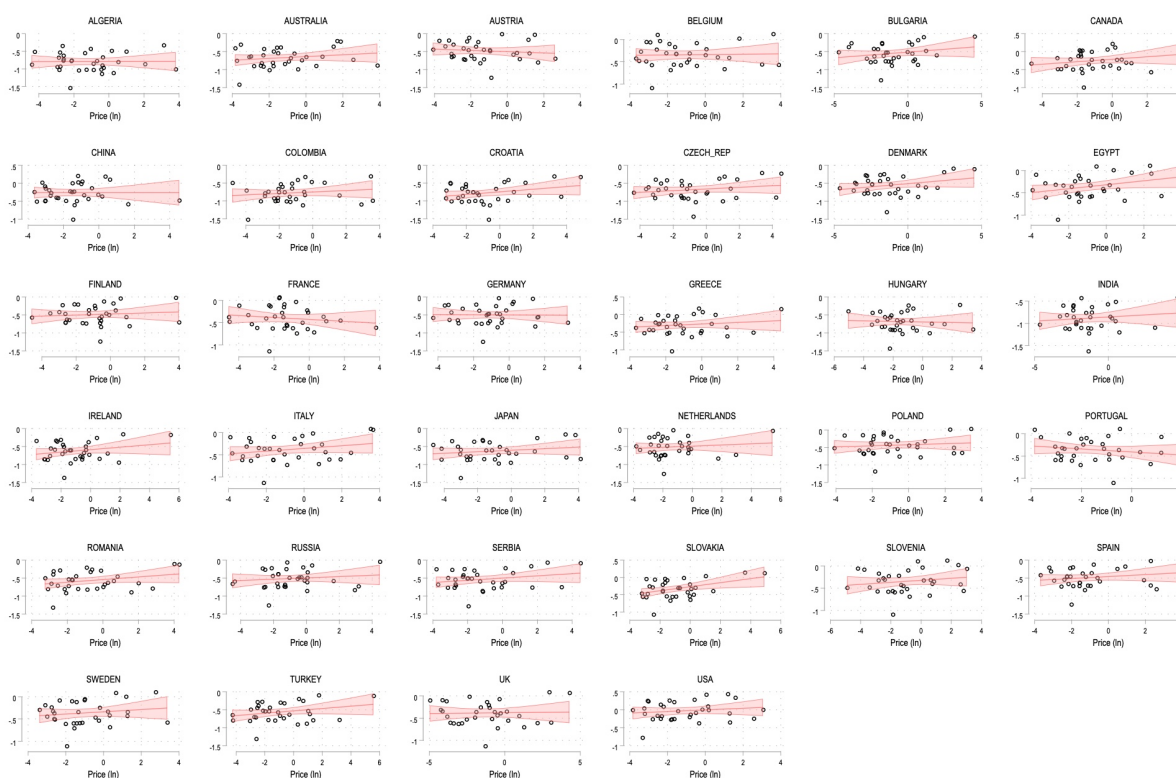


Figure 19. Relationship Between Price Elasticities and Price of Molecules by Country

Figure 19 provides within-country analyses of the relation between the price elasticities of molecules and molecule prices, while it also displays a cross-country comparison. Each subgraph shows the prediction for the price elasticities of molecules available in the country in question from a linear regression of price elasticities on average prices (log) at the molecule level. The results exhibit some degree of variation. Austria, France, and Portugal have a clear negative slope, meaning that pharmaceutical prices at the molecule level do not inversely vary with price elasticities. Algeria, Belgium, Germany, and Hungary show a rather neutral relation between price elasticities and prices. Australia, Bul-

garia, Canada, Croatia, Denmark, Egypt, and Slovakia have a somewhat strong positive slope, indicating an inverse relation between price elasticities and price, while the rest of the countries do not have a very strong positive slope. Overall, Figure 19 provides moderate evidence on the existence of within-country Ramsey pricing at the molecule level in many national pharmaceutical markets. Country-wise differences in pharmaceutical markets can be addressed by appropriate consideration of various determinants, including national health systems, insurance coverage, availability of generic products, reference pricing and parallel trade.

3.5.3 Ramsey Pricing? Income and Price Elasticity Relation

In principle, cross-country differential pricing based on Ramsey principles can only be implemented if the price elasticities of demand exhibit some degree of variation across countries. One major reason for price elasticities to vary is cross-country income inequalities. The marginal utility of money is expected to be higher in MLIC, which causes high price sensitivity, *i.e.* price elastic demand than it is expected to be in HIC. Consistently, the literature has underlined that pharmaceutical prices may vary inversely with GDP per capita income (Maskus, 2001; Sherer and Watal, 2002; Lichtenberg, 2011), assuming a negative relation between price elasticities and income (Danzon et al., 2011). However, this assumption has not been tested empirically so far. As a first attempt, Figure 20 provides clear evidence of the inverse relationship between the price elasticities of pharmaceuticals and GDP per capita income. Figure 21 indicates that the inverse relation persists in both HIC and MLIC, while MLIC have a slightly stronger relation.

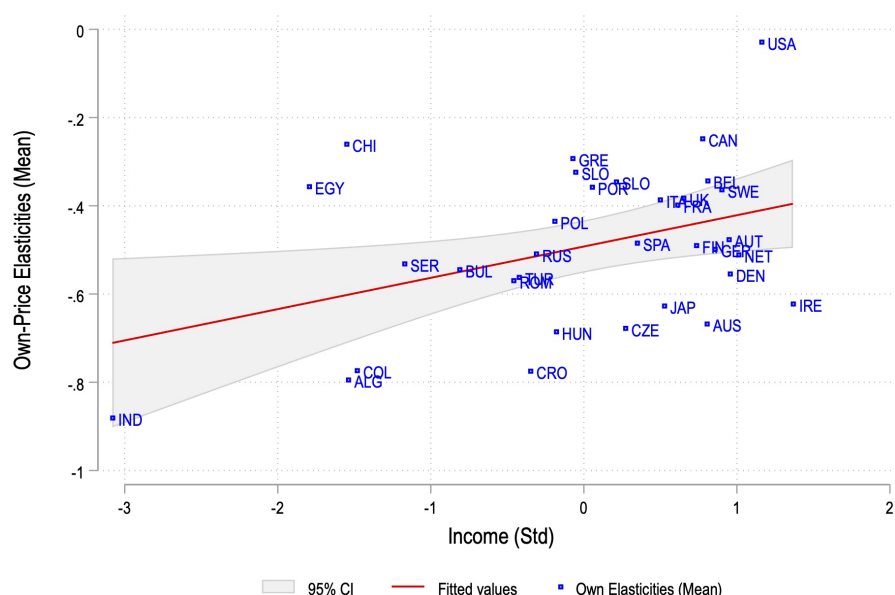


Figure 20. Relationship Between Price Elasticities and GDP Per Capita (PPP) Income

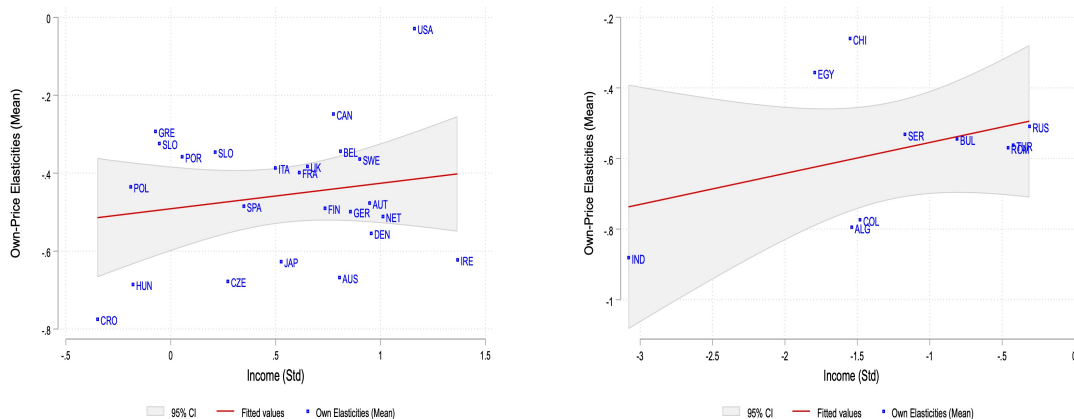


Figure 21. Relationship Between Price Elasticities and GDP Per Capita (PPP) Income for (a) High Income Countries (HIC); Middle and Low Income Countries (MLIC)

3.6 Sensitivity Analysis

3.6.1 Different Second Stage Models

As the first robustness check, we analyse if the price elasticity estimates are robust to different second-stage model choices. Figure 22 displays the results. Following Semenova et al. (2021), we estimate heterogeneous elasticities with three distinct methods: cross-validated Lasso (first panel), Debiased Lasso (middle panel) and OLS (last panel). Each point on the histograms represents one panel unit, *i.e.* country-molecule pair. Each bar is a collection of estimated price elasticities for country-molecule pairs. Consequently, pairs located in the same vertical bin have the same or very close estimated elasticities. Y-axis indicates how many times a particular elasticity value is estimated. The total number of estimated price elasticities is 1122 (33×34), which is the number of units, *i.e.* molecule-country pairs.

Notice that Figures 22(a) and 22(b) have identical distributions. The only difference between them is the grouping level. Figure 22(a) colours elasticity estimates at the country level, while Figure 22(b) colours at the molecule level to display the heterogeneity of elasticity estimates in both levels.

OLS and Debiased Lasso methods produce the same results. Cross-validated Lasso produces slightly different results, while the overall picture does not change. The results are pretty robust to different model choices.

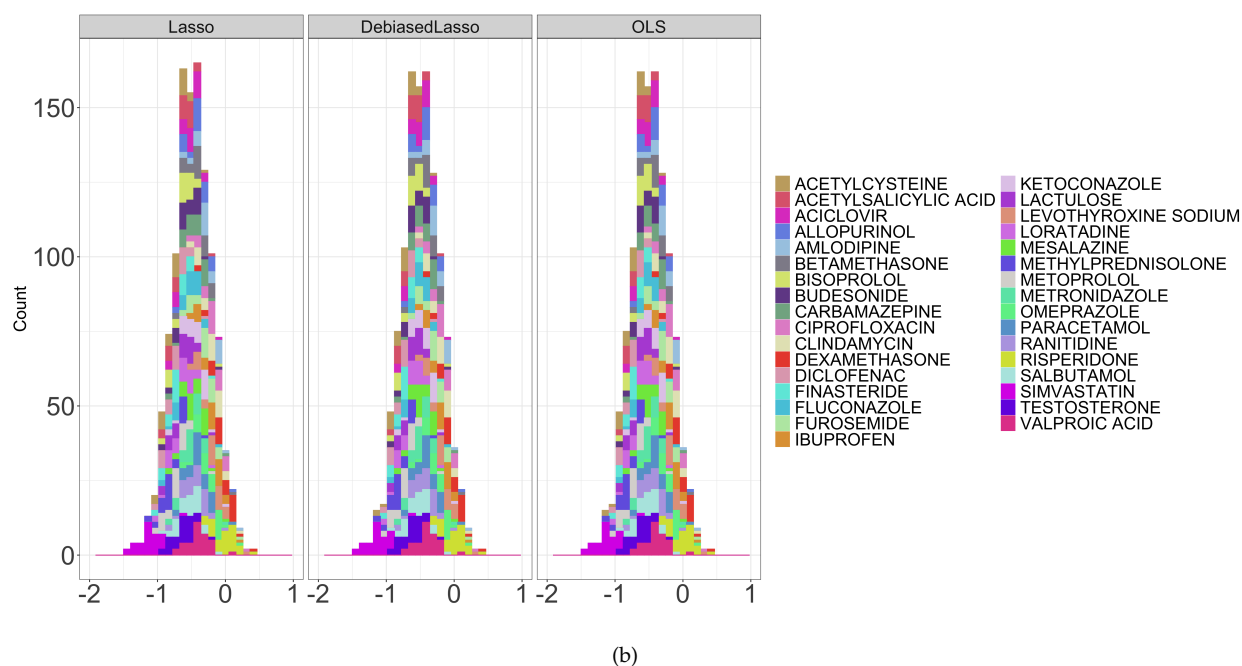
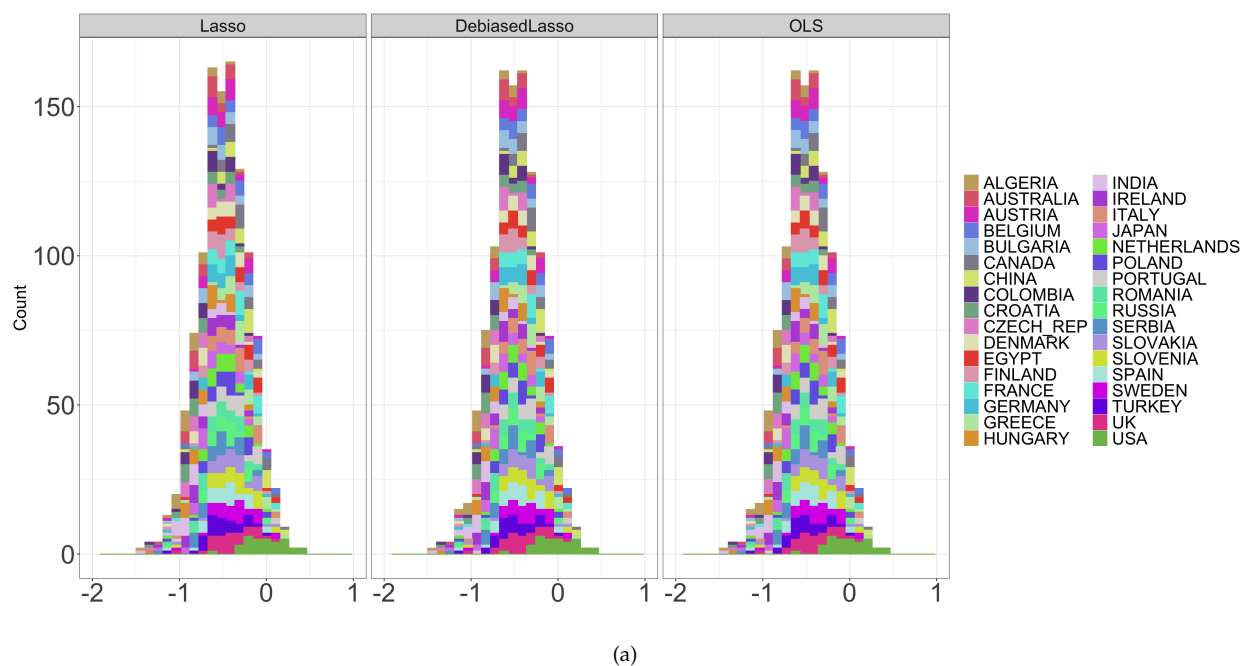


Figure 22. Different Second Stage Models

3.6.2 Heterogeneity in Treatment Effect: Generic Pharmaceuticals

Danzon et al. (2011) argue that generic markets are more competitive than originator markets because originators’ price has to cover the fixed cost resulting from pharmaceutical R&D along with marginal cost. Moreover, originators might have patent protection that prevents new entries into the market and allows firms to price discriminate across and within countries. On the other hand, generics’ prices generally reflect marginal production costs that tend to be uniform across countries. These aspects are highly likely to introduce different patterns of price elasticities for generic products. Therefore, we replicate the previous analyses only for generic pharmaceutical products as a robustness check.

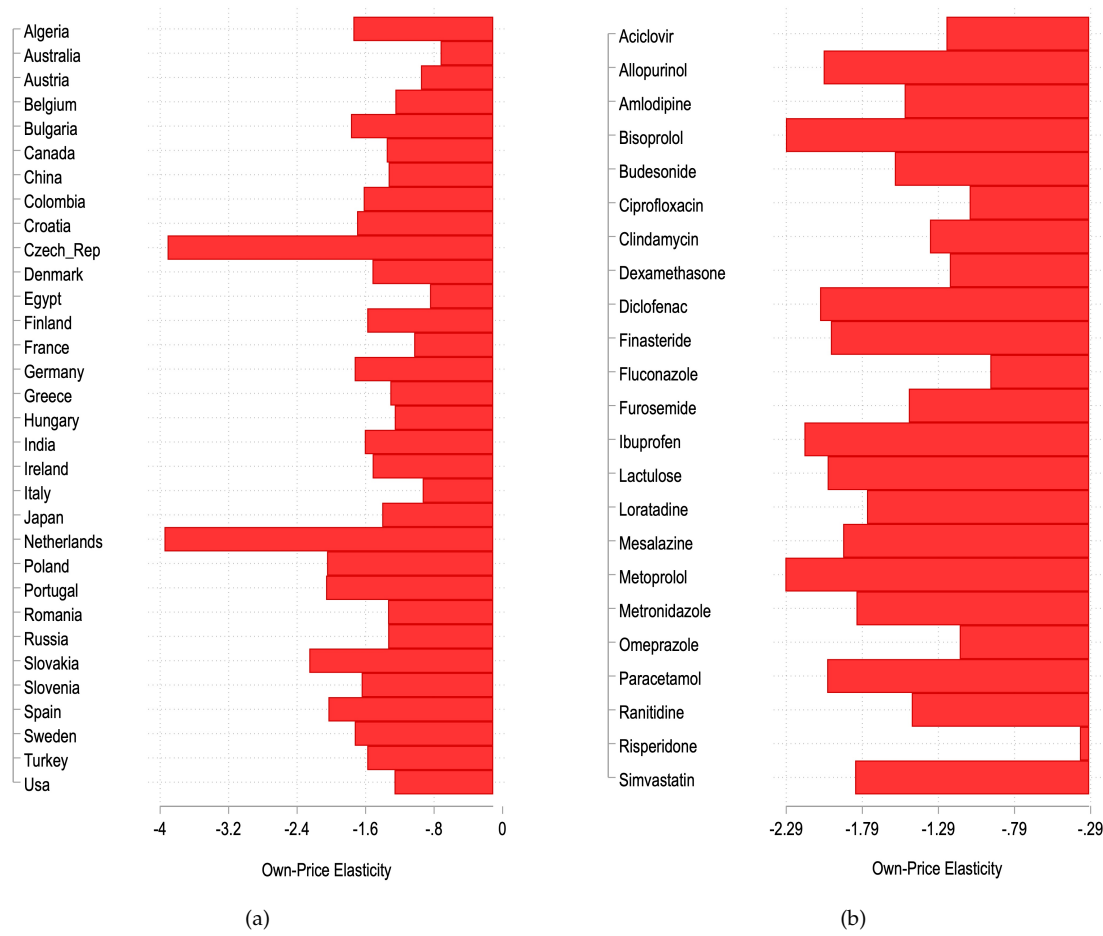


Figure 23. Price Elasticities by Country and Molecule for Generic Pharmaceuticals

Figure 23 displays the results with substantial differences across countries and molecules. Figure 23(a) shows that average price elasticities for generic products are enormously higher for Czech Republic and Netherlands, while other countries exhibit rather low price elasticities. Compared to the average price elasticities for the full range of pharmaceuticals (Figure 15), generic products are substantially more price-elastic, both at the country and molecule level. This result is intuitive because generic pharmaceuticals are

off-patent and thus can have many substitutes within-country and globally, leading to higher price sensitivity.

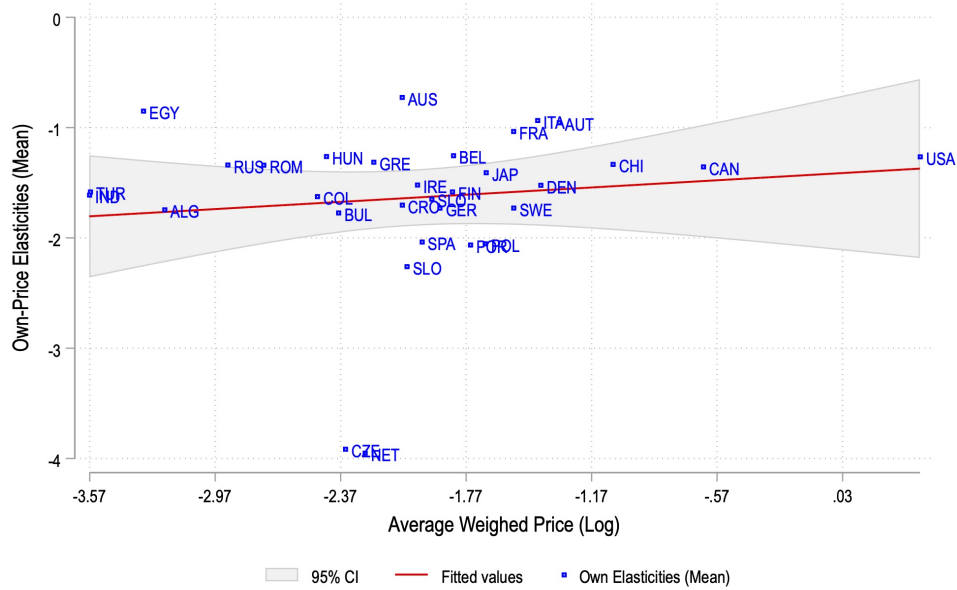


Figure 24. Relationship Between Price Elasticities and Price for Generic Pharmaceuticals.



Figure 25. Relationship Between Price Elasticities and GDP Per Capita (PPP) Income for Generic Pharmaceuticals

The relation between price elasticity and the price of generics is presented in Figure 24. A slight inverse relation is visible, yet, pretty weak compared to the full range of pharmaceutical products (Figure 17). Figure 25 displays a slightly positive, almost neutral relationship between price elasticities of generics and GDP per capita income. These results suggest that cross-country Ramsey pricing is questionable for generic pharmaceuticals.

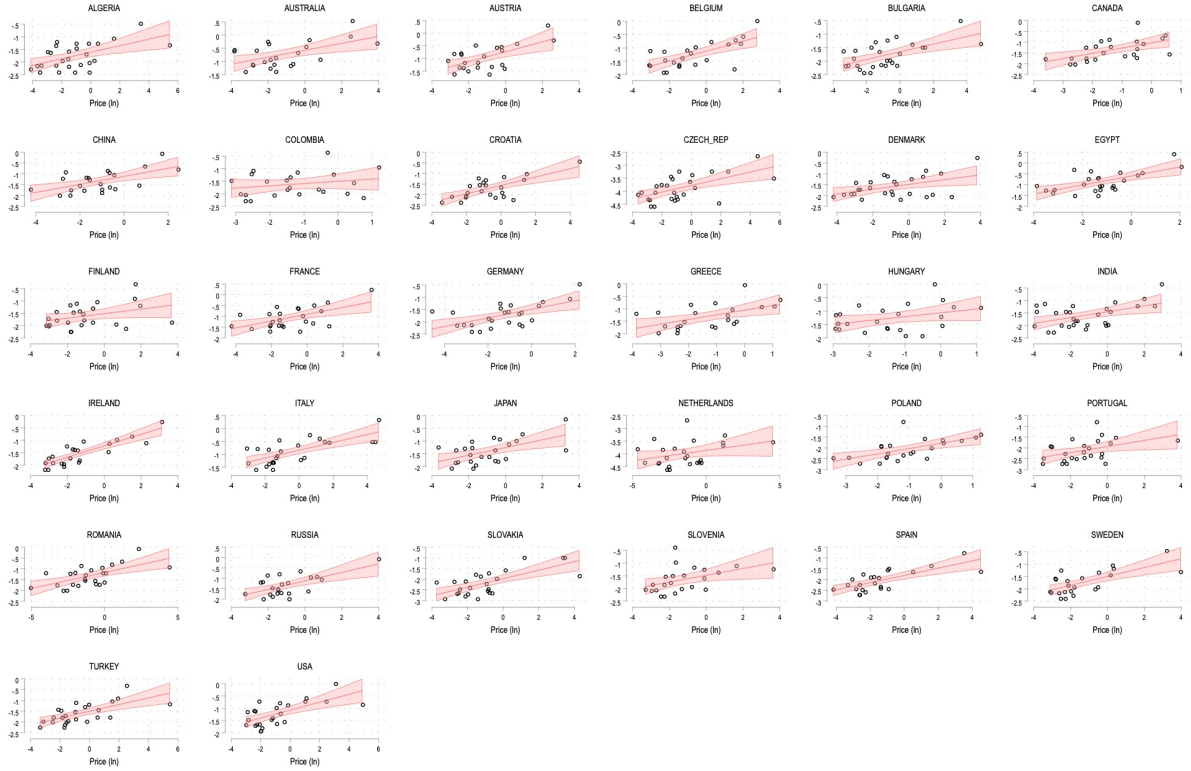


Figure 26. Price Elasticities by Molecule-Country Pairs for Generic Pharmaceuticals

Interestingly, Figure 26 shows that the inverse elasticity rule holds in national pharmaceutical markets; thus, within-country Ramsey pricing exists for generic products. Unlike the full range of pharmaceutical products (Figure 19), each sub-graph exhibits a clear positive slope, indicating a strong negative relationship between price elasticity and the price of generics in each national market.

3.6.3 DML with Cross-Sectional Setting

As another robustness check, we estimate the demand model by applying the classical DML algorithm introduced by Chernozhukov et al. (2018). The algorithm works in a cross-sectional setting. Consequently, some elements we employ in the demand model introduced in Section 3.4 are removed, including unit specific variables (ξ_i^E and ξ_i^P), lags of price and sales.

The overall results reported in previous sections stay robust. Price elasticity and price are inversely related, as seen in Figures 27 and 28, which are very similar to those estimated with the extended DML (Figures 17 and 18).

Figures 29 and 30 show a very strong correlation between price elasticities and GDP per capita income, indicating that price elasticities of pharmaceuticals are inversely re-

lated to income. Compared to the estimates of the extended DML (Figures 20 and 21), the classical DML produce stronger relation. However, despite differences, the overall results of the classical DML affirm the existence of Ramsey pricing in cross-national pharmaceutical markets.

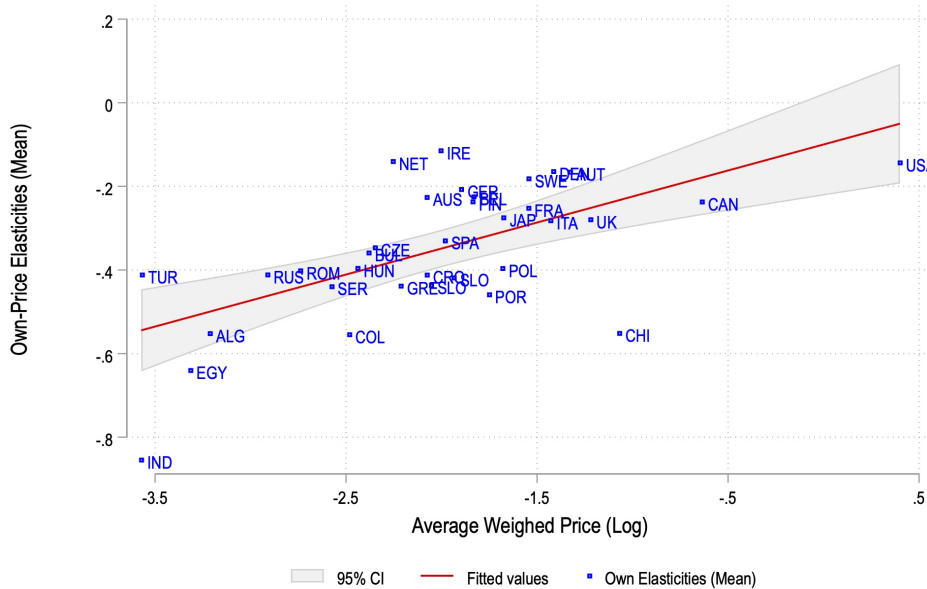


Figure 27. DML Estimates: Relationship Between Price Elasticities and Price

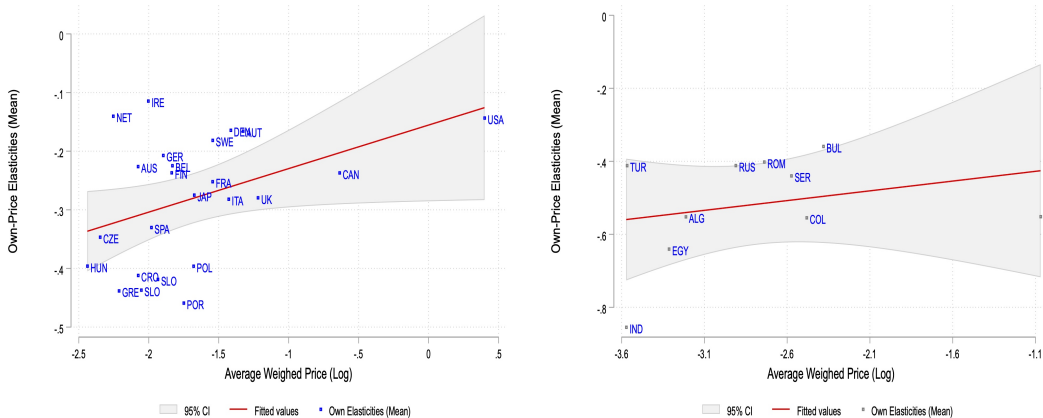


Figure 28. DML Estimates: Relationship Between Price Elasticities and Price for (a) High Income Countries (HIC); Middle and Low Income Countries (MLIC)

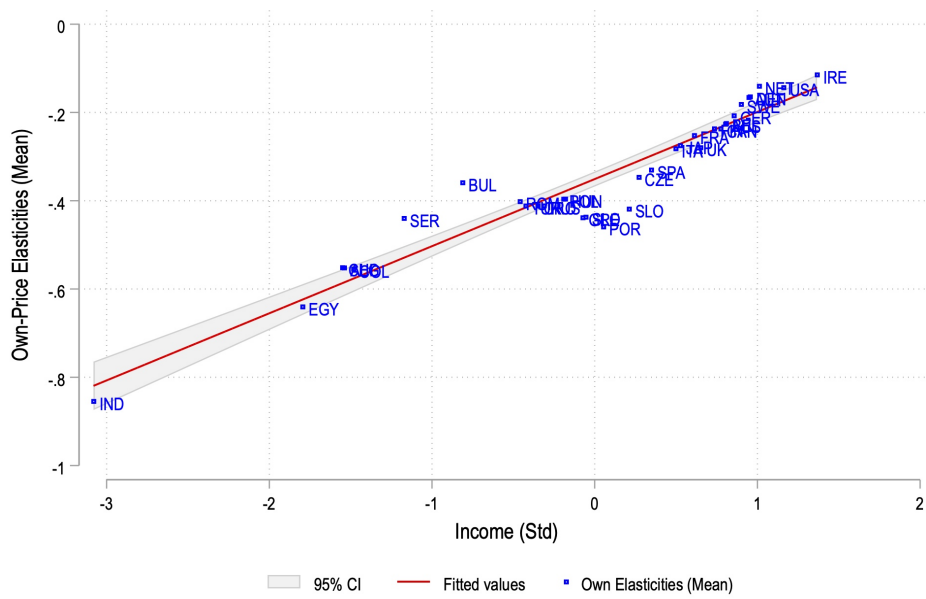


Figure 29. DML Estimates: Relationship Between Price Elasticities and GDP Per Capita (PPP) Income

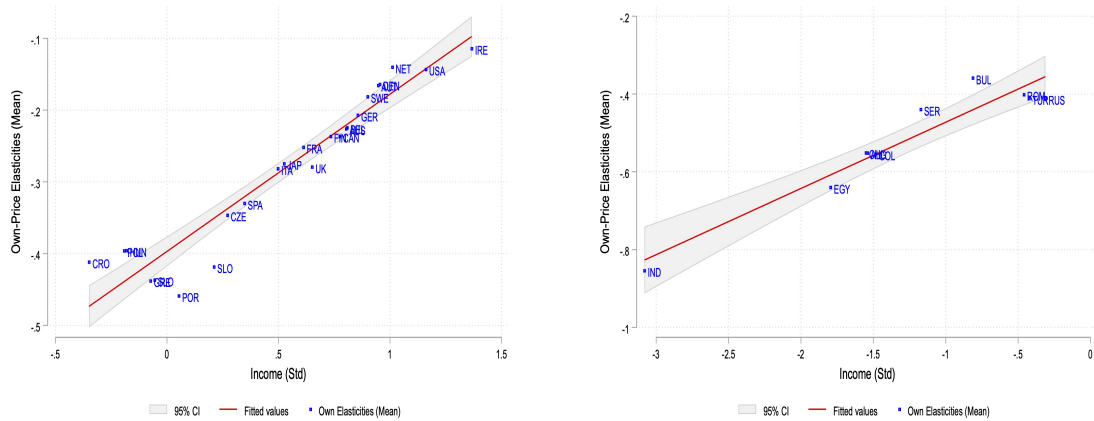


Figure 30. DML Estimates: Relationship Between Price Elasticities and GDP Per Capita (PPP) Income for (a) High Income Countries (HIC); Middle and Low Income Countries (MLIC)

3.6.4 Placebo test

As the last robustness check, we introduce a placebo test to examine the effectiveness of the DML algorithm employed to estimate the price elasticities of demand. Specifically, we randomise the input data and re-estimate own and cross price elasticities to show that the algorithm does not produce spurious results.

As the first step, we construct the placebo data set by including molecules serving different needs (unrelated ATCs) and countries that are relatively unrelated (*e.g.* Germany and Ecuador). As the second step, we perform randomisation to re-assign the standard units of different molecules in different countries. Starting from the original database, we make a permutation of the standard units at the molecule and country levels. In this way, we associate the standard units of molecule i in country c to the standard units of molecule j in country k . The randomisation step ensures the absence of any relationship between the movements of the prices. Therefore, it is reasonable to expect mostly insignificant elasticity estimates. Finally, we apply the DML algorithm to the placebo data set. Figure 31 displays the results. To facilitate the visualisation, we report the matrix of molecules' elasticities ($\mathbf{E} \in \mathbb{R}^{m \times m}$ in the form of heat map. By construction, the diagonal elements $e_{i,i}$ of $\mathbf{E}$ represent own-price elasticities, whereas non-diagonal elements $e_{i,j}, i \neq j$ represent the cross price elasticity of i with respect to j . More formally:

$$E_{i,j} := \begin{cases} \text{Own-elasticity}, & \text{if } i = j, \\ \text{Cross-elasticity}, & \text{otherwise.} \end{cases}$$

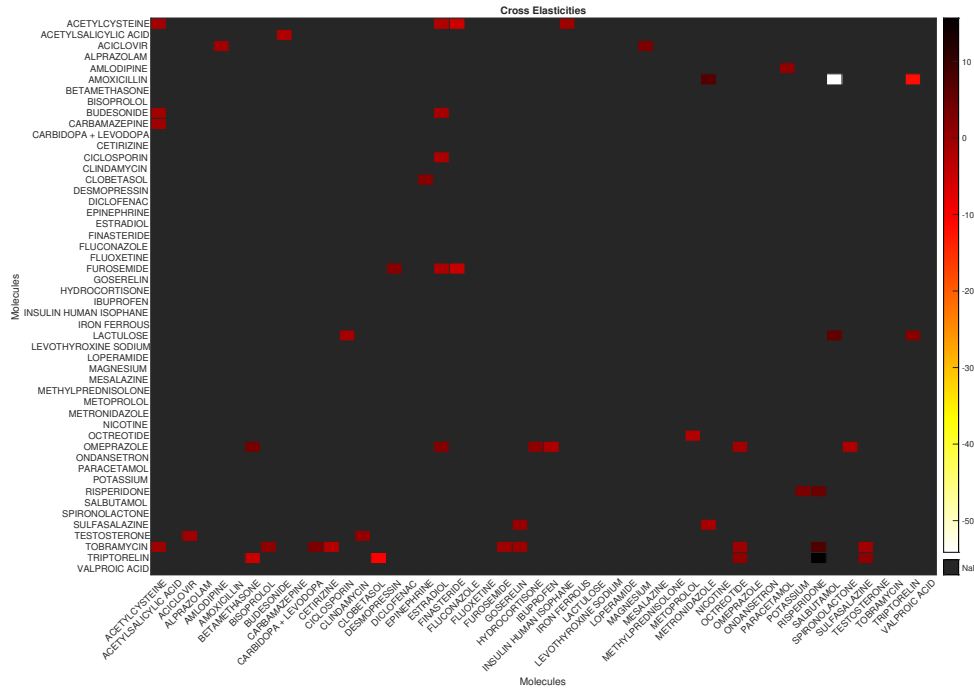


Figure 31. Heat Map of Placebo Price Elasticities Between the Molecules of Ecuador and Germany

Zero values have been set to black (NaN). The results show that the 96% of the elasticity estimates are not significantly different from zero when countries with no relevant trade relationship have interacted with each other (both on the main diagonal, *i.e.* own elasticities, and outside-of-diagonal, *i.e.* cross elasticities).

The placebo test has been repeated for several country pairs, yielding very similar results. For the sake of brevity, we only report the results for Germany and Ecuador.

3.7 Conclusion

In this study, we present the first empirical attempt to analyse Ramsey pricing of pharmaceuticals by testing the validity of the inverse elasticity rule. Previous studies either theoretically analysed Ramsey pricing (Danzon, 1997; Danzon et al., 2015; Barros and Martinez-Giralt, 2008) or indirectly assessed its existence by looking at the relationship between pharmaceutical prices and income (Maskus, 2001; Sherer and Watal, 2002). However, Ramsey pricing can be accurately analysed only by considering the association between price elasticities and prices, *i.e.* the inverse elasticity rule. To this end, we first estimate the price elasticity of demand. It is well known that previous demand analyses—that are within-country and at the insured patient level—considered pharmaceutical products as isolated entities, assuming no cross-price effects of substitutes and complements. Moreover, those studies suffered from omitted variables bias and endogeneity of prices. To overcome these methodological issues, we employ a recently developed machine learning method (DML) to estimate price elasticities.

DML with heterogeneous treatments is performed in two stages. In the first stage, we define and estimate the reduced forms of expectation functions of treatment and outcome, and then we save residuals. In the second stage, we regress the residuals of outcome on the residuals of treatment along with covariates using Lasso, Debiased Lasso, and OLS. DML has two main advantages: (1) non-self prices are included as controls to account for cross-price effects, (2) previous realisations of the demand system—48 lags of price and 48 lags of standard units in our case—included to account for unobserved unit effects. These strong properties of DML allow us to accurately estimate the cross-national price elasticities of demand.

According to the results, average price elasticities range from the highest (-0.88) for India to the lowest (-0.03) for the USA, while molecules have a higher variation from -1.23 to 0.02 . Our results support the long-standing idea that pharmaceutical demand tends to be inelastic.

A cross-country assessment of the inverse elasticity rule shows that prices vary inversely with price elasticities, both in HIC and MLIC, indicating the existence of Ramsey pricing in cross-country pharmaceutical markets. However, within-country Ramsey pricing is questionable when we look at pharmaceutical markets at the molecule level. Prices do not inversely vary with price elasticities in some countries (Austria, France, and Portugal), while others show a neutral relationship (Algeria, Belgium, Germany, and Hungary). Australia, Bulgaria, Canada, Croatia, Denmark, Egypt, and Slovakia have a somewhat strong inverse relation. The rest of the countries tend to have an inverse relation as well, yet not very strong.

We also analyse the relationship between GDP per capita and the price elasticities of pharmaceutical demand at the country level. We find evidence of the inverse relation between price elasticities and income in HIC and MLIC, supporting Ramsey pricing.

We perform four sensitivity analyses to build robustness for our results. First, we estimate the second stage with three different models: Lasso, Debiased Lasso, and OLS. We show that our results are robust to alternative model choices. Second, we replicate the analysis only for generic pharmaceutical products. We find that Ramsey pricing does not exist in cross-country markets for generic pharmaceutical products, while within-country Ramsey pricing exists. Third, we re-estimate the demand model using classical DML

(Chernozhukov et al., 2018) in a cross-sectional setting. The results are pretty similar to those we estimate with extended DML (Semenova et al., 2021) in a panel setting. Lastly, we perform a placebo test to show the effectiveness of the DML algorithm.

Conclusion

The present thesis includes three independent chapters that explore various empirical methods.

The first chapter analyses the effect of religion on innovation by addressing the inherent endogeneity of religion. Four variables measure religiosity: self-reported degree of being religious, frequency of attending religious activities, frequency of praying, and religiosity index. Innovation is measured by innovation attitudes which are personal traits that are favourable or unfavourable to innovation, presenting an individual's propensity to innovate. Innovation attitudes are broken down into positive (favourable to innovation) and negative (unfavourable to innovation) innovation attitudes. We build instruments for each measure of religiosity by computing the average religiosity of people of the same sex, age range, and religious affiliation who live in countries with the same dominant religious denomination.

We start by estimating a linear model using OLS. The results demonstrate that each measure of religiosity is significantly and positively correlated with the average negative innovation attitudes, implying that religiosity has negative effects on innovativeness. Regarding the average positive innovation attitudes, only the frequency of attendance to religious services is significantly and negatively correlated. The frequency of praying and degree of being religious are positively correlated with the average positive innovation attitudes, suggesting that these two religiosity measures positively affect innovativeness. We then apply the instrumental variables method. The significant and positive correlation remains robust between the average negative innovation attitudes and each measure of religiosity, along with increased correlation coefficients. But the overall picture changes with the average positive innovation attitudes. Religiosity index and frequency of praying turn out to be negatively and significantly correlated, while the frequency of attending religious services remains negative and significant. The degree of being religious is no longer significant, suggesting that the OLS results for the average positive innovation attitudes may be driven by endogeneity.

We perform a couple of sensitivity analyses to see if the model is robust. First, we estimate the baseline specification with a different dependent variable. Second, we consider different model specifications to ensure that weak instruments do not cause the results. Third, we reduce the main sample in a couple of ways and re-estimate the baseline specifications. Finally, we consider two possible violations of the exogeneity condition. The overall results stay robust, and the negative effect of religiosity on innovativeness remains valid.

We discuss three possible causal channels from religion to innovation: time allocation, the fear of uncertainty, and conventional roles reinforced by religion.

The second chapter employs skill relatedness and economic complexity methods to explore the importance of regional capabilities, in the form of workplace skills, on industrial specialisation and regional diversification. We first analyse the skill space of Italy's industry mix to explore the underlying interplay between different skill types.

Data-driven techniques we use reveal that workplace skills form two highly polarised clusters into social-cognitive and technical-physical skills. We then investigate whether the concepts of skill relatedness and skill complexity enhance or mitigate the regional diversification process of industries. We use a linear probability model with three-way fixed effects to quantify the effect of skill relatedness density and skill complexity on the probability of (1) forming a new regional comparative advantage in terms of the number of employed people in a given industry (entry) or (2) losing an established comparative advantage (exit). The independent variables, skill relatedness density and skill complexity, are constructed for three different aggregation levels: (1) full skills, (2) social-cognitive skills and (3) technical-physical skills. The results show a highly significant and positive relationship between higher skill relatedness density and the probability of entry, regardless of the skill type. Even though the correlation coefficients are pretty close among the skill types, similarity to technical-physical skills yields a slightly higher probability of entry. When it comes to exit, all estimates show a highly significant and negative association between skill relatedness density and the probability of losing an established comparative advantage, regardless of the skill type.

The main findings are robust to a couple of sensitivity analyses. Firstly, we consider different models. The baseline specifications are re-estimated with logistic and probit models to compare their average marginal effects with the linear probability model. Secondly, we re-identified the dependent variables, entry and exit, with a different data set and comparative advantage criterion. Instead of revealed comparative advantage in the number of employed people, we used revealed comparative advantage in the number of active enterprises. Thirdly, we decreased the time span of the comparative advantage to one year and estimated the baseline specification with both RCA criteria. Lastly, we used provinceyear and industryyear fixed effects instead of their separate inclusion. Each analysis supports the main findings and affirms their robustness.

The third chapter examines Ramsey pricing of pharmaceuticals by using a recently developed machine learning method: double/debiased orthogonal machine learning with heterogeneous treatments in a dynamic panel setting. The method incorporates previous realisations of the demand system, such as price lags and sales lags, accounting for unobserved unit heterogeneity and serial correlation. Moreover, we control for cross-price effects. These aspects of the demand model allow us to approximate to true own-price elasticities of pharmaceuticals. The overall results show that all countries have average price elasticities less than zero, ranging from the highest (0.88) for India to the lowest (0.03) for the USA. In contrast, molecules have a higher variation from 1.23 to 0.02.

The validity of the inverse elasticity rule is analysed by considering the relationship between pharmaceutical prices and the price elasticities of pharmaceutical demand at the country level. The results indicate that prices vary inversely with price elasticities both in high-income and middle-low income countries, meaning that the inverse elasticity rule holds. Moreover, we examine the relationship between country income and the price elasticities of pharmaceutical demand. According to the results, price elasticity has a negative association with country income, indicating the existence of Ramsey pricing in cross-country pharmaceutical markets.

We perform four sensitivity analyses to assess the robustness of results. First, we estimate the second stage with three different models: Lasso, Debiased lasso, and OLS. We show that the results are robust to alternative model choices. Second, we replicate the analysis only for generic pharmaceutical products. We find that Ramsey pricing does not exist in cross-country markets for generic pharmaceutical products, while within-country Ramsey pricing exists. Third, we re-estimate the demand model using classical DML in a

cross-sectional setting. The results are pretty similar to those we estimate with extended DML in a panel setting. Lastly, we perform a placebo test to display the effectiveness of the DML algorithm.

Appendix: Religiosity and Innovation Attitudes

A. Variables Index and Data Summary

Innovation Variables

All the innovation variables below are based on self-reported assessments on the following statements, where the options are; *very much like me* (1), *like me* (2), *somewhat like me* (3), *a little like me* (4), *not like me* (5), *not like me at all* (6). All variables are recoded as to be increasing in a scale from 0 (*not like me at all*) to 1 (*very much like me*) for the sake of easier interpretation.

We distinguish innovation attitudes as *positive innovation attitudes* (PIA) and *negative innovation attitudes* (NIA). PIA include creativity, being different, being free, and being adventurous, while NIA include following rules and following traditions.

adventurous: being adventurous, based on the statement "*He looks for adventures and likes to take risks. He wants to have an exciting life.*"

creative: creativity, based on the statement "*Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.*"

different: being different, based on the statement "*He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life.*"

free: being free, based on the statement "*It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.*"

innegav: average negative innovation attitudes. Computed as the simple mean of *rules*, *traditions*.

inposav: average positive innovation attitudes. Computed as the simple mean of *creative*, *different*, *free*, *adventurous*.

innosum: summary innovation attitudes. Computed as the sum of positive innovation attitudes minus the sum of negative innovation attitudes: $\text{insum} = (\text{creative} + \text{different} + \text{free} + \text{adventurous}) - (\text{traditions} + \text{rules})$. Rescaled to be between 0 (low propensity to innovate) and 1 (high propensity to innovate).

rules: following rules, based on the statement "*He believes that people should do what they're told. He thinks people should follow rules at all times, even when no-one is watching.*"

traditions: following traditions, based on the statement "*Tradition is important to him. He tries to follow the customs handed down by his religion or his family.*"

Religion Variables

attendance: frequency of attendance to religious activities, based on the question "*Apart from special occasions such as weddings and funerals, about how often do you attend religious services nowadays?*" where the options are; *every day* (1), *more than once a week* (2), *once a week* (3), *at least once a month* (4), *only on special holy days* (5), *less often* (6), *never* (7). The variable is rescaled to be between 0 (*never*) and 1 (*every day*) to ease the interpretation.

belonging: religious affiliation, belonging to a religion or denomination, a dummy variable based on the question "*Do you consider yourself as belonging to any particular religion or denomination?*" 0 is *no*, and 1 is *yes*.

belongingp: past belonging to a religion or domination, a dummy variable based on the question "*Have you ever considered yourself as belonging to any particular religion or denomination?*" 0 is *no*, and 1 is *yes*.

degree: the degree of being religious, based on the question "*Regardless of whether you*

belong to a particular religion, how religious would you say you are?" where 0 is *not religious at all*, 10 is *very religious*. The variable is rescaled to be between 0 (*not religious at all*) and 1 (*very religious*) to ease the interpretation.

denomination: belonging to a particular denomination, based on the question "*If you consider yourself as belonging to any particular religion or denomination, which one is it?*" where the options are; *Roman Catholic* (1), *Protestant* (2), *Eastern Orthodox* (3), *Other Christian denominations* (4), *Jewish* (5), *Islamic* (6), *Eastern religions* (7), *Other non-Christian religions* (8), *not declare which religion* (9), *not belong to a religion* (10).

denominationp: past belonging to a particular denomination, based on the question "*If you have ever considered yourself as belonging to any particular religion or denomination, which one was it?*" where the options are the same as those of *denomination*.

pray: frequency of praying, based on the question "*Apart from when you are at religious services, how often, if at all, do you pray?*" where the options are the same as those of *attendance*. The variable is rescaled to be between 0 (*never*) and 1 (*every day*) to ease the interpretation.

religiosity: religiosity index, computed as the principal component of *degree*, *attendance*, and *pray*. Rescaled to be between 0 *not religious on average*, and 1 *very religious on average*.

Instrumental Variables

iv_attendance: instrumental variable for *attendance*. Frequency of attendance to religious activities of individual *i*, instrumented by the average frequency of attendance to religious activities of people who have the same age, gender, and religious affiliation with *i* and live in a country with the same dominant denomination as the country of *i*. Takes values between 0 (*low frequency of attendance to religious activities*) and 1 (*high frequency of attendance to religious activities*).

iv_degree: instrumental variable for *degree*. The degree of religiosity of individual *i*, instrumented by the average degree of religiosity of people who have the same age, gender, and religious affiliation with *i* and live in a country with the same dominant denomination as the country of *i*. Takes values between 0 (*not religious*) and 1 (*very religious*).

iv_pray: instrumental variable for *pray*. Frequency of praying of individual *i*, instrumented by the average frequency of praying of people who have the same age, gender, and religious affiliation with *i* and live in a country that has the same dominant denomination as the country of *i*. Takes values between 0 (*low frequency of pray*) and 1 (*high frequency of pray*).

iv_religiosity: instrumental variable for *religiosity*. religiosity index of individual *i*, instrumented by the average of religiosity index of people who have the same age, gender and religious affiliation with *i* and live in a country with the same dominant denomination as the country of *i*. Takes values between 0 (*low religiosity*) and 1 (*high religiosity*).

Control Variables

age: age of respondent.

age2: square of age of respondent.

bornc: born-in country of respondent. A standard dummy variable, 0 is *no* which means that the respondent was born in a different country than the country where she takes the survey; and 1 is *yes*, meaning that the respondent was born in the same country where she takes the survey.

child: whether respondent lives with children at home or not. A standard dummy variable with 0 is *no* which means that the respondent does not live with children, and 1 is *yes* meaning that the respondent lives with children.

education: completed years of education of respondent, based on the question "About how many years of education have you completed, whether full-time or part-time? Please report these in full-time equivalents and include compulsory years of schooling."

fathere: completed level of education of respondent's father. The same classification problem, as of *mothere*, stand for *fathere* as well. The variable is, therefore, constructed by grouping two variables (*escedf*, *edulvlf*) from ESS data. The values are the same as those of *mothere*.

gender: a standard dummy variable for the gender of respondent, 0 is *male*, and 1 is *female*.

health: subjective general health of respondent where the values are: *very bad* (1), *bad* (2), *fair* (3), *good* (4), *very good* (5).

mothere: completed level of education of respondent's mother. The variable is constructed by grouping two variables (*escedm*, *edulvma*) from ESS data. The variable *edulvma* is classified by ISCED 1997 and only exists for ESS rounds 1,2,3,4; while *escedm* is classified by ISCED 2011 and only exists for ESS rounds 4,5,6,7,8. *mothere* thus a grouped version¹⁹ of these two variables and the values are: *less than lower secondary education completed* (1), *lower secondary education completed* (2), *upper secondary education completed* (3), *post-secondary non-tertiary education completed* (4), *tertiary education completed* (5).

paidwork: paid work status of respondent, based on the question "Have paid work in last seven days?" where 0 is *not marked* which means respondent does not have a paid job, and 1 is *marked* meaning that respondent has a paid job. The variable is controlled with the help of the variable *crpdwk* (control for paid work).

partner: whether respondent lives with a husband/wife/partner at home or not. A standard dummy variable with 0 is *no* which means that the respondent does not live with a partner, and 1 is *yes* meaning that the respondent lives with a partner.

pwbefore: whether respondent has ever had a paid job or not. A standard dummy variable with 0 is *no* which means that respondent has never had a paid job before; 1 is *yes* meaning that respondent has had a paid job before.

Fixed Effects

country: country of the respondent. The ESS has observations for 36 European countries, but we drop Israel since it is not located in continental Europe.

denomination: please refer to Religion Variables.

essround: survey dummies for eight rounds (2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016) of the ESS.

income1 & *income2*: both variables stand for net income of household from all sources. *income1* only exists for the ESS rounds 1,2,3; while *income2* exists for the ESS rounds 4,5,6,7,8 due to a change of classification. But they cannot be grouped because *income1* takes values with (12) ranges, while *income2* with (10) deciles. Thus we use them separately.

¹⁹For detailed information about ISCED classifications please refer to the guide *International Standard Classification of Education (ISCED)* by Eurostat. The related link:
<https://ec.europa.eu/eurostat/statistics-explained/pdfscache/44322.pdf>

occupation: occupation of respondent. Constructed by grouping the two occupation variables with different classifications (*iscoco* and *isco08*) from the ESS data. *iscoco* exists for the ESS rounds 1,2,3,4,5 and takes values of *International Standard Classification of Occupations 1988* (ISCO-88); while *isco08* only exists for the ESS rounds 6,7,8 and takes values of *International Standard Classification of Occupations 2008* (ISCO-08). The values of *occupation* are relabeled in accordance with ISCO-08²⁰ classification. Occupation variable includes 582 different values (between 0 and 9999) that refer to different occupation labels. We consider only the main categories, which are 9, as fixed effects; armed forces occupations (0-1000), managers (1000-2000), professionals (2000-3000), technicians, and associate prof. (3000-4000), clerical support workers (4000-5000), services and sales workers (5000-6000), skilled agricultural, forestry, and fishery workers (6000-7000), plant and machine operators and assemblers (7000-8000), elementary occupations (8000-9000), no answer (9000-9999).

Other Variables

devote: Based on the statement *"It is important to him to be loyal to his friends. He wants to devote himself to people close to him."* Takes the same values as *safe*.

family: Based on the statement *"A person's family ought to be his or her main priority in life."* Takes the same values as *lookafter*.

government: Based on the statement *"It is important to him that the government ensures his safety against all threats. He wants the state to be strong so it can defend its citizens."* Takes the same values as *safe*.

lgbt: Based on the statement *"Gay men and lesbians should be free to live their own life as they wish."* Takes the same values as *lookafter*.

lookafter: Based on the statement *"Society would be better off if everyone just looked after themselves."* where the options are *agree strongly* (1), *agree* (2), *neither agree or disagree* (3), *disagree* (4), *disagree strongly* (5). Recoded as to be increasing in a scale from 0 (*disagree strongly*) to 1 (*agree strongly*) for the sake of easier interpretation.

menwork: Based on the statement *"When jobs are scarce men should have more right to a job than women."* Takes the same values as *lookafter*.

safe: Based on the statement *"It is important to him to live in secure surroundings. He avoids anything that might endanger his safety."* where the options are *very much like me* (1), *like me* (2), *somewhat like me* (3), *a little like me* (4), *not like me* (5), *not like me at all* (6). Recoded as to be increasing in a scale from 0 (*not like me at all*) to 1 (*very much like me*) for the sake of easier interpretation.

securejob: Based on the statement *"How important do you think each of the following would be if you were choosing a job: a secure job"* where the options are *not important at all* (1), *not important* (2), *neither important nor unimportant* (3), *important* (4), *very important* (5). Recoded as to be increasing in a scale from 0 (*not important at all*) to 1 (*very important*) for the sake of easier interpretation.

success: Based on the statement *"Being very successful is important to him. He hopes people will recognize his achievements."* Takes the same values as *safe*.

trust: Based on the statement *"Using this card, generally speaking, would you say that most*

²⁰The correspondences between the classifications of ISCO-88 and ISCO-08 are taken from the guide: *International Standard Classification of Occupations: ISCO-08*. (2012) Volume I: Structure Group Definitions and Correspondence Tables by International Labour Organization.

people can be trusted, or that you cannot be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you cannot be too careful and 10 means that most people can be trusted." Normalized to be between 0 (cannot be too careful) and 1 (most people can be trusted)

unemployed: A dummy based on the question "Have you ever been unemployed and seeking work for a period of more than three months?" 0 is no, and 1 is yes.

womenwork: Based on the statement "A woman should be prepared to cut down on her paid work for the sake of her family." Takes the same values as *lookafter*.

Table A1. The ESS Data, Countries, and Dominant Denominations

COUNTRY	Ess 1	Ess 2	Ess 3	Ess 4	Ess 5	Ess 6	Ess 7	Ess 8
Albania (AL)						•		
Muslim (57%)								
Austria (AT)	•	•	•				•	•
Catholic (62%)								
Belgium (BE)	•	•	•	•	•	•	•	•
Not belong (56%)								
Bulgaria (BG)			•	•	•	•		
Orthodox (51%)								
Croatia (HR)				•	•			
Catholic (76%)								
Cyprus (CY)			•	•	•	•		
Orthodox (75%)								
Czech Republic (CZ)	•	•		•	•	•	•	•
Not belong (78%)								
Denmark (DK)	•	•	•	•	•	•	•	•
Protestant (52%)								
Estonia (EE)		•	•	•	•	•	•	•
Not belong (74%)								
Finland (FI)	•	•	•	•	•	•	•	•
Protestant (55%)								
France (FR)	•	•	•	•	•	•	•	•
Not belong (51%)								
Germany (DE)	•	•	•	•	•	•	•	•
Not belong (38%)								
Great Britain (GE)	•	•	•	•	•	•	•	•
Not belong (53%)								
Greece (GR)	•	•		•	•			
Orthodox (89%)								
Hungary (HU)	•	•	•	•	•	•	•	•
Not belong (45%)								
Iceland (IS)		•				•		•
Not belong (57%)								
Ireland (IE)	•	•	•	•	•	•	•	•
Catholic (72%)								
Italy (IT)	•					•		•
Catholic (71%)								
Kosovo (XK)						•		
Muslim (88%)								
Lithuania (LT)					•	•	•	•
Catholic (79%)								
Latvia (LV)				•				
Not belong (53%)								
Luxemburg (LU)	•	•						
Catholic (53%)								
Netherlands (NL)	•	•	•	•	•	•	•	•
Not belong (61%)								
Norway (NO)	•	•	•	•	•	•	•	•
Not belong (47%)								
Poland (PL)	•	•	•	•	•	•	•	•
Catholic (90%)								
Portugal (PT)	•	•	•	•	•	•	•	•
Catholic (80%)								
Romania (RO)				•				
Orthodox (81%)								
Russia (RU)			•	•	•	•		•
Orthodox (46%)								
Slovakia (SK)		•	•	•	•	•		
Catholic (63%)								
Slovenia (SI)	•	•	•	•	•	•	•	•
Catholic 50%								
Spain (ES)	•	•	•	•	•	•	•	•
Catholic (65%)								
Sweden (SE)	•	•	•	•	•	•	•	•
Not belong (68%)								
Switzerland (CH)	•	•	•	•	•	•	•	•
Not belong (34%)								
Turkey (TR)		•		•				
Muslim (96%)								
Ukraine (UA)		•	•	•	•	•		
Orthodox (58%)								

Notes: Surveyed countries across the ESS waves are reported. First column displays countries and their alphabetical codes. Dominant denomination and its percentage among respondents are presented below each country.

Table A2. Summary Statistics

	obs.	mean	sd	min	max
<i>creative</i>	284,142	0.68	0.24	0	1
<i>different</i>	284,142	0.60	0.27	0	1
<i>free</i>	284,142	0.76	0.22	0	1
<i>adventurous</i>	284,142	0.42	0.29	0	1
<i>traditions</i>	284,142	0.67	0.27	0	1
<i>rules</i>	284,142	0.58	0.27	0	1
<i>inposav</i>	284,142	0.62	0.18	0	1
<i>innegav</i>	284,142	0.62	0.22	0	1
<i>innosum</i>	284,142	0.54	0.14	0	1
<i>belonging</i>	284,142	0.62	0.48	0	1
<i>belonginge</i>	107,060	0.25	0.43	0	1
<i>denomination</i>	284,142	5.11	4.06	1	10
<i>denominatone</i>	25,829	1.66	1.12	1	8
<i>religiosity</i>	284,142	0.37	0.27	0	1
<i>iv_religiosity</i>	284,142	0.47	0.23	0	1
<i>degree</i>	284,142	0.48	0.30	0	1
<i>iv_degree</i>	284,142	0.48	0.18	0	1
<i>attendance</i>	284,142	0.27	0.25	0	1
<i>iv_attendance</i>	284,142	0.35	0.19	0	1
<i>pray</i>	284,142	0.40	0.41	0	1
<i>iv_pray</i>	284,142	0.40	0.24	0	1
<i>age</i>	284,142	47.63	18.29	15	114
<i>age2</i>	284,142	2603	1818	225	12996
<i>education</i>	282,293	12.42	4.09	0	56
<i>gender</i>	284,087	0.53	0.50	0	1
<i>health</i>	284,142	3.78	0.91	1	5
<i>paidwork</i>	284,142	0.55	0.50	0	1
<i>pwbefore</i>	284,142	0.90	0.29	0	1
<i>child</i>	283,749	0.38	0.49	0	1
<i>partner</i>	284,142	0.53	0.50	0	1
<i>bornc</i>	283,964	0.92	0.27	0	1
<i>fathere</i>	284,142	2.55	1.44	1	5
<i>mothere</i>	284,142	2.37	1.37	1	5
<i>essround</i>	284,142	4.58	2.17	1	8
<i>occupation</i>	253,195	4.74	2.50	0	9
<i>income1</i>	68,04	6.17	2.65	1	12
<i>income2</i>	150,443	5.30	2.78	1	10
<i>lgbt</i>	273,645	0.69	0.30	0	1
<i>womenwork</i>	114,915	0.55	0.29	0	1
<i>menwork</i>	148,614	0.35	0.31	0	1
<i>devote</i>	283,564	0.81	0.18	0	1
<i>success</i>	283,209	0.57	0.27	0	1
<i>lookafter</i>	32,966	0.36	0.28	0	1
<i>family</i>	33,114	0.80	0.19	0	1
<i>unemployed</i>	283,170	0.26	0.44	0	1
<i>trust</i>	283,516	0.50	0.25	0	1
<i>safe</i>	283,522	0.74	0.24	0	1
<i>securejob</i>	62,677	0.84	0.20	0	1
<i>government</i>	282,017	4.69	1.18	1	6
<i>N</i>	284,142				

Notes: Summary statistics for all variables are reported. All are weighted by gweight.

Table A3. Means by Country

	<i>religiosity</i>	<i>degree</i>	<i>attendance</i>	<i>pray</i>	<i>creative</i>	<i>different</i>	<i>free</i>	<i>advent.</i>	<i>trad.</i>	<i>rules</i>	<i>inposav</i>	<i>innegav</i>	<i>inmosum</i>
AL	0.49	0.73	0.24	0.50	0.70	0.69	0.76	0.42	0.77	0.67	0.64	0.72	0.53
AT	0.41	0.49	0.30	0.43	0.71	0.60	0.79	0.44	0.65	0.52	0.64	0.59	0.56
BE	0.32	0.47	0.17	0.30	0.67	0.63	0.77	0.43	0.66	0.56	0.62	0.61	0.55
BG	0.34	0.43	0.27	0.32	0.59	0.60	0.71	0.47	0.76	0.67	0.59	0.71	0.49
CH	0.40	0.51	0.25	0.45	0.74	0.64	0.84	0.42	0.63	0.50	0.66	0.57	0.59
CY	0.60	0.69	0.44	0.70	0.78	0.63	0.80	0.47	0.81	0.55	0.67	0.68	0.55
CZ	0.18	0.24	0.14	0.15	0.68	0.58	0.73	0.41	0.64	0.63	0.60	0.64	0.52
DE	0.34	0.44	0.23	0.36	0.70	0.60	0.80	0.35	0.62	0.51	0.61	0.57	0.55
DK	0.27	0.41	0.19	0.22	0.73	0.58	0.77	0.48	0.66	0.61	0.64	0.63	0.55
EE	0.24	0.35	0.19	0.19	0.60	0.59	0.75	0.41	0.59	0.53	0.59	0.56	0.54
ES	0.36	0.44	0.25	0.39	0.71	0.61	0.78	0.40	0.66	0.58	0.62	0.62	0.54
FI	0.37	0.51	0.21	0.38	0.67	0.63	0.76	0.43	0.59	0.60	0.62	0.60	0.55
FR	0.28	0.41	0.17	0.26	0.68	0.61	0.69	0.36	0.54	0.43	0.59	0.49	0.56
GB	0.31	0.40	0.20	0.33	0.69	0.62	0.76	0.46	0.62	0.53	0.63	0.58	0.56
GR	0.60	0.68	0.42	0.70	0.74	0.65	0.79	0.48	0.81	0.64	0.67	0.73	0.54
HR	0.55	0.62	0.42	0.62	0.64	0.51	0.72	0.33	0.74	0.55	0.55	0.64	0.49
HU	0.31	0.41	0.20	0.33	0.71	0.65	0.79	0.42	0.70	0.51	0.64	0.61	0.56
IE	0.52	0.52	0.43	0.63	0.71	0.64	0.77	0.49	0.67	0.57	0.65	0.62	0.56
IS	0.36	0.53	0.18	0.36	0.66	0.58	0.76	0.50	0.52	0.49	0.62	0.51	0.58
IT	0.50	0.59	0.37	0.54	0.69	0.66	0.75	0.39	0.75	0.65	0.62	0.70	0.52
LT	0.43	0.54	0.35	0.40	0.58	0.52	0.69	0.45	0.67	0.52	0.56	0.60	0.51
LU	0.33	0.43	0.25	0.31	0.70	0.66	0.74	0.40	0.64	0.60	0.62	0.62	0.54
LV	0.31	0.38	0.23	0.31	0.68	0.68	0.82	0.52	0.71	0.53	0.67	0.62	0.58
NL	0.33	0.46	0.19	0.33	0.71	0.63	0.79	0.47	0.63	0.60	0.65	0.62	0.56
NO	0.28	0.38	0.19	0.25	0.69	0.54	0.70	0.45	0.61	0.64	0.60	0.63	0.52
PL	0.62	0.64	0.53	0.71	0.63	0.62	0.77	0.41	0.77	0.71	0.61	0.74	0.50
PT	0.48	0.54	0.34	0.56	0.65	0.57	0.70	0.40	0.64	0.55	0.58	0.59	0.52
RO	0.58	0.68	0.39	0.67	0.66	0.61	0.67	0.43	0.74	0.70	0.59	0.72	0.49
RU	0.32	0.44	0.21	0.31	0.63	0.55	0.73	0.44	0.70	0.58	0.59	0.64	0.51
SE	0.24	0.33	0.18	0.20	0.70	0.57	0.73	0.43	0.57	0.52	0.61	0.54	0.56
SI	0.36	0.47	0.29	0.32	0.71	0.70	0.81	0.46	0.69	0.58	0.67	0.64	0.57
SK	0.49	0.59	0.37	0.52	0.67	0.62	0.73	0.40	0.71	0.67	0.60	0.69	0.51
TR	0.66	0.70	0.40	0.87	0.71	0.68	0.76	0.48	0.81	0.75	0.65	0.78	0.52
UA	0.43	0.50	0.32	0.49	0.55	0.53	0.69	0.37	0.69	0.61	0.53	0.65	0.48
XK	0.56	0.73	0.33	0.62	0.82	0.77	0.83	0.52	0.85	0.73	0.73	0.79	0.57
Total	0.37	0.48	0.27	0.40	0.68	0.60	0.76	0.42	0.67	0.58	0.62	0.62	0.54

Notes: Country means of dependent and independent variables are reported. All are weighted by gweight. See Appendix A for definitions.

B. First Stage Results

Table B1. Religiosity and Innovation Attitudes: First Stage Results of Table 4

	(1) <i>religiosity</i>	(2) <i>religiosity</i>	(3) <i>religiosity</i>	(4) <i>degree</i>	(5) <i>degree</i>	(6) <i>attendance</i>	(7) <i>attendance</i>	(8) <i>pray</i>	(9) <i>pray</i>
<i>iv_religiosity</i>	0.582*** (0.043)	0.596*** (0.043)	0.600*** (0.047)						
<i>iv_degree</i>				0.644*** (0.070)	0.648*** (0.073)				
<i>iv_attendance</i>						0.775*** (0.065)	0.784*** (0.068)		
<i>iv_pray</i>								0.719*** (0.052)	0.707*** (0.056)
<i>gender</i>	0.007 (0.004)	0.005 (0.004)	0.004 (0.004)	0.013 (0.005)	0.012 (0.005)	-0.005 (0.004)	-0.007 (0.004)	0.022*** (0.007)	0.022*** (0.008)
<i>age</i>	-0.000 (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.001)	-0.001** (0.000)	-0.002*** (0.000)	-0.002*** (0.001)	-0.003*** (0.001)
<i>age2</i>	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.001)	0.001 (0.001)
<i>paidwork</i>		-0.003 (0.002)	-0.002 (0.002)	-0.002 (0.002)	0.000 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.007 (0.004)	-0.005 (0.004)
<i>pubbefore</i>		-0.276** (0.133)	-0.384*** (0.085)	-0.280 (0.204)	-0.450*** (0.132)	-0.315 (0.131)	-0.424*** (0.084)	-0.218*** (0.062)	-0.241*** (0.064)
<i>partner</i>		0.018*** (0.002)	0.017*** (0.002)	0.016*** (0.002)	0.015*** (0.002)	0.024*** (0.002)	0.023*** (0.002)	0.011 (0.004)	0.010 (0.004)
<i>health</i>		0.003*** (0.001)	0.002** (0.001)	0.003* (0.001)	0.002 (0.001)	0.009*** (0.001)	0.009*** (0.001)	-0.006*** (0.002)	-0.008*** (0.002)
<i>child</i>		0.015*** (0.002)	0.015*** (0.002)	0.013*** (0.002)	0.014*** (0.002)	0.011*** (0.002)	0.010*** (0.002)	0.023*** (0.003)	0.023*** (0.003)
<i>bornc</i>			-0.069*** (0.004)		-0.061*** (0.004)		-0.049*** (0.005)		-0.108*** (0.007)
<i>fathere</i>			0.000 (0.001)		-0.001 (0.001)		0.001 (0.001)		0.002 (0.001)
<i>mothere</i>			-0.001 (0.001)		0.001 (0.001)		0.000 (0.001)		-0.005*** (0.001)
<i>N</i>	234,528	228,327	202,300	231,306	204,788	231,733	205,176	229,606	203,384

Notes: The first stage results are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table B2. Religiosity and Innovation Attitudes: First Stage Results of Table 5

	(1) <i>religiosity</i>	(2) <i>religiosity</i>	(3) <i>religiosity</i>	(4) <i>degree</i>	(5) <i>degree</i>	(6) <i>attendance</i>	(7) <i>attendance</i>	(8) <i>pray</i>	(9) <i>pray</i>
<i>iv_religiosity</i>	0.581*** (0.043)	0.595*** (0.043)	0.600*** (0.047)						
<i>iv_degree</i>				0.644*** (0.070)	0.648*** (0.073)				
<i>iv_attendance</i>						0.775*** (0.065)	0.784*** (0.068)		
<i>iv_pray</i>								0.718*** (0.052)	0.707*** (0.056)
<i>gender</i>	0.007 (0.004)	0.005 (0.004)	0.004 (0.004)	0.013 (0.005)	0.012 (0.005)	-0.005 (0.004)	-0.007 (0.004)	0.022*** (0.007)	0.022*** (0.008)
<i>age</i>	-0.000 (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.001** (0.000)	-0.002*** (0.000)	-0.002*** (0.001)	-0.003*** (0.001)
<i>age2</i>	0.000 (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.001)	0.001 (0.001)
<i>paidwork</i>		-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)	0.000 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.007 (0.004)	-0.005 (0.004)
<i>pubbefore</i>		-0.276** (0.133)	-0.384*** (0.085)	-0.280 (0.204)	-0.450*** (0.132)	-0.314 (0.131)	-0.424*** (0.084)	-0.218*** (0.062)	-0.241*** (0.064)
<i>partner</i>		0.018*** (0.002)	0.016*** (0.002)	0.017*** (0.002)	0.014*** (0.002)	0.024*** (0.002)	0.023*** (0.002)	0.011 (0.004)	0.010 (0.004)
<i>health</i>		0.002 (0.001)	0.003*** (0.001)	0.002* (0.001)	0.002 (0.001)	0.009*** (0.001)	0.009*** (0.001)	-0.006*** (0.002)	-0.008*** (0.002)
<i>child</i>		0.015*** (0.002)	0.016*** (0.002)	0.015*** (0.002)	0.014*** (0.002)	0.012*** (0.002)	0.010*** (0.002)	0.023*** (0.003)	0.023*** (0.003)
<i>bornc</i>			-0.069*** (0.004)		-0.069*** (0.004)		-0.049*** (0.005)		-0.108*** (0.007)
<i>fathere</i>			0.000 (0.001)		-0.001 (0.001)		0.001 (0.001)		0.002 (0.001)
<i>mothere</i>			-0.001 (0.001)		-0.001 (0.001)		0.000 (0.001)		-0.005*** (0.001)
<i>N</i>	234,387	228,192	202,191	231,156	204,668	231,582	205,055	229,462	203,268

Notes: The first stage results are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the following fixed effects: country, survey year, country-survey year, religious denomination, occupation, and income level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

C. Detailed OLS Results

Table C1. OLS Estimates: Religiosity and Creativity

<i>creative</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.035*** (0.005)	0.035*** (0.005)	0.033*** (0.005)						
<i>degree</i>				0.032*** (0.004)	0.031*** (0.004)				
<i>attendance</i>						0.005 (0.005)	0.002 (0.005)		
<i>pray</i>								0.025*** (0.004)	0.024*** (0.004)
<i>gender</i>	-0.015*** (0.003)	-0.014*** (0.003)	-0.015*** (0.003)	-0.013*** (0.003)	-0.015*** (0.003)	-0.011*** (0.003)	-0.013*** (0.003)	-0.014*** (0.003)	-0.016*** (0.003)
<i>age</i>	-0.001 (0.001)	-0.000 (0.001)	0.001 (0.001)	-0.000 (0.001)	0.001 (0.001)	-0.000 (0.001)	0.001 (0.001)	-0.000 (0.001)	0.001 (0.001)
<i>age2</i>	-0.000 (0.000)	-0.000* (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000** (0.000)	-0.000* (0.000)	-0.000*** (0.000)
<i>education</i>	0.005*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)
<i>paidwork</i>		-0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)	0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)
<i>pwbefore</i>		-0.252** (0.103)	-0.193* (0.106)	-0.253** (0.105)	-0.192* (0.108)	-0.261*** (0.101)	-0.205* (0.105)	-0.257** (0.101)	-0.200* (0.104)
<i>partner</i>		0.004* (0.003)	0.005* (0.003)	0.005* (0.003)	0.005** (0.003)	0.005* (0.002)	0.005** (0.003)	0.005* (0.003)	0.005** (0.003)
<i>health</i>		0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.001)	0.014*** (0.002)	0.014*** (0.001)
<i>child</i>		-0.010*** (0.002)	-0.013*** (0.002)	-0.010*** (0.002)	-0.013*** (0.002)	-0.010*** (0.002)	-0.012*** (0.002)	-0.011*** (0.002)	-0.013*** (0.002)
<i>bornnc</i>			-0.011** (0.004)		-0.011*** (0.004)		-0.013*** (0.004)		-0.011** (0.004)
<i>fathere</i>			0.007*** (0.001)		0.008*** (0.001)		0.008*** (0.001)		0.007*** (0.001)
<i>mothere</i>			0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)
<i>cons</i>	0.654*** (0.013)	0.843*** (0.104)	0.757*** (0.107)	0.839*** (0.106)	0.750*** (0.109)	0.859*** (0.101)	0.779*** (0.106)	0.848*** (0.101)	0.764*** (0.106)
<i>N</i>	233,137	226,965	201,219	229,884	203,663	230,311	204,054	228,212	202,280
<i>Adj. R²</i>	0.076	0.079	0.081	0.080	0.082	0.080	0.082	0.079	0.081

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C2. OLS Estimates: Religiosity and Being Different

<i>different</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.034*** (0.006)	0.036*** (0.006)	0.032*** (0.006)						
<i>degree</i>				0.032*** (0.004)	0.029*** (0.005)				
<i>attendance</i>						0.015*** (0.005)	0.013** (0.006)		
<i>pray</i>								0.021*** (0.004)	0.018*** (0.004)
<i>gender</i>	0.002 (0.003)	0.006* (0.003)	0.007** (0.003)	0.006* (0.003)	0.007** (0.003)	0.008** (0.003)	0.008*** (0.003)	0.006* (0.003)	0.007** (0.003)
<i>age</i>	-0.006*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)
<i>education</i>	0.003*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
<i>paidwork</i>		-0.002 (0.003)	-0.000 (0.003)	-0.002 (0.003)	-0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)	-0.002 (0.003)	0.000 (0.003)
<i>pwbefore</i>		-0.124 (0.094)	-0.181* (0.106)	-0.126 (0.094)	-0.181* (0.108)	-0.132 (0.095)	-0.189* (0.106)	-0.130 (0.095)	-0.189* (0.105)
<i>partner</i>		-0.004 (0.003)	-0.004 (0.003)	-0.005 (0.003)	-0.004 (0.003)	-0.005 (0.003)	-0.004 (0.003)	-0.004 (0.003)	-0.003 (0.003)
<i>health</i>		0.022*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)
<i>child</i>		-0.020*** (0.003)	-0.022*** (0.003)	-0.020*** (0.003)	-0.022*** (0.003)	-0.019*** (0.003)	-0.021*** (0.003)	-0.020*** (0.003)	-0.022*** (0.003)
<i>bornnc</i>			-0.016*** (0.004)		-0.016*** (0.004)		-0.018*** (0.004)		-0.017*** (0.004)
<i>fathere</i>			0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)
<i>mothere</i>			0.002* (0.001)		0.003** (0.001)		0.003** (0.001)		0.002* (0.001)
<i>cons</i>	0.763*** (0.014)	0.773*** (0.095)	0.822*** (0.107)	0.768*** (0.095)	0.814*** (0.109)	0.784*** (0.096)	0.835*** (0.107)	0.781*** (0.095)	0.833*** (0.106)
<i>N</i>	233,367	227,211	201,405	230,121	203,840	230,543	204,225	228,456	202,464
<i>Adj. R²</i>	0.074	0.079	0.081	0.080	0.082	0.080	0.081	0.080	0.081

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C3. OLS Estimates: Religiosity and Being Free

<i>free</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	-0.044*** (0.005)	-0.038*** (0.005)	-0.041*** (0.005)						
<i>degree</i>				-0.024*** (0.004)	-0.024*** (0.004)				
<i>attendance</i>						-0.048*** (0.004)	-0.050*** (0.005)		
<i>pray</i>								-0.008*** (0.003)	-0.010*** (0.003)
<i>gender</i>	-0.009*** (0.003)	-0.011*** (0.003)	-0.011*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)
<i>age</i>	-0.003*** (0.001)	0.000 (0.001)	0.001* (0.001)	0.001 (0.001)	0.001* (0.001)	0.000 (0.001)	0.001* (0.001)	0.001 (0.001)	0.001** (0.001)
<i>age2</i>	0.000*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
<i>education</i>	0.005*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)
<i>paidwork</i>		-0.003 (0.003)	0.000 (0.003)	-0.002 (0.003)	0.001 (0.003)	-0.002 (0.003)	0.000 (0.003)	-0.002 (0.003)	0.000 (0.003)
<i>pwbefore</i>		0.044 (0.144)	0.143 (0.130)	0.046 (0.145)	0.145 (0.129)	0.037 (0.143)	0.134 (0.129)	0.053 (0.146)	0.155 (0.128)
<i>partner</i>		-0.041*** (0.003)	-0.040*** (0.003)	-0.042*** (0.003)	-0.040*** (0.003)	-0.041*** (0.003)	-0.040*** (0.003)	-0.042*** (0.003)	-0.040*** (0.003)
<i>health</i>		0.010*** (0.001)	0.010*** (0.002)	0.011*** (0.001)	0.011*** (0.002)	0.011*** (0.001)	0.011*** (0.002)	0.011*** (0.001)	0.011*** (0.002)
<i>child</i>		-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.018*** (0.003)
<i>bornc</i>			-0.005 (0.004)		-0.004 (0.004)		-0.005 (0.004)		-0.003 (0.004)
<i>fathere</i>			0.003*** (0.001)		0.003*** (0.001)		0.003*** (0.001)		0.003*** (0.001)
<i>mothere</i>			0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)
<i>cons</i>	0.791*** (0.012)	0.667*** (0.145)	0.543*** (0.131)	0.662*** (0.146)	0.535*** (0.130)	0.669*** (0.144)	0.546*** (0.130)	0.648*** (0.147)	0.517*** (0.129)
<i>N</i>	233,529	227,361	201,521	230,303	203,980	230,708	204,350	228,610	202,580
<i>Adj. R²</i>	0.063	0.073	0.075	0.072	0.075	0.073	0.076	0.072	0.074

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C4. OLS Estimates: Religiosity and Being Adventurous

<i>adventurous</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	-0.017*** (0.006)	-0.011* (0.006)	-0.010 (0.006)						
<i>degree</i>				-0.002 (0.005)	-0.003 (0.005)				
<i>attendance</i>						-0.013** (0.006)	-0.011* (0.006)		
<i>pray</i>								-0.006* (0.003)	-0.006 (0.004)
<i>gender</i>	-0.075*** (0.003)	-0.072*** (0.003)	-0.071*** (0.003)	-0.072*** (0.003)	-0.071*** (0.003)	-0.072*** (0.003)	-0.072*** (0.003)	-0.072*** (0.003)	-0.071*** (0.003)
<i>age</i>	-0.012*** (0.000)	-0.008*** (0.000)	-0.007*** (0.000)	-0.008*** (0.000)	-0.007*** (0.000)	-0.008*** (0.000)	-0.007*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.002*** (0.000)	0.002*** (0.000)	0.001* (0.000)	0.002*** (0.000)	0.001** (0.000)	0.002*** (0.000)	0.001* (0.000)	0.002*** (0.000)	0.001* (0.000)
<i>paidwork</i>		0.002 (0.003)	0.004 (0.003)	0.003 (0.003)	0.004 (0.003)	0.002 (0.003)	0.004 (0.003)	0.002 (0.003)	0.004 (0.003)
<i>pwbefore</i>		-0.124 (0.117)	-0.045 (0.117)	-0.124 (0.118)	-0.045 (0.116)	-0.127 (0.118)	-0.048 (0.116)	-0.122 (0.118)	-0.042 (0.115)
<i>partner</i>		-0.037*** (0.003)	-0.035*** (0.003)	-0.037*** (0.003)	-0.035*** (0.003)	-0.037*** (0.003)	-0.035*** (0.003)	-0.037*** (0.003)	-0.035*** (0.003)
<i>health</i>		0.020*** (0.001)	0.019*** (0.001)	0.021*** (0.001)	0.019*** (0.001)	0.021*** (0.001)	0.019*** (0.001)	0.021*** (0.001)	0.019*** (0.001)
<i>child</i>		-0.028*** (0.003)	-0.029*** (0.003)	-0.029*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)
<i>bornc</i>			-0.007* (0.004)		-0.007* (0.004)		-0.007* (0.004)		-0.007* (0.004)
<i>fathere</i>			0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)
<i>mothere</i>			0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)
<i>cons</i>	0.797*** (0.013)	0.791*** (0.119)	0.678*** (0.119)	0.785*** (0.119)	0.673*** (0.118)	0.789*** (0.119)	0.676*** (0.118)	0.786*** (0.119)	0.672*** (0.117)
<i>N</i>	233,274	227,128	201,318	230,048	203,764	230,469	204,149	228,378	202,380
<i>Adj. R²</i>	0.163	0.173	0.175	0.172	0.174	0.172	0.174	0.172	0.174

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C5. OLS Estimates: Religiosity and Following Traditions

<i>traditions</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.278*** (0.006)	0.274*** (0.006)	0.274*** (0.006)						
<i>degree</i>				0.214*** (0.006)	0.212*** (0.006)				
<i>attendance</i>						0.191*** (0.006)	0.191*** (0.006)		
<i>pray</i>								0.120*** (0.003)	0.120*** (0.003)
<i>gender</i>	0.001 (0.003)	0.002 (0.003)	0.002 (0.003)	0.006** (0.003)	0.006* (0.003)	0.014*** (0.003)	0.013*** (0.003)	0.005 (0.003)	0.004 (0.003)
<i>age</i>	0.002*** (0.000)	0.001* (0.000)	0.001 (0.001)	0.001* (0.000)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)
<i>age2</i>	-0.000 (0.000)	0.000* (0.000)	0.000* (0.000)	0.000** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000** (0.000)
<i>education</i>	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>		-0.001 (0.003)	0.002 (0.003)	-0.001 (0.003)	0.001 (0.003)	-0.002 (0.003)	0.001 (0.003)	-0.001 (0.003)	0.001 (0.003)
<i>pwbefore</i>		-0.184*** (0.038)	-0.142*** (0.022)	-0.200*** (0.042)	-0.154*** (0.020)	-0.201*** (0.030)	-0.166*** (0.022)	-0.233*** (0.028)	-0.218*** (0.034)
<i>partner</i>		0.027*** (0.002)	0.024*** (0.002)	0.029*** (0.002)	0.026*** (0.002)	0.027*** (0.002)	0.024*** (0.002)	0.031*** (0.002)	0.027*** (0.002)
<i>health</i>		0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.006*** (0.001)	0.005*** (0.001)
<i>child</i>		0.002 (0.002)	0.001 (0.003)	0.004 (0.002)	0.003 (0.003)	0.005* (0.002)	0.004 (0.003)	0.004* (0.002)	0.003 (0.002)
<i>bornc</i>			-0.019*** (0.004)		-0.026*** (0.004)		-0.029*** (0.004)		-0.025*** (0.004)
<i>fathere</i>			-0.000 (0.001)		0.000 (0.001)		-0.000 (0.001)		-0.000 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>cons</i>	0.490*** (0.012)	0.676*** (0.040)	0.678*** (0.024)	0.682*** (0.044)	0.687*** (0.023)	0.738*** (0.032)	0.759*** (0.023)	0.771*** (0.028)	0.807*** (0.033)
<i>N</i>	233,556	227,390	201,539	230,321	203,994	230,747	204,378	228,640	202,599
<i>Adj. R²</i>	0.229	0.230	0.232	0.225	0.226	0.212	0.214	0.213	0.214

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C6. OLS Estimates: Religiosity and Following Rules

<i>rules</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.081*** (0.006)	0.079*** (0.006)	0.075*** (0.006)						
<i>degree</i>				0.051*** (0.006)	0.048*** (0.006)				
<i>attendance</i>						0.068*** (0.006)	0.066*** (0.006)		
<i>pray</i>								0.033*** (0.004)	0.031*** (0.004)
<i>gender</i>	-0.019*** (0.003)	-0.018*** (0.003)	-0.017*** (0.003)	-0.015*** (0.003)	-0.015*** (0.003)	-0.014*** (0.003)	-0.014*** (0.003)	-0.016*** (0.003)	-0.016*** (0.003)
<i>age</i>	0.000 (0.000)	-0.001** (0.001)	-0.002*** (0.001)	-0.001** (0.001)	-0.002*** (0.001)	-0.001** (0.001)	-0.002*** (0.001)	-0.001** (0.001)	-0.002*** (0.001)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.003*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)
<i>paidwork</i>		0.006** (0.003)	0.007** (0.003)	0.006** (0.003)	0.007** (0.003)	0.006** (0.003)	0.007** (0.003)	0.006** (0.003)	0.007** (0.003)
<i>pwbefore</i>		-0.302*** (0.046)	-0.273*** (0.057)	-0.310*** (0.047)	-0.280*** (0.058)	-0.304*** (0.047)	-0.274*** (0.058)	-0.317*** (0.045)	-0.295*** (0.057)
<i>partner</i>		0.023*** (0.002)	0.022*** (0.002)	0.023*** (0.002)	0.022*** (0.002)	0.022*** (0.002)	0.021*** (0.002)	0.024*** (0.002)	0.023*** (0.003)
<i>health</i>		0.003** (0.001)	0.003** (0.001)	0.003** (0.001)	0.004** (0.001)	0.003* (0.001)	0.003** (0.001)	0.004** (0.001)	0.004*** (0.001)
<i>child</i>		0.003 (0.002)	-0.000 (0.003)	0.004 (0.003)	0.000 (0.003)	0.004 (0.002)	0.001 (0.003)	0.004 (0.003)	0.000 (0.003)
<i>bornc</i>			-0.043*** (0.005)		-0.045*** (0.005)		-0.045*** (0.005)		-0.045*** (0.005)
<i>fathere</i>			-0.001 (0.001)		0.001 (0.001)		-0.001 (0.001)		-0.002 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>cons</i>	0.529*** (0.012)	0.833*** (0.047)	0.870*** (0.058)	0.844*** (0.048)	0.884*** (0.060)	0.844*** (0.048)	0.884*** (0.059)	0.862*** (0.046)	0.907*** (0.059)
<i>N</i>	232,561	226,405	200,723	229,292	203,139	229,694	203,509	227,635	201,768
<i>Adj. R²</i>	0.130	0.132	0.138	0.130	0.135	0.131	0.137	0.130	0.136

Notes: Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C7. Summary Statistics for Religious Belonging

	obs	mean	std dev	min	max
Believers					
<i>religiosity</i>	221,530	0.51	0.24	0	1
<i>degree</i>	220,275	0.61	0.23	0	1
<i>attendance</i>	220,549	0.36	0.25	0	1
<i>pray</i>	216,972	0.57	0.39	0	1
Never-believers					
<i>religiosity</i>	107,004	0.14	0.16	0	1
<i>degree</i>	105,964	0.22	0.24	0	1
<i>attendance</i>	106,424	0.08	0.14	0	1
<i>pray</i>	105,700	0.10	0.23	0	1
Once-believers					
<i>religiosity</i>	32,922	0.18	0.19	0	1
<i>degree</i>	32,744	0.29	0.26	0	1
<i>attendance</i>	32,872	0.10	0.15	0	1
<i>pray</i>	32,641	0.16	0.29	0	1

Table C8. Disaggregate OLS Results for Religious Belonging

	Believers		Never-believers		Once-believers	
	(<i>inposav</i>)	(<i>innegav</i>)	(<i>inposav</i>)	(<i>innegav</i>)	(<i>inposav</i>)	(<i>innegav</i>)
<i>religiosity</i>	-0.009** (0.004)	0.177*** (0.005)	0.049*** (0.009)	0.194*** (0.014)	0.046*** (0.010)	0.176*** (0.014)
	121,761	121,710	59,187	59,137	20,968	20,961
	0.155	0.185	0.137	0.164	0.130	0.152
<i>degree</i>	0.000 (0.004)	0.154*** (0.005)	0.020*** (0.006)	0.102*** (0.009)	0.021*** (0.007)	0.103*** (0.010)
	123,543	123,486	59,758	59,703	21,075	21,068
	0.156	0.178	0.137	0.160	0.129	0.149
<i>attendance</i>	-0.019*** (0.004)	0.123*** (0.006)	0.033*** (0.011)	0.178*** (0.014)	0.024** (0.012)	0.177*** (0.018)
	123,722	123,662	59,906	59,854	21,129	21,122
	0.156	0.170	0.138	0.159	0.128	0.147
<i>pray</i>	0.000 (0.003)	0.081*** (0.003)	0.038*** (0.007)	0.075*** (0.009)	0.030*** (0.006)	0.055*** (0.009)
	122,329	122,273	59,611	59,559	21,043	21,036
	0.155	0.171	0.138	0.153	0.130	0.141

Notes: OLS estimates for alternative measures of religiosity are reported. Robust standard errors clustered at the level of instruments are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

D. Detailed IV Results

Table D1. IV Estimates: Religiosity and Creativity

<i>creative</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	-0.142 (0.094)	-0.105 (0.088)	-0.173* (0.090)						
<i>degree</i>				-0.070 (0.132)	-0.171 (0.133)				
<i>attendance</i>						-0.126* (0.074)	-0.174** (0.074)		
<i>pray</i>								-0.023 (0.056)	-0.071 (0.057)
<i>gender</i>	-0.005 (0.006)	-0.005 (0.005)	-0.003 (0.005)	-0.007 (0.007)	-0.003 (0.007)	-0.008** (0.003)	-0.009*** (0.003)	-0.009 (0.006)	-0.005 (0.006)
<i>age</i>	-0.001 (0.001)	-0.000 (0.001)	0.000 (0.001)	-0.000 (0.001)	0.000 (0.001)	-0.000 (0.001)	0.000 (0.001)	-0.000 (0.001)	0.001 (0.001)
<i>age2</i>	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)
<i>education</i>	0.005*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)
<i>paidwork</i>		-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.000 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
<i>pwbefore</i>		-0.291*** (0.094)	-0.273*** (0.103)	-0.281*** (0.098)	-0.283*** (0.109)	-0.301*** (0.091)	-0.281*** (0.101)	-0.267*** (0.101)	-0.224** (0.106)
<i>partner</i>		0.007** (0.003)	0.008*** (0.003)	0.006* (0.003)	0.008** (0.003)	0.007** (0.003)	0.009*** (0.003)	0.005** (0.003)	0.006** (0.003)
<i>health</i>		0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.015*** (0.002)	0.015*** (0.002)	0.014*** (0.002)	0.014*** (0.001)
<i>child</i>		-0.008*** (0.003)	-0.009*** (0.003)	-0.009*** (0.003)	-0.010*** (0.003)	-0.008*** (0.002)	-0.010*** (0.003)	-0.010*** (0.003)	-0.011*** (0.003)
<i>bornnc</i>			-0.025*** (0.008)		-0.023** (0.009)		-0.022*** (0.006)		-0.021*** (0.007)
<i>fathere</i>			0.007*** (0.001)		0.007*** (0.001)		0.008*** (0.001)		0.008*** (0.001)
<i>mothere</i>			0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)		0.004*** (0.001)
<i>N</i>	233,137	226,965	201,219	229,884	203,663	230,311	204,054	228,212	202,280
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,325	2,363	2,141	825	749	2,707	2,499	2,017	1,731
<i>widstat</i>	186	192	159	79	74	141	130	195	162

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table D2. IV Estimates: Religiosity and Being Different

<i>different</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	-0.123 (0.090)	-0.097 (0.083)	-0.131 (0.081)						
<i>degree</i>				0.000 (0.108)	-0.069 (0.113)				
<i>attendance</i>						-0.106 (0.068)	-0.124* (0.066)		
<i>pray</i>								-0.070 (0.052)	-0.102** (0.051)
<i>gender</i>	0.012* (0.006)	0.014** (0.006)	0.016*** (0.006)	0.008 (0.007)	0.013* (0.007)	0.011*** (0.004)	0.012*** (0.004)	0.016** (0.007)	0.021*** (0.006)
<i>age</i>	-0.006*** (0.001)	-0.005*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.004*** (0.001)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.003*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
<i>paidwork</i>		-0.003 (0.003)	-0.001 (0.003)	-0.002 (0.003)	-0.000 (0.003)	-0.003 (0.003)	-0.001 (0.003)	-0.003 (0.003)	-0.001 (0.003)
<i>pwbefore</i>		-0.162 (0.102)	-0.245** (0.106)	-0.135 (0.100)	-0.225* (0.112)	-0.169* (0.103)	-0.248** (0.105)	-0.150 (0.099)	-0.220** (0.109)
<i>partner</i>		-0.002 (0.003)	-0.001 (0.003)	-0.004 (0.004)	-0.002 (0.004)	-0.002 (0.003)	-0.001 (0.003)	-0.003 (0.003)	-0.002 (0.003)
<i>health</i>		0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.024*** (0.002)	0.024*** (0.002)	0.022*** (0.002)	0.022*** (0.002)
<i>child</i>		-0.018*** (0.003)	-0.020*** (0.003)	-0.019*** (0.003)	-0.020*** (0.003)	-0.018*** (0.003)	-0.020*** (0.003)	-0.018*** (0.003)	-0.019*** (0.003)
<i>bornnc</i>			-0.027*** (0.008)		-0.022** (0.009)		-0.024*** (0.006)		-0.029*** (0.008)
<i>fathere</i>			0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)		0.005*** (0.001)
<i>mothere</i>			0.002* (0.001)		0.003** (0.001)		0.003** (0.001)		0.002 (0.001)
<i>N</i>	233,367	227,211	201,405	230,121	203,840	230,543	204,225	228,456	202,464
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,329	2,363	2,135	833	753	2,693	2,485	2,015	1,729
<i>widstat</i>	185	191	157	81	74	142	130	193	161

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table D3. IV Estimates: Religiosity and Being Free

<i>free</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	-0.304*** (0.071)	-0.332*** (0.071)	-0.366*** (0.068)						
<i>degree</i>				-0.289** (0.122)	-0.342*** (0.120)				
<i>attendance</i>						-0.325*** (0.060)	-0.338*** (0.060)		
<i>pray</i>								-0.177*** (0.046)	-0.208*** (0.045)
<i>gender</i>	0.006 (0.005)	0.007 (0.005)	0.008* (0.005)	0.004 (0.007)	0.006 (0.007)	-0.005 (0.003)	-0.006* (0.004)	0.007 (0.006)	0.009* (0.005)
<i>age</i>	-0.003*** (0.001)	-0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)	0.001 (0.001)
<i>age2</i>	0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
<i>education</i>	0.004*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)
<i>paidwork</i>		-0.004 (0.003)	-0.001 (0.003)	-0.003 (0.003)	0.000 (0.003)	-0.003 (0.003)	-0.001 (0.003)	-0.004 (0.003)	-0.001 (0.003)
<i>pwbefore</i>		-0.038 (0.131)	0.015 (0.148)	-0.029 (0.135)	0.002 (0.167)	-0.049 (0.131)	0.011 (0.146)	0.014 (0.140)	0.104 (0.126)
<i>partner</i>		-0.037*** (0.003)	-0.035*** (0.003)	-0.037*** (0.004)	-0.036*** (0.004)	-0.035*** (0.003)	-0.034*** (0.003)	-0.040*** (0.003)	-0.039*** (0.003)
<i>health</i>		0.011*** (0.001)	0.011*** (0.002)	0.012*** (0.002)	0.011*** (0.002)	0.014*** (0.002)	0.014*** (0.002)	0.009*** (0.002)	0.009*** (0.002)
<i>child</i>		-0.012*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)
<i>bornc</i>			-0.027*** (0.006)		-0.023*** (0.009)		-0.019*** (0.005)		-0.024*** (0.006)
<i>fathere</i>			0.003*** (0.001)		0.003*** (0.001)		0.003*** (0.001)		0.004*** (0.001)
<i>mothere</i>			0.004*** (0.001)		0.005*** (0.001)		0.005*** (0.001)		0.004*** (0.001)
<i>N</i>	233,529	227,361	201,521	230,303	203,980	230,708	204,350	228,610	202,580
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,333	2,374	2,148	836	761	2,729	2,522	2,018	1,731
<i>widstat</i>	187	193	159	84	78	143	130	192	159

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table D4. IV Estimates: Religiosity and Being Adventurous

<i>adventurous</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.004 (0.071)	-0.013 (0.062)	-0.074 (0.061)						
<i>degree</i>				0.051 (0.085)	-0.048 (0.085)				
<i>attendance</i>						-0.021 (0.057)	-0.057 (0.056)		
<i>pray</i>								-0.019 (0.041)	-0.066 (0.041)
<i>gender</i>	-0.076*** (0.005)	-0.072*** (0.005)	-0.068*** (0.005)	-0.075*** (0.006)	-0.069*** (0.006)	-0.072*** (0.003)	-0.071*** (0.003)	-0.071*** (0.006)	-0.065*** (0.006)
<i>age</i>	-0.012*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	0.002*** (0.000)	0.002*** (0.000)	0.001 (0.000)	0.002*** (0.000)	0.001 (0.000)	0.002*** (0.000)	0.001** (0.000)	0.002*** (0.000)	0.001* (0.000)
<i>paidwork</i>		0.002 (0.003)	0.003 (0.003)	0.003 (0.003)	0.004 (0.003)	0.002 (0.003)	0.004 (0.003)	0.002 (0.003)	0.003 (0.003)
<i>pwbefore</i>		-0.124 (0.118)	-0.071 (0.122)	-0.109 (0.124)	-0.065 (0.127)	-0.130 (0.118)	-0.068 (0.120)	-0.125 (0.117)	-0.058 (0.115)
<i>partner</i>		-0.037*** (0.003)	-0.034*** (0.003)	-0.038*** (0.003)	-0.035*** (0.003)	-0.037*** (0.003)	-0.034*** (0.003)	-0.037*** (0.003)	-0.035*** (0.003)
<i>health</i>		0.020*** (0.001)	0.019*** (0.001)	0.020*** (0.001)	0.019*** (0.001)	0.021*** (0.001)	0.020*** (0.001)	0.021*** (0.001)	0.019*** (0.001)
<i>child</i>		-0.028*** (0.003)	-0.028*** (0.003)	-0.029*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.028*** (0.003)	-0.027*** (0.003)
<i>bornc</i>			-0.011* (0.006)		-0.009 (0.006)		-0.009* (0.005)		-0.013** (0.006)
<i>fathere</i>			0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)
<i>mothere</i>			0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)		0.006*** (0.001)
<i>N</i>	233,274	227,128	201,318	230,048	203,764	230,469	204,149	228,378	202,380
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,323	2,367	2,141	846	765	2,706	2,500	2,011	1,726
<i>widstat</i>	187	194	160	82	77	143	131	192	159

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table D5. IV Estimates: Religiosity and Following Traditions

<i>traditions</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.120** (0.059)	0.162*** (0.055)	0.192*** (0.060)						
<i>degree</i>				0.215*** (0.074)	0.260*** (0.084)				
<i>attendance</i>						0.175*** (0.047)	0.187*** (0.050)		
<i>pray</i>								0.089** (0.041)	0.111** (0.045)
<i>gender</i>	0.010** (0.005)	0.009* (0.005)	0.007 (0.005)	0.006 (0.005)	0.003 (0.006)	0.014*** (0.003)	0.013*** (0.003)	0.008 (0.006)	0.005 (0.006)
<i>age</i>	0.002*** (0.000)	0.001 (0.001)	0.001 (0.001)	0.001* (0.000)	0.001 (0.001)	0.001 (0.000)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)
<i>age2</i>	-0.000 (0.000)	0.000** (0.000)	0.000** (0.000)	0.000** (0.000)	0.000* (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000** (0.000)
<i>education</i>	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
<i>paidwork</i>		-0.001 (0.003)	0.001 (0.003)	-0.001 (0.003)	0.001 (0.003)	-0.002 (0.003)	0.001 (0.003)	-0.001 (0.003)	0.001 (0.003)
<i>pwbefore</i>		-0.215*** (0.032)	-0.174*** (0.033)	-0.200*** (0.047)	-0.132*** (0.043)	-0.206*** (0.032)	-0.168*** (0.031)	-0.240*** (0.030)	-0.220*** (0.036)
<i>partner</i>		0.029*** (0.002)	0.025*** (0.003)	0.029*** (0.003)	0.025*** (0.003)	0.028*** (0.003)	0.024*** (0.003)	0.031*** (0.002)	0.027*** (0.002)
<i>health</i>		0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003** (0.001)	0.006*** (0.001)	0.005*** (0.001)
<i>child</i>		0.004 (0.003)	0.003 (0.003)	0.004 (0.003)	0.002 (0.003)	0.005** (0.002)	0.004 (0.003)	0.005* (0.003)	0.003 (0.003)
<i>bornc</i>			-0.025*** (0.006)		-0.023*** (0.007)		-0.029*** (0.005)		-0.026*** (0.006)
<i>fathere</i>			-0.000 (0.001)		0.000 (0.001)		-0.000 (0.001)		-0.000 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>N</i>	233,556	227,390	201,539	230,321	203,994	230,747	204,378	228,640	202,599
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,322	2,364	2,140	847	768	2,705	2,502	2,000	1,718
<i>widstat</i>	188	196	162	84	79	143	131	192	160

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table D6. IV Estimates: Religiosity and Following Rules

<i>rules</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>religiosity</i>	0.164** (0.066)	0.199*** (0.064)	0.225*** (0.070)						
<i>degree</i>				0.158* (0.090)	0.176* (0.098)				
<i>attendance</i>						0.240*** (0.059)	0.244*** (0.063)		
<i>pray</i>								0.090** (0.045)	0.123** (0.049)
<i>gender</i>	-0.024*** (0.005)	-0.025*** (0.005)	-0.025*** (0.005)	-0.022*** (0.006)	-0.023*** (0.007)	-0.018*** (0.003)	-0.018*** (0.004)	-0.023*** (0.006)	-0.026*** (0.006)
<i>age</i>	0.000 (0.000)	-0.001* (0.001)	-0.001** (0.001)	-0.001* (0.001)	-0.001** (0.001)	-0.001 (0.001)	-0.001** (0.001)	-0.001** (0.001)	-0.001** (0.001)
<i>age2</i>	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
<i>education</i>	-0.004*** (0.000)	-0.004*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)
<i>paidwork</i>		0.007** (0.003)	0.008*** (0.003)	0.006** (0.003)	0.007** (0.003)	0.007** (0.003)	0.008** (0.003)	0.006** (0.003)	0.008** (0.003)
<i>pwbefore</i>		-0.268*** (0.055)	-0.215*** (0.061)	-0.280*** (0.064)	-0.223*** (0.075)	-0.250*** (0.060)	-0.198*** (0.063)	-0.304*** (0.044)	-0.271*** (0.054)
<i>partner</i>		0.021*** (0.003)	0.020*** (0.003)	0.021*** (0.003)	0.020*** (0.003)	0.019*** (0.003)	0.017*** (0.003)	0.023*** (0.002)	0.022*** (0.003)
<i>health</i>		0.003* (0.001)	0.003** (0.001)	0.003* (0.001)	0.003** (0.001)	0.001 (0.002)	0.002 (0.002)	0.004*** (0.001)	0.005*** (0.002)
<i>child</i>		0.001 (0.003)	-0.003 (0.003)	0.002 (0.003)	-0.001 (0.003)	0.002 (0.003)	-0.002 (0.003)	0.002 (0.003)	-0.002 (0.003)
<i>bornc</i>			-0.033*** (0.006)		-0.037*** (0.007)		-0.036*** (0.005)		-0.035*** (0.007)
<i>fathere</i>			-0.001 (0.001)		-0.001 (0.001)		-0.002 (0.001)		-0.002 (0.001)
<i>mothere</i>			-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)		-0.004*** (0.001)
<i>N</i>	232,561	226,405	200,723	229,292	203,139	229,694	203,509	227,635	201,768
<i>idp</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>cdf</i>	2,313	2,356	2,140	811	747	2,723	2,515	2,007	1,727
<i>widstat</i>	188	195	160	80	75	143	130	194	164

Notes: Robust standard errors clustered at the instruments level are in parentheses. All regressions include the full set of fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Appendix: Workplace Skills as Regional Capabilities: Relatedness, Complexity, and Industrial Diversification

Table A1. ICP Categories

1.Knowledge	(B1) Administration and Management, (B2) Office Work, (B3) Economics and Accounting, (B4) Sales and Marketing, (B5) Services to Customers, (B6) Human Resources Management, (B7) Production and Processing, (B8) Food Production, (B9)IT and Electronics, (B10) Engineering and Technology, (B11) Technical Design, (B12) Building and Construction, (B13)Mechanical, (B14) Mathematics, (B15) Physics, (B16) Chemistry, (B17) Biology, (B18) Psychology, (B19) Sociology and Anthropology, (B20) Geography, (B21) Medicine and Dentistry, (B22) Therapy and Counseling, (B23) Education and Training, (B24) Italian Language, (B25) Foreign Language, (B26) Fine Arts, (B27)History and Archaeology, (B28) Philosophy and Theology, (B29) Civil Protection and Public Safety, (B30) Legislation and Institutions, (B31) Telecommunications, (B32) Communication and Media, (B33) Transportation
2.Skills	
2.1 Basic Skills	(C1) Reading Comprehension, (C2) Active Listening, (C3) Writing, (C4) Speaking, (C5) Mathematics, (C6) Science, (C7) Critical Thinking, (C8) Active Learning, (C9) Learning Strategies, (C10) Monitoring
2.2 Social Skills	(C11) Social Perceptiveness, (C12) Coordination, (C13) Persuasion, (C14) Negotiation, (C15) Instructing, (C16) Service Orientation
2.3 Complex Problem	(C17) Complex Problem Solving
2.4 Technical Skills	(C18) Operations Analysis, (C19) Technology Design, (C20) Equipment Selection, (C21) Installation, (C22) Programming, (C23) Quality Control Analysis, (C24) Operation Monitoring, (C25) Operation and Control, (C26) Equipment Maintenance, (C27) Troubleshooting, (C28) Repairing
2.5 Systems Skills	(C29) Systems Analysis, (C30) Systems Evaluation, (C31) Judgment and Decision Making
2.6 Resource Management Skills	(C32) Time Management, (C33) Management of Financial Resources, (C34) Management of Material Resources, (C35) Management of Personnel Resources
3.Attitudes	
3.1 Cognitive	(D1) Oral Comprehension, (D2) Written Comprehension, (D3) Oral Expression, (D4) Written Expression, (D5) Fluency of Ideas, (D6) Originality, (D7) Problem Sensitivity, (D8) Deductive Reasoning, (D9) Inductive Reasoning, (D10) Information Ordering, (D11) Category Flexibility, (D12) Math Reasoning, (D13) Number Facility, (D14) Memorisation, (D15) Speed of Closure, (D16) Flexibility of Closure, (D17) Perceptual Speed, (D18) Spatial Orientation, (D19) Visualisation, (D20) Selective Attention, (D21) Time Sharing
3.2 Psychomotor	(D22) Arms-Hand Steadiness, (D23) Manual Dexterity, (D24) Finger Dexterity, (D25) Control Precision, (D26) Multilimb Coordination, (D27) Response Orientation, (D28) Rate Control, (D29) Reaction Time, (D30) Wrist-Finger Speed, (D31) Speed of Limb Movement
3.3 Psychical	(D32) Static Strength, (D33) Explosive Strength, (D34) Dynamic Strength, (D35) Trunk Strength, (D36) Stamina, (D37) Extent Flexibility, (D38) Dynamic Flexibility, (D39) Gross Body Coordination, (D40) Gross Balance Body Equilibrium
3.4 Sensory	(D41) Near Vision, (D42) Far Vision, (D43) Visual Colour Discrimination, (D44) Night Vision, (D45) Peripheral Vision, (D46) Depth Perception, (D47) Glare Sensitivity, (D48) Hearing Sensitivity, (D49) Auditory Attention, (D50) Sound Localisation, (D51) Speech Recognition, (D52) Speech Clarity
4.Work Activities	
4.1 Information Input	(G1) Getting Information, (G2) Identifying Objects, Actions, and Events, (G3) Monitor Processes, Materials or Surroundings, (G4) Inspecting Equipment, Structures or Material, (G5) Estimate the Quantifiable Characteristics of Products, Events, or Information
4.2 Mental Process	(G6) Judging the Qualities of Things, Services or People, (G7) Evaluating Information to Determine Compliance with Standards, (G8) Processing Information, (G9) Analysing Data or Information, (G10) Making Decisions and Solving Problems, (G11) Thinking Creatively, (G12) Updating and Using Relevant Knowledge, (G13) Developing Objectives and Strategies, (G14) Scheduling Work and Activities, (G15) Organising, Planning, and Prioritising Work
4.3 Work Output	(G16) Performing General Physical Activities, (G17) Handling and Moving Objects, (G18) Controlling Machines and Processes, (G19) Interacting With Computers, (G20) Operating Vehicles, Mechanised Devices, or Equipment, (G21) Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment, (G22) Repairing and Maintaining Mechanical Equipment, (G23) Repairing and Maintaining Electronic Equipment, (G24) Documenting/Recording Information
4.4 Interacting With Others	(G25) Interpreting the Meaning of the Information for Others, (G26) Communicating with Supervisors, Peers, or Subordinates, (G27) Communicating with Persons Outside Organisation, (G28) Establishing and Maintaining Interpersonal Relationships, (G29) Assisting and Caring for Others, (G30) Selling or Influencing Others, (G31) Resolving Conflicts and Negotiating with Others, (G32) Performing for or Working in Directly with the Public, (G33) Coordinating the Work and Activities of Others, (G34) Developing and Building Teams, (G35) Training and Teaching Others, (G36) Guiding, Directing, and Motivating Subordinates, (G37) Train and Nurture Other People, (G38) Provide Consultation and Advice to Others, (G39) Performing Administrative Activities, (G40) Staffing Organisational Units, (G41) Monitoring and Controlling Resources

Author's own elaboration on ICP 2013²¹ and O*NET data descriptors²².

Table A2. Detected Communities

Cluster 1: Social-Cognitive	Critical Thinking, Active Learning, Active Listening, Administration and Management, Analysing Data or Information, Assisting and Caring for Others, Category Flexibility, Communicating with Persons Outside Organisation, Communicating with Supervisors, Peers, or Subordinates, Communication and Media, Complex Problem Solving, Coordinating the Work and Activities of Others, Deductive Reasoning, Developing Objectives and Strategies, Developing and Building Teams, Documenting/Recording Information Economics and Accounting, Education and Training, Establishing and Maintaining Interpersonal Relationships, Fine Arts, Flexibility of Closure, Fluency of Ideas, Food Production, Foreign Language, Geography, Getting information, Guiding, Directing, and Motivating Subordinates, History and Archaeology, Human Resources Management, IT and Electronics, Identifying Objects, Actions, and Events, Inductive Reasoning, Information Ordering, Instructing, Interacting With Computers, Interpreting the Meaning of the Information for Others, Italian Language, Judging the Qualities of Things, Services or People, Judgement and Decision Making, Learning Strategies, Legislation and Institutions, Making Decisions and Solving Problems, Management of Financial Resources, Management of Personnel Resources, Medicine and Dentistry, Memorisation, Monitoring, Monitoring and Controlling Resources, Negotiation, Number Facility, Office Work, Oral Comprehension, Oral Expression, Organising, Planning, and Prioritising Work, Originality, Performing Administrative Activities, Performing for or Working in Directly with the Public, Persuasion, Philosophy and Theology, Problem Sensitivity, Processing Information, Provide Consultation and Advice to Others, Psychology, Reading Comprehension, Resolving Conflicts and Negotiating with Others, Sales and Marketing, Scheduling Work and Activities, Selling or Influencing Others, Service Orientation, Services to Customers, Social Perceptiveness, Sociology and Anthropology, Speaking, Speech Clarity, Speech Recognition, Speed of Closure, Staffing Organisational Units, Telecommunications, Therapy and Counseling, Thinking Creatively, Time Management, Time Sharing, Training and Teaching Others, Updating and Using Relevant Knowledge, Writing, Written Comprehension, Written expression
Cluster 2: Technical-Physical	Mathematics, Science, Biology, Building and Construction, Arms-Hand Steadiness, Auditory Attention, Chemistry, Civil Protection and Public Safety, Control Precision, Controlling Machines and Processes, Coordination, Depth Perception, Drafting, Laying Out, and Specifying Technical Devices Parts and Equipment, Dynamic Flexibility, Dynamic Strength, Engineering and Technology, Equipment Maintenance, Equipment Selection, Estimate the Quantifiable Characteristics of Products, Events or Information, Evaluating Information to Determine Compliance with Standards, Explosive Strength, Extent Flexibility, Far Vision, Finger Dexterity, Glare Sensitivity, Gross Balance Body Equilibrium, Gross Body Coordination, Handling and Moving Objects, Hearing Sensitivity, Inspecting Equipment, Structures or Material, Installation, Management of Material Resources, Manual Dexterity, Math Reasoning, Mathematics, Mechanics, Monitor Processes, Materials or Surroundings, Multilimb Coordination, Near Vision, Night Vision, Operating Vehicles, Mechanised Devices, or Equipment, Operation Monitoring, Operation and Control, Operations Analysis, Perceptual Speed, Performing General Physical Activities, Peripheral Vision, Physics, Production and Processing, Programming, Quality Control Analysis, Rate Control, Reaction Time, Repairing, Repairing and Maintaining Electronic Equipment, Repairing and Maintaining Mechanical Equipment, Response Orientation, Selective Attention, Sound Localisation, Spatial Orientation, Speed of Limb Movement, Stamina, Static Strength, Systems Analysis, Systems Evaluation, Technical Design, Technology Design, Train and Nurture Other People, Transportation, Troubleshooting, Trunk Strength, Visual Colour Discrimination, Visualisation, Wrist-Finger Speed

Table A3. Summary Statistics for Dependent Variables

	2013-2016	2014-2017	2015-2018	2016-2019	Total
Entry (with <i>RCA</i>)					
0	36,854	37,044	37,212	37,255	148,365
1	2,679	2,569	2,365	2,308	9,921
Exit (with <i>RCA</i>)					
0	13,826	13,985	14,137	14,202	56,150
1	2,709	2,470	2,354	2,303	9,836
Entry (with <i>RCA</i> ¹)					
0	32,986	33,119	33,118	32,951	132,174
1	3,160	3,044	3,129	3,215	12,548
Exit (with <i>RCA</i> ¹)					
0	16,742	16,876	16,852	16,866	67,336
1	3,180	3,029	2,969	3,036	12,214

Table A4. Summary Statistics for Dependent Variables

	obs.	mean	sd	min	max
<i>three years</i>					
<i>Entry (RCA)</i>	158,286	0.062	0.24	0	1
<i>Exit (RCA)</i>	65,986	0.15	0.36	0	1
<i>Entry (RCA</i> ¹ <i>)</i>	144,722	0.09	0.28	0	1
<i>Exit (RCA</i> ¹ <i>)</i>	79,550	0.15	0.36	0	1
<i>one year</i>					
<i>Entry (RCA)</i>	236,337	0.038	0.191	0	1
<i>Exit (RCA)</i>	98,787	0.091	0.287	0	1
<i>Entry (RCA</i> ¹ <i>)</i>	216,001	0.055	0.228	0	1
<i>Exit (RCA</i> ¹ <i>)</i>	119,123	0.099	0.298	0	1

Notes: Summary statistics for dependent variables are reported.

Table A5. Correlation Matrix

	<i>F. ASRD</i>	<i>SC ASRD</i>	<i>TP ASRD</i>	<i>F. Comp.</i>	<i>SC Comp.</i>	<i>TP Comp.</i>	<i>Reg. Diversity</i>	<i>Ind. Ubiquity</i>	<i>Churn</i>	<i>Bus. Growth</i>	<i>GDP (log)</i>	<i>Pop. Density</i>	<i>Educ.</i>
<i>Full ASRD</i>	1.0000												
<i>SC ASRD</i>	0.9747	1.0000											
<i>TP ASRD</i>	0.9754	0.9469	1.0000										
<i>Full Complexity</i>	0.2633	0.2381	0.2593	1.0000									
<i>SC Complexity</i>	0.2282	0.2191	0.2389	0.6995	1.0000								
<i>TP Complexity</i>	0.2181	0.2092	0.2108	0.6664	0.6958	1.0000							
<i>Reg. Diversity</i>	0.4847	0.4857	0.4710	0.0985	0.0941	0.0899	1.0000						
<i>Ind. Ubiquity</i>	0.0580	0.0884	0.0773	0.2221	0.1530	0.1161	0.0000	1.0000					
<i>Churn</i>	-0.3160	-0.3145	-0.3080	-0.0242	-0.0816	-0.0442	-0.0043	0.1488	1.0000				
<i>Bus. Growth</i>	0.0446	0.0341	0.0259	0.0361	0.0254	0.0678	0.0647	-0.0483	0.2458	1.0000			
<i>GDP (log)</i>	0.6688	0.6766	0.6494	0.1418	0.1703	0.1432	0.1066	0.0000	-0.4751	-0.0166	1.0000		
<i>Pop. Density</i>	0.4310	0.4296	0.4114	0.1048	0.0888	0.0854	0.3897	-0.0000	-0.0283	0.0773	0.2569	1.0000	
<i>Education</i>	0.1181	0.1088	0.1124	0.0235	0.0135	0.0172	0.0435	0.0008	0.0618	0.1105	0.0807	0.1472	1.0000

Table A6. Variance Inflation Factors (VIF)

	Table 1	Table 1	Table 2	Table 2	Table 3	Table 3
	Col. 3	Col. 6	Col. 3	Col. 6	Col. 3	Col. 6
<i>Full ASRD</i>	2.84	3.13				
<i>Full Complexity</i>	1.12	1.16				
<i>SC ASRD</i>			3.02	2.80		
<i>SC Complexity</i>			1.08	1.07		
<i>TP ASRD</i>					2.54	2.72
<i>TP Complexity</i>					1.06	1.07
<i>GDP (log)</i>	2.37	2.29	2.53	2.20	2.24	2.17
<i>Pop. Density</i>	1.28	1.48	1.29	1.48	1.27	1.47
<i>Education</i>	1.04	1.06	1.04	1.06	1.04	1.06
<i>Churn</i>	1.48	1.49	1.48	1.50	1.48	1.49
<i>Bus. Growth</i>	1.12	1.08	1.12	1.08	1.13	1.08
<i>Reg. Diversity</i>	1.52	1.72	1.55	1.69	1.47	1.67
<i>Ind. Ubiquity</i>	1.10	1.21	1.09	1.21	1.08	1.17
<i>Mean VIF</i>	1.54	1.63	1.58	1.57	1.48	1.54

Table A7. Entry and Exit Models: Logistic and Probit Estimates

	Entry						Exit					
	(1) Logistic	(2) Logistic	(3) Logistic	(4) Probit	(5) Probit	(6) Probit	(7) Logistic	(8) Logistic	(9) Logistic	(10) Probit	(11) Probit	(12) Probit
<i>Full ASRD</i>	0.03253*** (0.00289)			0.01679*** (0.00138)			-0.02130*** (0.00305)			-0.01181*** (0.00165)		
<i>Full Complexity</i>	0.00469*** (0.00044)			0.00248*** (0.00022)			-0.00472*** (0.00046)			-0.00258*** (0.00025)		
<i>SC ASRD</i>		0.03351*** (0.00257)			0.01743*** (0.00126)			-0.02370*** (0.00279)		-0.01307*** (0.00149)		
<i>SC Complexity</i>		0.00682*** (0.00044)			0.00353*** (0.00022)			-0.00721*** (0.00046)		-0.00403*** (0.00025)		
<i>TP ASRD</i>			0.03855*** (0.00296)			0.02007*** (0.00141)			-0.02181*** (0.00303)		-0.01218*** (0.00166)	
<i>TP Complexity</i>			0.00574*** (0.00049)			0.00301*** (0.00025)			-0.00558*** (0.00057)		-0.00310*** (0.00030)	
<i>GDP (log)</i>	-0.26323 (0.41859)	-0.30919 (0.42067)	-0.34681 (0.41935)	-0.13812 (0.20469)	-0.16270 (0.20554)	-0.17445 (0.20522)	-0.17795 (0.43528)	-0.15335 (0.43658)	-0.15867 (0.43479)	-0.11436 (0.24033)	-0.10489 (0.24075)	-0.10378 (0.24013)
<i>Pop. Density</i>	0.07460 (0.74941)	0.08823 (0.75386)	0.18145 (0.75093)	0.04347 (0.36682)	0.05448 (0.36856)	0.08585 (0.36775)	-0.58826 (0.78513)	-0.55737 (0.78710)	-0.63379 (0.78414)	-0.29405 (0.43358)	-0.27422 (0.43410)	-0.31549 (0.43314)
<i>Education</i>	-0.03054* (0.01575)	-0.03381** (0.01583)	-0.02684* (0.01576)	-0.01657** (0.00769)	-0.01859** (0.00773)	-0.01541** (0.00770)	-0.02139 (0.01679)	-0.01947 (0.01689)	-0.02497 (0.01679)	-0.01217 (0.00916)	-0.01148 (0.00919)	-0.01443 (0.00915)
<i>Churn</i>	-0.01209* (0.00693)	-0.01165* (0.00696)	-0.00918 (0.00694)	-0.00684** (0.00337)	-0.00665** (0.00338)	-0.00512 (0.00337)	0.01068 (0.00727)	0.01148 (0.00731)	0.01038 (0.00731)	0.00644 (0.00402)	0.00685* (0.00403)	0.00617 (0.00403)
<i>Bus. Growth</i>	0.01824*** (0.00496)	0.01720*** (0.00498)	0.01773*** (0.00498)	0.00967*** (0.00244)	0.00920*** (0.00245)	0.00919*** (0.00245)	-0.01194** (0.00561)	-0.01047* (0.00562)	-0.01257** (0.00561)	-0.00658** (0.00313)	-0.00563* (0.00314)	-0.00700** (0.00313)
<i>Reg. Diversity</i>	-0.01984*** (0.00255)	-0.02009*** (0.00257)	-0.01984*** (0.00256)	-0.00954*** (0.00125)	-0.00972*** (0.00125)	-0.00952*** (0.00125)	0.02373*** (0.00267)	0.02418*** (0.00267)	0.02396*** (0.00266)	0.01317*** (0.00146)	0.01344*** (0.00146)	0.01335*** (0.00145)
<i>Ind. Ubiquity</i>	-0.06302*** (0.00441)	-0.06256*** (0.00442)	-0.06248*** (0.00442)	-0.03448*** (0.00228)	-0.03421*** (0.00228)	-0.03426*** (0.00228)	0.06533*** (0.00503)	0.06481*** (0.00503)	0.06529*** (0.00503)	0.03784*** (0.00278)	0.03752*** (0.00278)	0.03777*** (0.00278)
<i>Full FE</i>	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
<i>N</i>	155,583	155,583	155,583	155,583	155,583	155,583	65,649	65,649	65,649	65,649	65,649	65,649

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Appendix: Ramsey Pricing of Pharmaceuticals? A Debiased Machine Learning Analysis

A. Extended DML Estimates with Instrument

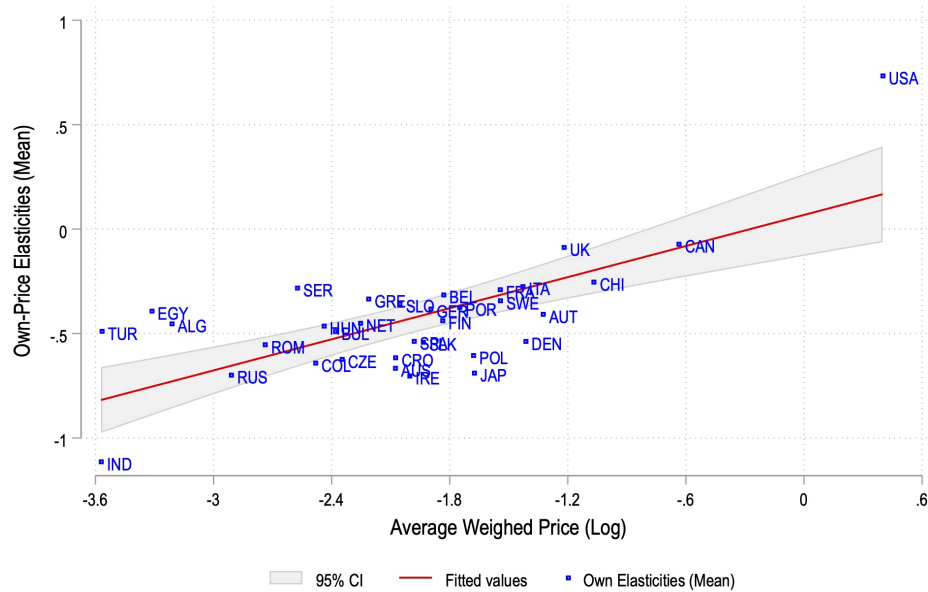


Figure A1. Estimates with Instrument: Relationship Between Price Elasticities and Price

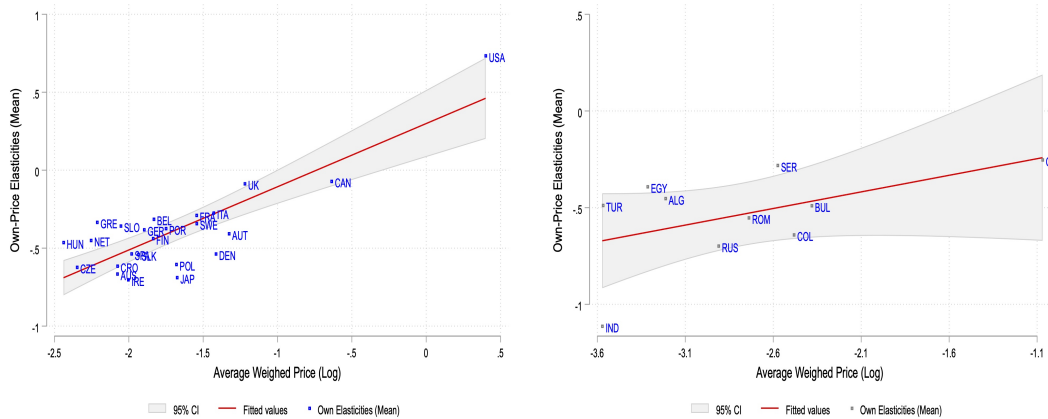


Figure A2. Estimates with Instrument: Relationship Between Price Elasticities and Price for (a) High Income Countries (HIC); Middle and Low Income Countries (MLIC)

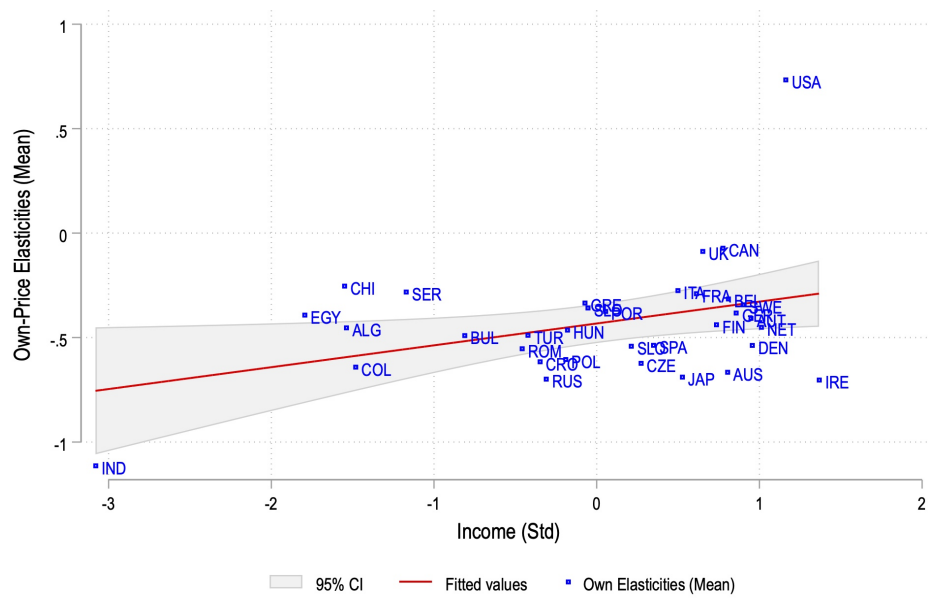


Figure A3. Estimates with Instrument: Relationship Between Price Elasticities and GDP Per Capita (PPP) Income

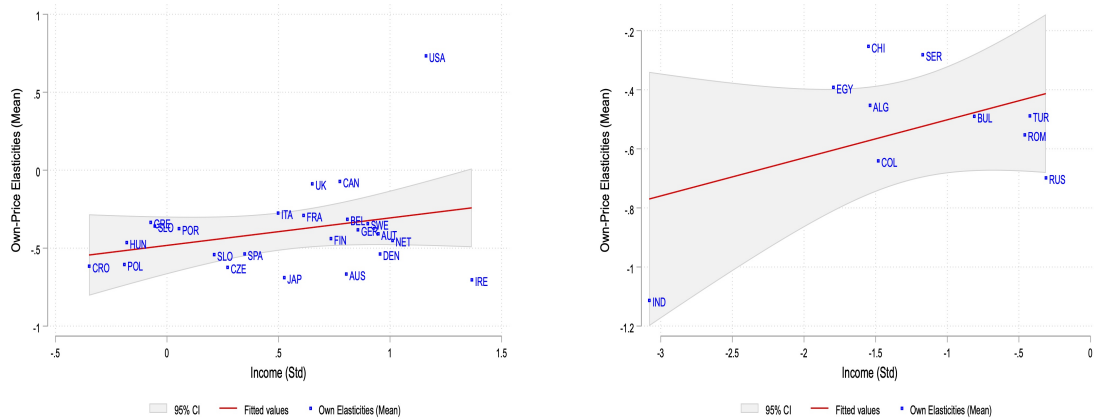


Figure A4. Estimates with Instrument: Relationship Between Price Elasticities and GDP Per Capita (PPP) Income for (a) High Income Countries (HIC); Middle and Low Income Countries (MLIC)

Bibliography

- Adhikari, B. K. and A. Agrawal (2016). "Religion, Gambling Attitudes and Corporate Innovation". In: *Journal of Corporate Finance* 37, pp. 229–248.
- Akcigit, U., J. Grigsby, and T. Nicholas (2017). "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age". In: *Harvard Business School Working Paper* 17–063.
- Alabdulkareem, A., M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, and I. Rahwan (2018). "Unpacking the polarization of workplace skills". In: *Science Advances* 4, pp. 1–9.
- Amabile, TM. (1996). "Creativity and Innovation in Organization". In: *Harvard Business School Background Note* 396–239.
- Anderson, K. A. (2017). "Skill networks and measures of complex human capital". In: *Pnas* 114(48), pp. 12720–12724.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables". In: *Journal of the American Statistical Association* 91.434, pp. 444–455. ISSN: 01621459. URL: <http://www.jstor.org/stable/2291629>.
- Antonietti, R. and R. Boschma (2021). "Social Capital, Resilience, and Regional Diversification in Italy". In: *Industrial and Corporate Change* 30(1), pp. 762–777.
- Antonietti, R., L. Cattani, F. Gambarotto, and G. Pedrini (2021). "Education, Routine, and Complexity- Biased Knowledge Enabling Technologies: Evidence from Emilia- Romagna, Italy". In: *Gran Sasso Science Institute Discussion paper* No. 2021-07.
- Assouad, A. and K. P. Parboteeah (2018). "Religion and innovation. A country institutional approach". In: *Journal of Management, Spirituality & Religion* 15.1, pp. 20–37. DOI: 10.1080/14766086.2017.1378589.
- Azzi, C. and R. G. Ehrenberg (1975). "Household Allocation of Time and Church Attendance". In: *Journal of Political Economy* 83.1, pp. 27–56.
- Bahar, D., S. Rosenow, E. Stein, and R. Wagner (2018). "Export Take-Offs and Acceleration: Unpacking Cross-Sector Linkages in the Evolution of Comparative Advantage". In: *Iza Discussion Paper Series* 12061.
- Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (Feb. 2015). *Demand Estimation with Machine Learning and Model Combination*. Working Paper 20955. National Bureau of Economic Research. URL: <http://www.nber.org/papers/w20955>.
- Balland, P. A. (2016). "EconGeo: Computing Key Indicators of the Spatial Distribution of Economic Activities". In: .
- Balland, P. A. and R. Boschma (2019). "Smart Specialisation: beyond patents". In: *Report for European Commission* volume.
- Balland, P. A., R. Boschma, J. Crespo, and D. Rigby (2019). "Smart specialization policy in the EU: Relatedness, knowledge complexity and regional diversification". In: *Regional Studies* 53(9), pp. 1252–1268.
- Balland, P. A., T. Broekel, D. Diodato, E. Giuliani, R. Hausmann, N. O'Clery, and D. Rigby (2022). "The new paradigm of economic complexity". In: *Research Policy* 51.3, p. 104450. DOI: <https://doi.org/10.1016/j.respol.2021.104450>.
- Balland, P. A. and D. Rigby (2017). "The Geography of Complex Knowledge". In: *Economic Geography* 93(1), pp. 1–23. URL: <https://doi.org/10.1080/00130095.2016.1205947>.
- Balland, P. A., D. Rigby, and R. Boschma (2015). "The technological resilience of US cities". In: *Cambridge Journal of Regions, Economy and Society* 8(2), pp. 167–184. URL: <https://doi.org/10.1093/cjres/rsv007>.
- Barro, R. J. and Jong-Wha Lee (2011). "A New Data Set of Educational Attainment in the World". In: *Journal of Development Economics* 104, pp. 184–198.
- Barro, R. J. and R. M. McCleary (2006). "Religion and Economy". In: *Journal of Economic Perspectives* 20.2, pp. 49–72.

- Barros, P. P. and X. Martinez-Giralt (2008). "On international cost-sharing of pharmaceutical RD". In: *International Journal of Health Care Finance and Economics* 8.4, pp. 301–312.
- Baum, C. F., M. E. Schaffer, and S. Stillman (2007). "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing". In: *The Stata Journal* 7.4, pp. 465–506. DOI: 10.1177/1536867X0800700402. eprint: <https://doi.org/10.1177/1536867X0800700402>. URL: <https://doi.org/10.1177/1536867X0800700402>.
- Baumol, W. J. and D. F. Bradford (1970). "Optimal Departures From Marginal Cost Pricing". In: *The American Economic Review* 60.3, pp. 265–283. ISSN: 00028282. URL: <http://www.jstor.org/stable/1817977>.
- Becker, G.S. (1965). "A Theory of the Allocation of Time". In: *The Economic Journal* 75.299, pp. 493–517.
- Becker, S. O. and L. Woessmann (2009). "Was Weber Wrong? A Human Capital Theory of Protestant Economic History". In: *The Quarterly Journal of Economics* 124.2, pp. 531–596.
- Belloni, A. and V. Chernozhukov (2013). "Least squares after model selection in high-dimensional sparse models". In: *Bernoulli* 19.2, pp. 521–547. URL: <https://doi.org/10.3150/11-BEJ410>.
- Benabou, Roland, Davide Ticchi, and Andrea Vindigni (2013). "Forbidden Fruits: The Political Economy of Science, Religion, and Growth". In: *Princeton University William S. Dietrich II Economic Theory Center Research Paper* 065.
- (2015). "Religion and Innovation". In: *American Economic Review* 105.5, pp. 346–351.
- Berry, S., J. Levinsohn, and A. Pakes (1995). "Automobile Prices in Market Equilibrium". In: *Econometrica* 63.4, pp. 841–890. URL: <http://www.jstor.org/stable/2171802>.
- Bigerna, S. and C. A. Bollino (2016). "Ramsey prices in the Italian electricity market". In: *Energy Policy* 88, pp. 603–612. ISSN: 0301-4215. DOI: <https://doi.org/10.1016/j.enpol.2015.06.037>.
- Blondel, V. D., J. Guillaume, R. Lambiotte, and E. Lefebvre (Oct. 2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. URL: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- Boschma, R. (2017). "Relatedness as driver behind regional diversification: A research agenda". In: *Regional Studies* 51(3), pp. 351–364.
- Boschma, R., P. A. Balland, and D. F. Kogler (2015). "Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010". In: *Industrial and Corporate Change* 24(1), pp. 223–250.
- Boschma, R., R. Eriksson, and U. Lindgren (2014). "Labour market externalities and regional growth in Sweden: the importance of labour mobility between skill-related industries". In: *Regional Studies* 48(10), pp. 1669–1690.
- Boschma, R., A. Minondo, and M. Navarro (2013). "The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness". In: *Economic Geography* 89(1), pp. 29–51.
- Botticini, Maristella and Zvi Eckstein (2005). "Jewish Occupational Selection: Education, Restrictions, or Minorities?" In: *Journal of Economic History* 65.4, pp. 922–948.
- Breschi, S., F. Lissoni, and F. Malerba (2003). "Knowledge-relatedness in firm technological diversification". In: *Research Policy* 32, pp. 69–87.
- Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics". In: *The Quarterly Journal of Economics* 132.3, pp. 1261–1318. ISSN: 0033-5533. URL: <https://doi.org/10.1093/qje/qjx013>.
- Caines, C., F. Hoffmann, and G. Kambourov (2017). "Complex-Task biased technological change and the labor market". In: *International Finance Discussion Papers* 1192.
- Camps, Susanna and Pilar Marques (2014). "Exploring How Social Capital Facilitates Innovation: The Role of Innovation Enablers". In: *Technological Forecasting and Social Change* 88, pp. 325–348.
- Chandra, A., J. Gruber, and R. McKnight (2014). "The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts". In: *Journal of Health Economics* 33, pp. 57–66. ISSN: 0167-6296. URL: <https://www.sciencedirect.com/science/article/pii/S0167629613001409>.
- Chang, S. J. (1996). "An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance 1981–89". In: *Strategic Management Journal* 17, pp. 587–611.
- Chen, Yangyang, E. J. Podolski, S. G. Rhee, and M. Veeraraghavan (2014). "Local Gambling Preferences and Corporate Innovative Success". In: *Journal of Financial and Quantitative Analysis* 49.1, pp. 77–106.
- Chen, Yangyang, E. J. Podolski, and Madhu Veeraraghavan (2017). "National Culture and Corporate Innovation". In: *Pacific Basin Finance Journal* 43, pp. 173–187.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1, pp. 1–68. URL: <https://doi.org/10.1111/ectj.12097>.

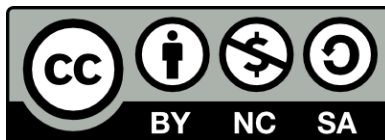
- Cinnirella, Francesco and Jochen Streb (2017). "Religious Tolerance as Engine of Innovation". In: *CESifo Working Papers* 6797.
- Correia, Sergio (2016). *A Feasible Estimator for Linear Models with Multi-Way Fixed Effects*. 2016 Stata Conference. Stata Users Group.
- Cutler, D. M. and J. Gruber (1996). "Does Public Insurance Crowd Out Private Insurance". In: *The Quarterly Journal of Economics* 111.2, pp. 391–430. URL: <http://www.jstor.org/stable/2946683> (visited on 05/11/2022).
- Danzon, M. P. (1998). "The Economics of Parallel Trade". In: *Pharmaeconomics* 13.3, pp. 293–304.
- (2019). "Drug Pricing and Value in Oncology". In: *Regulatory and Economic Aspects in Oncology* 213.
- Danzon, M. P., A. W. Mulchay, and A. Towse (2011). "Pharmaceutical Pricing in Emerging Markets: Effects of Income, Competition and Procurement". In: *NBER Working Paper* No. 17174.
- Danzon, M. P. and A. Towse (2003). "Differential Pricing for Pharmaceuticals: Reconciling Access, RD and Patents". In: *International Journal of Health Care Finance and Economics* 3, pp. 183–205.
- Danzon, M. P., A. Towse, and J. Mestre-Ferrandiz (2015). "Value-Based Differential Pricing: Efficient Prices for Drugs in a Global Context". In: *Health Economics* 24.3, pp. 294–301.
- Danzon, P. M. (1997). "Price Discrimination for Pharmaceuticals: Welfare Effects in the US and the EU". In: *International Journal of the Economics of Business* 4.3.
- (2000). "Making sense of drug prices". In: *Regulation: the Cato review of business and government* 21.1.
- (2018). "Differential Pricing of Pharmaceuticals: Theory, Evidence and Emerging Issues". In: *Pharmacoeconomics* 36.12, pp. 1395–1405.
- Diodato, D. and A. Weterings (2015). "The resilience of regional labour markets to economic shocks: exploring the role of interactions among firms and workers". In: *Journal of Economic Geography* 15(4), pp. 723–742.
- Dziallas, Marisa and Knut Blind (2019). "Innovation Indicators Throughout the Innovation Process: An Extensive Literature Analysis". In: *Technovation* 80–81, pp. 3–29.
- Einav, L., A. Finkelstein, and M. Polyakova (2018). "Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D". In: *American Economic Journal: Economic Policy* 10.3, pp. 122–53. URL: <https://www.aeaweb.org/articles?id=10.1257/pol.20160355>.
- Elekes, Z., A. Baranowska-Rataj, and R. Eriksson (2021). "Local access to skill-related high-income jobs facilitates career advancement for low-wage workers". In: *Papers in Evolutionary Economic Geography* 21.36, pp. 365–369.
- Ellis, R. P., B. Martins, and W. Zhu (2017). "Health care demand elasticities by type of service". In: *Journal of Health Economics* 55, pp. 232–243.
- Esteban, Joan, Gilat Levy, and Laura Mayoral (2015). "Liberty, Religiosity and Effort". In: — (2018). "Personal Liberties, Religiosity, and Effort". In: *Barcelona GSE Working Paper Series* 1031.
- (2019). "Personal Liberties, Religiosity, and Effort". In: *European Economic Review* 120, p. 103320.
- Farinha, T.F., P. A. Balland, A. Morrison, and R. Boschma (2019). "What drives the geography of jobs in the US? Unpacking relatedness". In: *Industry and Innovation* 26 (9), pp. 988–1022.
- Farjoun, M. (1994). "Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups". In: *Organization science* 5(2), pp. 185–199.
- (1998). "The independent and joint effects of the skill and physical bases of relatedness in diversification". In: *Strategic Management Journal* 19(7), pp. 611–630.
- Feser, E. J. (2003). "What Regions Do Rather than Make: A Proposed Set of Knowledge-based Occupation Clusters". In: *Urban Studies* 40(10), pp. 1937–1958.
- Finke, Roger and Laurence R. Iannaccone (1993). "Supply-Side Explanations for Religious Change". In: *The Annals of the American Academy of Political and Social Science* 527.1, pp. 27–39.
- Finke, Roger and Rodney Stark (1992). *The Churching of America 1776–1990: Winners and Losers in Our Religious Economy*. Rutgers University Press.
- Fitjar, R. D. and B. Timmermans (2017). "Regional skill relatedness: towards a new measure of regional related diversification". In: *European Planning Studies* 25(3), pp. 516–538. URL: <https://doi.org/10.1080/09654313.2016.1244515>.
- Ford, G. S., W. J. Baumol, and J. G. Sidak (1993). "Toward Competition in Local Telephony". In: *Southern Economic Journal* 62, p. 1109.
- Gabe, T. and J. R. Abel (2011). "Agglomeration of Knowledge". In: *Urban Studies* 48(7), pp. 1353–1371.
- Gill, Anthony and Erik Lundsgaarde (2004). "State Welfare Spending and Religiosity". In: *Rationality and Society* 16.4, pp. 399–436.
- Gilman, B. H. and J. Kautter (2008). "Impact of Multitiered Copayments on the Use and Cost of Prescription Drugs among Medicare Beneficiaries". In: *Health Research and Educational Trust* 43.3, pp. 478–495.

- Glynn, M. A. (1996). "Innovative Genius: A Framework for Relating Individual and Organizational Intel- ligences to Innovation". In: *The Academy of Management Review* 21, pp. 1081–1111.
- Goldman, D. P., G. F. Joyce, J. J. Escarce, J. E. Pace, M. D. Solomon, M. Laouri, P. B. Lansman, and S. M. Teutsch (2004). "Pharmacy Benefits and the Use of Drugs by the Chronically III". In: *JAMA The Journal of the American Medical Association* 291.19, pp. 2344–2350.
- Goldstone, John (1987). "Cultural Orthodoxy, Risk, and Innovation: The Divergence of East and West in the Early Modern World". In: *Sociological Theory* 5.2, pp. 119–135.
- Gorodnichenko, Yuriy and Gerard Roland (2010). "Culture, Institutions and the Wealth of Nations". In: *IZA Discussion Paper Series* 5187.
- Greene, W. H. (2012). "Econometric analysis". In: 7th ed.
- Guimaraes, Paulo and Pedro Portugal (2010). "A simple feasible procedure to fit models with high-dimensional fixed effects". In: *The Stata Journal* 10.4, pp. 628–649.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2003). "People's Opium? Religion and Economic Atti- tudes". In: *Journal of Monetary Economics* 50, pp. 225–282.
- (2006). "Does Culture Affect Economic Outcomes?" In: *Journal of Economic Perspectives* 20.2, pp. 23–48.
- Guo, Q., S. Zhu, and R. Boschma (2020). "Networks of export markets and export market diversification". In: *Industrial and Corporate Change* volume, pp. 1–17.
- Hausmann, R. and B. Klinger (2007). "The structure of the product space and the evolution of comparative advantage". In: *Working Paper Cambridge, MA: Center for International Development, Harvard University* 146.
- Hellevik, O. (2009). *Linear versus logistic regression when the dependent variable is a dichotomy*. URL: <https://EconPapers.repec.org/RePEc:spr:qualqt:v:43:y:2009:i:1:p:59-74>.
- Hidalgo, C. A. and R. Hausmann (2009). "The building blocks of economic complexity". In: *Proceedings of the National Academy of Sciences* 106(26), pp. 10570–10575.
- Hidalgo, C. A., B. Klinger, A. L. Barabasi, and R. L. Hausmann (2007). "The Product Space Conditions the Development of Nations". In: *Science* 317.317 (5837), pp. 482–487.
- Hidalgo, C. A. et al. (2018). "The principle of relatedness". In: *Springer Proceedings in Complexity*, pp. 451–457.
- Hofstede, Geert (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. 2nd ed. Sage Publications.
- (2011). "Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture". In: *Online Readings in Psychology and Culture* 2.1. DOI: <https://doi.org/10.9707/2307--0919.1014>.
- Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov (2010). *Cultures and Organizations: Software for the Mind*. 3rd ed. McGraw-Hill.
- Huang, D., D. Lu, and J. Luo (2016). "Corporate innovation and innovation efficiency: does religion mat- ter?" In: *Nankai Business Review International* 7(2), pp. 150–191. URL: <https://ideas.repec.org/a/eme/nbripp/v7y2016i2p150-191.html>.
- Iannaccone, L. R. (1998). "Introduction to the Economics of Religion". In: *Journal of Economic Literature* 36.3, pp. 1465–1495.
- (2008). "Looking Backward: A Cross-National Study of Religious Trends". In: *George Mason University Working Paper*.
- Iyer, S., A. Shrivastava, and R. Ticku (Jan. 2017). *Holy Wars? Temple desecrations in Medieval India*. Cambridge Working Papers in Economics 1705. Faculty of Economics, University of Cambridge.
- Iyer, Sriya (2016). "The New Economics of Religion". In: *Journal of Economic Literature* 54.2, pp. 395–441.
- Jack, W. and J. O. Lanjouw (2005). "Financing Pharmaceutical Innovation: How Much Should Poor Coun- tries Contribute?" In: *The World Bank Economic Review* 19.1, pp. 45–67.
- Kahn, K. B. (2018). "Understanding Innovation". In: *Business Horizons* 61, pp. 453–460.
- Kalliny, M. and A. V. Hausman (2007). "The Impact of Cultural and Religious Values on Consumer's Adop- tion of Innovation". In: *Academy of Marketing Studies Journal* 11, p. 125.
- Kalo, Z., L. Annemans, and L. P. Garrison (2013). "Differential pricing of new pharmaceuticals in lower income European countries". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 13.6. PMID: 24219049, pp. 735–741. DOI: 10.1586/14737167.2013.847367. URL: <https://doi.org/10.1586/14737167.2013.847367>.
- Kirkman, B. L., K. B. Lowe, and C. B. Gibson (2006). "A quarter century of Culture's Consequences: A review of empirical research incorporating Hofstede's cultural values framework". In: *Journal of Inter- national Business Studies* 37, pp. 285–320.
- Kogler, D. F. (2017). "Relatedness as driver of regional diversification: a research agenda, a commentary". In: *Regional Studies* 51(3).3, pp. 365–369. DOI: 10.1080/00343404.2016.1276282.

- Kogler, D. F., D. L. Rigby, and I. Tucker (2013). "Mapping Knowledge Space and Technological Relatedness in U.S. Cities". In: *European Planning Studies* 21(9), pp. 1374–1391.
- Kostis, P. C., K. I. Kafka, and P. E. Petrakis (2018). "Cultural Change and Innovation Performance". In: *Journal of Business Research* 88, pp. 306–313.
- Krenz, A. (2014). "Agglomeration of knowledge: A regional economic analysis for the German economy". In: *Cege Discussion Papers*, No. 206 volume.
- Kumar, Alok (2009). "Who Gambles in the Stock Market?" In: *The Journal of Finance* 114.4, pp. 1889–1933.
- Kumar, Alok, Jeremy Page, and O. G. Spalt (2011). "Religious Beliefs, Gambling Attitudes, and Financial Market Outcomes". In: *Journal of Financial Economics* 102, pp. 671–708.
- Landsman, P. B., W. Yu, X. Liu, S. M. Teutsch, and M. L. Berger (2005). "Impact of 3-tier pharmacy benefit design and increased consumer cost-sharing on drug utilization". In: *The American Journal of Managed Care* 11.10, pp. 621–628.
- Lee, GK. and MB. Lieberman (2010). "Acquisition vs. internal development as modes of market entry". In: *Strategic Management Journal* 31, pp. 140–158.
- Li, S. and D. Chollet (2006). "Price and Income Elasticity of the Demand for Health Insurance and Health Care Services: A Critical Review of the Literature". In: URL: <https://www.mathematica.org/publications/price-and-income-elasticity-of-the-demand-for-health-insurance-and-health-care-services>.
- Lichtenberg, F. (2011). "Pharmaceutical Companies' Variation Of Drug Prices Within And Among Countries Can Improve Long-Term Social Well-Being". In: *Health Affairs* 30.8, pp. 1539–1544.
- LoTurco, A. and D. Maggioni (2020). "The knowledge and skill content of production complexity". In: *Research Policy*.
- Maleug, D. A. and M. Schwartz (1994). "Parallel imports, demand dispersion, and international price discrimination". In: *Journal of International Economics* 34, pp. 167–195.
- Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, and Ar. Leibowitz (1987). "Health Insurance and the Demand for Health Care". In: *American Economic Review* 77.3, pp. 251–277.
- Martin-Cejas, R. R. (1997). "Airport pricing systems in europe and an application of Ramsey pricing to Spanish airports". In: *Transportation Research Part E: Logistics and Transportation Review* 33.4, pp. 321–327. ISSN: 1366-5545. DOI: [https://doi.org/10.1016/S1366-5545\(97\)00032-X](https://doi.org/10.1016/S1366-5545(97)00032-X).
- Maskus, K. E. (2001). "Parallel Imports in Pharmaceuticals Implications for Competition and Prices in Developing Countries". In:
- Melnick, G. A., J. Zwanziger, A. Bamezai, and R. Pattison (1992). "The effects of market structure and bargaining position on hospital prices". In: *Journal of Health Economics* 11.3, pp. 217–233. ISSN: 0167-6296. DOI: [https://doi.org/10.1016/0167-6296\(92\)90001-H](https://doi.org/10.1016/0167-6296(92)90001-H).
- Mewes, L. and T. Broekel (2020). "Technological Complexity and Economic Growth of Regions". In: *Papers in Evolutionary Economic Geography (PEEG)* 2049.
- Miller, A. S. and J. P. Hoffmann (1995). "Risk and Religion: An Explanation of Gender Differences in Religiosity". In: *Journal for the Scientific Study of Religion* 34.1, pp. 63–75.
- Miller, A. S. and Rodney Stark (2002). "Gender and Religiousness: Can Socialization Explanations Be Saved?" In: *American Journal of Sociology* 107.6, pp. 1399–1423.
- Montesor, S. and F. Quatraaro (2017). "Regional Branching and Key Enabling Technologies: Evidence from European Patent Data". In: *Economic Geography* 93(4), pp. 367–396.
- Morel, C. M., A. McGuire, and E. Mossialos (2011). "The Level Of Income Appears To Have No Consistent Bearing On Pharmaceutical Prices Across Countries". In: *Health Affairs* 30.8, pp. 1545–1552.
- Moser, Petra (2012). "Innovation without Patents: Evidence from World's Fairs". In: *Journal of Law and Economics* 55, pp. 43–74.
- (2013). "Patents and Innovation: Evidence from Economic History". In: *Journal of Economic Perspectives* 27, pp. 23–44.
- Muneepeerakul, R., J. Lobo, S. T. Shuttters, A. Gomez-Lievano, and M. R. Qubbai (2013). "Urban economies and occupation space: Can they get 'there' from 'here'?" In: *PLoS ONE* 8(9).
- Neffke, F., M. Hartog, R. Boschma, and M. Henning (2018). "Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification". In: *Economic Geography* 94(1), pp. 23–48. URL: <https://doi.org/10.1080/00130095.2017.1391691>.
- Neffke, F., M. Henning, and R. Boschma (2011). "How do regions diversify over time? Industry relatedness and the development of new growth paths in regions". In: *Economic Geography* 87(3), pp. 237–265.
- Neffke, F. and MS. Henning (2008). "Revealed Relatedness: Mapping Industry Space". In: *Papers in Evolutionary Economic Geography* Urban and Regional research centre Utrecht.
- (2013). "Skill-relatedness and firm diversification". In: *Strategic Management Journal* 34, pp. 297–316.

- Neffke, F., A. Otto, and A. Weyh (2017). "Inter-industry labor flows". In: *Journal of Economic Behavior Organization* 142, pp. 275–292.
- Outterson, K. (2005). "Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets". In: *Yale Journal of Health Policy, Law & Ethics* 193.
- Perret, J. K. (2014). "Growth and Innovation in Contemporary Russia". In: *Schumpeter Discussion Papers* 2014-006.
- Petrakis, P. E., P. C. Kostis, and D. G. Valsamis (2015). "Innovation and Competitiveness: Culture as a Long-term Strategic Instrument During the European Great Recession". In: *Journal of Business Research* 68.7, pp. 1436–1438.
- Petrou, P. and S. VANDOROS (2016). "Pharmaceutical price comparisons across the European Union and relative affordability in Cyprus". In: *Health Policy and Technology* 5.4, pp. 350–356.
- Preble, J. and R. Hoffman (2012). "General management of innovation: Lessons from the Shaker community". In: *Journal of Management History* 18, pp. 24–45. DOI: 10.1108/17511341211188637.
- Rahmati, R., A. Tafti, J. Westland, and C. Hidalgo (2021). "When All Products Are Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms". In: *MIS Quarterly* 45, pp. 1025–1058. DOI: 10.25300/MISQ/2021/15384.
- Ramsey, F. P. (1927). "A Contribution to the Theory of Taxation". In: *The Economic Journal* 37.145, pp. 47–61. ISSN: 00130133, 14680297. URL: <http://www.jstor.org/stable/2222721>.
- Recio-Román, Recio-Menéndez, and Román-González (2019). "Religion and Innovation in Europe: Implications for Product Life-Cycle Management". In: *Religions* 10, p. 589. DOI: <https://doi.org/10.3390/rel10100589>.
- Reisinger, M., L. Sauri, and H. Zenger (2019). "Post-TRIPS Options for Access to Patented Medicines in Developing Nations". In: *Journal of Health Economics* 66, pp. 163–179.
- Rigby, D. L. (2015). "Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes". In: *Regional Studies* 49(11), pp. 1922–1937. URL: <https://doi.org/10.1080/00343404.2013.854878>.
- Rogers, E. M. (1983). *Diffusion of Innovations*. The Free Press New York.
- Ruiter, Stijn and Frank Tubergen (2009). "Religious Attendance in Cross-National Perspective: A Multilevel Analysis of 60 Countries". In: *American Journal of Sociology* 115.3, pp. 863–895.
- Samuelsson, Kurt (1961). *A Critique of Max Weber*. Trans. by E. Geoffrey French. Harper Torchbooks.
- Sanchez-Borras, M., C. Nash, P. Abrantes, and A. Lopez-Pita (2010). "Rail access charges and the competitiveness of high speed trains". In: *Transport Policy* 17.2, pp. 102–109. ISSN: 0967-070X. DOI: <https://doi.org/10.1016/j.tranpol.2009.12.001>.
- Scheve, Kenneth and David Stasavage (2006). "Religion and Preferences for Social Insurance". In: *Quarterly Journal of Political Science* 1, pp. 255–286.
- Schmidt, F., M. Egler, and R. G. Geursen (2001). "Ramsey Pricing as an Explanation and Justification for Price Differences in the European Pharmaceutical Market". In: *Drugs Made in Germany* 44.3.
- Semenova, V., M. Goldman, V. Chernozhukov, and M. Taddy (2021). *Estimation and Inference on Heterogeneous Treatment Effects in High-Dimensional Dynamic Panels*. arXiv: 1712.09988 [stat.ML].
- Shane, Scott (1992). "Why Do Some Societies Invent More than Others?" In: *Journal of Business Venturing* 7.1, pp. 29–46.
- (1993). "Cultural Influences on National Rates of Innovation". In: *Journal of Business Venturing* 8.1, pp. 59–73.
- (1995). "Uncertainty Avoidance and the Preference for Innovation Championing Roles". In: *Journal of International Business Studies* 26.1, pp. 47–68.
- Shepherd, W. G. (1992). "Ramsey Pricing: Its Uses and Limits". In: *Utilities Policy* 2.4, pp. 296–298.
- Sherer, F. M. and J. Watal (2002). "Post-TRIPS Options for Access to Patented Medicines in Developing Nations". In: *Journal of International Economic Law* 5.4, pp. 913–939.
- Shore-Sheppard, Lara D. (Dec. 1996). "The Precision of Instrumental Variables Estimates With Grouped Data". In: 753. URL: <https://ideas.repec.org/p/pri/indrel/374.html>.
- Sondergaard, M. (1994). "Hofstede's Consequences: A Study of Reviews, Citations and Replications". In: *Organization Studies* 15, pp. 447–456.
- Steenkamp, J. B., F. Hofstede, and M. Wedel (1999). "A Cross-National Investigation into the Individual and National Cultural Antecedents of Consumer Innovativeness". In: *Journal of Marketing* 63, pp. 55–69.
- Tabellini, Guido (2005). "Culture and Institutions: Economic Development in the Regions of Europe". In: *Cesifo Working Paper* 1492.

- Tanner, A. N. (2014). "Regional Branching Reconsidered: Emergence of the Fuel Cell Industry in European Regions". In: *Economic Geography* 90(4), pp. 403–427. URL: <https://onlinelibrary.wiley.com/doi/10.1111/ecge.12055>.
- (2016). "The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases". In: *Journal of Economic Geography* 16(3), pp. 611–635.
- Taylor, M. Z. and Sean Wilson (2012). "Does Culture Still Matter? The Effects of Individualism on National Innovation Rates". In: *Journal of Business Venturing* 27, pp. 234–247.
- Timmermans, B. and R. Boschma (2014). "The effect of intra- and inter-regional labour mobility on plant performance in Denmark: the significance of related labour inflows". In: *Journal of Economic Geography* 14(2), pp. 289–311.
- Towse, A., A. Cole, and B. Zamora (2018). "The Debate on Indication-Based Pricing in the U.S. and Five Major European Countries". In: *OHE Consulting Report*.
- Towse, A., M. Pistollato, J. Mestre-Ferrandiz, Z. Khan, S. Kaura, and L. Garrison (2015). "European Union Pharmaceutical Markets: A Case for Differential Pricing?" In: *International Journal of the Economics of Business* 22.2, pp. 263–275.
- VanEck, N. J. and L. Waltman (2009). "How to normalize cooccurrence data? An analysis of some well-known similarity measures". In: *Journal of the American Society for Information Science and Technology* 8, pp. 1635–1651.
- Vaus, David De and Ian McAlliste (1987). "Gender Differences in Religion: A Test of the Structural Location Theory". In: *American Sociological Review* 52, pp. 472–481.
- Wang, C. L. and P. K. Ahmed (2004). "The Development and Validation of the Organisational Innovativeness Construct Using Confirmatory Factor Analysis". In: *European Journal of Innovation Management* 7.4, pp. 303–313.
- Waning, B., W. Kaplan, A. C. King, D. A. Lawrence, H. G. Leufkens, and M. P. Fox (2009a). "Global strategies to reduce the price of antiretroviral medicines: Evidence from transactional databases". In: *Bulletin of the World Health Organisation* 87, pp. 520–528.
- (2009b). "Global strategies to reduce the price of antiretroviral medicines: evidence from transactional databases". In: *Bulletin of the World Health Organization* 87.7, pp. 520–528.
- Weber, Max (2001). *The Protestant Ethic and the Spirit of Capitalism*. 2nd ed. Routledge London.
- Whittle, A. (2019). "Local and Non-Local Knowledge Typologies: Technological Complexity in the Irish Knowledge Space". In: *European Planning Studies* 27(4), pp. 661–677.
- Whittle, A. and D. F. Kogler (2019). "Related to what? Reviewing the literature on technological relatedness: Where we are now and where can we go?" In: *Pap Reg Sci.*, pp. 1–17.
- William, L. K. and S. J. McGuire (2010). "Economic Creativity and Innovation Implementation: The Entrepreneurial Drivers of Growth? Evidence from 63 Countries". In: *Small Business Econ* 34.4, pp. 391–412.
- Xiao, J., R. Boschma, and M. Anderson (2018). "Industrial diversification in Europe: The differentiated role of relatedness," in: *Economic Geography* 94(5), pp. 514–49.
- Yadav, P. (2010). "Differential pricing for pharmaceuticals". In: *UK Department for International Development*. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67672/diff-pricing-pharma.pdf.
- Yeung, K., A. Basu, R. N. Hansen, and S. D. Sullivan (2018). "Price elasticities of pharmaceuticals in a value based-formulary setting". In: *Health Economics* 27.11, pp. 1788–1804.



Unless otherwise expressly stated, all original material of whatever nature created by Duygu Buyukyazici and included in this thesis, is licensed under a Creative Commons Attribution Noncommercial Share Alike 3.0 Italy License.

Check on Creative Commons site:

<https://creativecommons.org/licenses/by-nc-sa/3.0/it/legalcode/>

<https://creativecommons.org/licenses/by-nc-sa/3.0/it/deed.en>

Ask the author about other uses.