Three Essays in the Economics of Migration and Education

Dissertation

zur Erlangung des akademischen Grades eines

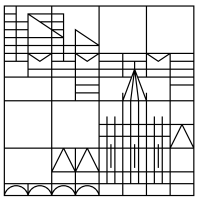
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

vorgelegt von

Maurizio Strazzeri

an der

Universität Konstanz



Sektion Politik – Recht – Wirtschaft
Fachbereich Wirtschaftswissenschaften

Konstanz, 2021

Tag der mündlichen Prüfung: 29.01.2021

1. Referent: Prof. Dr. Guido Schwerdt

2. Referentin: Jun. Prof. Dr. Luna Bellani

3. Referent: Prof. Dr. Erik Hornung

This dissertation is dedicated to the memory of my father,

Silvio Strazzeri.

Danksagung

Diese Dissertation konnte nur durch die Unterstützung anderer entstehen, bei denen ich mich an dieser Stelle bedanken möchte.

Zunächst möchte ich mich bei meinem Betreuer Guido Schwerdt bedanken. Seine Hinweise und Ratschläge waren maßgeblich für meine fachliche Weiterentwicklung in der empirischen Wirtschaftsforschung und haben meine Vorstellung von der wissenschaftlichen Behandlung ökonomischer Fragestellungen enorm beeinflusst.

Ich richte meinen Dank auch an die Universität Konstanz und an alle Beteiligten der Graduate School of Decision Sciences. Mir wurde eine exzellente Forschungsumgebung geboten, von der ich stark profitiert habe. Der regelmäßige Austausch mit ausgezeichneten Wissenschaftlern und Wissenschaftlerinnen half mir, die akademische Welt kennenzulernen, und war die Voraussetzung dafür, mich in ihr zu entfalten.

Ich möchte mich an dieser Stelle bei Heike Knappe, Jutta Obenland und Justine Overall bedanken, die durch zahlreiche Erinnerungsnachrichten und Verlängerungen von Deadlines die Bewältigung der administrativen Hürden deutlich einfacher gemacht haben.

Einen großen Dank richte ich an meine Bürokollegen, die mich über die letzten Jahre begleitet haben und sich nie über das Chaos beklagt haben, dass sich oft durch meine Anwesenheit im Büro verbreitet hat. Zu nennen sind hier: Nona Bledow, Yiyi Chen, Moritz Janas, Theresa Künzler, Sandra Morgenstern, Ruchira Suresh und Anika Zadruzynsky.

Ich bedanke mich bei meinem Kollegen Enzo Brox, der stets für fachliche Diskussionen Zeit gefunden hat. Er war bei meinen erfolgreichen und weniger erfolgreichen Forschungsprojekten ein wichtiger Ansprechpartner, der mich mit seiner Kreativität inspirieren konnte.

Zum Schluss möchte ich mich bei Freunden und Verwandten bedanken, die mich - bewusst oder unbewusst - bei der Fertigstellung dieser Dissertation unterstützt haben. Vielen Dank Sandy, Gerhard, Nina, Lucy, Miikka, Franz, Mama, Julian und Méline.

Contents

1 Introduction

1.1 Zusammenfassung

Diese Dissertation setzt sich zusammen aus drei Aufsätzen im Bereich der Migrations- und Bildungsökonomie. In meinem ersten Aufsatz untersuche ich den Einfluss von Internetnutzung auf die Migrationsentscheidung von Individuen in Nigeria. Meine empirische Untersuchung nutzt Variation in der Internetnutzung von Individuen, die durch die Verlegung von Tiefseekabeln in Westafrika erzeugt wurde. Meine Ergebnisse zeigen, dass Internetnutzung Migration substantiell begünstigt. Dieses Ergebnis ist konsistent mit dem ökonomischen Standardmodell zur Migrationsentscheidung, in dem Individuen den Nutzen und die Kosten der Migration miteinander vergleichen und die Internetnutzung die Kosten der Migration senkt. Weitere Untersuchungen zeigen, dass der Effekt besonders stark ist im Hinblick auf extra-kontinentale Migration sowie für Individuen aus dem Bereich der unteren Vermögensverteilung. Diese Ergebnisse lassen darauf schließen, dass die Reduktion der Kosten der Migration von Merkmalen der Zielländer und der potentiellen Migranten abhängen. Außerdem zeigen meine Schätzergebnisse, dass die Veränderung in den Migrationsraten, die durch die vermehrte Internetnutzung entstehen, mit höheren Geldsendungen in den Herkunftsregionen korrelieren, die potentiell die wirtschaftliche Entwicklung fördern.

In meinem zweiten Aufsatz untersuche ich den Einfluss von unterschiedlichem Schutzstatus auf die Arbeitsmarktintegration von Geflüchteten in Deutschland. Ich fokussiere mich dabei auf eine neue Geflüchtetenkohorte aus Syrien und Irak, die zwischen den Jahren 2013 und 2016 eingereist ist. Meine empirische Untersuchung nutzt Variation in Schutzstatus, die durch eine veränderte Auffassung des Bundesamt für Migration und Flüchtlinge im Hinblick auf die Vergabe des Schutzstatus nach der Genfer Flüchtlingskonvention für syrische und irakische Geflüchtete im März 2016 erzeugt wurde. Umfragedaten, die spezifische Informationen bezüglich des Asylprozesses der Geflüchteten beinhalten, erlauben es mir, die Politikveränderung in einem Fuzzy Regression Discontinuity Design zu untersuchen. Meine

Ergebnisse zeigen, dass subsidiärer Schutz die Arbeitsmarktintegration stark beeinträchtigt und das Arbeitseinkommen und die Beschäftigungswahrscheinlichkeit reduzieren. Dieses Ergebnis ist konsistent mit den Modellen aus der Migrationsliteratur, die nahelegen, dass die subjektiv empfundene Aufenthaltsdauer die Integrationsbereitschaft sowie die Arbeitsmarktintegration bestimmen. Weiterführende Schätzergebnisse weisen jedoch darauf hin, dass nachfrageseitige Faktoren ebenso eine Rolle spielen könnten. Zusammenfassend zeigen meine Ergebnisse einen politischen und ökonomischen Zielkonflikt in der Asylpolitik auf, der durch die Vergabe von temporärem Schutzstatus und erfolgreicher Arbeitsmarktintegration entsteht.

In meinem dritten Aufsatz, welcher in Zusammenarbeit mit Guido Schwerdt und Erik Hornung entstanden ist, untersuchen wir den Einfluss von Fasten während des Monats Ramadan auf die schulischen Leistung von Jugendlichen. Unsere empirische Untersuchung nutzt Variation in der durchschnittlichen Anzahl der Stunden, die Gläubige während des Monats Ramadan fasten und sich aufgrund des Mondkalenders jährlich verändern. Wir verknüpfen länderspezifische Fastenstunden mit zwei sich wiederholenden Querschnittdatensätzen, die Ergebnisse aus internationalen Bildungstests beinhalten. Unsere Schätzergebnisse zeigen, dass (1) Fastenstunden positiv korreliert sind mit den Leistungen in Bildungstests in Ländern mit einer muslimischen Mehrheit, aber nicht in anderen, und (2) Fastenstunden die Leistungsunterschiede zwischen Schülern, die nicht aus Ländern mit muslimischer Mehrheit kommen, und Schülern, die aus Ländern mit muslimischer Mehrheit kommen, in 8 europäischen Ländern mit hohen Immigrationsraten verringern. Unsere Ergebnisse sind konsistent mit zwei unterschiedlichen Erklärungsansätzen. Erstens könnte Fasten während des Monats Ramadan die Charakter-Fähigkeiten beeinflussen, die sich positiv bei der Formierung von Humankapital und schulischen Leistungen widerspiegeln (z.Bsp. Geduld, Selbstkontrolle, Durchhaltevermögen). Zweitens könnte, basierend auf der Theorie der sozialen Identität, die Ausübung von identischen Ritualen den Zusammenhalt in der Klasse festigen, die die schulischen Leistungen begünstigt. Weiterführende Ergebnisse lassen vermuten, dass unsere Hauptergebnisse durch eine Stärkung des sozialen Zusammenhalts in der Klasse erklärt werden.

1.2 Summary

This dissertation is a collection of three empirical essays in the economics of migration and education. In my first essay, I investigate the role of Internet usage in the migration decision. The study of the determinants of migration has a long history in the economics literature (Sjaastad, 1962; Borjas, 1987), and recent empirical contributions have linked migration decisions to advances in telecommunication technology (Lu et al., 2016) and media usage (Farré and Fasani, 2013). However, while there is abundant anecdotal and qualitative evidence that the Internet is extensively used by migrants, there has been no empirical study that investigates whether Internet usage increases the propensity to migrate. I address this gap in the literature and provide an analysis of the effect of Internet usage on the migration decision based on micro data from Nigeria. Studying this effect in Nigeria provides an interesting case as Nigeria has experienced a massive increase in the share of Internet users over the last decade and plays due to it's economic and demographic peculiarities an important role in African migration. In my empirical analysis, I follow Hjort and Poulsen (2019) and exploit variation in Internet usage generated by the arrival of submarine Internet cables in Western Africa. Overall, my results indicate that Internet usage substantially increases the likelihood of migration. I argue that this effect is consistent with a standard model of migration where individuals compare benefits and costs associated with moving, and Internet usage reduces the cost of migrating by lowering search and information costs or psychological costs. Further, I show that the effect is particularly large for migration out of Africa and for individuals from the lower part of the wealth distribution, which suggests differential changes in the reduction of migration costs induced by the Internet. In the final part of the paper, I complement my finding by investigating potential implications of increased migration rates due to Internet usage, and show that they are associated with higher remittances and economic development in the origin country.

In my second essay, I study the effect of protection statuses on refugees' labor market integration. The process of the economic assimilation of immigrants has received considerable attention in the economic literature on migration. In an early contribution, Chiswick (1978)

hypothesized that, upon arrival, immigrants lack the skills specific to their new country that would allow them to fully benefit from their human capital. Immigrants have high incentives to invest in such skills, and dynamic models of human capital suggest that the shorter the time span in which immigrants can benefit from these investments, the lower are the incentives to invest in country-specific human capital (e.g., Dustmann, 1999; Cortes, 2004). This suggests that refugees' protection statuses are an important element for labor market integration as they affect the expected permanence of stay in the host country. I study the effect of refugees' protection status on early labor market outcomes for a recent cohort of Syrian and Iraqi refugee migrants who entered Germany between the years 2013 and 2016. I exploit a change in the assessment of the Federal Agency responsible for asylum claims to grant full refugee status in accordance with the Geneva convention in March 2016. Using unique survey data from the years 2018 and 2019 that provide detailed information about the asylum procedure, I exploit this policy change in a fuzzy regression discontinuity design. My estimation results indicate a substantial negative effect of subsidiary protection status on earnings and employment, which is consistent with the findings from the literature on the assimilation of economic migrants. However, further analysis suggests that labor demand side factors could also explain my findings. In sum, this essay highlights a political and economic trade-off in asylum policies between granting temporary residence and labor market outcomes.

My third essay, which is joint work with Guido Schwerdt and Erik Hornung, investigates the impact of Ramadan fasting on educational outcomes. The study of the economics of religion is a relatively new field in the economics literature, and has been introduced to the economics community by the seminal work of Iannaccone (1998) (for a survey, see Iyer, 2016). To study the effect of Ramadan fasting on education outcomes, we follow Campante and Yanagizawa-Drott (2015) and use variation in annual fasting hours across time and countries. Combining country-specific daily fasting hours with two repeated cross-sectional international student achievement test surveys, we find that (i) fasting hours are positively associated with student performance in Muslim majority countries, but not in other countries, and (ii) fasting hours reduce performance gaps between students without and with Muslim

majority country origin in eight European immigration countries. Overall, our results are consistent with two potential causal mechanisms. First, religious practices such as Ramadan fasting might affect character skills - e.g., patience, self-control, and perseverance - that exert positive externalities on human capital formation and educational performance. Second, based on the theory of social identity, the performance of identical rituals might create a common identity among students, which potentially increases student performance. While our data does not allow to directly test one of the two channels, further estimation results suggest that our baseline results are likely driven by the creation of a common identity among students due to Ramadan fasting.

Internet usage and migration decisions: Evidence from Nigerian micro data*

Maurizio Strazzeri†

University of Konstanz
Department of Economics
Graduate School of Decision Sciences

Abstract

This paper investigates the role of Internet usage in the migration decision using micro-level data from Nigeria. Internet usage reduces migration costs such as search and information costs or psychological costs, which suggests that having access to the Internet increases the probability to migrate. My empirical analysis exploits variation in Internet usage induced by the arrival of submarine Internet cables in Western Africa. Results indicate a large positive effect of Internet usage on migration. The effect is particularly strong for migration out of Africa and is larger for individuals from the lower part of the wealth distribution.

*I thank Luna Bellani, Enzo Brox, Sebastian Findeisen, Tommy Krieger, Stephan Maurer, Guido Schwerdt, and seminar audiences at the University of Konstanz and the ESPE conference 2018 (Antwerp), 2019 (Bath), and the EEA-ESEM conference 2019 (Manchester) for helpful comments and suggestions.

†Email: maurizio.strazzeri@uni.kn

2.1 Introduction

The recent increase in the number of migrants from low and middle income countries to richer countries has led to a renewed interest in the determinants of migration among policy makers and economists alike. It has been noted that advances in modern information and communication technologies might partly explain these migration flows as they reduce the cost of information diffusion and provide potential migrants with information about migration opportunities that were previously unknown (Czaika and de Haas, 2014; Ortega and Peri, 2015).[1] Indeed, the large-scale adoption of the Internet around the world had tremendous impact on the amount, quality, and variety of information that individuals can obtain at relatively low cost. It also facilitated communication with peers over long distances by providing access to modern communication technologies such as emails, social media platforms, or instant messaging technologies. This might have important consequences for complex forward-looking migration decisions that require extensive information-gathering activities and are largely affected by communication via migration networks (Carrington et al., 1996; Munshi, 2003; McKenzie and Rapoport, 2007; Hanson and McIntosh, 2010). Qualitative studies provide evidence that the Internet is used by migrants to inform themselves about, e.g., potential destination countries, migration routes, and immigration regulations, or to maintain connections with peers in the origin country (e.g., Dekker and Engbersen, 2014; Zijlstra and Liempt, 2017), and it has been shown that Internet behavior in origin countries can be used to predict future migration flows (Böhme et al., 2020). However, less is known about whether the Internet is only used as a substitute for other means to implement migration decisions or whether Internet usage directly affects migration decision, i.e., Internet usage makes international migration more likely.

In this paper, I provide the first systematic analysis of the effect of Internet usage on migration decisions based on micro data from Nigeria. The analysis focuses on the years 2010 to

[1] Other potential explanations that have been discussed in literature are economic development and rising income levels in developing countries (Clemens, 2014), climate change which disproportionally affects low income countries (Beine and Parsons, 2015; Cattaneo and Peri, 2016; Missirian and Schlenker, 2017), or demographic imbalances due to a particularly young population in many developing countries (Ortega and Peri, 2015).

2016, where - at the beginning - Internet usage was extremely rare but became more common over time, which provides valuable variation in first-time Internet usage. However, any study of Internet usage and migration is confronted with considerable selection problems as various socio-economic background characteristics that influence the migration decision - such as age, education or income level - affect simultaneously the likelihood of Internet usage. To overcome such endogeneity concerns, I follow Hjort and Poulsen (2019) and exploit time and cross-sectional variation in Internet speed generated by the arrival of large submarine Internet cables from Europe that presumably facilitated Internet usage for individuals living in Nigeria. These submarine cables were connected to the terrestrial cable network of Nigeria between the years 2010 and 2012 and brought much faster Internet and higher Internet traffic capacities to locations close to the terrestrial cable network but not to others. My preferred empirical strategy exploits the arrival of the submarine Internet cables by comparing individuals located close to the terrestrial cable network before and after the arrival of submarine cables with individuals who are located further away during the same time period.

The empirical analysis is based on the comprehensive Nigerian geo-coded General Household Survey (GHS) panel which provides extensive information about individuals' Internet usage behavior. Additionally, the time dimension of the Nigerian GHS panel allows to reconstruct individuals' migration decisions as remaining household members state the whereabouts of former survey participants that moved out of the household. To exploit the variation generated by the arrival of submarine Internet cables, I match the linked information about migration decisions and previous Internet usage with detailed maps of the terrestrial cable network in Nigeria prior to the arrival of the submarine Internet cables.

I start my empirical analysis with various reduced form difference-in-difference estimates using binary or continuous measures of the distance to the terrestrial cable network. I show that individuals located in communities close to the terrestrial cable network respond with a larger increase in migration rates than those individuals located in more remote locations after the arrival of the submarine Internet cables. The change in migration rates for the period before and after the arrival of fast Internet is 0.14 pp lower for individuals located in commu-

nities twice as far as the comparison group (approx. -17.7 % relative to mean migration rates in locations close to the cable network). Estimation results using a binary treatment variable suggest that the effect is particularly large for individuals located in a 5 kilometers (km) radius around the terrestrial cable network, which is in line with the hypothesis that individuals are particularly affected if they live close to the terrestrial cable network. Moreover, the absolute effect on migration is twice as large for younger individuals in the age bracket of 20 to 35. As younger individuals are more likely to respond to faster Internet at the extensive (Internet take-up) and intensive margin (more frequent Internet usage), this results suggests that the association between the availability of fast Internet and migration rates is driven by changes in Internet behavior.

To further explore this channel, I show that the arrival of faster Internet is indeed associated with an increase in Internet usage at the extensive and intensive margin. With respect to the extensive margin, being located within a 5 km radius around the terrestrial cable network is associated with an increase in the probability to use the Internet by around 5 pp for the entire sample and 10 pp for younger individuals after the arrival of the submarine Internet cables in comparison to individuals located in more remote locations. Under the assumption that the exclusion restriction is satisfied - which I extensively test in the paper - these results indicate a large effect of Internet usage on migration decisions. Two-stage least squares estimates suggests that Internet usage increases the probability to migrate by 10 pp.

The positive effect of Internet usage on migration is consistent with a standard model of migration where individuals compare benefits and costs associated with moving (Sjaastad, 1962; Borjas, 1987), and Internet usage lowers the cost of migrating. Internet usage might reduce migration costs by lowering search and information costs or psychological costs that "incur due to the reluctance of individuals to leave familiar surroundings, family, and friends" (Sjaastad, 1962). It might also be the case that the exposure to foreign media that is more prevalent on the Internet changes individuals' preferences for one country in comparison to the origin country such as preferences for a particular climate or lifestyle. All these channels suggest that the effect of Internet usage on migration decisions varies across observable characteris-

tics of destination countries and potential migrants. My empirical set-up allows to assess such heterogeneous effects and I highlight two important consequences of higher migration rates that are caused by increased Internet usage.

First, I show that the effect of Internet usage on migration decisions is more important for migration out of Africa in comparison to migration within Africa, which suggests that the reduction in migration costs induced by Internet usage is larger for extra-continental destination countries. This is in line with the hypotheses that Internet usage lowers search and information costs and changes the set of information available to potential migrants, which might be particularly important for migration out of Africa, or that the exposure to foreign media transmitted via the Internet affect individuals' preferences. Second, I show that the effect of Internet usage on migration differs by individuals' wealth, and individuals from the lower part of the wealth distribution respond with a higher increase in migration rates due to Internet usage. Assuming that individuals' wealth is associated with skills (Angelucci, 2015), this result highlights the self-selection and resulting skill distribution of immigrants caused by the spread of the Internet. A potential explanation for the differential change in migration costs due to Internet usage might be that the migration related information provided by the Internet is more valuable for individuals at the lower part of the skill distribution. Additionally, these individuals are more often confronted with financial constraints that prevent them from migrating (Angelucci, 2015; Bazzi, 2017), which suggests that reduction in migration cost due to Internet usage relaxes these previous constraints and low-skilled migration becomes more likely.

In the final part of this paper, I complement my main findings with an investigation of possible feedback effects of the increased migration rates due to higher Internet usage. The impact of international migration on the development in sending countries has received increasing attention in the literature (Rapoport and Docquier, 2006; Yang, 2011; Docquier and Rapoport, 2012). In particular, many scholars have argued that remittances sent by migrants are an important element for the well-being of the household members left behind. My empirical investigation supports this hypothesis at least for some measures of economic develop-

ment. I show that households in locations that saw a relative increase in migration rates due to the arrival of the submarine Internet cable are also more likely to report to have received remittances in the following period. However, while this increase in remittances is accompanied by increased investment in secondary education of those left behind, I do not find any positive effects on household wealth. While I cannot rule out direct effects of the arrival of fast Internet on remittances, e.g., by providing a better infrastructure for money transfers, I tentatively interpret these findings as positive feedback effects of the Internet-induced increase in migration rates on the economic development in Nigeria.[2]

In sum, this paper provides novel insights into the effect of Internet usage on migration decisions, and highlights potential consequences of the spread of the Internet on the direction of international migration flows, self-selection of migrants, and the skill distribution of immigrants in receiving countries. Therefore, this paper contributes to the vast empirical literature on the determinants of migration (e.g., Hatton, 2005; Mayda, 2010; Ortega and Peri, 2013) and, in particular, to studies that investigate migration costs and the skill distribution of migrants. These studies exploit changes in migration costs due to, e.g., pre-existing migration networks (Munshi, 2003; McKenzie and Rapoport, 2007), immigration and border controls (Angelucci, 2012; Allen et al., 2018), or cultural and linguistic differences between destination and origin countries (Belot and Ederveen, 2012; Adsera and Pytlikova, 2015). In line with the application of this paper, Feigenberg (2020) highlights differential changes in migration costs for different types of potential migrants due to the United States-Mexico border fence construction that disproportionally reduces migration from low skilled migrants. This paper also relates to the literature that highlights wealth and income constraints which might deter low skilled migration (Dustmann and Okatenko, 2014; Bazzi, 2017; Cai, 2020). For instance, Angelucci (2015) shows that poor households in Mexico that experience an exogenous increase in income are more likely to migrate which worsens skills among Mexican migrants in the United States. This paper also contributes to the literature on the effect of media exposure on various socio-

[2] In a recent study, Lee et al. (2020) show that Internet based mobile technology significantly increases urban-to-rural remittances, which suggests that direct effects of the arrival of fast Internet on remittances might be a concern here.

economic outcomes.[3] In a related study, Farré and Fasani (2013) link TV usage to internal migration decisions in Indonesia and find, contrary to this study, that TV usage reduces the likelihood of migrating longer distances within Indonesia. The authors explain their finding by arguing that Indonesians, on average, over-estimate the returns to internal migration when they only have limited access to television. On the other hand, Braga (2007) finds that Albanians who were exposed to Italian television are more likely to migrate internationally. She argues that exposure to Italian television increased the availability of information about various lifestyles of societies in the Western world which might have affected the aspirations of migrants. Two other studies link advances in communication technologies to migration decisions. Lu et al. (2016) show that the installation of landline phones in rural China intensified internal migration by providing better access to information about job opportunities due to stronger migration networks and reducing the psychological costs of migrating. Aker et al. (2011) also stress the importance of information provision via mobile phones in migration decisions and show that adult education programs in Niger in which participants learned to use mobile phones led to an significant increase in seasonal migration.

The rest of the paper is organized as follows. In Section 2.2, I provide background information on international migration in Nigeria (Section 2.2.1) as well as the Nigerian Internet infrastructure and the arrival of the large submarine Internet cables starting in the year 2010 (Section 2.2.2). In Section 2.3, I present my data set. In Section 2.4, I introduce the main identification strategy. The baseline results and various robustness test are in Section 2.5, and in Section 2.6, I provide a discussion of the heterogeneous effects of Internet usage on migration. In Section 2.7, I discuss potential feedback effects of increased migration due to Internet usage, and Section 2.8 concludes.

[3]Media exposure has been linked to, e.g., educational outcomes (Gentzkow and Shapiro, 2008), crime rates (Dahl and DellaVigna, 2009), and fertility decisions (La Ferrara et al., 2012). For an extensive survey on media exposure, see DellaVigna and La Ferrara (2015).

2.2 Background

2.2.1 International migration in Nigeria

Nigeria is a multinational state in Western Africa that was formed under British colonial rule in the beginning of the nineteenth century and is today inhabited by hundreds of different ethnic groups who speak more than 500 different languages. Commonly referred to as the "Giant of Africa," it is is the most populous country in Africa with a population of more than 200 million people, which represents around one-sixth of the continent's population. It has experienced a massive population growth, having had only 45 million inhabitants in the 1960s, and due to its young population - more than 50 % of Nigerians are below the age of 18 - and high fertility rates it is projected to grow even further. The United Nations Department of Economic and Social Affairs estimates that by 2050 Nigeria's population will surpass that of the United States, and will increase to more than 700 million people by the end of the 21th century (UN DESA, Population Division, 2019). Nigeria has an abundance of natural resources and the biggest oil and natural gas resources on the African continent. It's economy and public finances depend heavily upon oil revenues. As a result, economic growth has been exceptionally volatile, especially during the oil price crises in the 1970s. Since Nigeria's independence in 1960 to the first general election in 1999, the Nigerian government was almost exclusively under the rule of military dictators. It has experienced a number of national conflicts, many of which have arisen from ethnic or religious conflicts. While a number of social and economic indicators such as life expectancy at birth and years of schooling have improved over the last two decades (UNDP, 2019), many Nigerians still suffer from extreme poverty as more than 50 % of people are below the international poverty line set by the Worldbank (WorldPovertyClock, 2020).

Due to its massive size and it's economic and demographic peculiarities, Nigeria has played an important role in African migration, and will likely do so in the years to come.[4] Over the past decades, from its independence to today, Nigeria experienced a "reverse migration transition" as it transformed from a major immigration country among Western African countries to

[4]This paragraph draws heavily on De Haas (2007) who provides an excellent short overview on the history of international migration in Nigeria.

a net emigration country (Black et al., 2004). In particular during the 1970s when Nigeria saw an increase in its oil revenues which led to higher incomes among middle class households, it attracted a large number of labor migrants from neighbouring countries. During this time, most Nigerians who left the country did so only for study or business reasons, and primarily went to the United Kingdom or the United States. Although this type of high skilled migration to Anglo-Saxon countries continued for the following decades, cross-border migration in Nigeria became more diverse, intense, and permanent during the 1980s when decreases in oil prices led to a long economic downturn and political repression and violence became more widespread (Hernández-Coss and Egwuagu Bun, 2007). High demand for less skilled workers in Western European countries such as Spain, Italy, Germany and France, the Gulf states, and more wealthy economies in Africa (e.g., South Africa, Botswana, Gabon) attracted more and more workers from Nigeria (Black et al., 2004). Increasing immigration restrictions among European countries implemented over time did not lower these migration flows but instead amplified the amount of irregular migration. While until the 1990s the majority of migrants used air links, the means of traveling changed and a considerable amount of migration from Nigeria to Europe is by now trans-Saharan (De Haas, 2006).

Today, a significant number of Nigerians who reach European terrain apply for asylum in Italy, Germany, France or the UK. Today, in absolute numbers, Nigerians are the largest group of asylum seekers from Western African countries in Europe, and the number of Nigerian asylum seekers has been growing over the years, even though rejection rates among Nigerian asylum seekers are relatively high.[5] The inflow of migrants from Nigeria to Europe might not stop any time soon as the willingness to leave the country seems still to be considerably high. In a survey conducted by the PEW Research Center, more than two-thirds of the respondent stated that they would leave Nigeria if they had the means and opportunity to do so, and

[5]Between 2010 and 2015, around 86 thousand Nigerians applied for asylum in the European Union which represents around 33 % among all applicants from the 15 Economic Community of Western African States (ECOWAS) countries. The absolute change in asylum applications from Nigerians between 2010 and 2015 is the highest among all other ECOWAS countries (23 thousand), and the relative change to the year 2010 (357 %) is only larger for Gambians (995 %) and Senegalese (657 %). Own calculations based on data from the United Nations High Commissioner for Refugees. Around 6.6 % of asylum seekers from Nigeria who received the decision about their asylum application between January and October 2019 received some form of international protection in Germany (BAMF, 2019).

almost 40 % say they plan to leave Nigeria within the next five years (Connor, 2018). Similar results can be obtained using the last wave of the Afrobarometer conducted in the year 2017 where around 25 % of the Nigerian respondents reported to have considered to move to another country, of whom 48 % are planning to move within the next two years (36 %) or are currently making preparations to move (12 %), and around 80 % would most likely go to a country outside of the African continent.

2.2.2 Nigerian Internet infrastructure and the arrival of fast Internet

The recent increase in the number of Internet users in Sub-Saharan African countries provides an interesting setting for investigating the impact of the exposure to the Internet on migration behavior. From being basically non-existent in the year 2000, the share of Internet users began slowly to grow over the following years, with approximately 7 % of individuals using the Internet in the year 2010 to around 25 % in the year 2017.[6] As the largest country in Sub-Saharan Africa, much of this growth in Internet usage was driven by Internet users in Nigeria, where latest numbers suggest that already around 125 million people use the Internet.[7]

Internet usage in Nigeria - as in other Sub-Saharan Africa countries - depends heavily on submarine Internet cables, which are globally responsible for about 99 % of international communication traffic (Brake, 2019).[8] Submarine Internet cable are fiber glass cables that are laid on the sea bed and carry telecommunication signals across oceans. They provide the necessary link between end users in Nigeria and Internet content provider that are not hosted on the African continent or, in most cases, even in Nigeria.[9] Submarine Internet cables are particularly important for African Internet users as only the minority of Internet content is hosted locally, and, in many cases, even local content is hosted overseas due to lower costs (Kende and Rose, 2015).[10] The first submarine Internet cable that provided Nigeria with a connection

[6] In comparison, the share of Internet users in the European Union (United States) grow from 20 % (43) in the year 2000 to 68 % (70) in the year 2010 and 82 % (87) in the year 2017.

[7] Source: https://www.internetworldstats.com/africa.htm.

[8] While possible, only a minority of international communication is carried out via satellites, which is slower and more expensive. An overview on the economics of Internet infrastructure gives Greenstein (2020).

[9] There is almost no connection between landline networks in Africa, which implies that international Internet traffic in Africa needs to travel overseas (Hjort and Poulsen, 2019).

[10] Chavula et al. (2014) estimate that, on average, more than 75 % of traffic that originates from African univer-

to the Internet is the *SAT-3* cable, which begins in Sesimbra, Portugal, and was connected to the port of Lagos in the year 2001 and finally started to operate in 2002 (Stanley et al., 2018). With a capacity of 340 Gbit/s, it served as the only source of Internet connectivity at this time and damages to the *SAT-3* cable were responsible for tremendous internet blackouts. For instance, when in July 2009 the *SAT-3* cable was accidentally cut, Nigeria's Telecommunication operators were forced to use satellite links to maintain Internet connectivity which reduced available bandwidth in Nigeria by 70 %, causing problems to various sectors in the economy as well as individual end users.[11]

Starting in the year 2010, four new submarine Internet cables were connected from Europe to land-based stations in Western Africa, bringing more reliable Internet connectivity, larger Internet traffic capacities, and lower bandwidth prices to end users in Nigeria. The first cable called *Main One* was connected in July 2010 to a land-based station in Lagos, Nigeria, with a capacity of 1.28 Tbit/s at time of installation that can be extended to up to 4.96 Tbis/s.[12] The availability of larger Internet traffic capacities due to the arrival of the new submarine Internet cables increased Internet speed for end user in Africa by around 35 % (Hjort and Poulsen, 2019). Additionally, the *Main One* cable had a significant impact on bandwidth prices, contributing to "an immediate drop of 50 % on the price of bandwidth in Nigeria" (Stanley et al., 2018). Lower prices and faster and more reliable Internet connection suggests that Internet usage became more convenient in Nigeria after the arrival of the *Main One* cable, with potential positive effects on Internet usage and Internet usage frequency (Hjort and Poulsen, 2019). However, it is important for the identification strategy of this paper that not all Nigerians immediately benefited from the arrival the new submarine Internet cables. After being plugged in to a land-based station, these submarine Internet cables brought faster Internet traffic capacities only to locations that are connected to the national terrestrial cable network.[13] The connection between the terrestrial cable network and the end user, the so called "last mile" technology,

sities are transmitted outside the continent.

[11]BBC News, 30 July 2009. http://news.bbc.co.uk/2/hi/technology/8176014.stm (accessed: 2020-09-09).

[12]https://www.mainone.net/our-network-3/cable-system.

[13]Internet traffic within Africa is transmitted via the telephone cables that were built many decades back (Hjort and Poulsen, 2019). In Section 2.3, I provide a discussion of the diffusion of the terrestrial cable network in Nigeria.

might be *wireline* through fiber or copper cables, or *wireless* using cell towers or satellites. Both the type of the last mile technology and the distance to the terrestrial cable network determine the experienced Internet speed of end user (Greenstein, 2020). As Internet access via mobile phones and cell towers was very rare at that time, it is likely that only locations in close distance to the terrestrial cable network benefited from the arrival of the submarine Internet cables.[14] Technical considerations suggest that fast Internet might be available only within 500 meters distance to the terrestrial cable network for last mile technologies based on copper cable (Hjort and Poulsen, 2019). However, potential spill-over effects to adjacent locations are likely as most Internet users in Nigeria had access to the Internet only via Internet cafes.[15] I provide a discussion of such potential spill-over effects in Section 2.4.

2.3 Data and descriptive statistics

The main data source of this paper is the geo-coded Nigerian General Household Survey (GHS) panel which is administered by the National Bureau of Statistics of the Federal Republic of Nigeria.[16] The first wave of the GHS panel is a sample of 5,000 households with more than 27,000 individuals based on 500 enumeration areas (communities) who were interviewed in the years 2010, 2012, and 2016. For the empirical analysis, I use a balanced sample at the household level (4,407 households) and exclude individuals who were below the age of 15 or above the age of 65 in the first wave.

To construct the main explanatory variables, I use information from the *ICT Usage Section* of the GHS panel interview questionnaire. In particular, I use the question: "Do you have access to the Internet?" to construct a binary measure of Internet usage and the follow-up question: "How often do you use the Internet?" to construct an ordinal measure of Internet usage frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once

[14] 3G services have only been introduced in 2007 in Nigeria and the share of subscribers to 3G or 4G services in 2010 was close to zero (GSMA, 2015).

[15] Around 80 % or participants in the Nigerian General Household Survey panel who reported to have access to the Internet in the year 2010 say their main access to the Internet is through Internet cafes. For studies that highlight the importance of Internet cafes in Sub-Saharan Africa, see, e.g., Mwesige (2004) or Adetoro (2010).

[16] The data is publicly available and can be obtained from the website of the World Bank Microdata Library: http://microdata.worldbank.org.

a week, 3 = daily).

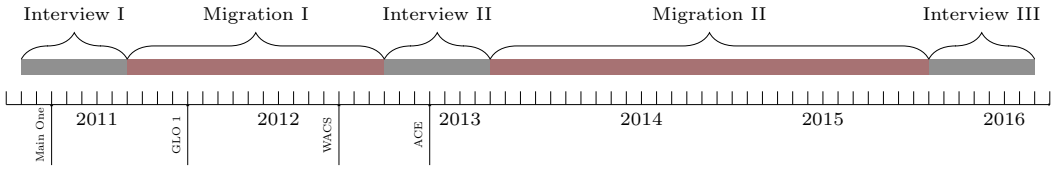
To construct the main outcome variable of this paper, I exploit the panel structure of the data. If a household member moves out of the household between two subsequent waves, the remaining household members report the whereabouts of the individual, i.e., the remaining household members state if an individual left the country or moved within Nigeria. I link the information about the migration decision provided by the remaining household members to the information the respondent provided in the previous wave. The main outcome variable *international migration* is equal to one if the individual moved out of Nigeria and zero otherwise. Note that, as the GHS panel does not provide information about migration behavior after the last (third) wave, the main empirical analysis will be based on the first and second wave of the GHS panel.

Basic background information such as age, gender, education, and other ICT usage such as TV usage and mobile phone usage is also available in the GHS panel. To obtain a measure of wealth, I follow Young (2013) in the context of developing countries, and use information about four housing conditions to obtain an ordinal measure of wealth, i.e., the wealth measure is equal to the number of housing conditions that are fulfilled (household conditions: (i) constructed floor made of other than dirt, sand or dung, (ii) flush toilet, (iii) tapped drinking water, (iv) electricity in house).

I link the data from the GHS panel with detailed maps of the terrestrial cable network prior to the arrival of the submarine Internet cables. I obtain the data for the terrestrial cable network in Nigeria from Hjort and Poulsen (2019) who use this information in a related study.[17] The GHS panel provides information about the location at the community level. I use this information as well as the information from the location of the terrestrial cable network to calculate the shortest distance between the community and the terrestrial cable. Figure A1 in the Appendix shows the terrestrial cable network (red solid lines) on a map of Nigeria as well as the communities included in the final GHS data set. Figure A1 illustrates the sparse diffusion of the terrestrial cable network in Nigeria, with many locations - in particular in Western Nigeria

[17] Hjort and Poulsen (2019) obtained the data from www.africabandwidthmaps.com and www.afterfibre.net.

Figure 1:
Time line of events: Arrival of submarine Internet cables and survey waves



Sources: Stanley et al. (2018), Hjort and Poulsen (2019), Nigerian GHS panel.

- being many hundreds of kilometres away from the terrestrial cable network. However, as shown in Figure A2 in the Appendix, which provides a magnified view on Southern Nigeria, there are also a large number of communities in the GHS panel that are relatively close to the terrestrial cable network.

In Figure 1, I provide an overview of the arrival of the submarine Internet cables in Nigeria as well as the interview periods of the Nigeria GHS panel.[18] As explained above, to analyze the effect of Internet usage on migration, I match information given by remaining household members from the subsequent interview to the socio-economic information from the previous wave. Ideally, for the empirical strategy outlined below, I would like to have at least one period before the arrival of fast Internet in Nigeria and some periods after. Unfortunately, as can be seen in Figure 1, the first submarine cable was connected already during the first interview period. If the boost in Internet speed already led to increased Internet usage in the first period in locations close to the terrestrial cable network, defining the first wave of the GHS panel as untreated in a difference-in-difference approach would contaminate the estimation results and bias the estimates towards zero. However, it seems unrealistic that the arrival of submarine Internet cables affected Internet usage behavior in such a short time period. Accordingly, and considering that in case of misclassification the difference-in-difference approach provides conservative estimates of the true effect, I define the first wave of the GHS panel as unaffected by the arrival of fast Internet in Nigeria throughout the paper.

In Table 1, I provide an overview of the selected covariates for the pooled data set by the

[18]Unfortunately, the GHS panel does not provide information about the exact day of interview.

Table 1:
Mean values of selected variables by Internet usage

	No Internet usage	Internet usage
Socio-economic		
Age	32.56	29.13
Female	0.53	0.35
Household member		
Head	0.26	0.23
Spouse	0.34	0.09
Son/Daugther	0.35	0.61
Other	0.05	0.08
Currently enrolled	0.22	0.45
Highest education		
No schooling	0.37	0.02
Some schooling	0.25	0.04
Secondary education	0.35	0.66
University degree	0.02	0.28
Number of wealth items		
0	0.25	0.02
1	0.31	0.06
2	0.28	0.31
3	0.14	0.53
4	0.02	0.09
Other ICT usage		
Television	0.51	0.97
Mobile phone	0.80	0.99
Location		
Urban	0.24	0.64
Distance next road	14.62	7.33
Internet usage frequency		
At least once a month	0.00	0.34
At least once a week	0.00	0.45
Daily	0.00	0.21
Outcome		
International Migration (in %)	0.18	1.23
Observations		
Total	20,328	1,298
Share in 2012	0.50	0.60

Note: Mean values of covariates by Internet usage. Number of wealth items based on housing conditions: Constructed floor (made of other than dirt, sand or dung), flush toilet, tapped drinking water, electricity in house.

binary measure of Internet usage. Table 1 illustrates - as expected - stark differences between Internet user and non-Internet users, while still showing at least some overlap across treatment status. On average, individuals who have access to the Internet are younger and less likely to be female, are more likely to be enrolled in school, and have a higher education at interview date. Unsurprisingly, usage of other ICT such as TV and mobile phones are more common among Internet users, and members of this group live in wealthier households which are located more

often in urban areas. Contrary to the experience in more developed countries, having access to the Internet does not mean daily usage in Nigeria in the years 2010 and 2012. Only one-fifth of the individuals of this group use the Internet daily, while around one-third access the Internet less than a week.

The lower part of Table 1 also illustrates stark differences in the propensity to migrate between Internet user and non-Internet user. Table 1 suggests that international migration is almost seven times as likely for Internet users than for non-Internet users. As shown in Table A1 in the Appendix, which reports OLS estimates of the binary variable measuring migration on Internet usage, this difference cannot be explained by the factors listed in Table 1 or unobserved location-specific heterogeneity across states or counties. However, the OLS estimates reported in Table A1 cannot be considered as causal effects of Internet usage on migration. The large differences in observable characteristics between both groups suggests that there are also considerable differences in unobservable variables between both groups that might be related to the decision to migrate, which, in turn, would bias these estimates.

2.4 Identification

2.4.1 Effect of the arrival of fast Internet on migration

The empirical approach in this paper exploits time and cross-sectional variation generated by the arrival of the first submarine Internet cable from Europe. Individuals living close to terrestrial cable network experienced significantly higher Internet speed after the arrival of the submarine Internet cable in comparison to the years before. On the contrary, individuals living in more remote locations were not affected by the arrival of the submarine Internet cable. If the availability of fast Internet affected Internet take-up and Internet usage frequency, which in turn influenced migration decisions, I expect that changes in migration rates before and after the arrival of the first submarine Internet cable are larger in locations close to the terrestrial cable network. More formally, to exploit the variation generated by the arrival of

fast Internet, I estimate the following fixed effect specification:

$$Migration_{i,c(i),t+k(t)} = \mu_{c(i)} + \beta_0 \mathbb{1}\left[t = 12\right] + \beta_1 \mathbb{1}\left[t = 12\right] * Distance_{c(i)} + v_{i,c(i),t}, \quad (1)$$

where $Migration_{i,c(i),t+k(t)}$ is a binary variable indicating whether individual i from community $c(i)$ moved to another country before time $t + k(t)$, where $k(t)$ represents the number of years between two successive waves of the Nigerian GHS panel. $Distance_{c(i)}$ is a measure of the distance of individual i's community $c(i)$ to the terrestrial cable network. $\mathbb{1}\left[\cdot\right]$ is an indicator function equal to one if the condition in the brackets is fulfilled, i.e., $\mathbb{1}\left[t = 12\right]$ is equal to one in year 2012 and zero otherwise.[19] $\mu_{c(i)}$ represents a set of community fixed effects and $v_{i,c(i),t}$ is an error term which captures all effects that influence $Migration_{i,c(i),t+k(t)}$ that are not caused by other factors included in Equation (1).

I estimate Equation (1) based on data for the years 2010 and 2012.[20] The inclusion of community fixed effects allows for any systematic time-invariant variation in migration behavior across locations due to, e.g., geographic characteristics, pre-existing migration networks, or income levels. On the other hand, the inclusion of a year dummy allows for systematic variation in migration behavior after the first and the second wave that affect all locations such as varying time windows for migration between the first and second wave and the second and third wave of the Nigerian GHS panel or time-specific economic shocks in origin or destination countries influencing migration behavior of all locations in the sample. Consequently, the coefficient of interest in Equation (1), β_1, can be interpreted as a difference-in-difference estimator. This means, β_1 measures the difference in migration rates between the time windows 2010-12 and 2012-15 between (i) communities that marginally differ in $Distance_{c(i)}$ for a continuous measure of $Distance_{c(i)}$ or (ii) connected ($Distance_{c(i)} = 1$) and unconnected locations ($Distance_{c(i)} = 0$) for a binary measure of $Distance_{i,c(i),t}$. While a binary measure facilitates the interpretation of β_1, it is challenging to determine the radius around the terrestrial cable network that defines treated and untreated locations. Based on technical con-

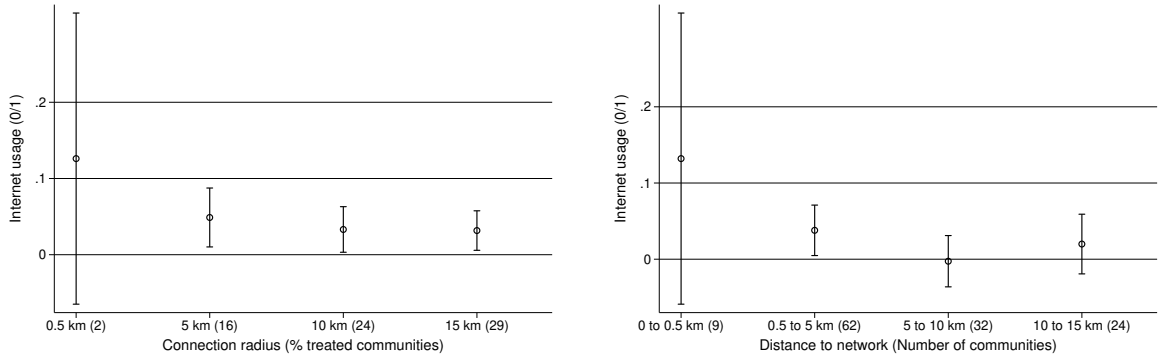
[19] All equations in this paper use the years since 2000 to refer to a given year.

[20] As explained in more detail above, information about the migration decision is always obtained from the following wave. See also Figure 1 and corresponding discussion in Section 2.3.

siderations, Hjort and Poulsen (2019) treat locations within a 0.5 km radius as connected to the terrestrial cable network. The authors argue that the 500 meters radius around the terrestrial cable network is a good proxy for the availability of fast Internet in a location as the transmission rate of "last-mile" technologies based on copper cables - which are prevalent in most African countries - become significantly lower beyond this threshold. However, the distance-connectivity relationship might differ tremendously for "last-mile" transmission via microwaves. Additionally, assuming that the availability of fast Internet is indeed restricted to buildings within a 0.5 km radius, significant spill-over effects to individuals living in adjacent locations might be likely as most Nigerian Internet users do not access the Internet via home based Internet connection. This is illustrated in Figure 2, which shows coefficient plots for regressions of Equation (1) when using individual i's Internet usage as dependent variable as well as a binary measure of distance based on various definitions of connected areas. The left plot of Figure 2 shows coefficient estimates for four separate regressions, where the binary variable takes on the value one if individual i is located either within a 0.5 km, 5 km, 10 km or 15 km radius around the terrestrial cable network. Indeed, the increase in Internet usage between connected and unconnected areas caused by the arrival of fast Internet is largest if I employ the most narrow definition for connected areas - even though insignificant possibly due to small sample size, and the effect becomes smaller if the definition becomes wider. To assess the spill-over effects, the right plot of Figure 2 shows coefficients of a similar regression but using a set of binary variables that exclusively define individuals' location (baseline: distance to terrestrial cable network larger than 15 km). This plot shows that I obtain a significant positive effect of the arrival of fast Internet also for individuals living within a corridor of 0.5 to 5 km around the terrestrial cable network, which supports the idea of positive spill-over effects to individuals living in adjacent location. However, this effect already disappears for individuals living in more remote locations.[21] To account for the unknown drop in transmission rate and potential spill-over effects, I use the logarithm of the distance to terrestrial cable

[21]The results shown in Figure 2 are qualitatively similar when using Internet usage frequency as dependent variable or restricting the sample to younger individuals, for which I expect even larger effects. See Figure A3 in the Appendix.

Figure 2:
Internet usage and distance to terrestrial cable network, 2010-12 change



Note: Plot on the left shows coefficient estimates for four separate regressions of Internet usage on an interaction term of a binary variable indicating if individual i is located in a community within the connection radius shown on the x-axis to the terrestrial cable network and an indicator variable for the year 2012. Plot on the right shows coefficient estimates for a regression of Internet usage on a set of binary variables indicating if individual i is located within a bin shown on the x-axis (baseline: Distance to terrestrial cable network larger than 15 km). All estimates include a year dummy for the year 2012 as well as community fixed effects. Number of observations: 21,626. 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster). Figure A3 in the Appendix shows equivalent plots for Internet usage frequency and a sample of younger individuals.

network for $Distance_c$. In Section 2.5.1, I show that this specification provides a reasonably good fit to the change in Internet usage and Internet usage frequency due to the arrival of fast Internet. Additionally, I report results for a binary measure of distance defining those locations as connected that are within a 5 km radium around the terrestrial cable network.

The arrival of fast Internet affected those individuals stronger who live closer to terrestrial cable network. Hence, if the arrival of faster Internet has a positive effect on Internet usage and subsequently on migration, I expect β_1 to be negative when using the continuous measure of distance and positive in the case of the binary measure. Moreover, I expect the effect to be significantly larger for younger individuals as they are more likely to change their Internet behavior as a response to faster Internet. For this reason, I also report results for the estimation of Equation (1) on a sub-sample of individuals between the age of 20 and 35.

If the common trend assumption is satisfied, β_1 measures the effect of the exposure to fast Internet on migration decisions for individuals located in connected areas. The common trend assumption states that migration rates in communities would have evolved similarly across locations between the time windows 2010-12 and 2012-15 if the submarine cables had not been

connected to Nigeria (see, e.g., Abadie, 2005). In Section 2.5.1, I provide an extensive discussion of possible violations of this parallel trend assumption, as well as of the robustness of my estimation strategy to various specification. I show that my estimation results are robust to (i) using a binary measure of treatment defining all locations as connected if they are within a 5 km radius around the terrestrial cable network, (ii) the exclusion of locations that are close or far away from treated locations, and (iii) the inclusion of various placebo treatments based on interactions of $\mathbb{1}[t = 12]$ with various pre-treatment level-differences between treated and untreated locations. Additionally, I assess pre-treatment trends of the outcome variable between connected and unconnected locations.

2.4.2 Effect of Internet usage on migration

As the Nigerian GHS panel provides information about Internet usage, I also show instrumental variable estimates of the effect of Internet usage on the decision to migrate, exploiting the plausible exogenous variation in Internet usage caused by the arrival of fast Internet. To ease the motivation of this approach, I assume to have a binary measure of distance to the terrestrial cable network that assigns individuals to locations that either receive fast Internet after the arrival of submarine Internet cables (connected locations) or not (unconnected locations). The share of Internet users increased due to the arrival of submarine Internet cables more in connected locations than in other unconnected locations. However, the arrival of fast Internet did not incentivize all individuals located in connected areas to use the Internet, and some individuals in unconnected locations might have started to use the Internet for other reasons. Assuming that the increase in Internet usage is larger in connected locations than in unconnected locations, such a fuzzy design still allows to estimate a local average treatment effect (LATE) for a subpopulation of my sample. Following De Chaisemartin and d'Haultfoeuille (2018), this subpopulation consists of *switchers* in connected locations, i.e., individuals located close to the terrestrial cable network (connected locations) that started to use the Internet only after the arrival of the submarine Internet cables. Under additional assumptions stated below, I can obtain the LATE for switchers in connected locations with the following two stage least

squares (2SLS) estimation:

$$Internet_{i,c(i),t} = \mu^F_{c(i)} + \beta^F_0 \mathbb{1}\left[t = 12\right] + \beta^F_1 \mathbb{1}\left[t = 12\right] * Distance_{c(i)} + e_{i,c(i),t}, \tag{2}$$

$$Migration_{i,c(i),t+k(t)} = \mu^S_{c(i)} + \beta^S_0 \mathbb{1}\left[t = 12\right] + \beta^S_1 \hat{Internet}_{i,c(i),t} + \epsilon_{i,c(i),t}, \tag{3}$$

where $Internet_{i,c(i),t}$ is a either a binary variable indicating Internet access of individual i in community $c(i)$ at time t or an ordinal measure of Internet usage frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). $\hat{Internet}_{i,c(i),t}$ is the predicted value of Internet usage or Internet usage frequency based on the estimated parameters of Equation (2). $e_{i,c(i),t}$ and $\epsilon_{i,c(i),t}$ are error terms intended to capture effects on the outcome variables that are not caused by factors included in Equation (2) or (3), respectively. All other variables are defined as before.

Equation (2) is the first stage equation and provides estimates of the impact of the arrival of fast Internet on Internet usage or Internet usage frequency. As in the reduced form equation explained above, Equation (2) includes a set of community fixed effects ($\mu^F_{c(i)}$) that allow for systematic time-invariant variation in $Internet_{i,c(i),t}$ across locations as well as a year dummy ($\mathbb{1}\left[t = 12\right]$) to control for an average time trend in Internet usage affecting all locations. The coefficient of interest in this 2SLS procedure is β^S_1 which measures the effect of Internet usage on the migration decision. Estimating β^S_1 by a 2SLS procedure as outlined in Equations (2) and (3) is equivalent to running two separate regressions of Equations (1) and (2) and dividing the estimated parameter β_1 from the reduced form equation by β^F_1 from the first stage equation.

De Chaisemartin and d'Haultfoeuille (2018) outline the assumptions when such a Wald difference-in-difference estimator represents an estimate of the LATE for switchers in connected locations in a fuzzy design. First and most importantly, the common trend assumption needs to be fulfilled. In this setting, the common trend assumption states that the migration rates would have evolved similarly if the share of Internet users had not expanded differently across locations. This is a more restrictive assumption than the one stated above as it assumes that there is no growth (or decline) in other observable or unobservable factors - except of In-

ternet usage - between the two time periods that influence migration rates while the previous common trend assumption allows for changes in factors that influence migration rates if they are caused by the arrival of fast Internet.[22] On the other hand, this assumption is less restrictive than requiring the common trend assumption from the reduced form model also on the treatment variable Internet usage. The share of Internet users are allowed to evolve differently across locations even in the absence of the arrival of fast Internet if the factors determining the different time trends - such as, e.g., previous share of Internet users - do not affect migration rates.

De Chaisemartin and d'Haultfoeuille (2018) further show that the identification of the LATE for switchers in connected locations requires that (i) the treatment effect of Internet usage on migration is stable over time for all individuals and (ii) the treatment effect of Internet usage on migration is homogeneous for switchers in connected and unconnected locations. In particular, the first assumption seems to be problematic as the information provided by the Internet changes over time which might also lead to heterogeneity of the effect of Internet usage on migration. However, the time periods considered here are rather small, suggesting that this might not be a major concern.

I provide a discussion of possible violations of the common trend assumption as well as the robustness of my estimation strategy to various specifications in Section 2.5.2. In addition to the sensitivity checks mentioned in Section 2.4.1, I provide a thorough discussion of possible violations of the exclusion restriction of the instrument used in the empirical analysis.

2.5 Results

2.5.1 Reduced form estimates

In this section, I provide a discussion of the results for the reduced form relationship between migration decisions and the arrival of fast Internet in Nigeria, and start with a visual inspection of the correlation between migration rates, Internet usage and the distance to the terrestrial

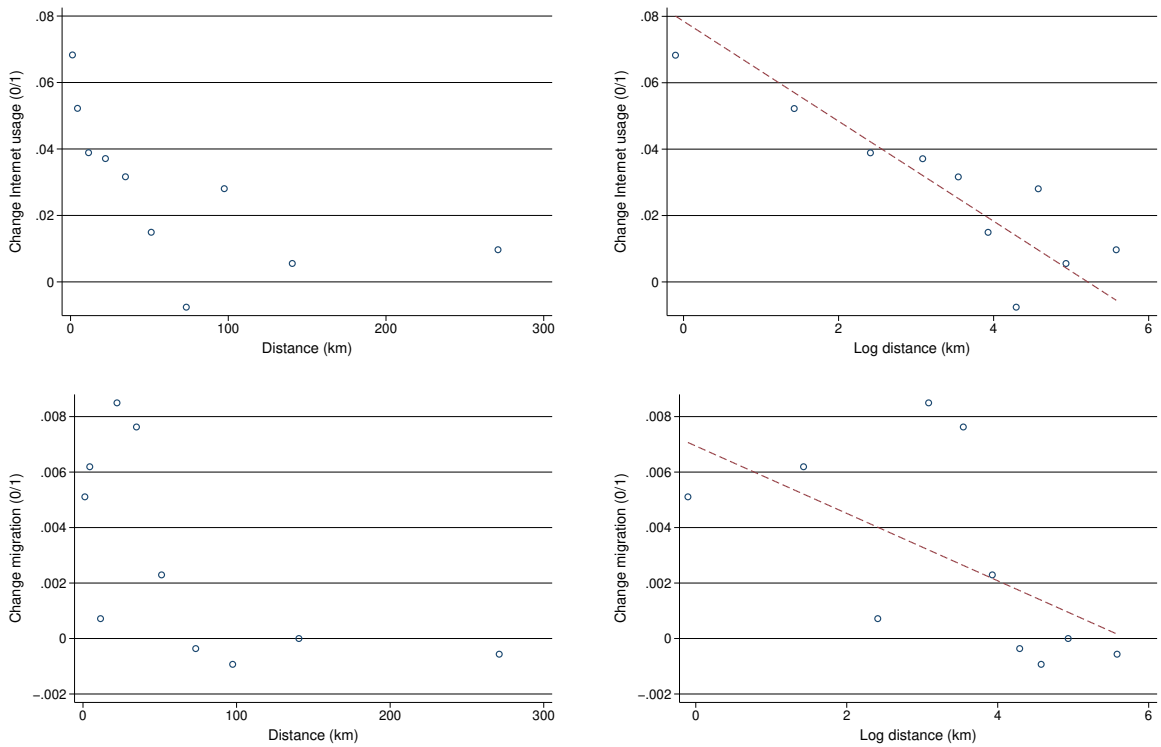
[22] Hence, the common trend assumption stated in this subsection involves an exclusion restriction on the excluded instrument, $\mathbb{1}[t = 12] * Distance_{c(i)}$, the interaction of the year dummy with the distance to the terrestrial cable network measure.

cable network. As the distance to the terrestrial cable network is measured at the community level, estimations of Equation (1) (and (2)) based on individual times year level data are very similar to regressions based on community times year level data where the outcome variable is the share of individuals who migrate (or use the Internet) in a community and year.[23] Additionally, having only two time periods for the main part of the analysis, the fixed effect specification gives equivalent results to a cross-sectional regression based on first differences. Therefore, plotting changes of mean migration rates or mean Internet usage in a community over the two time periods against the distance to the terrestrial cable network provides a clear visual representation of the reduced form relationships discussed above. Figure 3 shows binned scatter plots (10 equally sized bins) of the change in the average Internet usage (upper two plots) and migration rates (lower two plots) in a community between the treated year 2012 (after the arrival of the submarine cables) and the untreated year 2010 (before the arrival of the submarine cables) and the distance to the terrestrial cable network. The distance to the terrestrial cable network is measured in kilometres (km) in the plots on the left and in log in the plots on the right.

The arrival of fast Internet by the connection of the submarine Internet cables was transmitted by the terrestrial cable network and brought faster speed and traffic capacities to locations close to the pre-existing terrestrial cable network. In line with this idea, I expect larger changes in Internet usage and migration rates in communities closer to the cable network. Focusing first on the two plots on the left, Figure 3 depicts a striking relationship between the change in mean Internet usage and migration rates before and after the arrival of fast Internet in Nigeria and the distance to the cable network. Positive changes are more present in communities that are closer to the terrestrial cable network than in those further away. Interestingly, the relationship seems to be linear if we consider the distance in log instead of the absolute value in km, which motivates the use of log distance as a continuous measure of the distance to the terrestrial cable network in the analysis below. Further, Figure A4 in the

[23] In fact, both approaches are equivalent if weights that account for the number of observations in a community and year are applied. The results of this paper do not change if I run the entire empirical analysis based on community times year level data. I refrain from this approach because using individual level data allows me to easily control for individual-specific covariates.

Figure 3:
Internet usage, migration, and distance to terrestrial cable network, 2010-12 change



Note: Binned scatter plot (10 equally sized bins) of difference in community mean Internet usage (top) and migration (bottom) between 2012 and 2010 and distance to terrestrial cable network in kilometres (left) and logartihmized distance to terrestrial cable network (right). 435 communities included. Figure A4 in the Appendix shows equivalent plots for Internet usage frequency.

Appendix shows that the arrival of fast Internet did not only affect Internet take-up rates, but also the frequency with which individuals are using the Internet.

Table 2 reports the reduced form estimates of Equation (1). Each column of Table 2 reports the result of a separate regression of a binary variable indicating whether an individual migrated to another country and the log distance to the terrestrial cable network interacted with a year dummy for the post treatment period in the year 2012. In all regressions, I add community fixed effects as well as a dummy variable for the year 2012. In columns (2) and (4) I additionally control flexibly for a large set of control variables that is listed in Table 1. Columns (1) and (2) refer to the overall sample and columns (3) and (4) to a restricted sample including only individuals between the age of 20 and 35. For all specifications, Table 2 reports Huber-White (robust) standard errors clustered at the community level in parentheses, which

Table 2:
Reduced form estimation: Migration on distance to terrestrial network

	(1)	(2)	(3)	(4)
Log(Distance to network) * Year 12	-0.0014***	-0.0014***	-0.0024***	-0.0024***
	(0.0005)	(0.0005)	(0.0009)	(0.0009)
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

allows for arbitrary correlation of the error term within a community to take into account serial correlation of the error term, which would otherwise severely downward-bias standard errors (Bertrand et al., 2004).

All four specifications of Table 2 show a positive effect of the arrival of fast Internet on migration. The interaction term reported in Table 2 indicates that communities close to the terrestrial cable network experienced a larger increase in migration rates that those farther away. The effect is virtually unaffected by the inclusion of control variables. For the entire sample, I estimate that a 100 % increase in the distance to the terrestrial cable network reduces the migration rate by around 0.14 percentage points or around 17.7 % relative to the sample mean in connected areas in the year 2012 (Distance < 5 km; mean: 0.79 %). The effect almost doubles in absolute terms if I restrict the sample to younger individuals, which supports the hypothesis that younger individuals are more responsive in Internet take-up to the availability of fast Internet, which then results in a higher change in the probability to migrate. A 100 % increase in the distance to the terrestrial cable network reduces the migration rate by around 0.24 percentage points. However, taking into consideration the higher likelihood of migrating for younger individuals irrespective of the exposure to faster Internet, the effect is surprisingly similar at around 16.5 % relative to the sample mean in connected areas in the year 2012 (mean:

1.45 %).

I provide a number of robustness and sensitivity tests in the Appendix. I use the absolute value in km instead of the logarithm of the distance and my results are qualitatively identical (Table A2). For this specification, my estimates suggests that an increase in the distance to the terrestrial cable network by 10 km reduces the change in the migration rate induced by the arrival of the submarine Internet cables by around 0.026 percentage points or around 10 % relative to the sample mean for the entire sample.

Table A3 reports results for a binary measure of distance. Communities are defined as being connected if they are within a 5 km radius around the terrestrial cable network. Panel A of Table A3 reports estimation results for the entire sample and Panel B for the restricted sample of younger individuals. In the first column of Table A3, I estimate that the availability of fast Internet led to an increase in the migration rate by around 0.52 percentage points in communities close to the terrestrial cable network. In columns (2) to (4), I step-by-step exclude observations that are close to the connected area. If the point estimate changes by this procedure, this would indicate that the binary variable has been defined too narrowly. The downside of this procedure is that I possibly exclude individuals from locations that are more comparable to those in the treated areas. Nevertheless, it is comforting to see that the point estimate is hardly affected by the exclusion of these locations.

In Table A4, I exclude observations from remote communities. While there are arguments for including more remote locations in the sample as they are presumably less likely to be affected by the arrival of fast Internet than unconnected areas that are closer to the treated area, these locations might differ more from communities within the 5 km threshold around the terrestrial cable network. The estimates shown in Table A4 make clear that my findings are not driven by the inclusion of more remote areas. While some of the estimates become insignificant when excluding remote communities from my sample - likely due to the smaller sample size - the point estimates are remarkably unaffected.

A large part of the terrestrial cable network connects larger cities and runs through more urbanized areas. As a consequence, locations close to the terrestrial cable network and those

further away differ in terms of a number of characteristics. For instance, as shown in Table A5, which reports mean values of selected community characteristics, locations close to the terrestrial cable network are more likely to be in the large cities Lagos or Abuja or, in general, in urban areas. Unsurprisingly, these locations differ also with respect to socio-economic characteristics of the individuals living in the community such as the share of individuals who are college educated or were Internet user already before the submarine cables arrived. While the fixed effect specification applied in the paper takes into account level differences between treated and untreated communities, I might wrongly attribute growth in migration rates to the arrival of fast Internet if locations in these more urbanized areas experienced faster growth in migration rates, irrespective of whether they are close to the terrestrial cable network or not. To see if this was the case, I construct placebo treatment variables by interacting the selected community characteristics listed in Table A5 with a dummy variable for the year 2012 which indicates the post-treatment period - similar to the treatment variable in Equation (1). Table A6 reports estimation results when these placebo treatment variables are added to the estimation equation. The estimated effect of the arrival of fast Internet is basically unchanged and the estimated coefficients on the placebo treatments are mostly not significant. An exemption is the placebo treatment for the large cities Lagos and Abuja in the restricted sample for younger individuals. However, since the estimated placebo effect is negative - which indicates that locations in these cities saw a smaller growth in migration rates between 2010 and 2012 - the baseline estimates presented above might even be underestimated.

A standard approach in the literature to assess the validity of difference-in-difference estimations - as applied in this paper - is to evaluate the trend in the outcome variable by treatment status for the pre-treatment period. If migration rates in connected and unconnected communities followed a similar trend before the arrival of fast Internet in Nigeria, I would expect migration rates to evolve similarly for both groups if the submarine Internet cables would not have been connected. Unfortunately, the Nigerian GHS data does not provide more than one pre-treatment period to apply such a check. Moreover, I am not aware of any other data set that provides information about geographical location and migration rates at the community

level. However, to still be able to address concerns regarding the violation of the common trend assumption, I utilize information about remittances that remaining households receive. It is likely that migration rates in one period are correlated with the share of households that receive remittances in the following period. Hence, if migration rates in treated and untreated communities followed a similar trend before the arrival of fast Internet, I would expect that the difference in the share of households that receive remittances between households in connected and unconnected communities is similar in the years before the migration rates increased differently due to the connection of the submarine Internet cables. The time lag between migration and remittances allows me to evaluate the pre-treatment trend based on two periods as growth of migration rates in connected communities occurred only after the second wave in the year 2012. As a consequence, the difference in the share of households that receive remittances in connected and unconnected locations should not be different in the years 2010 and 2012. To assess if this is the case, Figure A5 shows an "event-study plot" for the differences in remittances between connected and unconnected locations by year. Indeed, as expected, I do not see a significant difference between the differences in remittances for locations close to the network and those further away in the year 2010 and 2012 but only in year 2015.

2.5.2 Instrumental variable estimates

Having established that the arrival of fast Internet in Nigeria led to an significant increase in migration rates among locations close to the terrestrial cable network in comparison to more remote locations, I move on to investigate the impact on Internet usage on migration in the structural model explained in Section 2.4.2, exploiting the variation in Internet usage induced by the arrival of submarine Internet cables. Table 3 reports first stage and second stage estimates of the effect of Internet usage and Internet usage frequency on migration rates for the overall sample (columns (1) and (2)) and the restricted sample of younger individuals (columns (3) and (4)). Columns (2) and (4) refer to estimations where I additionally include the set of control variables listed in Table 1 of Section 2.3. The first two rows of Table 3 show the first stage estimates for the respective endogenous variable, i.e., the effect of the

Table 3:
Instrumental variable estimation: Internet usage on migration

	(1)	(2)	(3)	(4)
First-stage estimates				
Internet usage				
Log(Distance to network) * Year 12	-0.014**	-0.014**	-0.023***	-0.021***
	(0.006)	(0.006)	(0.006)	(0.005)
F statistic	6.06	6.28	16.19	14.06
Internet usage frequency				
Log(Distance to network) * Year 12	-0.029***	-0.029***	-0.047***	-0.044***
	(0.009)	(0.009)	(0.010)	(0.010)
F statistic	9.72	10.26	21.58	18.89
Second-stage estimates				
Internet usage	0.096*	0.098*	0.105**	0.116**
	(0.056)	(0.056)	(0.048)	(0.054)
Internet usage frequency	0.047*	0.048**	0.050**	0.055**
	(0.024)	(0.024)	(0.022)	(0.024)
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

distance to the terrestrial cable network times the 2012 year dummy on Internet usage and the ordinal measure of Internet usage frequency, respectively. The third and forth row report the corresponding 2SLS estimate. In all specifications, I include a year fixed effect for the year 2012 as well as community fixed effects. Again, the reported robust standard errors in parentheses are clustered at the community level.

Focusing first on the first-stage estimates, Table 3 makes clear that the arrival of fast Inter-

net led to a significant change in the share of Internet users among communities that depends on the distance to the terrestrial cable network. As expected, locations closer to the terrestrial cable network saw a larger increase in the share of Internet users by the availability of fast Internet than more remote locations. This effect is significant for both, the entire sample as well as the restricted sample, and is not affected by the inclusion of control variables. My estimates indicate that an increase in the distance to the terrestrial cable network by 100 % reduces the share of Internet users by 1.4 percentage points for the entire sample and by 2.1 percentage points for younger individuals. The effect on the ordinal measure of Internet usage frequency is even larger. Here my estimates suggest that the arrival of fast Internet reduced the change in Internet usage frequency by around 2.9 or 4.4 percentage points of the ordinal measure of locations that are as twice as far away than other locations. Table 3 also reports the F-statistic on the reported instrument which is, in general, larger for the subsample of younger individuals and for the ordinal measure of Internet usage frequency, and in most instances above the rule-of-thumb-threshold of 10.

Turning next to the results of the second-stage estimation in the third and fourth row of Table 3, the reported results show a positive effect of Internet usage and Internet usage frequency on migration. Again, the estimates are hardly affected by the inclusion of socio-economic control variables. For the entire sample, I estimate that Internet usage increases the likelihood of migrating by around 9.8 percentage points. In terms of Internet usage frequency, my results suggest that an increase in Internet usage frequency by one increases the probability to migrate by 4.8 percentage points. Interestingly, the estimates do not differ for the unrestricted and restricted sample, which suggests that the effect is similar among older and younger individuals.

Again, I provide a number of sensitivity and robustness checks in the Appendix. Table A7 reports estimates when using a binary measure of distance to the terrestrial cable network instead of a continuous measure. As before, I define a location as being treated if it is within a 5 km radius around the terrestrial cable network. In the first column of Table A7, I use the entire sample for my estimations. Panel A refers to the full sample and Panel B shows the

same estimates for the restricted sample of young individuals. Focusing first on the first-stage estimates in column (1), Table A7 indicates that the arrival of fast Internet led to an increase in the share of Internet users by 5 percentage points among communities within the 5 km radius in comparison to those locations further away. To put this estimate into perspective, the share of individuals using the Internet in the year 2010 in the treated communities was at around 10 %, which suggests that the arrival of fast Internet increased Internet usage by 50 % among those individuals. Further, the corresponding second-stage estimates are remarkably similar to the baseline estimates in Table 3, which suggests that my baseline estimates are indeed driven by the boost in Internet usage among individuals located close to the terrestrial cable network. In columns (2) to (4), I step-by-step exclude observations that are close to the treated area as in the sensitivity analysis of the reduced form estimation. Again, I do not see large differences among the point estimates for the different samples, which confirms that the chosen 5 km corridor is not too narrow.

Table A8 reports estimates for the binary instrument when I exclude remote locations. Again, the idea is that locations far away from the treated communities are less comparable to the treated group. This might be of particular importance for Internet usage, where remote locations might lack any infrastructure that could help to access the Internet. It is remarkable that the point estimates of the first-stage effect do hardly differ when using the entire sample (column (1)) or the restricted sample that includes only observations within a 10 km radius around the terrestrial cable network. For this sample, the share of Internet users is very similar before the arrival of the submarine Internet cable (connected: 10 %, unconnected: 8 %; restricted: 13 % and 11 %). Again, this is very strong supporting evidence that the arrival of fast Internet was the driving force behind the increase in the share of Internet users.

In Table A9, I report results for the 2SLS when I additionally add the placebo treatment variables introduced above. The idea of this procedure is to allow for different growth paths among locations with particular characteristics more common in connected communities which results from the specific distribution of the terrestrial cable network in Nigeria. Focusing first on columns (1) to (5) where I step-by-step include separately the placebo-treatment variables,

Table A9 makes clear that the largest effect of these placebo treatments can be found for binary variables indicating whether an individual resides in Lagos or Abuja or in urban areas. In both instances, I see a drop for the first-stage estimate, which increases the effect of Internet usage on migration, in particular for the Lagos or Abuja place treatment variable as the reduced form effect became even larger in this case. In all other cases, the second-stage effect remains robust to the inclusion of the placebo treatment. In sum, this suggests that my baseline estimates might even be on the lower bound of the effect of Internet usage on migration decisions.

Finally, I address concerns regarding the exclusion restriction of the excluded instrument in the instrumental variable estimation. As explained above, to interpret the second-stage estimate of Internet usage on migration as causal, the exclusion restriction for the excluded instrument needs to be fulfilled. This implies for my difference-in-difference set-up that there was no growth (or decline) in other factors between 2010 and 2012 that differed between connected and unconnected locations that might have impacted migration rates. Indeed, in a related study using a similar identification strategy, Hjort and Poulsen (2019) show that the arrival of the submarine Internet cables positively influenced employment rates in locations close to the terrestrial cable network in comparison to those further away. This might be problematic for my analysis as employment possibilities might influence migration decisions. For instance, better employment possibilities might reduce the incentives to migrate as individuals might be able to obtain a higher income level at their home location. On the contrary, finding employment might lower liquidity constraints that kept potential migrants from migrating. However, Hjort and Poulsen (2019) conduct their analysis based on different data and on a larger set of Sub-Saharan countries. Further, Hjort and Poulsen (2019) do not show how the results are driven by particular countries, in particular how important the arrival of the submarine Internet cable was in Nigeria. Hence, to alleviate concerns that I missattribute the change in migration rates to an increase in Internet usage instead of better employment perspectives, I investigate to what extend the arrival of fast Internet changed employment status among individuals in connected areas in Nigeria. Columns (1) and (2) in Table A10 in the Appendix report reduced form estimates of Equation (1) when I use the employment status

of an individual as dependent variable. The size and the signs of the obtained point estimates suggest that there was a positive effect of the arrival of fast Internet on employment status in Nigeria, too. However, the point estimates are estimated very imprecisely and are significantly different from zero only for the entire sample and the continuous measure of distance to terrestrial cable network. In columns (3) and (5) I add individuals' employment status to the baseline estimation equation and obtain estimates for the effect of the arrival of fast Internet that are essentially the same as in Table 3 and Table A7 in the Appendix. This is comforting as it shows that the interpretation of my estimates does not change if they are conditional on employment status. Next, I exclude all locations within a radius of 5 km of the terrestrial cable network that saw an increase in the mean employment rate between 2010 and 2012.[24] If my estimates are not affected by this restriction, this suggests that employment growth is not driving my results. Columns (4) and (6) of Table A10 show estimation results when excluding these observations. Remarkably, while the estimate loses precision for the binary measure of distance, the point estimates are hardly affected by the exclusion of these locations. In sum, I conclude that employment growth caused by the arrival of fast Internet is unlikely to be explaining my results.

Additionally, I check whether there was a change in other factors between 2010 and 2012 that differed between connected and unconnected locations that might potentially have impacted migration rates. In doing so, I run reduced from regressions for Equation (1) on the full set of control variables as listed in Table 1. The estimation results of this procedure are reported in Table A11. The first two columns of Table A11 refer to estimations when I use the continuous measure of distance and the last two columns to estimations when the binary measure of distance is included. Table A11 makes clear that there were hardly any systematic changes in the difference of these variables between connected and unconnected locations over time. Almost all estimates are insignificant or inconsistently estimated for the two measures of distance.[25] If anything, Table A11 suggests that there was a slight change from individuals with

[24]To identify these locations, I first run a regression of employment status on a year dummy for the year 2012 and use the resulting residuals to calculate mean employment rates for each community. This allows me to abstract from any year specific effects on employment that affected all locations.

[25]As for the baseline analysis, the estimated effect should be in different directions for the continuous measure

no schooling to some schooling and a reduction in mobile phone usage among individuals in connected areas in comparison to those in unconnected areas, with the latter suggesting that there might have been a substitution between between different types of ICTs. Overall, the results do not suggest that any of these factors might be explaining the change in migration rates.

2.6 Discussion

2.6.1 Migration out of Africa

As the previous section has shown, the arrival of fast Internet in Nigeria has caused an increase in migration rates, which was likely driven by increased Internet usage of individuals living close to the terrestrial cable network. In this subsection, I investigate whether increased migration due to Internet usage in Nigeria rather affected migration within the continent or migration out of Africa. An often overlooked fact is that a large share of international migration in Africa is within the continent. Recent estimates suggest that only roughly one-half of Africans who cross borders move out of the continent (International Organization for Migration, 2018). Nonetheless, the share of migrants living outside of Africa has experienced a sharp increase since 1990, when the same number was only at around 30 %. It seems plausible that Internet usage has stronger effects on migration out of Africa. First, if Internet usage affects migration behavior by providing potential migrants with information about destination countries which are not available for non-Internet users, I would expect the effect of Internet usage on migration to be larger if destination countries are (culturally) more remote. Similarly, if the exposure to foreign media transmitted via the Internet - which is dominated by images of Western lifestyle - changed individuals' preferences, I would also expect the effect of Internet usage to be larger for extra-continental migration.

Table 4 shows Instrumental variable estimation results for different definitions of my dependent variable for the restricted sample of young individuals.[26] The first (second) row of

and the binary measure.

[26]Table A12 in the Appendix reports estimates for the entire sample. The results are qualitatively the same.

Table 4:
Instrumental variable estimation: Migration out of Africa

	(1)	(2)	(3)	(4)	(5)	(6)
Internet usage	0.116**	0.070*	0.016			
	(0.054)	(0.039)	(0.013)			
Internet usage frequency				0.055**	0.033*	0.008
				(0.024)	(0.017)	(0.006)
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable:						
All migration	Yes	No	No	Yes	No	No
Migration out of Africa	No	Yes	No	No	Yes	No
Migration within in Africa	No	No	Yes	No	No	Yes
Observations	8,963	8,957	8,957	8,963	8,957	8,957
Cluster	435	435	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable is a binary variable indicating if an individual migrated to (1) another country, (2) out of Africa, and (3) within Africa (migration out of Africa is coded 0 in this case). Please note that the number of observations declines due to missing information about the destination country. Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 shows result for the effect of Internet usage (Internet usage frequency) on various measures of migration. Columns (1) and (4) show the baseline results from above for Internet usage and Internet usage frequency, respectively, where the dependent variable is a binary variable indicating whether an individual moved to another country. In columns (2) and (5), the dependent variable is equal to one if the individual moved to a country outside of Africa and zero otherwise, and column (3) and column (4) refer to estimation results when the dependent variable is equal to one if the individual moved to another country within Africa and zero otherwise. Note that I do not have information about the destination country for some migrants, so the number of observations differ between the first and second or third column (forth and fifth or sixth column for Internet usage frequency).

Table 4 supports the idea that the effect of Internet usage on migration is larger for migra-

tion out of Africa. While I find positive effects of Internet usage on both outcome variables, the point estimate for migration within Africa is much smaller and not significant.

2.6.2 Internet usage, migration, and wealth

In this subsection, I provide estimation results to show whether the effect of Internet usage on migration depends on the wealth of potential migrants. Assuming that poverty is inversely related to skills (Angelucci, 2015), effect heterogeneity with respect to wealth implies that the spread of the Internet induces a change in the skill distribution of migrants. The self-selection and resulting skill distribution of immigrants is a major topic in the economic literature since Borjas (1987). While traditional approaches in the migration literature treated immigration as a change in the quantity of homogeneous labor supply, more recent work consider immigration within a framework of heterogeneous labor supply (e.g., Peri, 2016). A consequence of this approach is that the effect of immigration on the wages of natives depends on the skill distribution of immigrants and the elasticity of substitution across skill groups. Hence, it is important to know to what extent Internet usage facilitates migration of different skill groups. It is plausible to assume that the reduction in the costs of migrating due to Internet usage is larger for individuals from the lower part of the skill distribution. For instance, it is likely that the set of information that became available by using the Internet is larger for individuals from the lower part of the skill distribution, which suggest that responses due to Internet usage should also be larger for this group. Alternatively, studies have shown that migration from low-skilled individuals is often financially constrained (Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010; Bazzi, 2017). If access to the Internet reduces the cost of migrating, these previous constraints might be relaxed and low-skilled migration might become more common.

Table 5 reports estimates of the baseline 2SLS approach for subgroups defined by the relative household wealth for the sample of younger individuals.[27] I define individuals as being at the lower bound of the wealth distribution if the number of wealth items of their household

[27]Table A13 in the Appendix reports the same specifications for all individuals. Table A14 in the Appendix reports results when the endogenous variable and the excluded instrument is interacted with a binary variable indicating the low and high household wealth. The results are qualitatively similar to the one presented here.

Table 5:
Instrumental variable estimation: Relative wealth

	(1)	(2)	(3)
First-stage estimates			
Internet usage			
Log(Distance to network) * Year 12	-0.021***	-0.018**	-0.021***
	(0.005)	(0.007)	(0.007)
F statistic	14.06	6.36	10.02
Internet usage frequency			
Log(Distance to network) * Year 12	-0.044***	-0.041***	-0.048***
	(0.010)	(0.016)	(0.015)
F statistic	18.89	6.77	10.06
Second-stage estimates			
Internet usage	0.116**	0.209**	0.038
	(0.054)	(0.105)	(0.033)
Internet usage frequency	0.055**	0.091**	0.017
	(0.024)	(0.046)	(0.015)
Year 12 FE	Yes	Yes	Yes
Community FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes
Restricted: Low wealth	No	Yes	No
Restricted: High wealth	No	No	Yes
Observations	8,963	3,925	5,038
Cluster	435	414	432

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

is below the mean of the number of wealth items in the respective community in which they are living. Column (1) reports results for the overall sample and column (2) and column (3) report estimation results for the two subsamples. As for the baseline estimates, Table 5 shows results for two measures of Internet usage - Internet up-take and Internet usage frequency.

Interestingly, the first-stage effects do hardly differ between the two subgroups. The re-

ported F-statistic is slightly smaller for individuals at the lower bound of the wealth distribution, which might be the result of the smaller sample size, but the point estimates are very similar for low and high wealth individuals. However, the second-stage estimate differs considerably between both groups. While I do observe a positive effect of Internet usage for both subgroups, the effect is much smaller and not significantly different from zero for the sample of individuals with higher wealth. These results suggest that there is considerable effect heterogeneity of Internet usage on migration which might hint to a potential negative selection of migrants with respect to their skills.

2.7 Implications for remittances and economic development

The impact of international migration on the development in sending countries has received increasing attention in the literature. In particular, many scholars have argued that remittances sent by migrants are an important element for the well-being of the household members left behind. Rapoport and Docquier (2006) list a number or reasons why migrants send money back home, including altruism, repayment of loans to finance migration, or insurance and strategic motives. Such remittances contribute to the household income of the family members left behind and relax their liquidity constraints that might enable household members to purchase essential goods and escape poverty or to undertake investments in, e.g., businesses or children's education.[28]

If the exposure to fast Internet and the resulting increase in Internet usage affected migration rates in Nigeria, I would also expect a different development of the share of households that receive remittance depending on the distance to the terrestrial cable network after the migration decision. Similarly, remaining household members might use remittances to invest in various outcomes which might, again, lead to different development paths between locations close to the terrestrial cable network and those more remote. I empirically analyze if such

[28]The literature has examined the effect of remittances on a variety of outcomes. Common outcomes include income and measures of poverty, health, education, and asset ownership. For an overview, see, for instance, Gibson et al. (2011).

Table 6:
Feedback effects: Migration and economic development

	(1)	(2)	(3)	(4)
$\mathbb{1}$(Distance to network $<$ 5km) * Year 15	0.029**	-0.077	0.022	0.093*
	(0.014)	(0.056)	(0.021)	(0.055)
$\mathbb{1}$(Distance to network $<$ 5km) * Year 10	0.003	0.027	0.009	-0.028
	(0.005)	(0.046)	(0.024)	(0.063)
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Dependent variable:				
Remittances	Yes	No	No	No
Wealth items	No	Yes	No	No
Share HH member enrolled (age 10 to 18)	No	No	Yes	No
Share HH member enrolled (age 15 to 18)	No	No	No	Yes
Observations	10,414	10,414	8,031	3,196
Cluster	436	436	435	432

Note: Estimate of various outcome variables on interactions between a binary variable indicating whether a household is located within a 5 km radius around the terrestrial cable network and year dummies for the year 2010 and 2015. Number of observations is smaller in the third and fourth column as not all households have children in the depicted age bracket. Robust standard errors clustered at the community level in parentheses. Significance levels: * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01.

different growth paths are observable in my sample by estimating the following equation:

$$y_{h,c(h),t} = \mu_c + \sum_{j \in \{10,15\}} \{\beta_{0,j} \mathbb{1}[t=j] + \beta_{1,j} \mathbb{1}[t=j] * \mathbb{1}\left[Distance_{h,c(h)} < 5km\right]\} + u_{h,c(h),t}, \quad (4)$$

where $y_{h,c(h),t}$ is either a binary measure indicating whether household h in community $c(h)$ received remittances in year t or a measure of development. I estimate Equation (4) based on a sample at household level and include all three waves of the Nigerian GHS panel.

Table 6 reports estimation results of Equation (4) for various outcome variables. It is comforting to see that the interaction term is not significant before the arrival of fast Internet in Nigeria for all dependent variables. Further, the estimate shows that remittances increased sharply in connected areas relative to unconnected areas after the second migration period between the year 2012 and 2015. I do not find that these remittances are used to invest in wealth items. If anything, the effect of migration on wealth of families left behind might be even negative. One explanation could be that household members might have sold off assets to finance the migration costs of the migrant. Additionally, I do not find a significant effect on

the share of household members that are enrolled for the age bracket 10 to 18. This might be due to compulsory schooling in Nigeria for children up to the age of 15. Focusing on children above the age of compulsory schooling, I do find a positive and significant effect.

Overall, the results suggest that the increase in migration due to the arrival of fast Internet had a positive effect on children's education in locations close to the terrestrial cable network. However, the causal channel might also be different. The availability of faster Internet might have a direct effect on remittances even without an increase in migration because it facilitates bank account transfers (Lee et al., 2020). Further, the positive effect on children's education might also be driven by changes in other factors due to arrival of fast Internet such as better educational infrastructure.

2.8 Conclusion

This paper provides evidence on the effect of the arrival of fast Internet on migration rates in Nigeria. Following Hjort and Poulsen (2019) I exploit the arrival of submarine Internet cables from Europe to Nigeria that increased Internet speed for individuals located close to the terrestrial cable network but not for others. Using this time and cross-sectional variation in a difference-in-difference approach, I show that locations close to the terrestrial cable network saw a larger increase in migration rates than remote locations after the arrival of fast Internet. Further, I provide evidence that this effect is likely driven by increased Internet usage among individuals located in communities close to the cable network. The effect of Internet usage on migration is more relevant for migration out of Africa. I further highlight interesting effect heterogeneity with respect to household wealth, suggesting that the spread of the Internet might lead to a negative selection of migrants. Finally, I show that the effect of the arrival of fast Internet on migration rates is followed by an increase in remittances among locations close to the terrestrial cable network. These remittances might be responsible for economic development in these locations, as they are also correlated with higher school enrollment.

Assessing the Role of Asylum Policies in Refugees' Labor Market Integration:
The Case of Protection Statuses in the German Asylum System*

Maurizio Strazzeri†

University of Konstanz
Department of Economics
Graduate School of Decision Sciences

Abstract

I study the effect of refugees' protection status on labor market outcomes focusing on a recent cohort of Syrian and Iraqi refugees entering Germany between 2013 and 2016. My empirical analysis exploits a sudden and unpredictable change in the assessment of the Federal Agency responsible for asylum claims to grant full refugee status in accordance with the Geneva convention to refugees from these two countries in March 2016. Using data from the IAB-BAMF-SOEP survey of refugees and exploiting the policy change in a fuzzy regression discontinuity design, estimation results indicate a substantial negative effect of subsidiary protection status on earnings and employment.

*I thank seminar audiences at the University of Konstanz and MINES ParisTech for helpful comments and suggestions. I am grateful to Enzo Brox, Sebastian Findeisen, Stephan Maurer, Wanda Mimra, Guido Schwerdt, and Ahmed Tritah for useful suggestions that improved this version of the paper.

†Email: maurizio.strazzeri@uni.kn

3.1 Introduction

In the last decade, the European Union (EU) has experienced a sizable increase in the number of refugee migrants from outside the European continent (Dustmann et al., 2017). In light of the increasing evidence that refugee migrants perform particularly worse in labor markets in Western Europe (Fasani et al., 2018; Brell et al., 2020) - even in the medium and long run (Bratsberg et al., 2014) - with potential negative consequences for future generations (Bauer et al., 2013), policies that improve the labor market integration of refugee migrants have the potential to bring large benefits to refugee migrants as well as their host countries. The literature on the economic assimilation of economic migrants suggests that the duration and permanence of stay is an important determinant of immigrants' economic integration as investments in destination country-specific skills largely depend on the time period in which immigrants can benefit from their investments (Dustmann, 1993, 1999, 2000; Cortes, 2004; Dustmann and Görlach, 2016).[29] This might be particularly important for refugee migrants whose relocation decisions are not, or only to a minor extent, based on economic considerations, which makes them less economically selected than economic migrants, resulting in lower levels of host country-specific human capital upon arrival (Becker and Ferrara, 2019; Brell et al., 2020). However, refugee migrants are confronted with a considerable amount of uncertainty about their future settlement in the host country due to long waiting times for asylum claims or the lack of a clear perspective on permanent residence (Hainmueller et al., 2016; Dustmann et al., 2017). In particular, it has been noted that the various forms of protection that refugee migrants receive in the host country might be an influential factor for labor market integration as they typically differ in terms of the time frame they offer to refugee migrants to obtain permanent residence in the host country (Dustmann et al., 2017).

In this paper, I empirically investigate the link between different types of protection statuses and labor market outcomes, focusing on a recent cohort of refugee migrants from Syria and Iraq who received one of the two most prevalent protection statuses - refugee status in

[29] This result follows from a standard dynamic human capital model (e.g., Ben-Porath, 1967). Chiswick (1978) provided an early contribution to show that human capital differs across countries and newly arrived immigrants have an incentive to invest in destination country-specific human capital. For a survey, see Duleep (2015).

accordance with the Geneva convention or subsidiary protection status - in Germany.[30] The German asylum system provides an interesting case for the goal of this study as both statuses offer refugee migrants equivalent access to the labor market and social security system. However, both statuses differ considerably in terms of the requirements to receive permanent residence in Germany. While Geneva convention refugees can apply for permanent residence 5 years after arrival in Germany if they meet minor requirements such as sufficient knowledge of the German language (A2 level) and being able to cover at least 50 % of their costs of living, subsidiary protection refugees need to prove that they have an acceptable command of the German language (B1 level), the ability to cover all of their costs of living, *and* contributed at least 60 months to the German social security system.[31] Considering that granted protection statuses are regularly checked and can be revoked if the reasons for obtaining a protection status are not applicable any more, the additional requirements for subsidiary protection refugees might severely affect their perception of the likelihood of permanent residence in Germany, with potential consequences for their integration efforts and labor market outcomes. On the other hand, the meritocratic elements included in the requirements might counteract such negative effects as they give higher incentives to be employed for subsidiary protection refugees who are willing to stay in Germany (Schammann, 2019).

The empirical analysis of this paper exploits a sudden and unpredictable change in the assessment of the German Federal Agency for Migration and Refugees (German: Bundesamt für Migration und Flüchtlinge, BAMF) to grant full refugee status in accordance with the Geneva convention to refugee migrants from Syria and Iraq. Asylum seekers qualify for full refugee status in accordance with the Geneva convention if the cause of flight was due to individual persecution resulting from an innate trait or membership of a social group (see also the discussion in Section 3.2.1). While Syrian and Iraqi asylum seekers who received notification about their asylum claim in the year 2015 or the first three month of the year 2016 were almost entirely granted full refugee status in accordance with the Geneva convention, the assessment

[30] As explained in Section 3.2.1, the German asylum system offers more than one form of subsidiary protection where one form of subsidiary protection is termed "subsidiary protection."

[31] In general, German residents contribute to the social security system if they are employed and earn more than 450 Euros per month.

of BAMF changed suddenly in the following months and around one-fifth of the asylum seekers receiving notification in April 2016 were granted only subsidiary protection. The share of refugee migrants who were granted subsidiary protection status remained high in the following months. As refugee migrants cannot precisely influence the timing of the decision on their asylum claim, and applied for asylum many months before they received the notification about their protection status, the change in the assessment of the BAMF provides valuable, and plausibly exogenous, variation in the likelihood to receive subsidiary protection status for refugee migrant who receive notification about their asylum claim close to change in the decision-making practice of the BAMF.

My empirical analysis is based on the comprehensive longitudinal IAB-BAMF-SOEP survey of refugees, which provides extensive information about the asylum procedure and socioeconomic background for a sample of refugee migrants who entered Germany in the years 2013 to 2016. Most importantly for this study, the survey collects information about the year and month at which refugee migrants received the notification about their asylum claim as well as their current protection status and labor market outcomes. This allows me to exploit the change in the assessment of the BAMF in a regression discontinuity (RD) design, which generates causal estimates of the effect of the policy change on current subsidiary protection status and labor market outcomes under relatively weak identification assumptions (Hahn et al., 2001; Lee and Lemieux, 2010).

The results of my empirical investigation clearly indicate a substantial negative effect of subsidiary protection status on various labor market outcomes. Based on a sample of Syrian and Iraqi refugees who reported to have either protection status in accordance with the Geneva convention or subsidiary protection status, I estimate a significant decline in the probability of being in any employment, in full-time employment, as well as in monthly labor earnings two and one-half years after the policy change for refugee migrants who received notification after March 2016. The drop in employment by around 9 pp is almost entirely driven by the reduction in full-time employment, which suggests that the policy change had an effect on the employment probability as well as on the share of full-time employment among em-

ployed refugees. This is also reflected in the estimates for monthly labor income, where local linear regression at both sides of the threshold suggest a drop in monthly labor income of around 140 Euros for the entire sample and around 220 Euros among employed individuals. Additionally, I show that the change in the assessment of the BAMF to grant refugee status in accordance with Geneva convention to Syrian and Iraqi asylum seekers can still be seen two and one-half years after the policy change. Using current reported protection status as outcome variable, RD design estimates indicate that the exposure to the new policy regime of the BAMF increased the share of refugee migrants with subsidiary protection status by around 25 pp. Under the assumption that the exclusion restriction is satisfied, this result suggests that the effect of subsidiary protection status on labor market outcomes is four times as large as the previously discussed reduced form estimates.

My estimates of the effect of subsidiary protection status on labor market outcomes represent the local average treatment effect for the subgroup of compliers, i.e., those refugee migrants who have subsidiary protection status only due to the change in the assessment of the BAMF. It seems plausible to assume that the change in the assessment of the BAMF targeted a specific group of refugee migrants, which suggests that the group of compliers differs from other refugee migrants. To address such concerns, I complement my baseline results with a complier analysis, and show that the change in the decision-making practice by the BAMF affected, indeed, more strongly refugee migrants with background characteristics that are commonly attributed to improve labor market outcomes for refugee migrants - such as male gender, younger age, not married, and no children in the household.

The negative effect of subsidiary protection status on labor market outcomes is consistent with the proposed causal mechanism that subsidiary protection status reduces the perception of the likelihood of permanent residence in Germany, which, in turn, lowers the incentives to invest in country-specific human capital and the probability to be active in the labor market. Additionally, the size of the estimated effect suggests that the counteracting effect due to the meritocratic elements in the criteria to obtain permanent residence are only of minor importance. In the final part of the paper, I test this causal mechanism, and show that there

is no evidence that subsidiary protection status reduces refugees' investments in country-specific human capital. While my estimation results indicate that subsidiary protection status increases worries of refugee migrants that they cannot remain in Germany, it also positively affects participation in integration classes and hours spent studying German. These puzzling results suggest that the negative effect of subsidiary protection status on labor market outcomes might not necessarily be driven by labor supply side factors but instead by labor demand side factors.[32] For instance, if the employment of refugee migrants requires costly on-the-job training, firms prefer to hire refugee migrants with better prospects of staying to regain their investment costs. While the data at hand does not allow me to investigate this link, further research in this area might help understand the causal mechanism that explains the observed negative consequences of subsidiary protection status on labor market outcomes.

My empirical analysis provides quasi-experimental evidence based on micro data to confirm the existence of an economic and political trade-off in asylum policies as noted by Dustmann et al. (2017). Granting permanent residence status to refugee migrants presumably induces political costs, but provides economic and social benefits by reducing unemployment among refugee migrants in the society. This trade-off becomes particularly important if refugee migrants who are initially offered only temporary protection end up staying for longer in the host country because, e.g., the reasons for providing temporary protection do not change any time soon (Dustmann et al., 2017). The results of my paper suggest that optimal asylum policies should take into consideration the likelihood that the reasons for granting temporary protection status remain over a longer time period (5 to 10 years), and, in case the likelihood is high, to offer the same time frame to obtain permanent residence as for beneficiaries of full refugee status in accordance with the Geneva convention. To be aware of such an trade-off is also important in light of the efforts of the EU to harmonize the European asylum system, where to this day a large heterogeneity in granted protection statuses across countries exist (Dustmann et al., 2017), and similar criteria to grant protection might be desirable.[33]

[32]While the majority of the literature on the economic assimilation of immigrants focuses on supply side factors, it has been shown that the demand side is also an influential factor in explaining immigrants' employment (Åslund and Rooth, 2007; Azlor et al., 2020).

[33]For a discussion of the harmonization of European asylum policies and potential benefits, see Hatton (2015).

The paper contributes to the increasing literature studying the effect of asylum policies on labor market outcomes of refugees. Most closely to the question of this study, Fasani et al. (2018) exploit variation in refugees' exposure to high full refugee status recognition rates (measured by the ratio of the number of refugees receiving Geneva convention protection status to total number of decision made) across entry cohorts and within country or within entry cohort and across countries based on data from the European Labor Force Survey, and find that the exposure to high full refugee status recognition rates improves labor market outcomes. Another finding of Fasani et al. (2018) is that dispersal policies of refugees have a negative impact on labor market integration and, related to the question of dispersal policies, Brücker et al. (2020) show that residence requirements reduces employment rates among refugee migrants in Germany. Rosholm and Vejlin (2010) study the effect of a reduction of welfare payments for refugee migrants in Denmark and find that lower income transfers increase job-finding rates of refugees, and Hainmueller et al. (2016) analyze the effect of the length of asylum procedure and find that they are negatively associated with labor market performance in Switzerland. This paper further relates to studies that investigate other policies that are intended to facilitate refugees' labor market integration such as language and integration courses. Arendt et al. (2020) show that a Danish reform that expanded language classes for refugees positively affected employment and income, and Battisti et al. (2019) provide evidence that job-search assistance is conductive for employment prospects of refugees. More broadly, this study also relates to the literature on citizenship or legal status and labor market outcomes of immigrants. In particular, Gathmann and Keller (2018) find that faster access to citizenship increases employment outcomes of immigrants in Germany, and Devillanova et al. (2018) provide quasi-experimental evidence from an amnesty program in Italy and show that legal status of immigrants increase employment rates.

The rest of the paper has the following structure. In Section 3.2, I provide background information on the German asylum system (Section 3.2.1) and the change in the assessment of the BAMF to grant full refugee status to asylum seekers from Syria and Iraq (Section 3.2.2). In Section 3.3, I present my data set based on the IAB-BAMF-SOEP survey of refugees. In Section

3.4, I introduce the main identification strategy (Section 3.4.1) and provide a discussion of the validity of the RD design (Section 3.4.2). The baseline results are illustrated in Section 3.5. Robustness tests of the RD design (Section 3.6.1), and results of an alternative identification strategy (Section 3.6.2) can be found in Section 3.6. In Section 3.7, I discuss effects of subsidiary protection status on other integration efforts, and Section 3.8 concludes.

3.2 Institutional Background

3.2.1 Asylum system and protection statuses in Germany

The German Federal Agency for Migration and Refugees (German: Bundesamt für Migration und Flüchtlinge (BAMF)) acts in the field of duties of the German Federal Ministry of the Interior.[34] Part of its responsibilities is the management of the German asylum procedure, starting from the first registration of asylum seekers after entering the country until the final decision about the asylum application is made. Upon arrival and first registration, asylum seekers are distributed among a number of reception centers in Germany where they file an asylum application with the closest branch office of the BAMF. If Germany is responsible for the asylum application in accordance with the Dublin III regulation, the asylum applicant will be invited to attend an individual hearing.[35] The individual hearing is organized by the case worker of the BAMF that is responsible for the final decision of the asylum application, and might be attended by an interpreter and possibly a lawyer or another person if requested by the asylum seeker. During the individual hearing, asylum seekers state the reasons why they decided to flee and what kind of persecution they experienced. The individual hearing is of highest priority for asylum seekers as the information and evidence put forth in the hearing will be the main basis of decision-making for the case worker of the BAMF and - in case of an appeal against the decision of the BAMF - the only accepted evidence from the asylum seeker in front of a court.

[34]This subsection is based on online information from the BAMF (www.bamf.de) and Tiedemann (2014).

[35]The Dublin III regulations state that the asylum application must be processed in the country that the asylum seeker entered first. The Dublin III regulations are a EU law and applied to all EU member states except Denmark plus Island, Norway, and Switzerland.

Based on the information from the individual hearing and additional research on the credibility of the claim of the asylum seeker, the case worker of the BAMF makes a decision about the protection status of the individual. The case worker of the BAMF checks if one of the four protection status - (i) political asylum in accordance with the German constitution, (ii) refugee status in accordance with the Geneva convention, (iii) subsidiary protection status, (iv) suspension of deportation - can be granted to the applicant. If this is the case, the applicant will receive a positive notification about the asylum application as well as a temporary residence permit. If not, the applicant will receive a rejection letter and is obliged to leave Germany within 14 or 30 days depending on the reason for rejection. In case the rejected applicant does not obey to the obligation to leave Germany, he or she will be deported. If obstacles to deportation exist - such as missing documents - the rejected applicant might be tolerated to stay in Germany for a short period of time (usually around 3 month), after which the rejected applicant will be, again, obliged to leave the country.

As I will show below, the two most prominent protection status in Germany are protection status in accordance with the Geneva convention and subsidiary protection status. Political asylum in accordance with the German constitution is rarely granted as it requires asylum seeker to enter Germany from a country that is not considered a "safe" country. As all German neighbouring countries are considered "safe," political asylum can only be granted to asylum seekers entering Germany by plane. To receive protection in accordance with the Geneva convention, asylum seekers need to prove that they have been persecuted because of either their race, religion, nationality, political opinion or membership of a particular social group. It must be an innate trait (e.g., skin color, gender) or a characteristics of the individual that is so crucial for his or her identity or conscience (e.g., religion, sexuality, political opinion) that the individual cannot be forced to live without it. If asylum seekers do not fulfill the criteria to obtain full refugee status in accordance with the Geneva convention, they might obtain subsidiary protection if they fear death, torture or other inhuman treatment in their country of origin. In case of subsidiary protection, persecution does not need to relate to specific traits but can be the result of violence in the course of civil wars. Lastly, if asylum seekers cannot be

granted either of these three protection statuses, they might by suspended from deportation. This might be the case if deportation would lead to other human rights violations or the risk of life or freedom, e.g., if asylum seekers have a disease that cannot be treated in their country of origin.

It follows from this discussion that the main element that determines if asylum seekers obtain either refugee status in accordance with the Geneva convention or subsidiary protection status - the two protection status studied in this paper - is that asylum seekers can prove that individual persecution was due to an innate trait or a crucial characteristic of their identity. As I will discuss in the next subsection, the assessment of whether this criteria is fulfilled for the large number of asylum seekers from Syria or Iraq that entered Germany between 2014 and 2016, changed suddenly and noticeably in March 2016. The type of protection status that asylum seekers receive has important consequences. While both statuses allow for immediate access to the German labor market and social security system, they considerably differ in terms of the prospects to receive permanent residence in Germany.

Refugees in accordance with the Geneva Convention receive preferential treatment when applying for permanent residency.[36] With the notification about the decision about the asylum application, Geneva convention refugees receive an initial temporary residence permit for three years which can be prolonged for 2 additional years each time the residence permit expires and the reasons for granting the protection status are still applicable. Three years after arrival in Germany, Geneva refugee migrants can apply for a permanent residence permit if they can prove that they have a good command over the German language (C1 level) and are able to cover at least 75 % of their cost of living. Otherwise they can apply for permanent residency after five years if they show sufficient knowledge of German (A2 level) and are able to cover at least 50 % of their cost of living. Subsidiary protection refugees receive an initial residence permit of only one year which can be also prolonged for 2 additional years each time the residence permit expires and the reasons for granting the protection status are still

[36]Geneva convention refugees are treated in accordance with §26 Act on the Residence, Economic Activity and Integration of Foreigners in the Federal Territory (AufenthG). Subsidiary protection refugees are treated in accordance with §9 AufenthG.

applicable. Subsidiary protection refugees do not have a fast track to apply for permanent residency, but can apply also five years after arrival in Germany. However, the requirements they need to meet are more advanced as they need to show that they have acceptable command of the German language (B2 level), are able to cover all of their cost of living, and have contributed for at least 60 month to the German social security system.

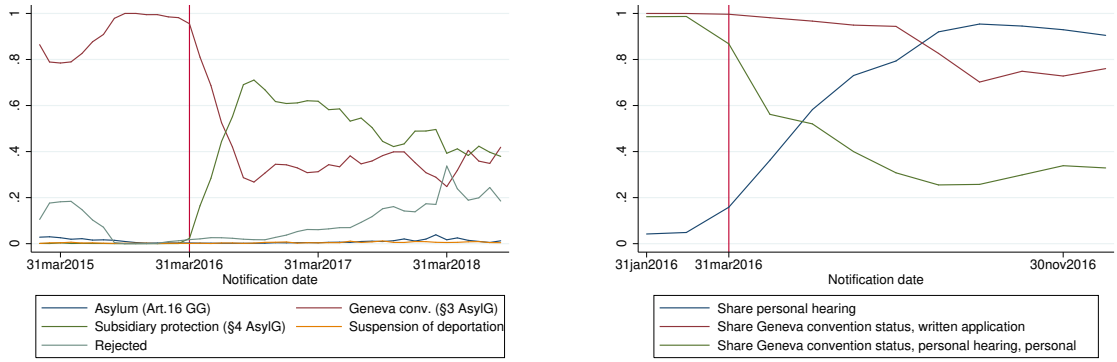
3.2.2 Increase in irregular migration and changes in asylum policies

In the years 2014 to 2016, an unprecedented number of refugee migrants entered the EU with the intention to apply for asylum. The majority of these migration flows originated from Syria, Afghanistan, Iraq, and Pakistan, and first encountered terrain of the EU in Greece by crossing the Eastern Mediterranean sea by boat.[37] Many of these refugee migrants wanted to continue their journey to Northern member states of the EU - such as Germany, France, the UK, and Sweden, by transiting countries of the Western Balkan. According to the Dublin III regulations asylum seekers are required to apply for asylum in the country which they first entered. However, Southern countries of the EU - and in particular Greece - were overwhelmed by the number of refugee migrants, and let the majority of these refugee migrants pass their Northern borders - many times without registration. The large number of refugee migrants crossing European borders revealed the weakness of the European asylum system and its unequal distribution of refugee migrants among member states, putting pressure on Northern European countries to accept a larger number of refugee migrants to apply for asylum in their territory. In particular, Germany unilaterally suspended the Dublin III regulations and started to process a considerable share of the asylum claims that resulted from these migration flows.

The German asylum system was not prepared for the large number of refugee migrants and Germany had to undertake major actions to increase its capacity to process the large number of asylum claims (Bertoli et al., 2020). In addition to transferring workers form other branches of the government to the BAMF (Grote, 2018), the German Federal Ministry of the Interior gave

[37] Based on data from Frontex, the European Agency for border control, Dustmann et al. (2017) show that around 38 % of illegal crossings to Europe between 2009 and 2015 were of individuals from Syria (Afghanistan: 20 %, Iraq: 5 %, Pakistan: 5 %).

Figure 4:
Protection status and notification date, Syrian asylum seekers



(a) Received protection statuses **(b)** Type of application

Note: Left plot illustrates the type of protection status received by month of notification date of Syrian asylum seekers. Source: Own calculations based on monthly published data from BAMF (data available upon request). Equivalent plot for Iraqi asylum seekers can be found in Figure B1b in the Appendix. Right plot shows for Syrian asylum seekers (i) the share of decisions made by month of notification date on basis of personal hearings (blue line), (ii) the share of asylum applicants that were granted Geneva convention status on basis of written applications by month of notification date (red line), and (iii) the share of asylum applicants that were granted Geneva convention status on basis of personal interviews (green line). Source: Deutscher Bundestag (2017).

order to the BAMF in autumn 2014 to decide asylum claims from Syrian and Iraqi refugee migrants on basis of written asylum applications instead of personal hearings, with the intention to speed up the asylum procedure for these refugee migrants. The order was retracted in the beginning of the year 2016, after which, the standard procedure of personal hearings gradually replaced written applications (Deutscher Bundestag, 2016). The change in the type of the asylum application had consequences for the protection status of Syrian and Iraqi refugee migrants. Figure 4a shows the type of protection status received by Syrian asylum seekers per month of notification about the asylum claim. As illustrated in Figure 4a almost all Syrian refugee migrants were granted refugee status in accordance with the Geneva convention if they received notification about their asylum claim before March 2016 (rejections not considered), but the share of refugee migrants who received subsidiary protection status increased noticeably afterwards.[38] At the same time, the share of decisions made on basis of personal hearings was increasing, as illustrated by the blue line in Figure 4b. More importantly, the

[38]While less pronounced, the same pattern can be seen for asylum applicants from Iraq as illustrated in Figure B1b in the Appendix.

assessment to grant full refugee status for decisions made on basis of personal hearings, as illustrated by the green line in Figure 4b, changed also significantly after March 2016, which indicates that the sudden increase in the share of Syrian refugee migrants with subsidiary protection was driven by the change in the type of application.[39] This suggests that the sudden change in the share of refugee migrants with subsidiary protection status was not due to compositional differences in the group of asylum applicants. Instead, this suggests that personal hearings led case worker to decide differently on the question of individual persecution.[40]

As I explain in more detail in Section 3.4, another important assumption of the identification strategy of this paper is that receiving notification after the change in the assessment of the BAMF in March 2016 only affected refugee migrants likelihood to obtain subsidiary protection status instead of Geneva convention protection. This rules out that there were any other changes in asylum law that affected refugee migrants differently, depending on whether they received notification shortly before or after the policy change. Indeed, in mid March 2016, the German government passed a number of reforms of the German asylum law to further increase the speed of asylum procedures. The legislative package included a series of laws that affected the live of refugee migrants in Germany such as reductions of benefits for asylum seekers and rejected asylum applicants, ease of terms to deport rejected asylum applicants, or penalties for refugees who violate residential obligations. Almost all of these changes were targeted on rejected asylum applicants, and did not depend on the notification date of the asylum claim, except of one. With the passage of the law, asylum seekers that receive subsidiary protection status were subject to a temporary ban on family reunification. Before the law was passed, Geneva convention refugees as well as subsidiary protection refugees were allowed to bring their spouse and children below the age of 18 to Germany after they were granted protection. Hence, refugee migrants who entered without their spouse or children and received notification after the policy change were not only affected by a higher likelihood

[39]Unfortunately, there is no data on the share of decisions made based on personal hearings for Iraqi asylum seekers.

[40]In Section 3.4.2, I provide a detailed discussion of the validity of the RD design, which involves a comparison of refugee migrants who received notification about their asylum claim shortly before and after March 2016, and based on observable characteristics, I do not find any differences among both subpopulations.

to obtain subsidiary protection but also by the ban of family reunification. To avoid that I wrongly attribute changes in labor market outcomes to subsidiary protection status instead of being affected by the ban on family reunification, I exclude these refugee migrants from my sample as explained below.[41]

3.3 Data

The main data source of the paper is the longitudinal IAB-BAMF-SOEP survey of refugees (SOEP refugee panel) which provides an excellent source to study questions regarding the labor market integration of refugees in Germany.[42] The SOEP refugee panel is based on a sample of individuals who entered Germany between 2013 and 2016 with the intention to apply for asylum, as well as their household members. All individuals of the sample above the age of 18 are interviewed annually, and the first wave was conducted in the year 2016. The empirical analysis of this paper is based on the latest wave of SOEP refugee sample from the year 2018. I restrict the SOEP refugee sample to Syrian or Iraqi individuals in working age (18 to 65) who applied for asylum and had received notification about the asylum application before the time of interview.

Based on this sample of 2,061 individuals, around 81 % of the respondents reported to have either currently protection status in accordance with the Geneva Convention or subsidiary protection status. The other roughly 20 % consist of refugee migrants who either did not obtain protection in Germany, received protection for humanitarian reasons (suspension of deportation), or already have permanent residency in Germany. To increase the precision of my first stage estimates based on the discontinuity induced by the policy change, I exclude these individuals from the data set. This results in a sample size of 1,683 refugee migrations who either have protection status in accordance with the Geneva Convention or subsidiary protection status.[43] As explained above, the new policy regime did not only affect the likelihood to

[41] An alternative approach would be to disentangle both effects in a difference in discontinuity design as proposed by Grembi et al. (2016). However, as only a relatively small number of refugee migrants entered Germany without their spouse and children, the approach does not provide fruitful results in my application.

[42] The data set can be ordered online at the research data center SOEP of the DIW: https://www.diw.de/soep. For detailed information about the study design, see Kroh et al. (2016).

[43] The main results of the paper are qualitatively not affected by the exclusion of these observations.

obtain subsidiary protection status but also included a ban on family reunification for refugee migrants with subsidiary protection status that received notification after the policy change. Around 71 % of the respondents are married of whom 18 % reported to have entered Germany without their spouse or children, and would have been affected by the ban if they obtained subsidiary protection status and received notification after the policy change. As it seems to be likely that being affected by the ban on family reunification also influences labor market outcomes, I also exclude these individuals from the sample, which results in a final data set with 1,470 observations.

The SOEP refugee panel provides detailed information about the asylum process. In particular, respondents were asked about the date (month and year) when they received notification about the decision of the asylum application, which allows to construct a variable that indicates whether an individual was affected by the policy change or not. The SOEP refugee panel provides also information about labor market outcomes. I use this information to construct two binary outcome variables that indicate whether an individual was (i) in any paid employed or (ii) in full-time employment at the time of the interview, as well as the reported monthly net labor income. Finally, I use the background information available in the SOEP refugee to construct an extensive set of control variables covering individual-specific characteristics such as age, gender, martial status, work experience before migrating, or time spent in Germany. This information is used to assess the validity of the RD design and is illustrated in Section 3.4.2.

3.4 Identification

3.4.1 Empirical strategy

Estimating the effect of subsidiary protection status instead of protection in accordance with the Geneva convention on labor market outcomes poses considerable difficulties. As explained in more detail above, granted protection statuses target specific groups of asylum seekers and are not randomly distributed among refugee migrants impeding causal estimates of the effect

of protection statuses on labor market outcomes. For instance, using cross-sectional variation in protection statuses among refugee migrants with one of the two protection statuses might lead to biased estimates of the true effect if there are individual-specific unobserved factors that explain labor market outcomes and the type of protection status simultaneously. These factors might be abundant in my setting and could relate to, e.g, the prevalence of economic motives to migrate, different experiences made when fleeing, or loss of valuable assets in the country of origin. In this paper, I overcome such endogeneity concerns by exploiting the discontinuity in the probability of receiving subsidiary protection status at the point in time when the BAMF changed its decision making practice. While before April 2016 basically all non-rejected applicants from Syrian and Iraqi were granted refugee status in accordance with the Geneva convention, this suddenly changed afterwards with a high and increasing share of refugees who only received subsidiary protection. Hence, the probability of receiving subsidiary protection changed noticeably for those receiving notification after the policy change in March 2016. I exploit this variation in the share of refugees with subsidiary protection status in a fuzzy regression discontinuity (RD) design using the date of notification about the asylum application as assignment variable. Under assumptions discussed in more detail below, a fuzzy RD design allows in my setting to identify the local average treatment effect (LATE) for a subgroup of refugee migrants by calculating the ratio between the estimated discontinuity of the labor market outcome variable and the jump in the share of refugee migrants with subsidiary protection status at the time of the policy change. The subgroup of refugee migrants for whom the LATE is identified consists of asylum seekers that (i) received notification about their asylum application at the time of the policy change and (ii) are compliers, i.e., refugee migrants who receive subsidiary protection status if they receive notification of their asylum application after the policy change but would receive protection status in accordance with the Geneva convention if they received notification before the policy change.

As suggested by Hahn et al. (2001) and Imbens and Lemieux (2008), I implement the fuzzy RD design by a two-stage least square (2SLS) estimation procedure using a binary variable indicating the policy change as the excluded instrument and the assignment variable as ex-

ogenous control variable. Formally, I estimate the following system of equations:

$$Sub_i = \alpha_0 + \alpha_1 \mathbb{1}\left[t_i > c\right] + \alpha_2 f(t_i - c) + \alpha_3 \mathbb{1}\left[t_i > c\right] f(t_i - c) + \eta_i \tag{5}$$

$$Y_i = \beta_0 + \beta_1 \hat{Sub_i} + \beta_2 f(t_i - c) + \beta_3 \mathbb{1}\left[t_i > c\right] f(t_i - c) + \epsilon_i \tag{6}$$

where Sub_i is binary variable indicating if individual i reported to have a subsidiary protection status in the last wave of the SOEP and $\hat{Sub_i}$ is the predicted values of Sub_i based on parameter estimates of Equation (5). $\mathbb{1}\left[t_i > c\right]$ is an indicator function equal to 1 if i's month of notification about his or her asylum application (t_i) was after the change in the decision making practice of the BAMF in March 2016 (c).[44] $f(t_i - c)$ is a function of the assignment variable, the distance between t_i and c, and Y_i is a measure of the labor market outcome of i as reported in the last wave of the SOEP. η_i and ϵ_i are error components capturing factors that influence the outcome variables Sub_i and Y_i and are not included in Equation (5) and (6), respectively.

I estimate Equation (5) and (6) based on a sample of refugee migrants from Syria and Iraq who reported in the last wave of the SOEP refugee panel to have subsidiary protection status or protection status in accordance with the Geneva convention. As standard in the literature, I employ local linear and polynomial regressions on both sides of the threshold and report results for various bandwidth selection choices. Following the suggestions by Imbens and Lemieux (2008) and Lee and Lemieux (2010), I use a rectangular kernel which is equivalent to standard linear regressions on both sides of the threshold.[45]

The parameter of interest in this paper is β_1 and represents the LATE for compliers at the threshold under the following two assumptions (Imbens and Angrist, 1994; Hahn et al., 2001).

[44] I treat individuals who received notification about the asylum application in March 2016 as individuals who received notification before the change in the decision making practice even though the discussion in Section 3.2.2 suggests that some individuals were already exposed to the new decision practice in March 2016. In Section 3.6.1, I show that my results are robust to excluding those observations in a Donut RD design.

[45] Imbens and Lemieux (2008, p. 625) state that "from a practical point of view, one may just focus on the simple rectangular kernel, but verify the robustness of the results to different choices of bandwidth" and Lee and Lemieux (2010, p. 319) write that "it is [...] simpler and more transparent to just estimate standard linear regressions [...] with a variety of bandwidths, instead of trying different kernels corresponding to particular weighted regressions that are more difficult to interpret." See also the discussion in Hinnerich and Pettersson-Lidbom (2014).

The first assumption is *monotonicity* at threshold date, i.e., receiving notification shortly after c did not cause some individuals to receive protection status in accordance with the Geneva convention who would have obtained a subsidiary protection status in case they received notification shortly before c. Based on the discussion in Section 3.2.2, this assumption seems to be fulfilled as the new policy regime seems to be more strict in terms of granting a protection status in accordance with the Geneva convention and subsidiary protection status was very rare in the old policy regime. The second assumption is the *exclusion restriction* at the threshold date, i.e., receiving notification shortly after c did not impact Y except through Sub. This assumption requires that (i) the exposure to the new policy regime is "as good as randomly assigned" close at the threshold date (*independence*) and (ii) the exposure to the new policy regime did not affect labor market outcomes through other channels than an increase in the share of individuals with subsidiary protection status (*exclusion*).[46] Independence is fulfilled if there is imprecise control over the assignment variable - which is a standard assumption in RD designs - and its assessment is part of the following subsection. However, even if the exposure to the new policy regime is as good as randomly assigned close to the threshold, the exclusion restriction is violated if the exposure to the new policy regime affected labor market outcomes through other channels than the reception of subsidiary protection status. For instance, refugees with subsidiary protection who entered Germany without their spouse and received notification shortly after the threshold were affected by the ban on family reunification for which specific labor supply responses might be expected. Since I exclude asylum seekers that are affected by the ban on family reunification, this channel should not affect the identification strategy in this paper. Additionally, as the discussion in Section 3.2.2 has shown, there were no other changes in asylum policies that might have affected only asylum seekers who received notification shortly after to policy change, which suggests that I can rule out any other channel that might have affected labor market outcomes of refugees close to the cutoff.

[46]By disentangling the assumption of independence and exclusion from the exclusion restriction, I follow Angrist and Pischke (2008) and Imbens and Lemieux (2008)

3.4.2 Validity of the RD design

Lee (2008) shows formally for a sharp RD design that if individuals cannot *precisely* control the assignment variable, the variation in the treatment variable is as good as randomly assigned for observations with similar values of the assignment variable and, particularly, those observations close to the cutoff. It follows for a fuzzy RD design that imprecise control of the assignment variable implies random assignment of the instrumental variable for observations close to the threshold. If refugees were able to *precisely* influence the timing of the notification date of their asylum application, and if refugees have a benefit to be treated in accordance with the old or new policy regime, it is likely that refugees on one side of the cutoff differ systematically from those on the other side. For instance, assume refugees with better labor market prospects might be better informed about asylum policies and know about the change in decision making practice and others do not. If those refugees with better labor market prospects prefer to avoid the new policy regime with a higher chance of receiving subsidiary protection status, they would put effort into receiving the notification about the asylum application before the threshold while the others would not. The result of this thought experiment would be that refugees on both sides of the cutoff differ with respect to their labor market prospects independently of the protection status they received.

However, this scenario seems to be rather unlikely due to the following aspects. First, the change in the decision making practice has never been publicly announced, which makes it implausible that even well-informed refugees knew about this policy change. Moreover, those refugees who received notification of their asylum application close to the cutoff arrived in Germany and applied for asylum several month before. This is illustrated in Figure B2 in the Appendix, which shows histograms of the arrival (left) and application month (right) relative to the policy change for refugee migrants who received notification 3 month before or after the policy change. Additionally, there is no recorded or anecdotal evidence that refugee migrants can influence the processing time of asylum applications.

Nonetheless, selective sorting around the threshold could still be possible. Assume that caseworker responsible for the asylum application knew about the policy change and were

selective about the refugees who would fall into the old policy regime by influencing the processing time of the application. If such a selection is correlated with factors that influence labor market outcomes, this would invalidate the RD design.

An intuitive approach to assess the prevalence of sorting is to investigate the density of the assignment variable (McCrary, 2008). The intuition is that strategic sorting implies an unexpectedly high number of decisions made shortly before (or after) the threshold, resulting in a discontinuity of the density of the assignment variable at the cutoff. Additionally, a discontinuity of the density of the assignment variable might point to selective attrition as, for instance, in DiNardo and Lee (2004). Selective attrition means that refugees who receive subsidiary protection because of the change in decision making practice are more likely to drop out of the sample (e.g, because they left Germany or they do not want to participate in the interview). This threatens the validity of the RD design, in particular if the reason for dropping out of the sample is correlated with labor market outcomes.

Figure B3 in the Appendix plots on the left-hand side the density of the assignment variable for the SOEP refugee panel sample used in the empirical analysis and the plot on the right-hand side illustrates the same distribution for the official register data. The vertical lines in Figure B3 indicate the threshold date at the End of March 2016. Both plots show a very similar density of the assignment variable which highlights the good quality of the SOEP survey. Further, as the graph on the left is based on the survey participants of the SOEP in 2018 and the graph on the right is based on actual decisions made by the BAMF in each month, the similarity between both densities suggests that selective attrition might not be of importance in this study. Visually inspecting the density of the assignment variables in Figure B3, one might see a discontinuity shortly after the cutoff starting in May 2016. However, focusing only on the month before and after the cutoff, the density seems to be rather smooth. Additionally, I formally test the null hypothesis that the discontinuity of the density of the running variable is equal to zero as proposed by McCrary (2008) and cannot reject the null hypothesis (bin size: .460, bandwidth: 12.306, log difference in height: -0.093, standard error: 0.104).

A second test to check if the instrumental variable is "as good as randomly assigned" close

to the cutoff is to compare pre-determined background characteristics of refugees who received notification about the asylum application before and after the threshold. While it is likely that those two groups differ in many dimensions for the overall sample, they should become more similar when restricting the sample to observations close to the cutoff. Table B1 in the Appendix shows mean values of selected pre-determined covariates for refugees who were not affected by the policy change ($t < c$) and those who were affected ($t > c$) as well as t-values of a two-sided mean comparison test. The first three columns refer to a sample that includes refugees who received notification about their application 18 months before or after the policy change and the last three columns further restrict the sample to three month before and after the policy change.

Focusing on the sample with a bandwidth choice of 18 months, Table B1 shows that refugees who received notification before the policy change are more likely to be male and slightly older than their counterparts who received notification after the threshold. Further, a higher percentage of those refugees had already acquired work experience before they moved to Germany and have spent, at the time of the SOEP interview, more time in Germany. The lower part of Table B1 shows also differences with respect to the outcome and treatment variables. In contrast, focusing on the last three columns in Table B1, the differences between both groups lose significance and the absolute difference between the mean values become much smaller - except for the treatment and outcome variables, which supports the hypothesis that selective sorting is not an issue in my setting.

As a final step, I check if pre-determined characteristics show a discontinuity at the threshold. If such pre-determined characteristics are not continuous around the threshold, I might wrongly attribute changes in labor market outcomes to changes in protection status if such discontinuities around the threshold were responsible for the changes in labor market outcomes. Table B2 in the Appendix shows RD estimates for various specifications (bandwidth choice and polynomial order) on various covariates.[47] For almost all covariates, I cannot reject

[47]Figures B4a to B4m in the Appendix shows the corresponding RD plots. In Figures B4n to B4p in the Appendix, I follow Bauernschuster and Schlotter (2015) and show RD plots for predicted values of the treatment and outcome variables based on separate regressions of these variables on the full set of control variables. If predicted variables show a discontinuity at the threshold, this would indicate that differences in observable characteristics

the null hypothesis that the estimated discontinuity is equal to zero for all specifications. If I find significant effects for some variables, these are not robust across all specifications. The most worrisome discontinuity can be found for the variable *month since migration*. However, the effect is relatively small compared to the sample mean, which suggest that the resulting bias should be negligible.

In sum, the fact that refugee migrants cannot affect the timing of the decision of the asylum application as well as the three tests of the independence assumption around the threshold due to imprecise control of the assignment variable suggests that selective sorting does not play a major role in my setting.

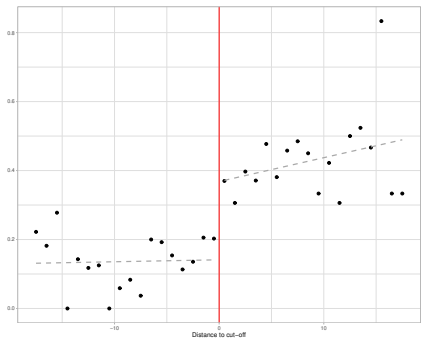
3.5 Results

Before discussing the estimates of the main identification strategy, I will first provide a visual inspection of the the relationship between the notification date of the asylum application, subsidary protection status, and the outcome variables. Figure 5 shows binned scatter plots between the notification date and subsidiary protection status as well as the three main outcome variables. Each dot in Figure 5 shows the mean value of the corresponding outcome by the month of the notification date. The red vertical line indicates the threshold date between March and April 2016, and the dashed lines are linear fits based on the mean values of each side of the threshold.

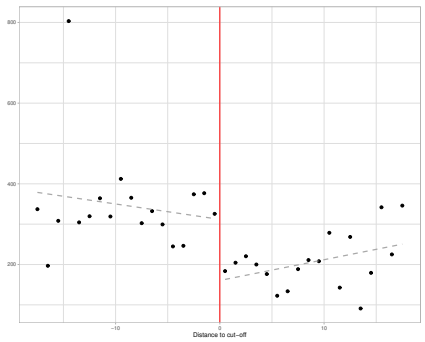
Figure 5a illustrates the discontinuity in the share of refugees who received subsidiary protection status after the policy change. The observations left of the threshold indicate that the share of refugees who report in the last wave of the SOEP to have subsidiary protection status is almost entirely below 20 % before the policy change. On the other hand, this share increased to more than 35 % directly after the policy change and remains significantly higher afterwards. However, contrary to what discussion of the administrative data of the BAMF in Section 3.2.2 suggested, the share of refugees with subsidiary protection is considerably above zero before the threshold. A possible explanation for the sizable mismatch might be

might be responsible for discontinuities of the treatment or outcome variables at the threshold. However, as shown in Figures B4n to B4p, this is not a concern here.

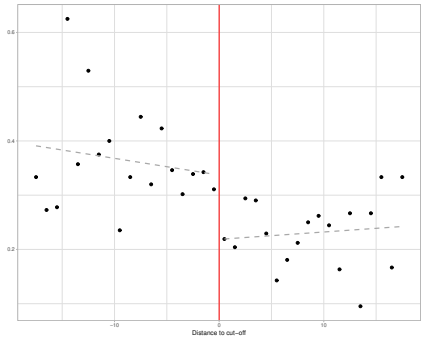
Figure 5:
RD plots, first-stage and outcome variables



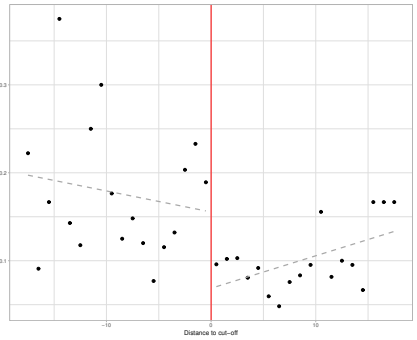
(a) Subsidiary protection



(b) Monthly labor earnings



(c) Any employment



(d) Full-time employment

Note: Mean of selected variables by value of the assignment variable with fitted lines on both sides of the threshold. Selected Bandwidth: 18 months.

that the administrative data illustrates the share of protection statuses issued in each month based on first-time decisions and the survey data refers to the protection status during the last wave of the SOEP. As rejected refugees and refugees who do not obtain a protection status in accordance with the Geneva Convention can take court action against the decision, which might result in receiving a protection status or receiving a better protection status, this might explain the discrepancies between the administrative data and the data from the SOEP.[48]

[48] Another explanation could be measurement error in the treatment and/or assignment variable due to misreporting. Measurement error in the treatment variable - which in case of a binary treatment variable would lead to an upward bias in a simple 2SLS procedure (Kane et al., 1999; Jiang and Ding, 2019) - seems to be unlikely as respondents are explicitly asked to check their German identification card which states the protection status on the backside. Measurement error of the assignment variable might be more important here as respondents might not remember the exact month of the notification of the asylum application. Measurement error of the assignment variable might lead to difficulties in identifying the LATE as the discontinuity in the assignment variable might vanish (see, e.g., Hullegie and Klein, 2010; Pei and Shen, 2016; Davezies and Le Barbanchon, 2017). However, as Figure 5 illustrates a sizeable discontinuity, I conclude that measurement error of the assignment variable is not a concern in this study.

Turning next to the relationship between the notification date and monthly labor earnings as shown in Figure 5b, again, a striking discontinuity around the cutoff can be observed. Average labor earnings were almost entirely above 300 Euros before the change in decision-making practice, which changed suddenly to around 200 Euros afterwards. A similar pattern - while less pronounced - can also be seen for the binary outcome variables *Any employment* in Figure 5c. While the average share of refugees with any employment is most of the time in the range between 30 and 40 % in the old policy regime, this pattern changes in the new policy regime where mean employment lies between 20 and 35% percent most of the time. A similar picture emerges when turning to the outcome variable *Full-time employment* in the Figure 5d. Here, the average share of refugees reporting to have full-time employment during the last wave of the survey drops significantly at the cutoff from 15 to 20% to around 10% or less after the policy change.

Table 7 reports results of the first-stage and reduced form estimates of the baseline instrumental variables estimation discussed in Section 3.4. The first row of Table 7 shows results for the estimates based on Equation (5) for various bandwidth and selections of the order of polynomial for the assignment variable. The second to fifth rows show the same estimation specification using the outcome variables as dependent variable instead of the treatment variable. Inference is based on Huber-White standard errors which are shown in parentheses.[49]

The estimates shown in Table 7 overall confirm the conclusions drawn from the visual inspection. The estimated discontinuity in the likelihood of being a beneficiary of subsidiary protection induced by the policy change is positive and sizable across all specifications. Referring to the estimate in column (1) where all observations are included, the estimation result suggests that the policy change led to an increase in the share of refugee migrants with sub-

[49] A large part of the literature uses standard errors clustered at the value of the running variable in RD designs with a discrete running variable as suggested by Lee and Card (2008) to account for model misspecification. Kolesár and Rothe (2018) show that such standard errors "do not guard against model misspecification, and that they have poor coverage properties." In particular, they show that clustered standard errors are substantially smaller than Huber-White standard errors in case of small to moderate bandwidths and that the actual coverage rate of confidence intervals based on clustered standard errors with nominal level 95 % might be as low as 58 %, while confidence intervals based on Huber-White standard errors have coverage much closer to 95 %. Since clustered standard errors are much smaller than Huber-White standard errors in my setting, I use Huber-White standard errors throughout the paper instead of clustered standard errors.

Table 7:
First-stage and reduced form RD estimates

	(1)	(2)	(3)	(4)	(5)
First stage estimation					
Subsidiary protection	0.24***	0.18***	0.21***	0.19***	0.15**
	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
F-statistic	40	12	27	17	5
Reduced form estimation					
Any employment	-0.09**	-0.11**	-0.10**	-0.08*	-0.05
	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
Full-time employment	-0.09***	-0.11***	-0.10***	-0.10***	-0.13***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)
Monthly earnings (excl 0)	-222.98**	-160.41	-214.69**	-227.33**	-248.20
	(90.95)	(127.90)	(98.82)	(115.23)	(157.01)
Monthly earnings	-142.74***	-158.98***	-148.44***	-145.30**	-152.76**
	(42.69)	(57.35)	(48.40)	(56.91)	(74.63)
Bandwidth selection	none	none	18	12	6
Polynomial order	1	2	1	1	1
Observations	1470	1470	1399	1238	782

Note: First stage and reduced form RD estimates for various polynomial orders and bandwidth selection choices. Each row shows estimation results for a separate outcome variable. Estimates for the outcome variable *Monthly earnings (exlc 0)* are based on a restricted sample of employed individuals. Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

sidiary protection status by around 24 pp. The estimated coefficient becomes smaller when using a higher polynomial order of the assignment variable or only observations around the threshold within a selected bandwidth. However, I can reject the null hypothesis that the estimated coefficient is equal to zero in all specifications.

The estimated effect of the policy change on the likelihood of being in any employment or full-time employment is negative throughout all measures and specifications. Interestingly, while the effect is slightly smaller in some specifications, the overall drop in employment by 9 percentage points (column 1) seems to be entirely driven by the drop in full-time employment. This result indicates that the policy change had an effect on the employment probability as well as on the share of full-time employed among all employed refugee migrants. The consequences of these two effects can also be seen in the change of monthly labor earnings as shown in the fourth and fifth row of Table 7. The fourth column of Table 7 shows estimation

Table 8:
OLS and fuzzy RD estimates

	OLS estimate	IV estimate	Control complier mean
Any employment	-0.07	-0.37**	0.64***
	(0.05)	(0.17)	(0.15)
Full-time employment	-0.09**	-0.40***	0.45***
	(0.04)	(0.13)	(0.12)
Monthly earnings (excl 0)	-250.87**	-770.57**	1327.70***
	(116.31)	(341.75)	(284.57)
Monthly earnings	-137.92***	-603.92***	761.01***
	(52.19)	(196.81)	(182.36)
Observations	396	1470	1470

Note: OLS (column 1) and 2SLS (column 2) estimates of the effect of subsidiary protection status on various labor market outcomes. Each row reports results for a separate outcome variable. The first column reports OLS results of the effect of subsidiary protection status on labor market outcomes based on subsample of observations close to the threshold (Bandwidth: 3 month). The second column reports instrumental variable estimates that corresponds to specification (1) in Table 7. The estimation of the corresponding mean of the control complier group follows suggestions by Cohodes (2020, p. 139-140). Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

results for a sample of employed individuals. If the policy change would not have affected the composition of - in general, better paid - full-time and non-full-time employment among refugee migrants, I would expect the effect to be zero in this case. However, the effect is large and significant throughout almost all specifications and suggests that monthly labor earnings dropped by around 220 Euros per month among employed refugee migrants. When using the entire sample, as shown in the fifth column of Table 7, I also obtain negative effects of the policy change on monthly labor earnings, which is in line with the estimated negative consequences of the policy change on the overall employment probability.

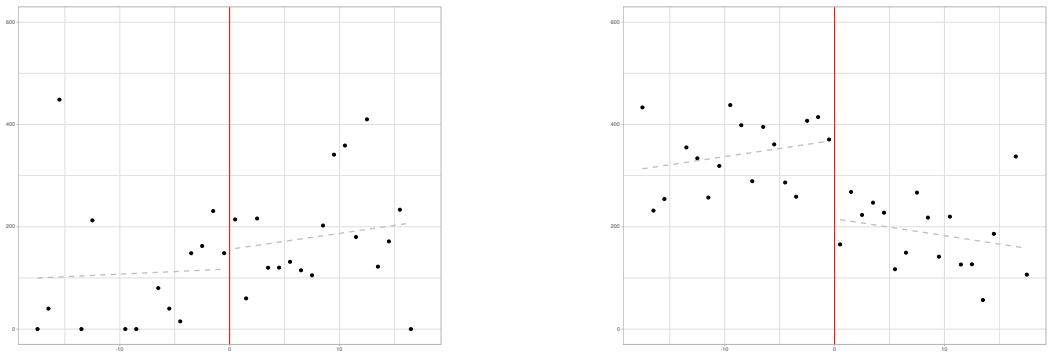
The second column of Table 8 reports corresponding 2SLS estimates for specification (1) from Table 7. As explained above, the 2SLS procedure identifies the average treatment effect for the group of compliers, i.e., those refugee migrants who obtained subsidiary protection status only due to the policy change of the BAMF. To facilitate the interpretation of the effect, Table 8 also reports the mean of the control complier group, which estimates the mean values of potential outcomes of not having subsidiary protection status for the group of compliers. Focusing first on the estimated treatment effect in the second column of Table 8, the 2SLS procedure reveals large and significant negative effects of having subsidiary protection status on

the likelihood of being employed as well as on monthly earnings. Subsidiary protection status reduces the likelihood of having any employment by 37 percentage points for the group of compliers, which implies that unemployment is twice as likely for those refugee migrants that receive subsidiary protection status due to the policy change.[50] Table 8 also makes clear that subsidiary protection status has an effect on the type of employment. While the share of full-time employment is at around 70 % (45/64) among employed refugee migrants for the untreated complier group, this number shrinks to 19 % ((45-40)/(64-37)) for those employed refugee migrants who received subsidiary protection status due to the policy change. The change in the composition of employment results in a significant drop in monthly labor earnings from around 1,330 Euros to 560 Euros among employed refugees in the complier group or from 760 Euros to 160 Euros among all refugees.

The first column of Table 8 reports coefficient estimates of a linear regression of each of the labor market outcomes on subsidiary protection status based on a subsample of individuals close to the policy change (bandwidth: 3 month). The OLS estimates give the average treatment effect of subsidiary protection status on labor market outcomes for individuals close to the threshold if the unconfoundedness assumption holds (Rosenbaum and Rubin, 1983), i.e., subsidiary protection status is not correlated with other variables that affect labor market outcomes such that treatment status is as good as randomly assigned for individuals close to the threshold. In general, this assumption is not fulfilled in fuzzy RD designs as individuals self-select into treatment based on incentives derived from the effect of the treatment variable on the outcome (Heckman et al., 1999). While self-selection seems to be not of a concern in this study as treatment status is determined by a third party, there might be still systematic differences between refugee migrants with subsidiary protection status and Geneva convention refugees at the threshold if the granting of subsidiary protection status is targeted at a specific subgroup of individuals that might have different labor market perspectives. For instance, the unconfoundedness assumption is violated if asylum seekers with more dominant economic motives of migration or better labor market perspectives have a higher likelihood of receiving

[50] Unemployment in control complier group: 1-0.64=0.36. Unemployment in treated complier group: 1-(0.64-0.37)=0.73.

Figure 6:
Testing external validity of fuzzy RD design



(a) Subsidiary protection refugees **(b)** Geneva convention refugees

Note: Mean of monthly labor income by value of the assignment variable with fitted lines on both sides of the threshold conditional on protection status. Figure on the left (right) includes only refugee migrants who reported to have subsidiary protection status (protection status in accordance with the Geneva Convention). Selected bandwidth: 18 month.

subsidiary protection status. Assuming that economic motives of migration affect labor market outcomes positively irrespective of protection status, OLS estimates will be upward biased, which implies under a constant treatment effect model, i.e., the effect of subsidiary protection does not vary across individuals, that IV estimates are larger than OLS estimates in absolute terms, which is in line with the results reported in Table 8.

In a heterogeneous treatment model, OLS and IV estimates might not only differ due to the violation of the uncounfoundedness assumption, it could also be the case that the average effect on compliers differs from the average effect on the other two subpopulation of *always-taker* and *never-taker* (Imbens and Angrist, 1994). Always-taker are refugee migrants who always receive subsidiary protection status irrespective of the policy regime to which they are exposed. On the contrary, never-taker are refugee migrants who receive Geneva convention protection status in the new and the old policy regime. A plausible procedure to assess the external validity of IV estimates is to compare average outcomes across compliance groups, i.e., of always-taker and treated complier and of never-taker and untreated complier (Angrist, 2004). If the average outcomes between these groups are not equal, this suggests that complier and always-taker (or never-taker) are substantially different and external validity of the IV estimates might be unlikely. I assess the external validity of the IV estimates in a fuzzy

RD design in Figure 6, following Bertanha and Imbens (2020), and plot discontinuities of the outcome variable *monthly labor income* at the threshold conditional on protection status. In Figure 6a, observations close but left of the threshold consist of the subgroup of always-taker, and observations close but right to the threshold consist of always-taker and treated complier. In Figure 6b, observations close but left of the threshold consist of never-taker and untreated complier, and observations close but right to the threshold consist of never-taker. The discontinuity in average monthly labor income at the threshold is very small in Figure 6a, indicating that there is no substantial difference between always-taker and treated compliers. On the other hand, the large discontinuity in Figure 6b suggests substantial differences in labor market outcomes between never-taker and untreated complier. Since average income of untreated complier and never-taker are considerably larger than those of never-taker alone, it follows that untreated complier performing much better than never-taker in terms of labor market outcomes. These results suggests that the IV results are not informative for never-taker, and are consistent with the notion that the subgroup of complier consists of refugee migrants with *a priori* better labor market perspectives.

To further characterize the subgroup of complier, I report in Table 9 split sample estimates of the first stage equation by various characteristics. If compliers have, on average, better labor market perspectives, I would expect first stage estimates to differ for characteristics that are commonly attributed to increase labor market performance. In sum, the results reported in Table 9 support this view. First-stage estimates are much larger for subsamples restricted to males than for females, and younger individuals who are not married or do not have children in their household in comparison to their counterparts. Interestingly, first-stage estimates do not differ with respect to education measured by being a college graduate but are slightly larger for individuals with prior work experience before migration than for individuals without prior work experience. The only result reported in Table 9 that does not support the view that compliers consists of individuals with better labor market perspective refers to the location, where the results of Table 9 suggest that the group of compliers is larger among refugee migrants located in East Germany than in West Germany.

Table 9:
Complier characteristics

	No	Yes
Sample restricted to:		
Female	0.26***	0.13*
	(0.05)	(0.07)
Age 30 or older	0.25***	0.19***
	(0.07)	(0.05)
Married	0.37***	0.14***
	(0.07)	(0.05)
Children in household	0.37***	0.12**
	(0.07)	(0.06)
Located in West Germany	0.28**	0.20***
	(0.11)	(0.04)
College graduate	0.21***	0.21**
	(0.05)	(0.09)
Without prior work experience	0.24***	0.18**
	(0.05)	(0.07)

Note: Split sample estimates of first-stage equation by subgroup. Estimates correspond to specification (1) in Table 7. Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.6 Robustness

3.6.1 Robustness of the RD design

In this section, I provide a number of robustness tests for the RD design estimates of this paper. First, I construct a placebo sample of refugees who do neither have a protection status in accordance with the Geneva convention nor subsidiary protection status. This sample consists mainly of rejected asylum seekers or accepted refugees who are accepted on humanitarian reasons (suspension of deportation). If the discontinuity in labor market outcomes at the threshold for Geneva convention refugees and refugees with subsidiary protection was caused by the increasing share of refugees with subsidiary protection, I would not expect to see a similar discontinuity at the threshold for the placebo sample. Figure B5 and Table B3 in the Appendix show the relationship between the notification date and labor market outcomes and the estimates of the RD design for the placebo sample, respectively. As expected, the plots in Figure B5 do not show a clear and sizable discontinuity at the threshold. Moreover, the

discontinuity is - if anything - positive. A similar conclusion can be drawn using the reduced form RD estimates in Table B3. In comparison with the estimates for the baseline sample, the placebo sample estimates are smaller and less precisely estimated across all specifications.

Second, I estimate the main results based on alternative definitions of the threshold. That is, instead of setting c equal to March 2016 in Equation (5) and (6), I use each month within an 18 month corridor around the original threshold in separate RD regressions as alternative cutoffs. If the result is indeed driven by the policy change, we would not expect to see similar large effects for alternative definitions of the threshold. Figure B6 shows coefficient plots of the estimated reduced form coefficients for the alternative cutoffs. Since the sample is relatively small, I see significant effects also for some of the alternative thresholds. However, the estimated effects become - except for one specification for the outcome variable *Any employment* - smaller. This is very reassuring for the identification strategy applied in this paper.

Third, I apply a Donut RD design. In a Donut RD design, the observations close to the threshold are excluded. The idea of this design is to avoid biased estimates due to sorting around the cutoff. While the discussion of the validity of the RD design above suggests that sorting is unlikely to be of importance in this study, it cannot be ruled out entirely. Hence, it is informative to what extent the results are driven by the observations close to the cut-off. Table B4 in the Appendix contrasts the baseline results from the section above with those obtained by a Donut RD design. As can be seen in Table B4, the estimates are hardly affected.

Finally, I additionally control for the full set of pre-determined control variables. While a valid RD design does not require the inclusion of covariates in the regression, it might increase precision of the estimates. On the other hand, if I do not find significant effects after including a set of control variables, this might hint to non-random allocation of the instrument around the cutoff, indicating that the baseline effects from the Section above might be caused by differences in pre-determined variables around the cutoff. Table B5 shows the reduced form estimate after controlling for the set of control variables and indicates that the estimation results are hardly affected.

3.6.2 Exploiting time of asylum application

Fuzzy RD designs enjoy great popularity in applied economic research as they provide estimates of the LATE under a set of mild assumptions that can be credibly tested and visualized (Bertanha and Imbens, 2020). However, a potential downside of the fuzzy RD design is that the identification of the LATE depends heavily on observations close to the threshold. This might be particularly problematic in case of survey data where the number of observations are, in general, rather small. To address concerns that my estimates of the LATE are only driven by an unreliable small number of observations at the threshold, I propose a second identification strategy to estimate the effect of subsidiary protection status on labor market outcomes which exploits the policy change in an alternative setting. This identification strategy is based on a comparison of refugee migrants who entered Germany at the same month and applied for asylum in the same month but received notification about the decision of the asylum application before and after the policy change. The basic idea of this identification strategy is that whether applicant cohorts from the same arrival and application month receive notification about the asylum application before or after the policy change depends solely on factors that are unrelated to labor market outcomes of refugees - such as the number of applications a caseworker has to process. Formally, I estimate the following system of equations by 2SLS:

$$Sub_i = \delta_{ma} + \alpha_1 \mathbb{1}\left[t_i > c\right] + \eta_i \tag{7}$$

$$Y_i = \gamma_{ma} + \beta_1 \hat{Sub}_i + \epsilon_i \tag{8}$$

where δ_{ma} and γ_{ma} are month of arrival times month of application fixed effects and all other variables are defined as above. In Equation (7) and Equation (8), the inclusion of month of arrival times month of application fixed effects allows for any systematic variation in subsidiary protection status and labor market outcomes across cohorts that arrived in Germany in the same month and applied for asylum at the same month. Consequently, the estimation of α_1 - which measures the effect of being notified about the protection status after c on the

probability of having a subsidiary protection status - and β_1 - which measures the effect of subsidiary protection status on labor market outcomes - is based on variation within cohorts that arrived in the same month and applied for asylum in the same month. Again, under the assumption that - conditional on the same arrival month and application month - receiving notification about the asylum application is as good as randomly assigned (independence) and does not affect labor market outcomes through other channels than protection status (exclusion), β_1 gives an estimate of the LATE for compliers.

Table B6 in the Appendix reports reduced form estimates of the effect of receiving notification about the asylum application after c on the probability of having subsidiary protection status (first row) as well as on labor market outcomes (second to fourth row) using month of application (column 1) and month of application times month of arrival (column 2) fixed effects. In column 3 of Table B6, I additionally add a set of control variables. Column 4 and 5 of Table B6 report results for the same specification using the placebo sample consisting of refugees without international protection as introduced above. Reported standard errors are clustered at the level of the arrival time application month. The first row of Table B6 makes clear that if decisions on asylum applications are made after March 2016, the likelihood of receiving subsidiary protection status is significantly higher even after flexibly controlling for the month of application, month of application times month of arrival, and adding control variables. The estimated results suggest that receiving notification after March 2016 led to an increase in the probability to have subsidiary protection status by 23 percentage points. The corresponding F-test on the excluded instrument is between 62 and 45 which underlines the relevance of the instrument. Turning to the effects on labor market outcomes, Table B6 illustrates as expected the negative effects of receiving notification after March 2016 on the probability of being employed as well as monthly earnings. It is comforting to see that I do not find any effects for the placebo sample, which suggests that time until receiving notification does not *per se* affect labor market outcomes.

Table B7 in the Appendix reports the corresponding 2SLS estimates (column 2), estimation results of a linear regression of each of the labor market outcomes on subsidiary protection

status controlling for arrival times application month fixed effects (column 1), and the estimation results obtained from the fuzzy RD design discussed above (column 3). While the 2SLS estimates from the fixed effect specification are smaller than the fuzzy RD estimates, they are still large and significant. The results suggest that subsidiary protection status reduces employment by 30 percentage points and monthly labor income by around 427 Euros. Similar to the results from the fuzzy RD design, the effect on employment seems to be largely driven by a reduction in full-time employment. Comparing the results with the OLS estimates in column 1, Table B7 illustrates, again, a large discrepancy, which suggests that the effect on complier might considerably larger as for other subpopulations.

Major concerns of this identification strategy are that the duration of the asylum procedure might have a direct negative effect on employment as suggested by Hainmueller et al. (2016), or that refugee migrants start only to look for employment after they receive the notification - which would reduce the time of job search for refugees receiving notification in the new policy regime in comparison to their counterpart - and might influence labor market outcomes directly. To assess if these concerns affect the estimation results, Table B8 in the Appendix shows estimation results of the IV strategy when additionally controlling linearly for (i) the time between application and receiving notification (column 2), (ii) the time between the notification and the interview date (column 3), or (iii) both (column 4).[51] The estimates reported in Table B8 show that including these control variables significantly reduces the power of the instrument and the IV results are estimated less precisely. However, as the point estimates become larger, this would suggest that both variables affect labor market outcomes positively.

In sum, the results shown in this section provide evidence that the RD design is robust to a number of checks and that the estimates of an alternative specification provide similar results to the fuzzy RD design which supports the obtained baseline estimates of a negative effect of subsidiary protection status on labor market outcomes.

[51] Please note that the interview month varies between respondents which helps identifying the parameters in this approach.

3.7 Discussion

The negative effect of subsidiary protection status on labor market outcomes can be explained by changes in labor supply. Subsidiary protection status likely reduces the expected length of stay in Germany, which potentially affects integration efforts of immigrants (Dustmann, 1993, 1997, 1999; Cortes, 2004). Investments in host country-specific human capital - such as language skills, schooling and training, obtaining knowledge about the host country's institutions and production methods - might be of particular importance for refugee migrants as their relocation decision is not entirely based on economic considerations but often the result of *ad hoc* decisions triggered by violence and individual persecution, making refugee migrants less economically selected than economic migrants.[52] Consequently, refugee migrants' host country-specific human capital is, in general, lower than that of economic migrants upon arrival, which translates into lower levels of wages and employment (Brell et al., 2020). Lower level of human capital suggests high incentives for refugee migrants to invest in country-specific human capital as the costs of investments due to, e.g., forgone wages are lower and the rate on return of the investment might be higher (Chiswick, 1978). On the other hand, the uncertainty that refugee migrants face in terms of length of stay in the host country might counteract incentives to invest in country-specific human capital as it affects the time span that allows to reap the gains of the costly investment. Based on these considerations, more insecure protection statuses such as subsidiary protection status lead to lower investments in country-specific human capital and might worsen labor market outcomes.

On the other hand, there might also be labor demand side effects, which can explain the negative effect of subsidiary protection status on labor market outcomes, consistent with the increasing literature that shows that employment of immigrants is affected by labor demand side factors (Åslund and Rooth, 2007; Azlor et al., 2020). Kosyakova and Brenzel (2020) provide anecdotal evidence that German firms think that the conditions to hire refugee migrants are not clearly outlined, which might create uncertainty about the duration of a potential employment of refugee migrants. If employment of refugee migrants is costly, firms prefer to hire

[52] For a survey of the adjustment of immigrants in labor markets, see Duleep (2015).

Table 10:
Fuzzy RD design estimates, perceived duration of stay and integration efforts

	Worries	Integration classes		Hours studying German	
	(1)	(2)	(3)	(4)	(5)
Subsidiary protection	0.59**	0.60***	0.65**	2.50***	2.33**
	(0.28)	(0.20)	(0.28)	(0.86)	(1.16)
Only unemployed	No	No	Yes	No	Yes
Observations	1454	1456	1060	1454	1061

Note: Fuzzy RD design estimates of the effect of subsidiary protection status on various outcomes measuring integration efforts. In column (1), the dependent variable is an ordinal response to the interview question: "Do you have worries that you cannot remain in Germany?" (1: no worries, 2: some worries, 3: a lot of worries). In column (2) and (3),the outcome variable is a binary variable whether a refugee migrant has attended a integration class in Germany. In column (4) and (5), the dependent variable is the number of hours an individual spends learning German per day. Estimates correspond to specification (1) in Table 7. Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

refugee migrants with better prospects of staying to regain their investment costs. This implies that firms might prefer to hire Geneva convention refugee migrants instead of subsidiary protection refugee migrants.

While the available data does not allow me to disentangle labor demand and labor supply side factors, I test in this subsection if subsidiary protection status also reduces refugees efforts to invest in host country-specific capital, which would be in line with the labor supply side explanation. Table 10 reports fuzzy RD design estimates for various outcomes measuring the refugees' uncertainty about the length of stay in Germany and investments in country-specific human capital. The dependent variable in column (1) is an ordinal measure of the answer to the interview question: "Do you have worries that you cannot remain in Germany?" ("no worries" is coded as 1, "some worries" is coded as 2, and "a lot of worries" is coded as 3). The dependent variable in columns (2) and (3) is a binary variable indicating whether a refugee migrant has attended an integration class in Germany, and the dependent variable in columns (4) and (5) is a variable measuring the number of hours a refugee migrants spent studying German, and are intended to measure refugees' investments in country-specific human capital. While the estimate in column (1) of Table 10 is consistent with the idea that subsidiary protection status increases the uncertainty about the length of stay in Germany for refugee

migrants, measures of investment in country-specific human capital are not negatively affected by subsidiary protection status. To test if this effect is driven by higher employment rates among refugee migrants with Geneva convention status, which might increase the opportunity cost to spend time in integration classes or studying German, I restrict the sample to those refugee migrants who are not employed at the day of interview in column (3) and (5). As the estimation results do not change for the restricted sample, I conclude that higher employment of Geneva convention refugees do not explain my findings. In sum, the results of Table 10 suggest that changes in refugees' labor supply due to subsidiary protection status do not explain the baseline findings of my paper, and labor demand side factors might be more important. A potential explanation for the positive effect on integration efforts could be that subsidiary protection refugees invest in host country-specific human capital to be able to prove in front of German authorities that they are willing to integrate in case their temporary residence permit might not be prolonged.

3.8 Conclusion

In this paper, I analyze the effect of refugees' protection status on labor market outcomes, focusing on a recent cohort of Syrian and Iraqi refugees. My empirical analysis exploits novel, plausible exogenous variation in the likelihood to receive subsidiary protection status due to a change in the assessment of the Federal Agency responsible for asylum claims to grant full refugee status in accordance with the Geneva convention. My results based on a fuzzy RD design suggest that subsidiary protection status has a substantial negative effect on labor earnings and employment probability, in particular, in the probability to be full-time employed. Further, I show in a detailed complier analysis that those refugee migrants who were affected by the policy change have *a priori* better labor market perspectives and have characteristics that are commonly attributed to improve labor market outcomes. My results are consistent with the causal mechanism that a reduction in the perception of permanent stay in the host country reduces refugees' willingness to invest in host country-specific human capital, which, in turn, reduces labor market performance. However, the results of the discussion section

show that refugees with subsidiary protection invest even more in country-specific human capital, which suggest that there might exist also demand side factors that explain my results. In sum, my empirical analysis confirms the existence of an economic and political trade-off in asylum policies as granting permanent residence presumably induces political costs but provides economic and social benefits by reducing unemployment.

Religious practice and student performance: Evidence from Ramadan fasting*

Erik Hornung†
University of Cologne
Department of Economics

Guido Schwerdt‡
University of Konstanz
Department of Economics

Maurizio Strazzeri§
University of Konstanz
Department of Economics
Graduate School of Decision Sciences

Abstract

Using two separate difference-in-differences frameworks, we investigate the impact of Ramadan fasting on educational outcomes exploiting arguably exogenous variation in annual fasting hours. Based on panel data from six waves of international TIMSS tests spanning 1995-2015, we find that fasting hours positively covary with student performance over time in Muslim majority countries, but not in other countries. Based on panel data from PISA for eight European immigration countries, we further show that performance gaps between students without and with Muslim majority country origin are lower when fasting is more intense. Both findings suggest that Islamic religiosity positively affects student performance.

*We thank seminar audiences at the University of Cologne, University of Konstanz, IZA, Ifo Institute, ASREC conference 2018 (Chapman), and the CESifo Area Conference on the Economics of Education 2018 for helpful comments and suggestions.

†Email: hornung@wiso.uni-koeln.de
‡Email: guido.schwerdt@uni.kn
§Email: maurizio.strazzeri@uni.kn

4.1 Introduction

Religious doctrines often contain cultural norms that either encourage or discourage secular education. While Protestantism and Judaism are associated with higher investment in human capital (see, e.g. Becker and Woessmann, 2009; Botticini and Eckstein, 2005, 2007; Chaudhary and Rubin, 2011), the opposite has been observed for Islam (see, e.g. Kuran, 2018). In many cases, the literature highlights a mechanism where human capital formation is encouraged through non-pecuniary or spiritual returns to investment. If religious norms shape attitudes towards education, should we expect more religious individuals to make differential human capital investments in accordance to their denomination? While religiosity itself is hard to measure, we can ascribe revealed preferences such as the observance of religious practice to higher levels of religiosity.

In this paper, we use the variation of average fasting hours during Ramadan to estimate the effect of the strictness of a specific Islamic religious practice on student's educational performance. The Islamic holy moth of Ramadan is one of the five pillars of Islam and an obligatory element for Muslim believers. During Ramadan, Muslims who reached puberty are required to refrain from drinking, smoking, and sexual activities from sunrise to sunset. As explained in more detail below, the month of Ramadan is set in accordance with the Islamic lunar calendar, which rotates over the solar Gregorian calendar and the seasons. If Ramadan falls into northern hemisphere summer, Muslim believers on the northern (southern) half of the earth fast more (less) hours than in winter. The change in fasting hours is further amplified by the distance to the equator. The resulting variation of daily fasting time for Muslim believers during Ramadan can be seen as exogenous in the context of schooling and therefore provides an ideal source to analyze the effect of religiosity as measured by the strictness of the religious practice of Ramadan fasting.

To exploit exogenous variation in religiosity approximated by the strictness of Ramadan fasting in the context of educational performance, we combine country-specific daily fasting hours with two repeated cross-sectional international student achievement test surveys,

TIMSS (Trends in International Mathematics and Science study) and PISA (Programme for International Student Assessment). Using TIMSS data for the period 1995 to 2015, we show that the length of daily fasting hours during the last Ramadan before the achievement test was taken, significantly increases the average performance of 8th graders (students around age 15) in Muslim majority countries, but not in non-Muslim majority countries. The use of results taken from tests that were executed *after* Ramadan allow us to abstract from the detrimental short-term consequences of fasting during Ramadan and focus on changes in educational performance that arise due to changes in religiosity in the long run. Results from our preferred specification suggest that an increase of average Ramadan fasting hours by 10 % increases math test scores by around 10 % of a standard deviation and science test scores by 13 %. These results are robust to the inclusion of a large set of student and school characteristics, and different categorizations of Muslim-majority countries.

Focusing on eight Western European countries that participated in the PISA study between 2003 and 2018, we find test scores to significantly increase in daily fasting hours for students whose families immigrated from Muslim-majority countries in comparison to students without Muslim-majority country origins. Such within-country specifications allow us to address concerns that TIMMS results are driven by spurious trends correlating with the seasonality of Ramadan fasting. An increase of average Ramadan fasting hours by 10 % significantly increases the difference in test scores between Muslim and non-Muslim students by 2.8 to 3.3 %, depending on the subject evaluated (mathematics, science, or reading). In sum, if we ascribe the strictness of a religious practice to higher levels of religious participation, our estimates reveal - based on two independent data sets - a positive effect of Islamic religiosity on student performance.

Using self-reported information on the attendance at religious services from the World Value Survey (WVS) covering the years between 1981 and 2014, we provide support for our interpretation that longer fasting hours during Ramadan early in life reflect religiosity. Indeed, we find that changes in fasting hours by the movement of Ramadan over the solar calendar are associated with a higher frequency of attendance at religious services among young Muslim

individuals relative to non-Muslim individuals within the same age group. Focusing on the age group 15 to 18, we find that an increase in average Ramadan fasting hours by 10 % increases the difference in weekly attendance of religious services between Muslim and non-Muslim students by 4.6 %, while we do not find such an effect for other age cohorts.

Theoretically, there are a range of mechanisms that may produce a positive relationship between changes in religiosity and changes in educational performance. Higher levels of religiosity correlate with certain character skills such as strength of motivation, ability to work on long-term plans, and socio-emotional regulations that may improve human capital formation in the long run (see Heckman, 2013). Religious practices such as Ramadan observance often require patience, self-control and perseverance. These traits can be further amplified within religious communities which also exercise social control over believers (Bankston and Zhou, 1995). If character skills exert positive externalities on human capital formation, we expect higher levels of religiosity to be positively related to educational performance. Another channel through which religiosity might influence educational performance is outlined in the theory of social identity popularized by Akerlof and Kranton (2002). This theory highlights the importance of the social environment of students in the context of schooling and predicts that the creation of a more unified school community increases the skill formation of students. Religious practices among students, i.e., the performance of identical rituals, are likely to affirm a common identity and hence increase educational performance if class composition is sufficiently homogeneous with respect to religious preferences.

We exploit the fact that we can identify the country of origin of immigrants in the PISA data to investigate the relative importance of the two possible channels mentioned above. The available information allows us to calculate the cohort share of Muslim students for all schools in our sample. If the common identity created by the performance of the religious practice among students is stronger the more students participate, we expect to see a larger increase in educational performance for Muslim students in schools with a high share of Muslims. Indeed, our estimations based on a triple-interaction specification reveal that the effect of Ramadan fasting between Muslim and non-Muslim students is larger in schools with a high

share of Muslim students in the cohort. Moreover, our estimates indicate that the overall positive effect of Ramadan fasting is mainly driven by Muslim students from schools with a high Muslim share. Although we are aware that other reason might have influenced our results - for instance, the presence of spill-over effects -, we tentatively interpret our estimates as evidence that religiosity impacts student performance by creating a common identity among students - and not by affecting their character skills.

Prior literature has highlighted that certain religious practices are costly and might crowd out other economically relevant activity. In a closely related study, Campante and Yanagizawa-Drott (2015) show that more intensive Ramadan fasting is related to lower GDP growth. Furthermore, Ramadan observers show lower performance during physically and mentally challenging tasks while fasting. More specifically, grades of university students have been shown to decline in the number of weeks a course overlaps with the month of Ramadan (see Oosterbeek and van der Klaauw, 2013). Consequently, the relationship between the intensity of religious practice and human capital formation is estimated to be negative in the short run. Furthermore, in utero exposure to Ramadan fasting has been shown to be detrimental to health outcomes, cognitive ability, and labor market performance in the long run (Almond and Mazumder, 2011; Majid, 2015).

In addition to the aforementioned literature on religiosity and economic outcomes, this paper mainly contributes to three other strands of literature. First, by providing causal estimates of the effect of the strictness of a religious practice on student performance, it adds to the scarce empirical literature that analyzes the effect of religiosity on educational outcomes. As we will discuss below, while theoretical contributions are abundant in both the economics and sociology literature, only Gruber (2005) provides causal estimates of the effect of religious participation on educational attainment. Focusing on Christian denominations in the US, he finds - similar to our results - positive effects of religiosity on educational performance. In most studies, the direction of causality is argued to run from education to religiosity. Glaeser and Sacerdote (2008) shows that education is positively related to the attendance of religious services but negatively related to religious beliefs in samples of individuals from

Western countries. The positive relationship between education and religiosity at the individual level is typically confirmed in the literature (for surveys see Iannaccone, 1998; Iyer, 2016). The predominant explanation is that education raises all forms of social interaction including attendance of religious services.[53]

Second, our paper relates to the literature of religious education and economic outcomes (Botticini and Eckstein, 2005, 2007; Becker and Woessmann, 2009). These studies highlight the importance of major reforms within Protestantism and Judaism that increased human capital formation, which in turn affected occupational choices, and economic prosperity. Our paper adds to this literature by indicating that not only reforms within religious denominations affect human capital formation but also changes in religiosity due to stricter religious practices. Finally, our paper relates to the increasing number of studies that focus on Islam and economic development. Studies in this area investigate the roots of the relative economic underdevelopment of the Islamic world. In general, this literature points at institutional differences with respect to norms that govern economic activity, e.g., inheritance systems, concepts of cooperations, or interest rate restrictions (Kuran, 2003, 2004; Rubin, 2011). However, other studies have stressed the importance of differences in educational and scientific advancement due to norms and other restrictions imposed by Muslim authorities that impeded educational outcomes of Muslims (Chaudhary and Rubin, 2011; Chaney, 2011, 2016). While we do not attempt to provide a direct test of this hypothesis, our results indicate that at least current Islamic religiosity cannot be considered as barrier to human capital formation.

The remaining paper is organized as follow. In Section 4.2 we discuss the existing literature on education and religiosity and the religious practice of Ramadan fasting and how it might affect student performance. Section 4.3 describes our data set and Section 4.4 explains our empirical identification strategy. Section 4.5 presents our main estimation results based on TIMSS and PISA data and Section 4.6 provides a discussion of potential mechanisms that might explain our findings. Finally, Section 4.7 concludes.

[53] In the context of developing countries, religious organizations may differentiate themselves through the strength of religious beliefs and provide educational services with differential quality (Iyer et al., 2014).

4.2 Background

4.2.1 Religiosity and education

In the social science, it has long been argued that scientific advancements would lower societies' attachment to religions. The main idea of the so called "secularization hypothesis" is that an increase in critical thinking reduces believes in supernatural forces. Early empirical investigations based on cross-country data question this view as many of these studies find a positive correlation between the level of education and various religious measures (for a survey, see Iannaccone (1998)). Based on these observations, McCleary and Barro (2006) hypothesize that educated individuals have a higher level of abstract thinking which is necessary to understand religious ideas. Another explanation proposed by Glaeser and Sacerdote (2008) assumes that more educated individuals obtain higher returns of social activities such as church going. However, recent studies which put emphasis on causality provide more often evidence that confirms the secularization hypothesis. When analyzing panel data for German cities between 1890 and 1930, Becker et al. (2017) find that education is negatively correlated with church attendance. Similarly, studies that exploit changes in compulsory schooling laws consistently identify negative effects of education on various measures of religiosity (Arias-Vazquez, 2012; Cesur and Mocan, 2013; Hungerman, 2014; Mocan and Pogorelova, 2017). In contrast, Brown and Taylor (2007) use individual level panel data of the National Child Development Survey and find a positive association between church attendance and education and Sander (2002) who instruments education with parental schooling does not find significant effects of education on the frequency of attending religious services.

Other scholars have argued that the direction of causality might run the opposite direction, i.e., that religiosity might influence educational outcomes. Motivated by the debate about educational achievements of students in Catholic schools (Coleman et al., 1982; Evans and Schwab, 1995; Neal, 1997), Fan (2008) provides a model in which religious participation is conductive to children's human capital formation. In this model, educational attainment of children is determined by their parental human capital as well as their social capital. As religious partic-

ipation positively influences the stock of social capital, parents have an incentive to allocate a positive amount of time to religious activities. The idea that parental concerns about the moral development of their children induces religious participation has been noted many times in sociological literature. For instance, Wilson (1978) states that parents consider the church as a place of character building for their children, which is why "couples with growing children have the highest rate of church attendance." Indeed, there seems to be a link between religious activities of children and their character skills, which has been shown in a number of empirical studies. Several religious measures (e.g., church attendance, importance of religious faith) are inversely related to juvenile drug use, delinquency, absence from school, and high levels of depression (Cochran and Akers, 1989; Evans et al., 1995; Freeman, 1985; Wright et al., 1993). Additionally, religiosity has been linked to youth political and civic involvement during adolescence (Smith, 1999). Such pro-social behavior has been repeatedly used in the education literature to measure non-cognitive skills of children (Heckman and Rubinstein, 2001; Lleras, 2008; Jackson, 2012). This suggests that religiosity influences this skill set, e.g., by providing moral directives, role models, and social control (Smith, 2003; Bankston and Zhou, 1995). Non-cognitive abilities or character skills are associated with improvements in student performance (Heckman and Rubinstein, 2001). The "strength of motivation, an ability to act on long-term plans, and the socio-emotional regulation needed to work with others" are skills necessary to perform well in school (Heckman, 2013). Hence, if character skills exert positive externalities on human capital formation, we should also expect religious practices to be positively related to educational performance.

Another channel through which Ramadan fasting might influence educational performance is outlined in the theory of social identity popularized by Akerlof and Kranton (2002). This theory highlights the importance of the social environment of students in the context of schooling and predicts that the creation of a more unified school community increases the skill formation of students. Religious practices among students, i.e., the performance of identical rituals, are likely to affirm a common identity and hence increase educational performance if class composition is sufficiently homogeneous with respect to religious preferences.

Despite the theoretical arguments that suggest a positive effect of religiosity on educational outcomes, there is almost no empirical research that causally addresses this question. An exemption is the study of Gruber (2005) who uses as an instrument for religious participation the historic share of the population in an area which is of an individual's religion. He motivates the use of this instrument by pointing to some literature that shows a correlation between higher religious density and attendance of religious services (e.g., Phillips, 1998). His results indicate minor but positive effects of religious participation on educational outcomes. For instance, an increase of religious participation by 100 % is associated with a raise in the years of schooling by roughly 0.5 years.

As the above discussion has shown, two main issues arise when studying the interplay between religiosity and education. First, an individual's religiosity cannot be observed. Different approaches to proxy religiosity have been applied in the literature. Most of them rely on variables that are assumed to be correlated with an individual's religiosity. Such variables could either be behavioral outcomes - for instance, the frequency of attending religious services - or self-stated assessments about religious beliefs. Related to the concept of using behavior outcomes to measure religiosity is the strictness of a religious practice. Orthodox Jews who conduct no business on the sabbath might be perceived as more religious than Jews who do not observe to this religious practice. In general, studies that apply behavioral and self-stated measures of religiosity do not find varying effects.

Besides the difficulty with respect to measurement, empirical studies focusing on the effect of religiosity on educational performance need to address endogeneity concerns. It is important to separate causal effects of religiosity on education from the impact of other correlated variables. By using the religious practice of Ramadan fasting, our identification strategy makes use of a behavioral measure of religiosity. Additionally, since the month of Ramadan is set in accordance of the lunar calendar, average fasting times vary over time and provide us with an ideal source to isolate the causal effect of strictness of a specific religious practice. However, it is not clear to what extent the variation in fasting hours capture differences in religiosity. To shed more light on this, we discuss in the next subsection in more detail the practice of

Ramadan fasting and its physiological and psychological consequences, which in turn could influence student performance.

4.2.2 Ramadan fasting, religious participation, and education

Ramadan is the ninth month of the Islamic calendar (*Hijri*) and is perceived as a sacred month by Muslim believers. During the month of Ramadan, Muslims around the world are required to fast (*Sawm*) from sunrise to sunset to commemorate the first revelation of the Quran to Muhammad. It is one of the five pillars of Islam and therefore a mandatory element of the Islamic religion. Besides abstention from food, Ramadan requires Muslims to refrain from drinking, smoking, and sexual activities during the day.

Ramadan fasting is an integral part of the life of more than a billion Muslim believers. A notable exemption from the practice of Ramadan fasting applies to young children, sick people, elderly, and breastfeeding women. Children are required to fast during Ramadan when they reach puberty. However, even younger children are often encouraged to practice Ramadan fasting in an appropriate way. Also, due to its social character with frequent family visits, adjusted meal times and regular community gatherings, it is likely that children enjoy to participate Ramadan fasting before the age of adolescence to imitate the behavior of their peers, feel "grown up," and be part of the religious society.

Ramadan fasting is a physiologically demanding task and direct negative consequences have been well documented in the literature. During the month of Ramadan, individuals reduce their level of activity, loose weight, suffer from headaches, and have more difficulties to concentrate (see, e.g., Afifi, 1997; Leiper and Molla, 2003; Meckel et al., 2008). While Toda and Morimoto (2004) suggest that these negative effects in terms of physical conditions do not persist beyond the month of Ramadan, they might translate into future outcomes by affecting different stages of the production process of various socio-economic outcomes. For example, Campante and Yanagizawa-Drott (2015) hypothesize that Ramadan fasting reduces worker productivity and show that GDP growth in Muslim majority countries is lower in years in which average fasting hours are longer. A number of studies find that prenatal exposure

to Ramadan fasting reduces birth weight, and increases the likelihood of being disabled as an adult (Almond and Mazumder, 2011; Van Ewijk, 2011; Almond et al., 2015; Majid, 2015). Oosterbeek and van der Klaauw (2013) show that the number of weeks an academic class falls in the month of Ramadan negatively affect the final grade of Muslim students.

However, when considering long-term socio-economic outcomes, psychological effects of Ramadan fasting due to changes in individual lifestyles and social climate gain in importance. Schielke (2009) reports in his anthropological study that during the month of Ramadan the social life within Muslim societies underlies enormous transformations. Mosques are filled with believers during night who perform the additional voluntary *tarawih* prayers. Wealthy citizens organize events with free food for the poor and needy at fast-breaking time (table of the merciful). Prewdawn (*suhur*) and fast-breaking (*iftar*) meals are social events and often celebrated with friends and family, and the end of the fasting period is marked by major festivals (*Eid al-Fitr*). Overall, the social climate is filled with social, moral, and pious commitment, which influences individual decision making and might extend beyond the month of Ramadan itself.

There is a close link between religious practices, beliefs, and religious participation. For instance, Clingingsmith et al. (2009) analyze the effect of the Hajj pilgrimage using a Pakistani lottery that decides who is in a given year allowed to participate in this religious practice. They find that eight month after the pilgrimage Hajjis are more likely to report that (i) they are regarded as religious persons, (ii) they increased the observances of other religious practices, (iii) they decreased participation in other localized practices, (iv) they changed their attitudes towards equality and harmony than the control group who was not allowed to participate in the Hajj due to the lottery outcome. Further, Campante and Yanagizawa-Drott (2015) who use the same identification strategy as our study report that a changes in average fasting hours during Ramadan affect reported subjective well-being, religious participation, beliefs, and attitudes of religiously committed Muslims. We might also expect that the strictness of a religious practice is an influential factor for religious participation especially in case of young people (religious socialization). As our student performance data is based on students at the

age of around 15, fasting hours before the test was taken proxy the strictness of the first Ramadan fasting they experience. To provide evidence about the link between the strictness of the religious practice of Ramadan fasting and religious participation for this age group, we exploit data from the World Value Survey (WVS, see Section 4.6).

4.3 Data

We first collect country-level data on fasting hours during Ramadan, which we need to implement our two separate difference-in-differences frameworks. To address our research question as comprehensively as possible, we then use data from all existing waves of the two largest international large-scale student assessments: the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). Finally, we combine fasting hours and educational data to construct our two main estimation samples.

4.3.1 Ramadan fasting

To calculate fasting hours during Ramadan, we follow the study of Campante and Yanagizawa-Drott (2015). The online data service of the Astronomical Applications Department of the US Naval Observatory provides information about sunrise and sunset times for any geographical coordinate on earth. We download yearly sunrise and sunset times for all countries included in this study by using the coordinates of the capital. Ramadan dates in terms of the Gregorian calendar are obtained using the calendar converter from Islamic Philosophy Online. Combining these two data sources, we calculate average fasting hours as the average daily daylight hours during Ramadan.

4.3.2 TIMSS data

Our first identification strategy uses the Trends in International Mathematics and Science Study (TIMSS). TIMSS was first conducted in 1995 and, including the latest wave of 2015, covers a time period of 20 years. Its goal is to provide data on the performance in mathematics and science of 4th and 8th graders around the world, which is comparable across countries as

well as over time. To capture direct effects of the observance of Ramadan fasting, we focus on the performance of 8th grader (age between 14 and 15), as 4th grader (between 10 and 11) are less likely to practice the religious activity due to their young age.

TIMSS is conducted on a regular 4 year cycle, which adds up to six distinct assessment tests between 1995 and 2015. The number of countries included in a survey varies between 32 and 44. Over the entire period, more than 70 countries participated in TIMSS. The participating countries in TIMSS are geographically very diverse. While most countries are either located in Europe (31) or Asia (26), we can also obtain information about countries in Africa (6), Americas (5), and Oceania (2). Especially the large set of countries from Middle East and Northern Africa facilitates our analysis since these countries have a high share of Muslim inhabitants.

To provide a valid and reliable measurement of student achievements, TIMSS employs a two-stage random sample design. In the first stage, a random sample of schools is drawn for each country. Based on this random sample of schools, one or more classes of 8th grader are selected in the second stage, of which all students are tested. To ensure national representativeness, sampling probabilities of schools are set in accordance with the rules of the TIMSS & PIRLS International Study Center, and supervised by the Canadian government agency Statistics Canada (LaRoche et al., 2016).

A valuable feature of TIMSS is its application of a concurrent calibration method that allows measurements of time trends in student achievements. The original scales for overall mathematics and science were established in the first wave in 1995 by setting the mean of the national average scores for all participating countries to 500 and a standard deviation of 100. The method of concurrent calibration enables TIMSS to link all successive surveys to the same metric. This process requires a large proportion of test items to be similar between two successive tests. The test items used in the previous assessment together with linear transformations allow the comparison of test scores between different waves of the survey (Yamamoto and Kulick, 2016).

Besides test scores of students, TIMSS collects a comprehensive set of background information about students, teacher, and schools. The students questionnaire contains information

about personal characteristics and family background of students. Teacher provide information about their personal characteristics and their behavior in class, and school principals assess the quality of school resources. While some questions are asked in all waves of TIMSS, some information is missing or not directly comparable across waves.

We combine all available waves of TIMSS to construct a repeated cross-sectional data set. With respect to background characteristics, we select those variables that are available in all waves. We refrain to use variables from the teacher questionnaire as one student might be linked to more than one science teacher, which makes it difficult to apply a consistent approach to merge the data. Since our identification strategy outlined below rests on within country variation over time, we restrict our sample to countries that participated at least twice in TIMSS. We also exclude countries or part of countries that participated with different names and demarcations. For example, in some waves, England and Scotland participated individually, in other waves they participated together. The same is true for states of the Former Republic of Yugoslavia. A complete list of all countries included in this paper based on TIMSS data can be found in Table C1 in the Appendix.

Since TIMSS does not collect any information about a student's religious denomination or parental background, our analysis is constrained to country-level variation. We use information provided by the PEW Research Group about the percentage of Muslim inhabitants in a country and add this information to our data set. Overall, we identify 18 countries with a Muslim majority: Kazakhstan, Lebanon, Malaysia, United Arab Emirates, Qatar, Bahrain, Oman, Indonesia, Syria, Egypt, Kuwait, Saudi Arabia, Palestine, Turkey, Jordan, Morocco, Iran, and Tunisia. The 3rd column of Table C1 in the Appendix reports the share of Muslims by country and Figure C1 in the Appendix displays the location of all Muslim majority countries on a world map.

Table 11 gives an overview of the main variables used in our first identification strategy by country status and years.[54] Our unbalanced repeated cross-sectional data set is based on more than 1.15 million student observation from 56 countries. As can be seen in Table 11, average

[54]Table C2 in the Appendix summarizes the large set of control variables.

Table 11:
Descriptive statistics: TIMSS, 1995 to 2015, main variables

	All	1995	1999	2003	2007	2011	2015
Muslim countries							
Test scores							
Math	415.10	402.70	426.47	405.99	401.18	415.63	429.35
	(92.23)	(53.39)	(94.50)	(84.17)	(90.71)	(95.25)	(96.85)
Science	435.67	440.19	430.22	429.62	434.01	436.72	443.02
	(93.02)	(65.12)	(89.60)	(81.95)	(90.97)	(95.12)	(104.87)
Fasting hours							
Year before test	12.22	11.17	10.51	10.74	11.79	12.98	13.97
	(1.45)	(0.28)	(1.10)	(0.69)	(0.20)	(0.45)	(0.83)
Number of observations							
Students	364,786	9,084	40,072	58,036	48,780	103,919	104,895
Countries	18	2	7	12	11	16	14
Non-Muslim countries							
Test scores							
Math	494.06	509.71	499.12	480.56	475.12	489.74	509.99
	(107.44)	(99.60)	(106.01)	(111.81)	(109.05)	(109.54)	(101.79)
Science	493.56	505.35	498.00	482.49	475.67	490.03	508.81
	(105.45)	(94.06)	(106.38)	(112.11)	(106.57)	(107.65)	(97.97)
Fasting hours							
Year before test	11.56	10.80	9.86	10.38	11.62	13.04	14.13
	(2.39)	(1.14)	(2.22)	(2.06)	(0.43)	(1.19)	(2.54)
Number of observations							
Students	805,168	159,141	132,409	146,815	77,697	146,190	142,916
Countries	38	21	29	30	18	26	24

Note: Mean values of main variables. Standard deviations are reported in parentheses. A country is categorized as a Muslim country if more than 50 % of its population is Muslim. Applied population weights are standardized to sum up to 1 for each country-year cluster.

mean fasting hours by country status varies significantly over the years. For Muslim majority countries, average daily fasting hours were only around 10.5 hours in 1999 and increased to almost 14 hours in 2015. This number masks a significant amount of heterogeneity across countries. For instance, in our sample average daily fasting hours vary between 9.4 to 14.9 hours for Turkey and only between 11.9 and 12.5 for Malaysia. Additionally, average test scores in Muslim majority countries are consistently lower than in non-Muslim majority countries, which highlights the importance to adjust our empirical strategy for time invariant country-specific factors that are relevant for potential level differences in test score across countries.

4.3.3 PISA data

Our second identification strategy is based on the Program for International Student Assessment survey (PISA). PISA is a worldwide study to assess and compare performance of 15-year-old students in mathematics, science, and reading. It was first conducted in 2000 and is repeated every 3 years. The seven publicly available waves of PISA contain information of 41 to 72 countries. In general, PISA and TIMSS are very similar in its set-up. However, two major differences need to be mentioned here.

First, instead of focusing on the grade of students, PISA's target population depends on age. Hence, while also relying on a two-stage sampling procedure, PISA chooses in the second stage randomly 35 15-year-old students in selected schools. Second, starting with the second wave, PISA collects information about the country of origin of students' parents for 8 Western European countries in the sample. These countries are: Austria, Belgium, Switzerland, Germany, Denmark, Finland, Great Britain, and Netherlands. We use this information to identify Muslim students at the individual level. We define a student as being Muslim if both parents are from a Muslim majority country. Unfortunately, countries that collect data on parental country of origin use different questionnaires to obtain this information. Additionally, countries of origin are sometimes grouped, making it difficult to assess the dominant religion of the region.

Table C3 in the Appendix lists all parental countries of origin by wave and country and shows which countries (or groups of countries) are selected as Muslim majority countries. Again, to define Muslim majority countries we use information provided by PEW Research group. Please note that for Denmark, Finland, and Great Britain, we cannot identify any Muslim students in some years. Overall, the share of Muslim students based on our definition seems to be rather small.[55] We assume that this is the result of grouping more than one country into one item in the PISA questionnaire. This implies that we potentially misclassify a significant share of students as non-Muslim even though their parents might have Muslim

[55]Table C4 in the Appendix shows the share of Muslim students by country and year for the 8 countries used in our analysis.

Table 12:
Descriptive statistics: PISA, 2003 to 2018, main variables

	All	2003	2006	2009	2012	2015	2018
Muslim students							
Test scores							
Reading	418.78	392.80	402.41	419.68	434.64	430.85	419.37
	(92.60)	(95.60)	(101.26)	(84.32)	(88.96)	(89.85)	(92.94)
Science	414.59	388.33	402.77	416.43	426.19	425.81	416.15
	(86.88)	(84.26)	(86.12)	(85.20)	(86.81)	(86.33)	(86.83)
Math	432.69	417.16	421.37	434.29	439.63	440.77	434.70
	(81.88)	(80.92)	(85.05)	(79.37)	(83.55)	(81.82)	(78.49)
Fasting hours							
Year before test	13.62	8.85	10.63	12.59	14.71	16.30	16.54
	(2.75)	(0.38)	(0.15)	(0.05)	(0.32)	(0.54)	(0.72)
School characteristics							
Share of Muslim students	0.20	0.19	0.24	0.23	0.19	0.20	0.19
	(0.18)	(0.14)	(0.23)	(0.20)	(0.15)	(0.16)	(0.16)
Number of observations							
Students	10,906	986	1,203	2,117	2,317	2,343	1,940
Countries	8	6	6	7	8	8	6
Non-Muslim students							
Test scores							
Reading	504.43	507.90	505.87	503.59	507.82	503.83	497.52
	(92.93)	(90.15)	(96.11)	(90.91)	(89.69)	(92.17)	(97.84)
Science	514.25	512.75	521.04	519.17	518.95	510.60	502.96
	(94.31)	(97.01)	(94.98)	(94.03)	(91.47)	(94.30)	(92.71)
Math	515.80	523.34	520.56	517.64	514.70	509.57	508.88
	(88.97)	(92.63)	(89.96)	(89.50)	(89.97)	(85.31)	(85.21)
Fasting hours							
Year before test	13.26	8.54	10.52	12.61	14.78	16.46	16.72
	(3.08)	(0.59)	(0.24)	(0.09)	(0.44)	(0.79)	(0.85)
School characteristics							
Share of Muslim students	0.02	0.02	0.02	0.03	0.03	0.03	0.02
	(0.06)	(0.05)	(0.05)	(0.07)	(0.06)	(0.06)	(0.05)
Number of observations							
Students	340,856	49,028	56,933	58,438	60,694	59,264	56,499
Countries	8	8	8	8	8	8	8

Note: Mean values of main variables. Standard deviations are reported in parentheses. A student is defined as being Muslim if his or her mother and father are from a Muslim majority country. The countries included are Austria, Belgium, Switzerland, Germany, Denmark, Finland, Great Britain, and Netherlands. Applied population weights are standardized to sum up to 1 for each country-year cluster.

majority country origin. For our empirical analysis, which relies on the comparison between Muslim and non-Muslim students, this might result in a downward bias of our estimates.

PISA collects various background characteristics at student, parent, and school level. When combining the 6 waves between 2003 and 2018 for the 8 countries mentioned above, we again

select an array of variables which are comparable across surveys. Table 12 gives an overview of our main variables used by individual status and year.[56] Similar to the TIMSS data, Table 12 shows a significant amount of variation in average daily fasting hours and a level effect in test scores between students with Muslim majority country background and those without. Additionally, Table 12 depicts a strikingly difference in the share of Muslim students in school between students from Muslim majority countries and other students, indicating a selection of Muslim students into same schools.

4.3.4 Combining fasting hours and educational data

Both international achievement test studies are matched with the information about average daily Ramadan fasting hours. Each student observations is matched with the average fasting hour of the country in which the test was taken. With respect to timing, we use fasting hours of the last Ramadan before the test was taken by a student. This can only be accurately done for TIMSS data in which the test dates are known. For the PISA data, in which test dates are not publicly available, we use the first day of the year in which the test was officially taken as test date.[57]

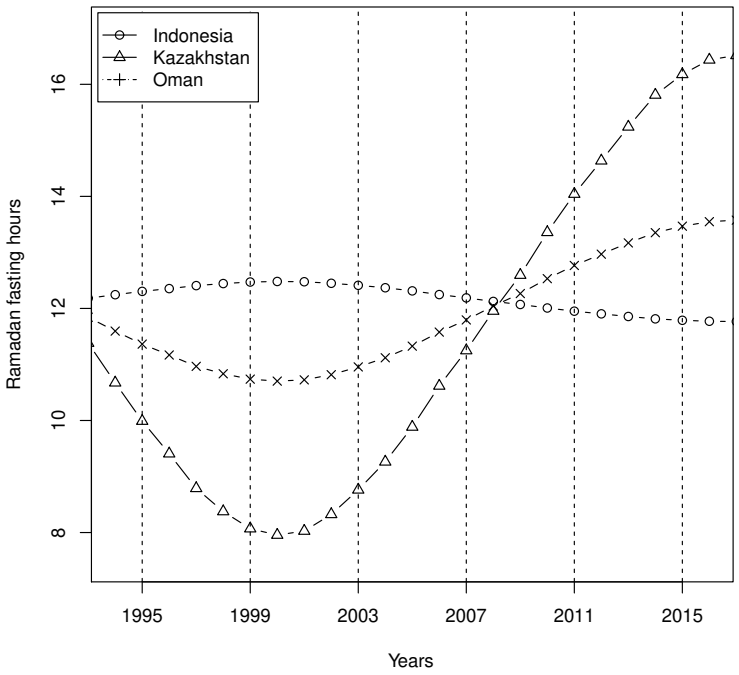
4.4 Empirical strategy

We exploit the different variation of average Ramadan fasting hours over time across countries. Since the month of Ramadan is set in accordance with the Islamic lunar calendar, it rotates over the Gregorian calendar and the seasons, which leads to a variation of daily Ramadan fasting hours over time. If Ramadan falls into northern hemisphere summer, Muslim believers on the northern (southern) half of the earth fast more (less) hours than in winter. The strength of the variation depends on the latitude. Countries which are located closer to the equator experience less variation of average Ramadan fasting hours over time.

Figure 7 shows the average fasting hours of 3 Muslim-majority countries in our sample. At the beginning of the time period depicted in Figure 7, Ramadan falls into the northern

[56] Table C5 in the Appendix summarizes our control variables.
[57] PISA tests take usually place in the first quarter of a test year.

Figure 7:
Average daily fasting hours during Ramadan, 3 countries, example



Note: Each line shows the average fasting hours during Ramadan, measured by the hours between sunrise and sunset in each country's capital. The dotted vertical lines refer to the test date of each TIMSS wave.

hemisphere winter. In this case, Kazakhstan and Oman fast on average less than 12 hours. The opposite is true for Indonesia, which capital is located in the southern hemisphere. The average fasting then increases over time for Muslim believers in Kazakhstan and Oman, while it decreases for Muslims in Indonesia. Figure 7 also shows how the fasting hours differ between countries. Since Oman is closer to the Equator than Kazakhstan, the variation of average Ramadan fasting hours in Oman is less pronounced. Note that the time period under analysis does not include a full rotation of the month of Ramadan over a year, which takes place roughly every 33 years. Hence, the increasing lines for Oman and Kazakhstan in Figure 7 will reach a peak followed by a decrease and show the exact same pattern again.

We use this exogenous variation of fasting hours to determine the effect of the strictness of a specific religious practice on educational outcomes. Longer fasting hours naturally enhances the physiological impact of Ramadan. Similarly, the change in daily routines and societal environment are likely stronger affected by an increase in average fasting hours. To exploit

this sort of variation we first estimate the equation:

$$Test_{ict} = \beta_1 Ramadan_{ct} + \delta_c + \gamma_t + \epsilon_{ict}, \tag{9}$$

where $Test_{ict}$ is the standardized achievement test score of student i in country c in year t, $Ramadan_{ct}$ is the logarithmized average daily number of fasting hours during the last Ramadan before the test was taken, δ_c and γ_t represent a set of dummy variables to capture country and year fixed effects, respectively.

We use country fixed effects to account for time-invariant differences across countries that capture geography and cultural factors that may be correlated with $Ramadan_{ct}$. Year fixed effects are intended to capture effects that affect student test scores over time but are constant across countries. To reduce residual variation, we further include a set of control variables (student characteristics, school characteristics).

We expect Ramadan fasting to have a stronger effect in Muslim majority countries. Hence, we first estimate (9) for countries with a Muslim majority population and expect β_1 be positive. To further investigate the linkage between Ramadan fasting and the share of Muslims in a country, we extend our sample to also include Muslim minority countries and add interaction terms in (9):

$$Test_{ict} = \beta_1 Ramadan_{ct} \ \%Muslim_c + \beta_2 Ramadan_{ct} + \delta_c + \gamma_t + \eta_{ct} + \epsilon_{ict}, \tag{10}$$

where $\%Muslim_c$ is the share of Muslim inhabitants in a country c, and η_{ct} a set of Muslim majority-by-year fixed effect which control for year effects that affect Muslim majority countries differently. In this specification, we expect β_2 to be close to zero, as no effect should be seen in countries with vanishingly small Muslim population. On the other hand, β_1 should again be positive. Adding the share of Muslim inhabitants in a country imposes the restriction of a linear effect. To provide an alternative specification and to allow for effect heterogeneity, we also replace $\%Muslim_c$ with dummy variables indicating the Muslim share by quartiles of the distribution of countries' Muslim share from our sample.

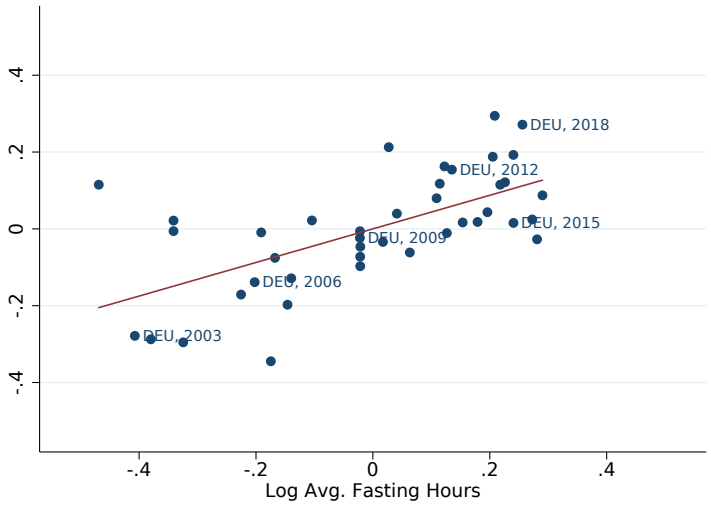
In our second identification strategy based on PISA data, we further exploit the fact that we know the country of origin of students' parents. We define a student as being Muslim if his or her parents are from a Muslim majority country. The identification strategy is based on the following estimation equation:

$$Test_{ict} = \beta_1 Ramadan_{ct} \;\; Muslim_{ict} + \beta_2 Ramadan_{ct} + \beta_3 Muslim_{ict} + \delta_c + \gamma_t + \epsilon_{ict}, \quad (11)$$

where $Muslim_{ict}$ is a dummy variable indicating whether a student is an first of second generation migrant from an Muslim majority country, and all other variables are defined as above. In some specifications, we replace δ_c and γ_t with country times year fixed effects, η_{ct}, to flexibly allow for country-specific time trends. In this specification, β_1 measures the difference of the marginal effect of Ramadan fasting between Muslim and non-Muslims, which we expect to be positive.

Figure 8 illustrates the variation in performance gaps between students whose parents are from Muslim majority countries and all other students that we exploit in the estimation of Equation (11). Figure 8 plots country-and year-specific performance gaps (for simplicity only gaps for Germany are labeled) between both types of students against average daily fasting hours. Both variables are adjusted by their country means. Figure 8 provides clear evidence for a positive correlation between performance gaps and fasting hours. It also highlights the nature of our identifying assumption. We can identify β_1 as long as year-and country-specific shocks to student performance that are unrelated to religious participation but still correlated with fasting hours affect first and second generation migrants from an Muslim majority country in a similar way as all other students. For instance, this assumption is violated if countries' educational spending is correlated with average fasting hours and its impact on student performance differs between students from Muslim majority countries and other students. We address this concern below in Section 4.5.

Figure 8:
Performance gaps and Ramadan fasting hours, science scores, PISA



Note: Performance gaps in science scores between students whose parents are from Muslim majority countries and all other students and logarithm of average fasting hours during Ramadan before test was taken. Both variables are adjusted by their country means. Observations for Germany are labelled with *DEU, year*. Equivalent plots for math and reading test scores can be found in the Appendix (Figure C2).

4.5 Results: Ramadan fasting and student performance

4.5.1 Evidence from TIMSS data

In this section we discuss our results with respect to overall effects of Ramadan fasting on student achievements. Estimates of the effect of Ramadan fasting are presented in Table 13 and Table 14 based on TIMSS and PISA data respectively. To account for the complex data structure, we apply student population weights throughout the entire analysis. These weights are adjusted to sum up to an equal number across country-year cluster to avoid biased estimates towards larger countries.[58] Additionally, we report cluster-robust standard errors at country-year level in all specifications.

Table 13 reports results from various specifications of Equation (9) and Equation (10) based on TIMSS data. All specifications are estimated separately for math and science test scores and include year-fixed effects. To facilitate the interpretation of our estimations, we rescale test

[58]Our results do not change if we apply students weights that do not account for the size of the student population. Additionally, the interpretation of our results are not affected if we use house weights which account for differences in sample size. Results are available upon request.

Table 13:
Estimation results, TIMSS, 1995 to 2015

	Only Muslim countries						All countries			
	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ramadan	2.25*	2.18**	1.40***	1.70***	1.05***	1.31***	-0.20**	-0.24**	-0.32***	-0.37***
	(1.22)	(0.95)	(0.36)	(0.41)	(0.35)	(0.40)	(0.10)	(0.10)	(0.12)	(0.14)
Ramadan x % Muslim							1.02***	1.04**		
							(0.37)	(0.46)		
Ramadan x $\mathbb{1}(q_{25} < \%\text{ Muslim} \leq q_{50})$									0.17	0.20
									(0.13)	(0.12)
Ramadan x $\mathbb{1}(q_{50} < \%\text{ Muslim} \leq q_{75})$									0.33**	0.30
									(0.14)	(0.19)
Ramadan x $\mathbb{1}(\%\text{ Muslim} > q_{75})$									1.09***	1.16***
									(0.35)	(0.39)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Muslim-by-year FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes
R squared	0.03	0.02	0.17	0.12	0.28	0.25	0.50	0.45	0.50	0.45
Observations	364,786	364,786	364,786	364,786	364,786	364,786	1,169,954	1,169,014	1,169,954	1,169,014
Cluster	62	62	62	62	62	62	210	210	210	210

Note: Estimation of standardized achievement test scores on logarithmized average fasting hours before the test was taken. A country is categorized as a Muslim country if more than 50 % of its population is Muslim. A list of the included control variables can be found in Table C2 in the Appendix. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country times year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

scores of each wave to the mean and standard deviation of the test scores of the first wave. Hence, the estimated coefficients can be interpreted in terms of one standard deviation of the distribution of test scores of the first TIMSS wave.

In columns (1) to (6) of Table 13, we restrict the sample to Muslim majority countries and add step-by-step country-fixed effects and other control variables. In all three specification, we see a positive effect of Ramadan fasting hours on test scores. The coefficient is sizable and significant in all specifications for math and science test scores. In columns (5) and (6), our preferred specification, our estimates suggest that an increase of average Ramadan fasting hours by 10 % increases math test scores by around 11 % of a standard deviation and science test scores by 13 %.

Extending our sample to Muslim minority countries and using different interactions of Ramadan fasting with the level of Muslims in a population, our previous results are largely confirmed. In columns (7) and (8), we show that, as expected, the effect of Ramadan fasting on student achievements increases with the share of Muslims in a population. Interestingly, we

find a small effect of Ramadan fasting on test scores for countries with a small Muslim population. In these countries the Ramadan coefficient might be picking up some small time effects caused by unobserved factors correlated with Ramadan fasting hours. However, the finding that these effects are small and, if anything, negative strengthens our finding of substantial positive effects of Ramadan fasting in Muslim majority countries.

To test if the interaction between Ramadan fasting and the share of Muslims in a country is non-linear, we use a second specification as shown in columns (9) and (10). Here we assume effect heterogeneity across groups of countries with different Muslim shares in the population. We create binary variables indicating whether the share of Muslims in a country is within the 25th and 50th, 50th and 75th, or above the 75th percentile of the distribution of country Muslim shares in our sample.[59] Interacting these dummies with average Ramadan fasting hours, we see a large and significant effect only for countries above the 75th percentile, suggesting that our baseline effects are mainly driven by countries with a high share of Muslims.

Our results are robust to a number of checks, which we report in the Appendix. One potential concern is that countries might choose test dates depending on average daily fasting hours in a given year. For instance, if fasting hours are long, Muslim majority countries might choose test dates that are sufficiently far away from the last day of Ramadan to avoid potential negative consequences from the physical activity of fasting. If this is the case, we might observe positive effects of fasting hours on test scores if, for instance, negative consequences from fasting on test scores are independent of fasting hours but dependent on the time between the last day of Ramadan and the test date. To test if this is the case, in Table C6 in the Appendix we report estimates when we control for the time between the the last day of Ramadan and the test date. While we do observe a positive relation between the time span between Ramadan and the test date and test scores - which indicates the existence of negative effects of fasting on test scores - our estimates remain sizable and significant. Secondly, we might be concerned that fasting hours might be correlated with other factors which, in turn, affect test scores. Indeed, in a related study Campante and Yanagizawa-Drott (2015) show that GDP per capita is

[59]The last column in Table C1 in the Appendix lists all countries by the quartiles used in this estimation.

affected by fasting hours in Muslim majority countries. If GDP per capita has a positive effect on test scores by, e.g., higher educational spending, we would obtain downward-biased estimates. Table C7 shows results of our baseline estimates when we additionally control for GDP per capita. Contrary to what we expected, our estimated effects become smaller and are less significant for the restricted sample of Muslim majority countries. A potential explanation for this result could be collinearity of GDP per capita and fasting hours, resulting in less precisely estimated coefficients. On the other hand, the effect becomes larger when we use the entire sample and interact fasting hours with the size of the Muslim population in a country. Finally, we address concerns regarding the size of our panel data. Our baseline results rely on a panel data set with 15 Muslim majority countries and 6 wave observations. To see if our results are driven by a particular country or wave, we piecewise delete one country or wave from our data set and estimate, again, our baseline specification. Table C8 and Table C9 show estimation results when piecewise deleting one wave or country, respectively. Overall, our estimates are hardly affected by this procedure, supporting the estimation strategy used in this paper.

4.5.2 Evidence from PISA data

In a next step, we test whether positive effects of Ramadan fasting can also be seen for first and second generation migrants in Western European countries. More specifically, we compare the performance of students with parents from Muslim majority countries with their counterparts whose parents are not from one of those countries. Table 14 reports results from various specifications of Equation (11) based on PISA data. All specifications are estimated separately for math, science, and reading test scores and include year and country-fixed effects.[60] The first row of Table 14 reports the coefficient of the interaction of (log) average fasting hours and a binary variable indicating whether a student is Muslim. In the context of a "difference-in-difference" specification, this coefficient measures the difference between the marginal effect of Ramadan fasting for Muslim and non-Muslim students.

[60] In our PISA sample, we also rescale test scores of each wave to facilitate interpretation. Due to the specific rescaling method applied in PISA, we use the mean and the standard deviation of the distribution of test scores of all OECD countries in each wave. Therefore, the interpretation of our estimates are in terms of an average standard deviation of the distribution of test scores of OECD countries in all PISA waves.

Table 14:
Estimation results, PISA, 2003 to 2018, baseline

	Science		Reading		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Muslim x Ramadan	0.41***	0.33***	0.36***	0.28**	0.34***	0.29***
	(0.12)	(0.12)	(0.12)	(0.12)	(0.10)	(0.10)
Ramadan	-0.18		-0.24		-0.68***	
	(0.23)		(0.18)		(0.16)	
Muslim	-1.36***	-1.15***	-1.19***	-1.00***	-1.13***	-0.99***
	(0.30)	(0.32)	(0.30)	(0.32)	(0.26)	(0.26)
Country & Year FE	Yes	No	Yes	No	Yes	No
Country x Year FE	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.29	0.29	0.29	0.30	0.27	0.28
Observations	351,762	351,762	351,762	351,762	351,762	351,762
Cluster	48	48	48	48	48	48

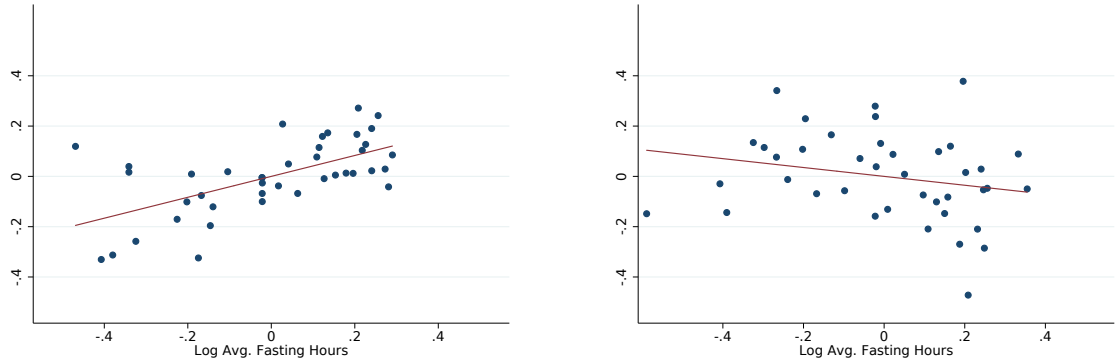
Note: Estimation of standardized achievement test scores on logarithmized average fasting hours. Muslim students are identified using the country of origin of their parents. A list of the included control variables can be found in Table C5 in the Appendix. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country times year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14 reveals that Ramadan fasting has a positive effect on student achievements for Muslims in comparison to non-Muslim students for all subjects. Our estimates in columns (1), (3), and (5) - where we control for large set of student and school characteristics and include country and year fixed effects - indicate that an increase of average fasting hours by 10 % increase the difference in test scores between Muslim and non-Muslim students by around 3.4 to 4.1 % of an average standard deviation.[61] In a more demanding specifications - columns (2), (4), and (6) - where we additionally include country by year fixed effects to capture any wave-specific shocks to students in each country - our estimates become smaller, but are still large and significant. Results from this specification indicate that an increase of average fasting hours by 10 % increases the difference in test scores between Muslim and non-Muslim students by 2.8 to 3.3 % of an average standard deviation.

Again, we provide a number of robustness tests. A potential concern might be that students from Muslim majority countries might be affected by increased government efforts to improve

[61] Our estimates are hardly affected by the inclusion of student and school characteristics.

Figure 9:
Performance gaps and Ramadan fasting hours, science scores, PISA



(a) Muslim immigrants vs natives **(b)** Non-Muslim immigrants vs natives

NOTE: Plot on the left (right) shows performance gaps in science scores between students whose parents are from Muslim majority countries (non-Muslim majority countries) and natives and logarithm of average fasting hours during Ramadan before test was taken. Both variables are adjusted by their country means. Equivalent plots for math and reading test scores can be found in the Appendix (Figure C3).

immigrants' educational outcomes. If these increased government efforts are correlated with fasting hours - which mainly increase in our sample for all countries - our identification strategy erroneously assigns positive effects of these government efforts to fasting hours. We address this concern by constructing a placebo group of immigrants from non-Muslim majority countries and check if we also obtain positive effects for this group. Figure 9 plots performance gaps between immigrants from Muslim majority countries and natives on the left and performance gaps between immigrants from non-Muslim majority countries on the right.[62] While we see a positive relationship between the performance gap and fasting hours for Muslim students, we do not see such a relationship for non-Muslim students. A similar conclusion can be drawn if we refer to the estimation results in Table C11 in the Appendix where we include an interaction term of fasting hours and a binary variable indicating non-Muslim immigrant students to our baseline specification. Our estimated results are hardly effected by the inclusion of this interaction term. Secondly, we, again, apply a piecewise deletion procedure to address concerns relating to the small sample size of our panel. Table C12 in the Appendix shows estimation results when we piecewise delete one wave from our sample and Table C13 in the

[62]Table C11 in the Appendix lists the selected countries to define the group of non-Muslim immigrant students based on parental country of origin.

Appendix shows the estimations results when excluding one country. Again, our estimated results are qualitatively not affected by this procedure, supporting our identification strategy.

4.6 Discussion

4.6.1 Ramadan fasting and religious participation

We now turn to our discussion about potential mechanisms of our estimated effect of Ramadan fasting on student performance. First, we investigate whether fasting hours during Ramadan early in life affect religious participation. If we find a positive association between fasting hours and religious participation, this might be a channel through which Ramadan fasting affects student performance. To evaluate the existence of this channel, we use data from the World Value Survey (WVS). The WVS has been conducted six times and covers a time period between 1981 and 2014.[63] The WVS includes several aspects of religious orientation and involvement as well as about participants' religious denomination. As dependent variable, we use information about self-reported attendance in religious services. We generate a binary variable ($Participation_{ict}$) indicating whether an individual attends religious services at least once per week. To assess whether religious socialization early in life affects religious participation, we estimate the following equation similar to our estimation strategy with the PISA data:

$$Participation_{ict} = \beta_1 Ramadan_{ct} \;\; Muslim_{ict} + \beta_2 Muslim_{ict} + \eta_{ct} + \epsilon_{ict}, \qquad (12)$$

where $Ramadan_{ct}$ is the logarithmized average daily number of fasting hours during the last Ramadan before the WVS interview, $Muslim_{ict}$ is a binary variable indicating whether i stated in the interview to be of a Muslim denomination, and η_{ct} are country times wave fixed effects. The coefficient of interest in Equation (13) is β_1, which measures the difference in religious participation between Muslims and non-Muslims induced by a change in fasting hours. Since

[63] The first (second, third, fourth, fifth, sixth) wave was conducted during the years 1981-1984 (1989-1993, 1994-1998, 1999-2004, 2005-2009, 2010-2014).

Table 15:
Estimation results, World Value Survey, 1981 to 2014

	(1)	(2)	(3)	(4)	(5)	(6)
Ramadan x Muslim	0.48*	0.46*	-0.01	-0.01	-0.18	-0.17
	(0.25)	(0.25)	(0.14)	(0.14)	(0.16)	(0.16)
Mean outcome if Muslim	0.43	0.43	0.42	0.42	0.47	0.47
Country x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Muslim FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Restricted to age:						
15-18	Yes	Yes	No	No	No	No
19-35	No	No	Yes	Yes	No	No
36-65	No	No	No	No	Yes	Yes
R squared	0.29	0.29	0.27	0.27	0.25	0.25
Observations	8,567	8,567	114,395	114,395	134,299	134,299
Cluster	185	185	197	197	197	197
Countries	68	68	68	68	68	68

Note: Estimation of religious participation on logarithmized average fasting hours. Dependent variable: Binary, attends at least once per week religious services. The baseline controls are education level, marital status, gender, social class, income level, and population size of town of residence. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

are interested in the effect of fasting on religious participation for younger individuals - and we expect the effect to be largest particularly for younger individuals - we split our sample in 3 different age groups: 15-18, 19-35, 36-65.

Our estimation results are shown in Table 15. As before, we report cluster-robust standard errors at country-year level in all specifications. Columns (1), (3), and (5) of Table 15 show our estimates without adding control variables and columns (2), (4), and (6) show estimation results if we include a set of control variables to Equation (13). As expected, we see sizable and significant effects of Ramadan fasting hours on religious participation for the youngest age group. Our estimation result in column (2) suggests that an increase of average Ramadan fasting hours by 10 % increases attendance in religious services by Muslim individuals in comparison to non-Muslims by around 4.8 p.p. Strikingly, this effect is only observable for the youngest age group. In light of the results by Campante and Yanagizawa-Drott (2015) who find a positive effect of the number of fasting hours on religious participation only for committed believers, this result suggest a higher share of committed Muslim believers among

younger individuals.

4.6.2 Common identity and Ramadan fasting

Next, we investigate whether the positive effect of Ramadan fasting is induced by creating a common identity among students. To address this question we hypothesize that the effect of Ramadan fasting for Muslim students is larger when they are surrounded by relatively more Muslim students. To empirically investigate whether positive effects of Ramadan fasting are induced by the creation of a common identity among students in school through practicing a common religious practice, we add information about the share of Muslim students in a school cohort to Equation (11):

$$\begin{aligned} Test_{ict} =& \beta_1 Muslim_{ict} + \beta_2 Share_{ict} + \beta_3 Muslim_{ict} Share_{ict} + \\ & \beta_4 Ramadan_{ct} Muslim_{ict} + \beta_5 Ramadan_{ct} Share_{ict} + \\ & \beta_6 Ram_{ct} Share_{ict} Mus_{ict} + \eta_{ct} + \epsilon_{ict}, \end{aligned} \tag{13}$$

where $Share_{ict}$ measures the percentage of Muslim students in school of student i in country c and year t. To ease the interpretation of our estimates, we define $Share_{ict}$ to be 1 if the share of Muslim students in the school of student i is above the median of the conditional distribution of Muslim students in schools of country c and year t given a positive share of Muslim students in school and zero otherwise.[64]

Based on our theoretical discussion in Section 4.2, we hypothesize that the effect of Ramadan fasting for Muslim students is larger when they are surrounded by relatively more Muslim students. Hence, we expect β_6 - which measures the difference between the difference of the marginal effect of fasting for Muslim and non-Muslim students in schools with a large share of Muslim students and schools with low share of Muslim students - to be positive. The sum of β_5 and β_6 measures the difference between the marginal effect of fasting for Muslim students in schools with a large share of Muslim students and those in schools with low share,

[64] We drop $Ramadan_{ict}$ from Equation (13) as we include country times year fixed effects, η_{ct}, to the equation to control for country-specific time trends.

Table 16:
Estimation results, PISA, 2003 to 2018, extension

	Science		Reading		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Ramadan x Share x Muslim		0.26*		0.33**		0.10
		(0.15)		(0.16)		(0.12)
Ramadan x Share	0.31***	0.27***	0.33***	0.30***	0.31***	0.29***
	(0.10)	(0.10)	(0.11)	(0.11)	(0.09)	(0.09)
Ramadan x Muslim		-0.07		-0.19		-0.01
		(0.12)		(0.15)		(0.12)
Muslim x Share		-0.73*		-0.96**		-0.29
		(0.40)		(0.43)		(0.33)
Share	-1.08***	-0.97***	-1.14***	-1.04***	-1.08***	-1.02***
	(0.26)	(0.25)	(0.28)	(0.28)	(0.24)	(0.23)
Muslim		0.05		0.40		-0.08
		(0.29)		(0.39)		(0.29)
Country x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.33	0.34	0.32	0.32	0.31	0.31
Observations	102,594	102,594	102,594	102,594	102,594	102,594
Cluster	41	41	41	41	41	41

Note: Estimation of standardized achievement test scores on logarithmized average fasting hours. The sample includes only schools with at least one Muslim student. Muslim students are identified using the country of origin of their parents. *Share* is a binary variable indicating a high share of Muslim students in the school cohort of a student. A list of the included control variables can be found in Table C5 in the Appendix. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

which we expect to be positive as well. β_5 estimates the same effect for non-Muslim students. Depending on how the common identity created by the religious practice for Muslim students affects non-Muslim students, β_5 might be positive, negative, or zero.

We estimate Equation (13) on a restricted sample of individuals who are in schools with a positive share of Muslims.[65] Table 16 reports our estimation results. All specifications are estimated, again, separately for math, science, and reading test scores and include our set of control variables and country times year fixed effects. In columns (1), (3), and (5), we first exclude the binary variable indicating a Muslim student and show a positive effect of average Ramadan fasting hours on test scores for students in schools with a higher Muslim share, which is what our hypothesis would suggest. An increase of average fasting hours by 10 %

[65] Table C14 in the Appendix shows results if we use the entire sample. Our results are qualitatively the same.

increases the difference in test scores between students in schools with a higher Muslim share and students in schools with a low Muslim share by 3.1 to 3.3 % of an average standard deviation.

We then add our Muslim dummy to our estimation equation to assess whether the performance of Muslim students in schools with a high Muslim share is affected more by Ramadan fasting hours than the performance of Muslim students in schools with a low Muslim share (columns (2), (4), and (6)). Indeed, our results suggests that this is the case as we estimate positive coefficients of the triple interaction term for all subjects, which also remain significant for science and reading test scores. An increase of average fasting hours by 10 % increases the difference of the Muslim-non-Muslim test score gap in schools with high share of Muslims and the Muslim-non-Muslim gap in schools with low Muslim share by 2.1 to 4.4 % of an average standard deviation. Further, our results suggest that the overall positive effect of Ramadan fasting hours are exclusively driven by Muslim students from schools with a high Muslim share as our estimate of the baseline coefficient of the interaction term *Ramadan x Muslim* is close to zero. This results indicates that Ramadan fasting hours affect student performance by creating a common identity among students. Interestingly, the coefficient on *Ramadan x Share* remains positive and significant when we at the triple interaction term. Potential explanations for this finding could be that we misclassify Muslim students as non Muslim students more often in case of schools with a high Muslim share, or that there are spill-over effects of the positive effect of Ramadan fasting for Muslim students to non Muslim students.

4.7 Conclusion

Using plausible exogenous variation induced by the rotating Islamic calendar and its interaction with country latitude, we provided causal evidence for a positive effect of the average number of Ramadan fasting hours on student performance based on data from two independent international student achievement tests. We also established that this is accompanied by increased religious participation among younger Muslim individuals, suggesting that religiosity might be the driving force explaining this result. Additionally, by showing that our effects

are mainly caused by Muslim students from schools with a high Muslim share, we provided evidence that religiosity impacts student performance by creating a common identity among students, and not by affecting their character skills. Overall, in light of the scarce empirical literature on the effect of religiosity on educational outcomes - in particular with respect to Islamic religiosity - our paper provides important insights to recent debates about impeded educational outcomes of Muslims due to norms and other restrictions imposed by Muslim authorities (Chaudhary and Rubin, 2011; Chaney, 2011, 2016).

5 Author's contribution

Paper 1:

Internet usage and migration decisions: Evidence from Nigerian micro data

The article is single-authored.

Paper 2:

Assessing the Role of Asylum Policies in Refugees' Labor Market Integration: The Case of Protection Statuses in the German Asylum System

The article is single-authored.

Paper 3:

Religious practice and student performance: Evidence from Ramadan fasting

The article is joint work with Erik Hornung and Guido Schwerdt. We share authorship. The main idea of the paper was developed by Erik Hornung and Guido Schwerdt. I prepared the data and conducted the empirical analysis. Section two to seven were written by me, but I benefited from comments of my co-authors. I contributed to the introduction, but large parts were written by Erik Hornung.

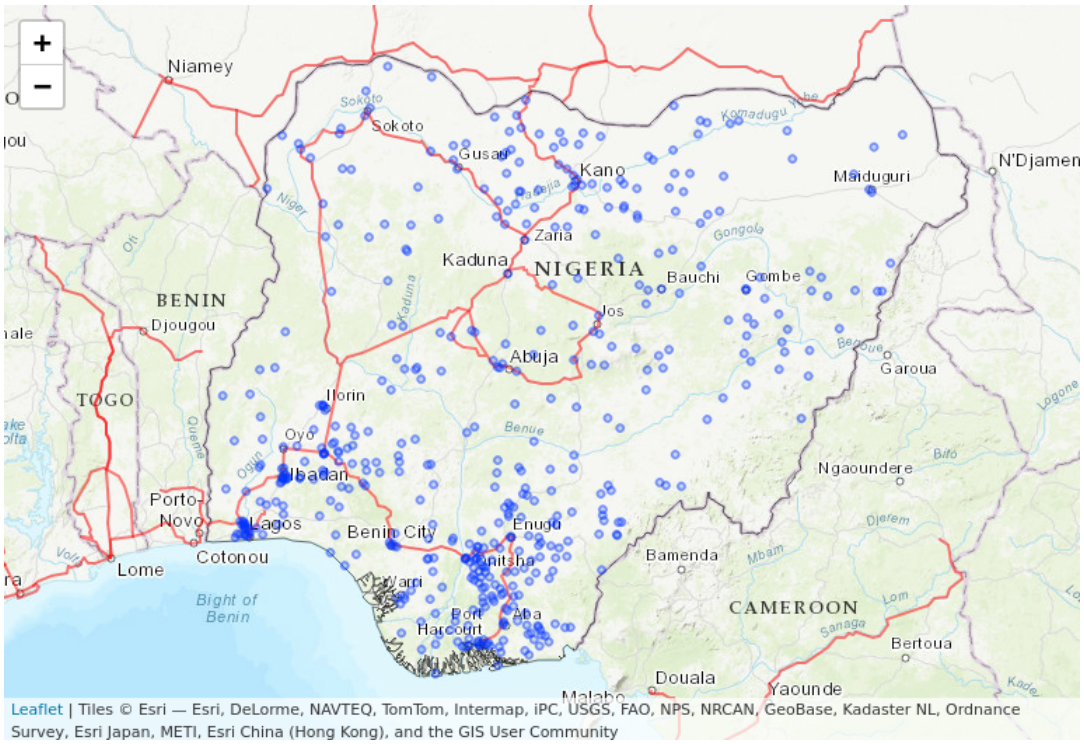
A Appendix of Essay 1

Table A1:
OLS: International migration (binary) on Internet usage and Internet usage frequency

	(1)	(2)	(3)	(4)	(5)
Panel A					
Internet usage	0.0103***	0.0082**	0.0082**	0.0085**	0.0080**
	(0.0039)	(0.0039)	(0.0039)	(0.0038)	(0.0037)
Panel B					
Internet usage frequency	0.0051***	0.0041**	0.0042**	0.0044**	0.0043**
	(0.0019)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Year 12 FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
State FE	No	No	Yes	No	No
County FE	No	No	No	Yes	No
Community FE	No	No	No	No	Yes
Observations	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435

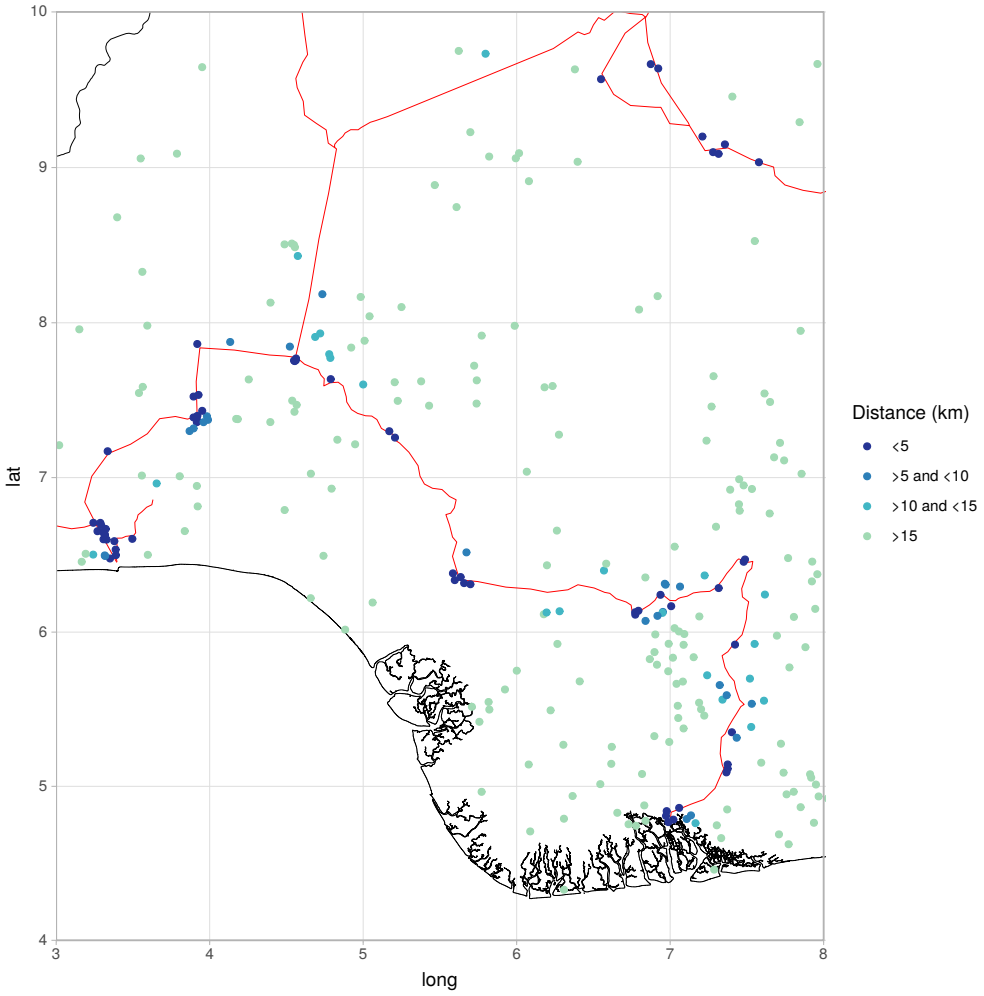
Note: Regression of a binary variable indicating whether an individual moved to another country on Internet usage and Internet usage frequency in the previous wave. Internet usage frequency is a ordinal measure (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1:
Nigeria, terrestrial cable network, and included communities



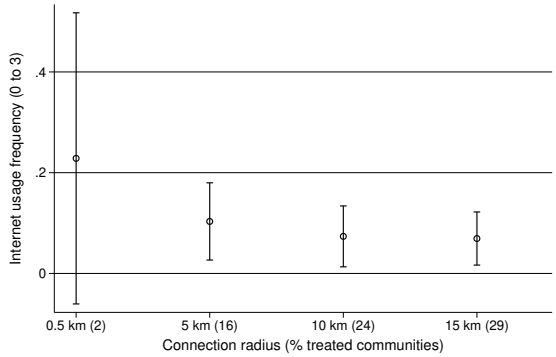
Notes: Red lines illustrate the diffusion of the terrestrial cable network in Nigeria and neighbouring countries. Blue dots indicate communities that are included in the final data set. Sources: Mapcruzin.com, Hjort and Poulsen (2019), Nigerian GHS panel.

Figure A2:
South Nigeria, terrestrial cable network, and included communities

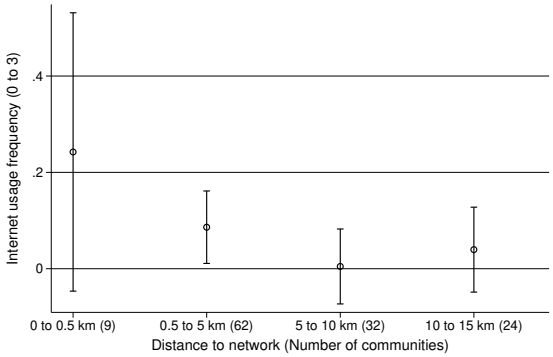


Notes: Red lines illustrate the diffusion of the terrestrial cable network in South Nigeria. Colored dots indicate communities included in the data set. Sources: Mapcruzin.com, Hjort and Poulsen (2019), Nigerian GHS panel.

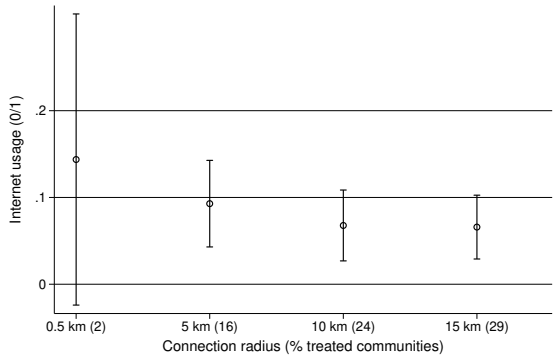
Figure A3:
Internet usage and distance to terrestrial cable network, 2010-12 change



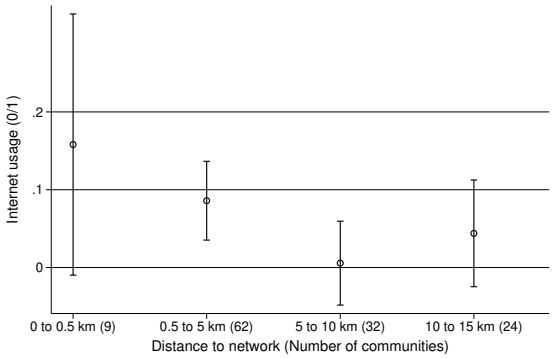
(a) Full sample, Internet usage frequency



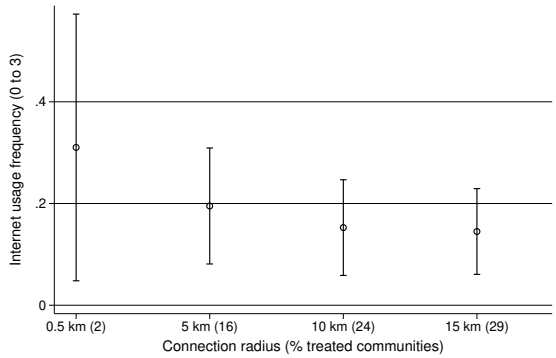
(b) Full sample, Internet usage frequency



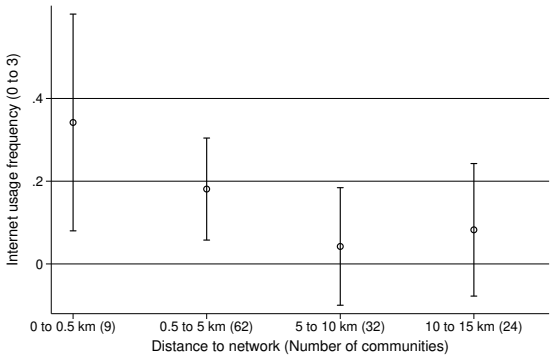
(c) Young individuals, Internet usage



(d) Young individuals, Internet usage



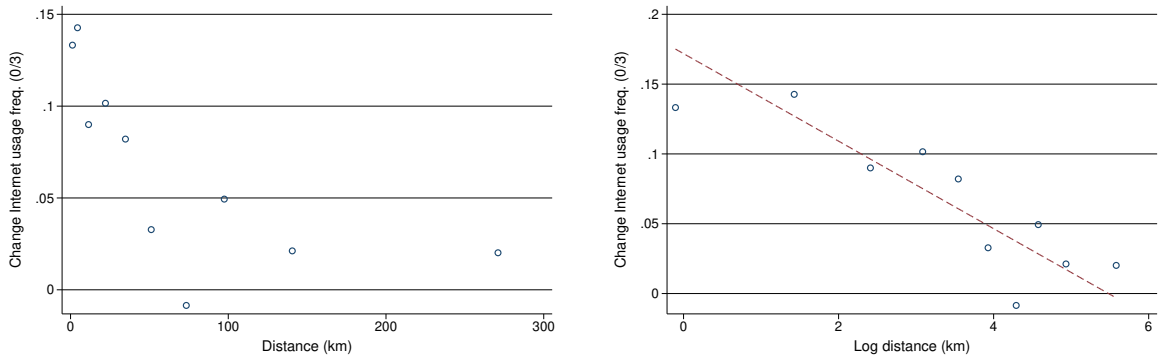
(e) Young individuals, Internet usage frequency



(f) Young individuals, Internet usage frequency

Note: Plots on the left show coefficient estimates for four separate regressions of Internet usage or Internet usage frequency on an interaction term of a binary variable indicating if individual i is located in a community within the connection radius shown on the x-axis to the terrestrial cable network times an indicator variable for the year 2012. Plots on the right show coefficient estimates for a regression of Internet usage or Internet usage frequency on a set of binary variables indicating if individual i is located within a bin shown on the x-axis (baseline: Distance to terrestrial cable network larger than 15 km). All estimates include a year dummy for the year 2012 as well as community fixed effects. Young individuals are between 20 and 35 at interview date. Number of observations: 21,626 (full sample), 8,963 (young individuals). 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster).

Figure A4:
Internet usage frequency and distance to terrestrial cable network, 2010-12 change



Note: Binned scatter plot (10 equally sized bins) of difference in community mean Internet usage frequency between 2012 and 2010 and distance to terrestrial cable network in kilometres (left) and logartihmized distance to terrestrial cable network (right). 435 communities included.

Table A2:
Reduced form estimation: Robustness, network distance not logarithmized

	(1)	(2)	(3)	(4)
Distance to network * Year 12	-0.0026***	-0.0027***	-0.0037***	-0.0037***
	(0.0008)	(0.0008)	(0.0012)	(0.0012)
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3:
Reduced form estimation: Robustness, binary measure

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
𝟙(Distance to network $<$ 5km) * Year 12	0.0052*	0.0053*	0.0053*	0.0062**
	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Observations	21,626	20,259	18,406	16,721
Cluster	435	403	361	327
Panel B: Age 20 to 35				
𝟙(Distance to network $<$ 5km) * Year 12	0.0104*	0.0106*	0.0108*	0.0117**
	(0.0055)	(0.0055)	(0.0055)	(0.0055)
Observations	8,963	8,408	7,682	6,986
Cluster	435	403	361	327
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance between 5 and 10 km	No	Yes	Yes	Yes
Distance between 10 and 20 km	No	No	Yes	Yes
Distance between 20 to 30 km	No	No	No	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4:
Reduced form estimation: Robustness, binary measure, excluding remote locations

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
𝟙(Distance to network $<$ 5km) * Year 12	0.0052*	0.0041	0.0046	0.0049
	(0.0029)	(0.0030)	(0.0034)	(0.0048)
Observations	21,626	15,944	6,492	4,639
Cluster	435	333	145	103
Panel B: Age 20 to 35				
𝟙(Distance to network $<$ 5km) * Year 12	0.0104*	0.0093*	0.0085	0.0085
	(0.0055)	(0.0056)	(0.0063)	(0.0080)
Observations	8,963	6,582	2,696	1,970
Cluster	435	333	145	103
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance $>$ 100 km	No	Yes	Yes	Yes
Distance $>$ 20 km	No	No	Yes	Yes
Distance $>$ 10 km	No	No	No	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5:
Community mean values by distance to the terrestrial cable network

	Distance > 5 km	Distance < 5 km	Mean diff.
Community			
located in states: Lagos, Abuja	0.02	0.17	-0.15***
			(0.03)
located in urban area	0.22	0.73	-0.52***
			(0.05)
with high share of educated individuals in 2010	0.23	0.38	-0.15**
			(0.06)
with high share of Internet user in 2010	0.21	0.49	-0.28***
			(0.06)

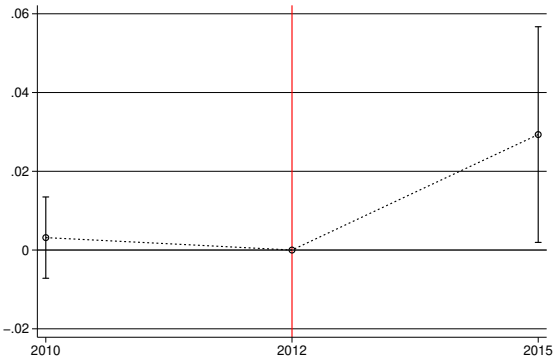
Note: Mean values and mean difference tests of selected community characteristics by distance to the terrestrial cable network. A community has a high share of Internet users (educated individuals) if the share of Internet users (college educated individuals) is in the highest quartile in the sample. Sample size: 435 (distance < 5 km: 71). Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6:
Reduced form estimation: Robustness, additional controls

	(1)	(2)	(3)	(4)	(5)
Panel A: Entire sample					
Log(Distance to network) * Year 12	-0.0015***	-0.0011**	-0.0011***	-0.0013***	-0.0011***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0004)
Lagos or Abuja * Year 12	-0.0034				-0.0044
	(0.0036)				(0.0038)
Urban * Year 12		0.0026			0.0016
		(0.0020)			(0.0021)
Internet usage year 10 * Year 12			0.0037*		0.0033
			(0.0021)		(0.0026)
Education year 10 * Year 12				0.0014	-0.0003
				(0.0023)	(0.0029)
Observations	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435
Panel B: Age 20 to 35					
Log(Distance to network) * Year 12	-0.0027***	-0.0021**	-0.0021***	-0.0021**	-0.0022***
	(0.0010)	(0.0008)	(0.0007)	(0.0008)	(0.0008)
Lagos or Abuja * Year 12	-0.0103***				-0.0124***
	(0.0039)				(0.0048)
Urban * Year 12		0.0030			0.0024
		(0.0038)			(0.0039)
Internet usage year 10 * Year 12			0.0040		-0.0001
			(0.0038)		(0.0045)
Education year 10 * Year 12				0.0064	0.0064
				(0.0041)	(0.0047)
Observations	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435
Included covariates (Panel A and B):					
Year 12 FE	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). *Lagos or Abuja* is a binary variable indicating whether an individual is located either in Logos or Abuja, *Urban* is a binary variable indicating whether an individual resides in an urban area. *Internet usage year 10* is a binary variable indicating whether an individual lives in a community where the share of Internet users in 2010 was in the highest quartile in the sample. *Education year 10* is a binary variable indicating whether an individual lives in a community where the share of college educated individuals in 2010 was in the highest quartile in the sample. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A5:
Robustness: Parallel pre-trends, remittances



Note: Plot shows estimated coefficients $\beta_{1,10}$ and $\beta_{1,15}$ of the equation:

$$Remittances_{h,c(h),t} = \mu_c + \sum_{j \in \{10,15\}} \{\beta_{0,j} \mathbb{1}[t = j] + \beta_{1,j} \mathbb{1}[t = j] * \mathbb{1}\left[Distance_{h,c(h)} < 5km\right]\} + \epsilon_{h,c(h),t},$$

where $Remittances_{h,c(h),t}$ is a binary variable indicating whether household h located in community $c(h)$ has received remittances within 12 month before the interview year t, and μ_c represents a set of community fixed effects. 2010 (2012, 2015) refers to the first (second, third) wave of the Nigerian GHS panel. Estimates are based on a sample at household level. Number of observations: 10,414. Plotted 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster). Point estimates and standard errors can be found in column (1) of Table 6 in Section 2.7.

Table A7:
Instrumental variable estimation: Robustness, binary instrument

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
First-stage estimates				
Internet usage				
𝟙(Distance to network < 5km) * Year 12	0.050**	0.050**	0.051**	0.053***
	(0.020)	(0.020)	(0.020)	(0.020)
F statistic	6.24	6.22	6.52	7.01
Internet usage frequency				
𝟙(Distance to network < 5km) * Year 12	0.105***	0.106***	0.108***	0.113***
	(0.039)	(0.039)	(0.039)	(0.039)
F statistic	7.23	7.29	7.63	8.37
Second-stage estimates				
Internet usage	0.104	0.106	0.101	0.101
	(0.072)	(0.073)	(0.070)	(0.068)
Internet usage frequency	0.049	0.050	0.047	0.047
	(0.033)	(0.033)	(0.032)	(0.031)
Observations	21,626	20,259	19,202	18,406
Cluster	435	403	379	361
Panel B: Restricted: Age 20 to 35				
First-stage estimates				
Internet usage				
𝟙(Distance to network < 5km) * Year 12	0.092***	0.092***	0.094***	0.096***
	(0.026)	(0.026)	(0.026)	(0.026)
F statistic	12.40	12.27	12.99	13.38
Internet usage frequency				
𝟙(Distance to network < 5km) * Year 12	0.194***	0.196***	0.200***	0.206***
	(0.058)	(0.058)	(0.058)	(0.058)
F statistic	11.15	11.36	11.94	12.72
Second-stage estimates				
Internet usage	0.112*	0.115*	0.111*	0.112*
	(0.062)	(0.063)	(0.061)	(0.060)
Internet usage frequency	0.053*	0.054*	0.052*	0.052*
	(0.031)	(0.031)	(0.030)	(0.029)
Observations	8,963	8,408	8,014	7,682
Cluster	435	403	379	361
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance between 5 and 10 km	No	Yes	Yes	Yes
Distance between 10 and 15 km	No	No	Yes	Yes
Distance between 15 and 20 km	No	No	No	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Binary variable indicating if distance to the terrestrial cable network is below 5 km times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are the same as for the baseline estimates (Table 3). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8:
Instrumental variable estimation: Robustness, binary instrument, excluding remote

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
First-stage estimates				
Internet usage				
𝟙(Distance to network < 5km) * Year 12	0.050**	0.042**	0.035	0.050*
	(0.020)	(0.020)	(0.023)	(0.027)
F statistic	6.24	4.46	2.43	3.63
Internet usage frequency				
𝟙(Distance to network < 5km) * Year 12	0.105***	0.091**	0.068	0.098*
	(0.039)	(0.039)	(0.046)	(0.055)
F statistic	7.23	5.25	2.16	3.23
Second-stage estimates				
Internet usage	0.104	0.097	0.132	0.097
	(0.072)	(0.084)	(0.127)	(0.105)
Internet usage frequency	0.049	0.045	0.068	0.050
	(0.033)	(0.039)	(0.068)	(0.055)
Observations	21,626	15,944	6,492	4,639
Cluster	435	333	145	103
Panel B: Restricted: Age 20 to 35				
First-stage estimates				
Internet usage				
𝟙(Distance to network < 5km) * Year 12	0.092***	0.082***	0.075**	0.100***
	(0.026)	(0.027)	(0.031)	(0.036)
F statistic	12.40	9.47	5.77	7.48
Internet usage frequency				
𝟙(Distance to network < 5km) * Year 12	0.194***	0.175***	0.136*	0.179**
	(0.058)	(0.059)	(0.072)	(0.088)
F statistic	11.15	8.69	3.55	4.13
Second-stage estimates				
Internet usage	0.112*	0.113	0.112	0.085
	(0.062)	(0.071)	(0.088)	(0.084)
Internet usage frequency	0.053*	0.053	0.062	0.047
	(0.031)	(0.035)	(0.054)	(0.050)
Observations	8,963	6,582	2,696	1,970
Cluster	435	333	145	103
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance > 100 km	No	Yes	Yes	Yes
Distance > 15km	No	No	Yes	Yes
Distance > 10km	No	No	No	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Binary variable indicating if distance to the terrestrial cable network is below 5 km times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are the same as for the baseline estimates (Table 3). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9:
Instrumental variable estimation: Robustness, additional controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Entire sample						
First-stage estimates						
Internet usage						
Log(Distance to network) * Year 12	-0.014**	-0.010**	-0.012**	-0.015**	-0.014**	-0.009**
	(0.006)	(0.004)	(0.006)	(0.006)	(0.005)	(0.004)
F statistic	6.28	5.74	4.08	6.25	6.11	4.22
Internet usage frequency						
Log(Distance to network) * Year 12	-0.029***	-0.019***	-0.022**	-0.028***	-0.027***	-0.014**
	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)	(0.007)
F statistic	10.26	7.73	5.39	8.59	9.47	3.94
Second-stage estimates						
Internet usage	0.098*	0.150*	0.094	0.074*	0.096*	0.124
	(0.056)	(0.084)	(0.065)	(0.043)	(0.055)	(0.079)
Internet usage frequency	0.048**	0.078*	0.051	0.040*	0.049**	0.079
	(0.024)	(0.040)	(0.033)	(0.021)	(0.025)	(0.051)
Observations	21,626	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435	435
Panel B: Restricted: Age 20 to 35						
First-stage estimates						
Internet usage						
Log(Distance to network) * Year 12	-0.021***	-0.016***	-0.016***	-0.021***	-0.019***	-0.013***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
F statistic	14.06	10.40	8.47	13.32	13.17	7.23
Internet usage frequency						
Log(Distance to network) * Year 12	-0.044***	-0.031***	-0.031***	-0.043***	-0.040***	-0.022**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
F statistic	18.89	10.74	10.07	17.76	17.37	5.92
Second-stage estimates						
Internet usage	0.116**	0.174**	0.127*	0.099**	0.106**	0.171**
	(0.054)	(0.079)	(0.067)	(0.044)	(0.051)	(0.084)
Internet usage frequency	0.055**	0.087**	0.067**	0.050**	0.052**	0.102*
	(0.024)	(0.040)	(0.034)	(0.021)	(0.024)	(0.052)
Observations	8,963	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435	435
Included covariates (Panel A and B):						
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	No
Lagos or Abuja * Year 12	No	Yes	No	No	No	Yes
Urban * Year 12	No	No	Yes	No	No	Yes
Internet usage year 10 * Year 12	No	No	No	Yes	No	Yes
Education year 10 * Year 12	No	No	No	No	Yes	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Control variables included are the same as for the baseline estimates (Table 3). *Lagos or Abuja* is a binary variable indicating whether an individual is located either in Logos or Abuja. *Urban* is a binary variable indicating whether an individual resides in an urban area. *Internet usage year 10* is a binary variable indicating whether an individual lives in a community where the share of Internet users in 2010 was in the highest quartile in the sample. *Education year 10* is a binary variable indicating whether an individual lives in a community where the share of college educated individuals in 2010 was in the highest quartile in the sample. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10:
Reduced form estimation: Robustness, exclusion restriction

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Entire sample						
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0079		0.0052*	0.0060		
	(0.0182)		(0.0029)	(0.0042)		
Log(Distance to network) * Year 12		-0.0083**			-0.0014***	-0.0017***
		(0.0037)			(0.0005)	(0.0006)
Observations	21,626	21,626	21,626	20,391	21,626	20,391
Cluster	435	435	435	406	435	406
Panel B: Age 20 to 35						
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0268		0.0105*	0.0117		
	(0.0279)		(0.0055)	(0.0078)		
Log(Distance to network) * Year 12		-0.0100			-0.0024***	-0.0027**
		(0.0061)			(0.0009)	(0.0012)
Observations	8,963	8,963	8,963	8,442	8,963	8,442
Cluster	435	435	435	406	435	406
Included covariates (Panel A and B):						
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional control: Employment status	No	No	Yes	No	Yes	No
Excluded communities (Panel A and B):						
Distance $<$ 5 km & Empl. growth $>$ 0	No	No	No	Yes	No	Yes
Dependent variable (Panel A and B):						
Employment status	Yes	Yes	No	No	No	No
Migration	No	No	Yes	Yes	Yes	Yes

Note: Dependent variable is either a binary variable indicating whether an individual was employed within the last 7 days or a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11:
Reduced form estimation: Robustness, exclusion restriction other

	(1)	(2)	(3)	(4)
Dependent variable:				
Age	0.0452	-0.0100	-0.0127	0.2536
	(0.0527)	(0.0455)	(0.2324)	(0.1728)
Female	-0.0006	-0.0015	0.0019	0.0062
	(0.0018)	(0.0029)	(0.0089)	(0.0152)
Household head	-0.0005	-0.0040	-0.0032	0.0039
	(0.0013)	(0.0025)	(0.0064)	(0.0117)
Spouse	-0.0001	-0.0017	0.0009	0.0037
	(0.0016)	(0.0027)	(0.0081)	(0.0129)
Son/Daughter	0.0016	0.0027	0.0034	0.0144
	(0.0021)	(0.0038)	(0.0105)	(0.0180)
Other household member	-0.0010	0.0030	-0.0011	-0.0220**
	(0.0012)	(0.0021)	(0.0053)	(0.0087)
Relation to HH head (ordinal)	0.0001	0.0126**	0.0045	-0.0335
	(0.0035)	(0.0063)	(0.0163)	(0.0279)
No schooling	0.0052	0.0061	-0.0365	-0.0339
	(0.0043)	(0.0044)	(0.0246)	(0.0217)
Some schooling	-0.0105**	-0.0118**	0.0391	0.0435
	(0.0048)	(0.0054)	(0.0264)	(0.0282)
Secondary education	0.0057	0.0093	-0.0088	-0.0182
	(0.0035)	(0.0059)	(0.0164)	(0.0256)
University degree	-0.0003	-0.0037	0.0062	0.0086
	(0.0017)	(0.0033)	(0.0075)	(0.0147)
Education (ordinal)	-0.0002	-0.0042	0.0401	0.0329
	(0.0069)	(0.0087)	(0.0326)	(0.0365)
Wealth items = 0	-0.0029	-0.0029	0.0290**	0.0162
	(0.0044)	(0.0050)	(0.0143)	(0.0162)
Wealth items = 1	0.0070	0.0101	-0.0389	-0.0352
	(0.0071)	(0.0078)	(0.0291)	(0.0322)
Wealth items = 2	-0.0080	-0.0074	0.0354	0.0240
	(0.0066)	(0.0076)	(0.0313)	(0.0357)
Wealth items = 3	0.0039	-0.0006	-0.0346	-0.0052
	(0.0058)	(0.0072)	(0.0291)	(0.0367)
Wealth items = 4	0.0000	0.0008	0.0090	0.0003
	(0.0030)	(0.0034)	(0.0182)	(0.0220)
Wealth items (ordinal)	0.0027	-0.0033	-0.0356	-0.0019
	(0.0109)	(0.0130)	(0.0509)	(0.0584)
Enrolled	-0.0019	-0.0068	-0.0017	0.0025
	(0.0034)	(0.0051)	(0.0162)	(0.0239)
TV usage	0.0013	0.0081	0.0053	-0.0057
	(0.0051)	(0.0059)	(0.0204)	(0.0222)
Mobile phone usage	0.0119**	0.0103*	-0.0081	0.0002
	(0.0060)	(0.0061)	(0.0268)	(0.0271)
Reported coefficient:				
Log(Distance to network) * Year 12	Yes	Yes	No	No
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	No	No	Yes	Yes
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	No	Yes	No	Yes
Observations	21,626	8,963	21,626	8,963
Cluster	435	435	435	435

Note: First column specifies the dependent variable of a regression on measures of distance to the terrestrial cable network times an indicator variable for the year 2012. Control variables included are listed in Table 1 (dependent variable excluded). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12:
Instrumental variable estimation: Migration out of Africa, full sample

	(1)	(2)	(3)	(4)	(5)	(6)
Internet usage	0.098*	0.049	0.016			
	(0.056)	(0.032)	(0.013)			
Internet usage frequency				0.048**	0.024*	0.008
				(0.024)	(0.014)	(0.006)
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable:						
All migration	Yes	No	No	Yes	No	No
Migration out of Africa	No	Yes	No	No	Yes	No
Migration within in Africa	No	No	Yes	No	No	Yes
Observations	21,626	21,612	21,612	21,626	21,612	21,612
Cluster	435	435	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable is a binary variable indicating if an individual migrated to (1) another country, (2) out of Africa, and (3) within Africa (migration out of Africa is coded 0 in this case). Please note that the number of observations declines due to missing information about the destination country. Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13:
Instrumental variable estimation: Relative wealth, full sample

	(1)	(2)	(3)
First-stage estimates			
Internet usage			
Log(Distance to network) * Year 12	-0.014**	-0.008**	-0.012**
	(0.006)	(0.004)	(0.005)
F statistic	6.28	4.28	5.63
Internet usage frequency			
Log(Distance to network) * Year 12	-0.029***	-0.018**	-0.029***
	(0.009)	(0.008)	(0.011)
F statistic	10.26	4.82	7.28
Second-stage estimates			
Internet usage	0.098*	0.273	0.032
	(0.056)	(0.175)	(0.032)
Internet usage frequency	0.048**	0.120	0.014
	(0.024)	(0.073)	(0.013)
Year 12 FE	Yes	Yes	Yes
Community FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Restricted: Low wealth	No	Yes	No
Restricted: High wealth	No	No	Yes
Observations	21,626	9,733	11,893
Cluster	435	421	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

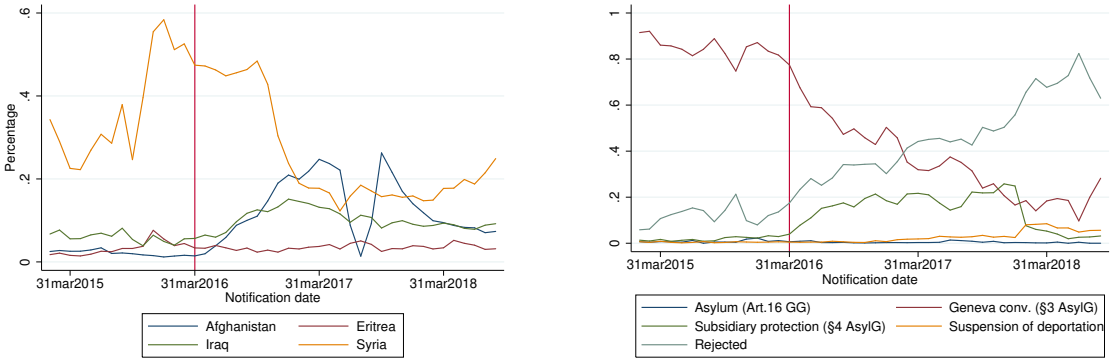
Table A14:
Instrumental variable estimation: Relative wealth, 2 endogenous variables

	(1)	(2)	(3)	(4)	(5)	(6)
First-stage estimate						
Log(Dist. to network) * Year 12	-0.021***	0.001	-0.045***	0.003		
	(0.007)	(0.002)	(0.014)	(0.004)		
Log(Dist. to network) * Year 12 * Low wealth	0.003	-0.019***	0.005	-0.044***		
	(0.010)	(0.006)	(0.021)	(0.014)		
Second-stage estimates						
Internet usage * Low wealth					0.149	
					(0.104)	
Internet usage					0.068	
					(0.050)	
Internet usage frequency * Low wealth						0.064
						(0.045)
Internet usage frequency						0.032
						(0.022)
F-statistic (First stage)					10.80	11.04
Endogenous variable:						
Internet usage	Yes	Yes	No	No	Yes	No
Internet usage frequency	No	No	Yes	Yes	No	Yes
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,963	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage times a binary variable indicating low wealth and Internet usage frequency and Internet usage frequency times an binary variable indicating low wealth on migration decisions. Excluded instruments: Log distance to the terrestrial cable network times an indicator variable for the year 2012 and an interaction with a binary variable indicating low wealth. Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Appendix of Essay 2

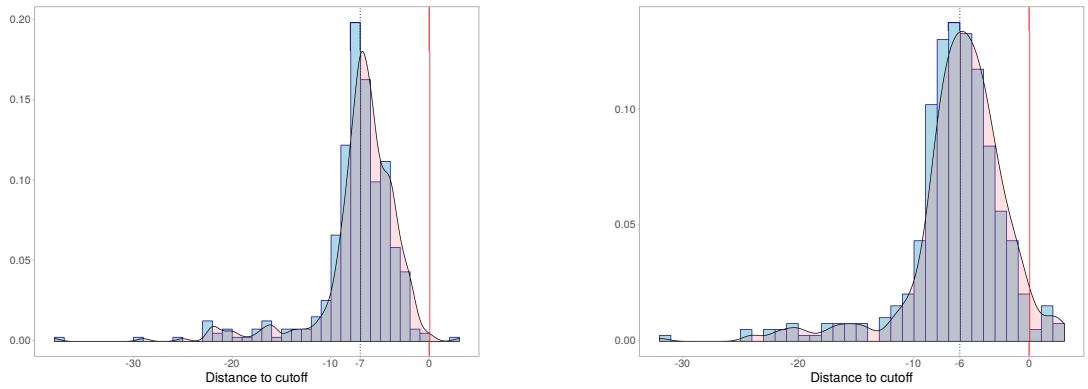
Figure B1:
Protection status and notification date



(a) Decisions by origin country (4 largest groups) **(b)** Received protection status (Iraqi)

Note: Left plot shows the share of decisions made by the BAMF for asylum seekers of the four largest groups of asylum seekers by month of notification date. Right plot illustrates the type of protection status received by month of notification date for Iraqi asylum applicants. Source: Own calculations based on monthly published data from BAMF (data available upon request).

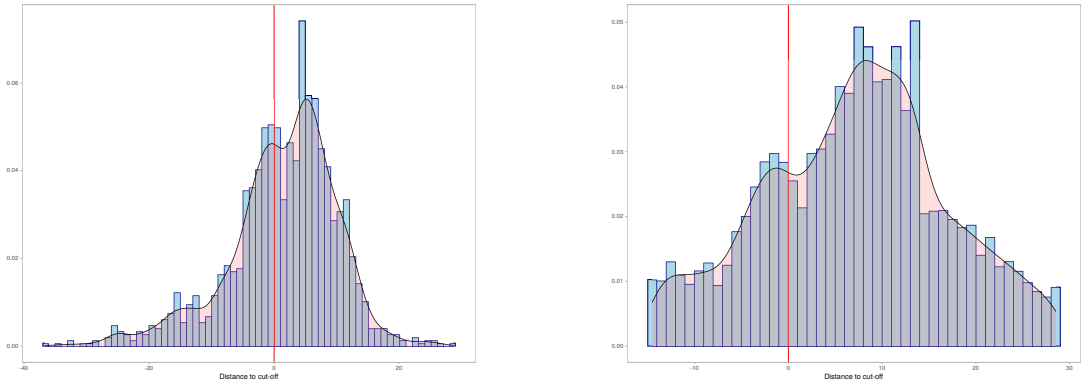
Figure B2:
Validity of RD design: arrival and application dates relative to policy change



(a) Arrival month **(b)** Application month

Note: Normalized histogram and Gaussian kernel density estimate of the month of arrival (left) and application for asylum (right) - both relative to the time of the policy change (between March and April 2016) - for refugee migrants who received notification within a 3 month corridor before and after the policy change. Number of observations: 396. The dashed vertical lines indicate the (rounded) mean value of each plotted variable and the red vertical lines indicate the change in BAMF's decision making policy.

Figure B3:
Validity of RD design: density of assignment variable



(a) SOEP Sample **(b)** Official asylum statistic

Note: Normalized histogram and Gaussian kernel density estimate of assignment variable month of notification about decision of asylum application (relative to cutoff). The red vertical lines indicate the change in BAMF's decision making policy. The graph on the left uses data from the SOEP. The graph on the right uses data from the official record of the BAMF.

Table B1:
Validity of RD design: mean differences, covariates and outcome

	BW: 18 month			BW: 3 month		
	$t < c$	$t > c$	t-val	$t < c$	$t > c$	t-val
Female	34	41	-2.6	35	36	-0.1
Age between						
18 and 35	54	61	-2.8	59	58	0.2
36 and 55	43	36	2.4	39	39	-0.1
55 and 65	4	3	1.2	2	2	-0.1
Married	64	67	-1.1	66	68	-0.6
No children in household	34	30	1.5	29	28	0.2
Age of youngest child in household between						
0 and 4	38	43	-1.8	43	42	0.2
5 and 10	18	18	-0.0	19	19	-0.0
11 and 15	10	8	0.8	9	11	-0.6
College graduate	23	20	1.2	22	23	-0.2
No work experience prior migration	33	39	-2.1	34	31	0.7
Work experience prior migration						
Self-employed or blue-collar worker	35	34	0.3	34	41	-1.3
White-collar worker	32	27	1.9	32	28	0.7
Located in East Germany	17	13	1.9	22	18	1.2
Years since migrating						
0 to 1	0	1	-1.6	0	0	1.0
2 to 3	76	97	-12.6	96	95	0.4
4 to 5	24	3	13.5	4	5	-0.7
Labor market outcomes						
Any employment	35	22	5.1	33	24	1.9
Full-time employment	17	9	4.6	21	10	3.0
Subsidiary protection	15	42	-10.6	18	36	-4.1
Observations	525	874		206	190	

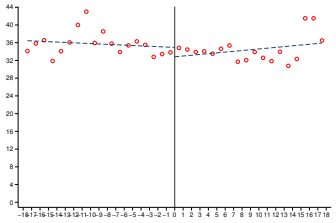
Note: Mean values of covariates (in percent) and t-values of mean-comparison test by value of the instrument for varying time spans around the cut-off.

Table B2:
Validity of RD design: RD estimates, covariates

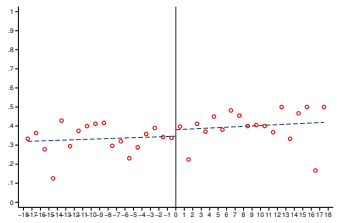
	$E[X]$	RD estimates				
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:						
Age (in years)	34.37	-0.15	2.19*	0.39	2.12**	1.66
Female	0.38	0.05	0.02	0.04	0.04	-0.03
Married	0.66	0.06	0.04	0.04	0.06	0.01
No children in household (below 16)	0.32	-0.08*	-0.02	-0.05	-0.04	0.06
Youngest child in household: 0-4	0.41	0.06	-0.00	0.06	0.01	-0.04
Youngest child in household: 5-10	0.18	0.02	-0.03	-0.02	-0.00	-0.03
Youngest child in household: 11-15	0.09	-0.01	0.04	0.02	0.03	0.02
College graduate	0.21	-0.00	0.00	-0.00	-0.02	0.03
No work experience prior migration	0.37	0.04	-0.01	0.03	-0.01	-0.00
Self-employed or blue-collar worker	0.35	0.05	0.06	0.05	0.08	0.04
White-collar worker	0.29	-0.09**	-0.06	-0.08*	-0.07	-0.04
Located in East Germany	0.14	-0.03	-0.04	-0.05	-0.06	-0.07
Months since migrating	39.49	0.92**	0.88*	0.83*	1.69***	0.79
Bandwidth selection	none	none	none	18	12	6
Polynomial order		1	2	1	1	1
Observations	1470	1470	1470	1399	1238	782

Note: Mean value of covariates and corresponding RD estimates. Significant estimates are indicated with stars based on Huber-White standard errors. See RD plots of covariates and predicted outcome variables in the Appendix.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

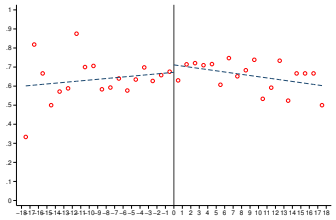
Figure B4:
Validity of RD design: RD plots, covariates and predicted outcomes



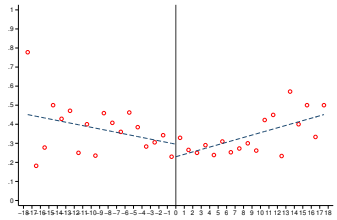
(a) Age



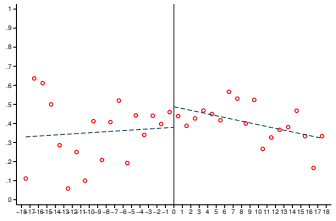
(b) Female



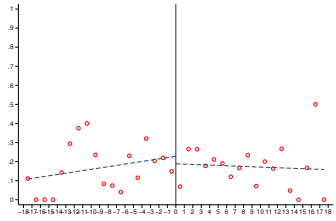
(c) Married



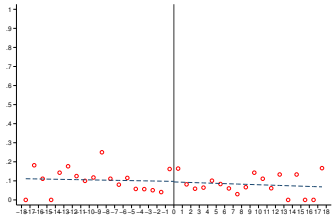
(d) No children in household



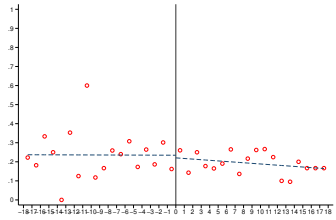
(e) Youngest child in household: 0 to 4



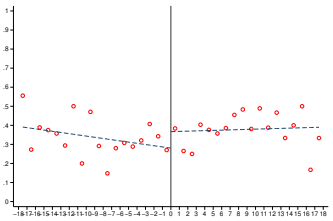
(f) Youngest child in household: 5 to 10



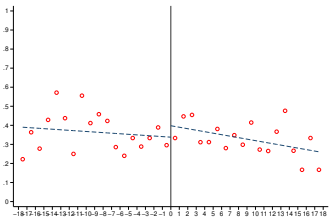
(g) Youngest child in household: 11 to 15



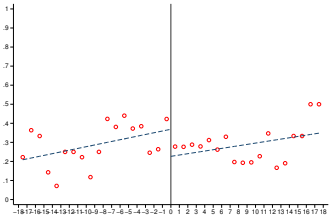
(h) College graduate



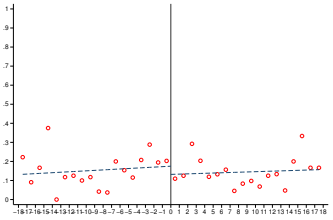
(i) No work experience prior migration



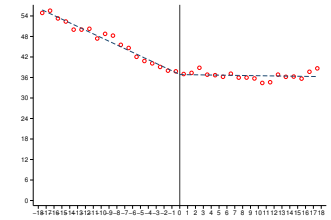
(j) Self-employed or blue-collar worker



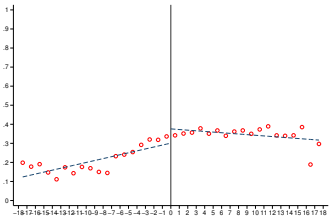
(k) White-collar worker



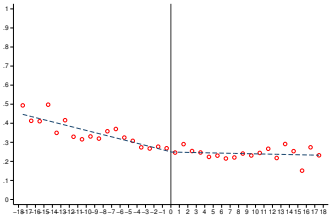
(l) Located in East Germany



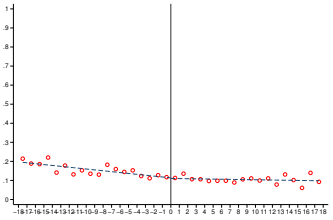
(m) Month since migrating



(n) Predicted subsidiary protection



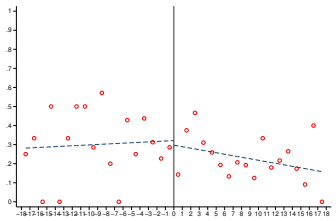
(o) Predicted employment (any)



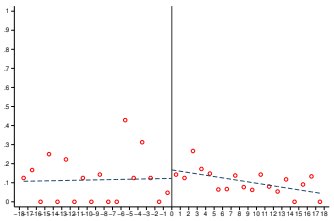
(p) Predicted employment (full-time)

Note: Mean of selected variables by value of the assignment variable with fitted lines on both sides of the cut-off.

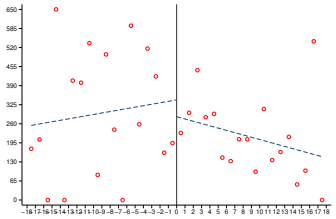
Figure B5:
Robustness: RD plots, placebo sample



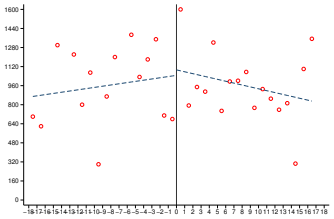
(a) Any employment



(b) Full-time employment



(c) Monthly earnings



(d) Monthly earnings (excl 0)

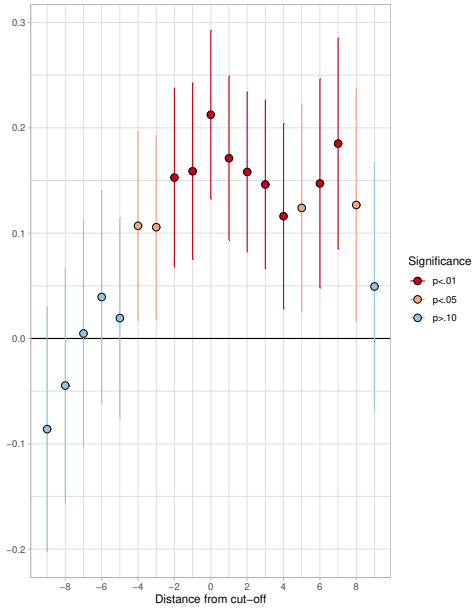
Note: Sample includes refugees who do not have an international protection status.

Table B3:
Robustness: Placebo RD estimates, reduced form

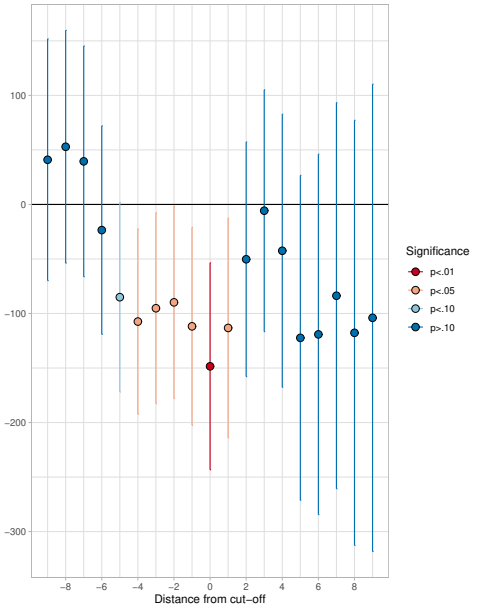
	(1)	(2)	(3)	(4)	(5)
Any employment	-0.05	0.04	-0.03	0.06	0.15
	(0.06)	(0.09)	(0.07)	(0.09)	(0.15)
Full-time employment	0.04	0.09	0.06	0.07	0.27**
	(0.05)	(0.07)	(0.05)	(0.06)	(0.11)
Net earnings (excl 0)	24.29	-24.33	-2.02	37.62	377.88
	(157.88)	(227.40)	(205.06)	(231.09)	(300.99)
Net earnings	-46.85	21.76	-35.96	60.10	252.93
	(77.84)	(109.63)	(91.61)	(106.60)	(159.45)
Bandwidth selection	none	none	18	12	6
Polynomial order	1	2	1	1	1
Observations	722	722	634	471	215

Note: Reduced form RD estimates for placebo sample. Huber-White standard errors are reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

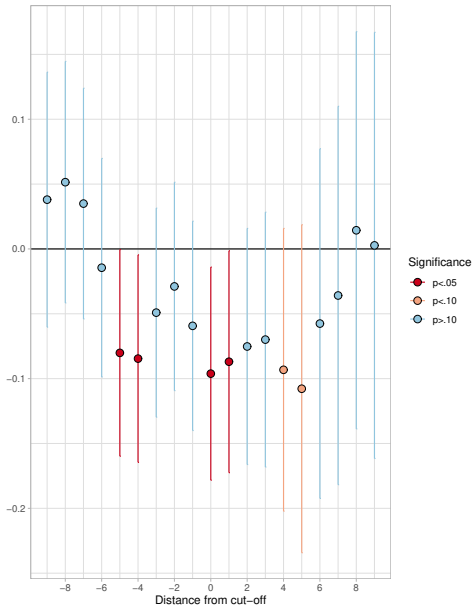
Figure B6:
Robustness: RD estimates, varying cut-off



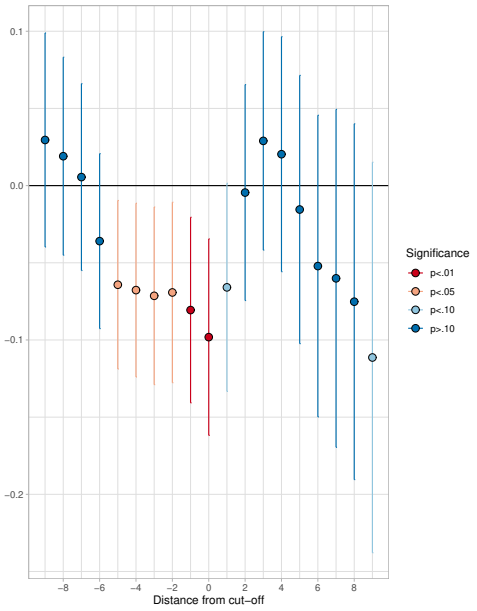
(a) Subsidiary protection



(b) Monthly earnings



(c) Any employment



(d) Full-time employment

Note: Plot of RD estimates and 95 % confidence interval for various cut-off based on baseline specification with first order polynomial and a selected bandwidth of 18 month.

Table B4:
Robustness: Donut RD estimates, 2SLS

	Donut IV estimate	IV estimate
Any employment	-0.33*	-0.37**
	(0.17)	(0.17)
Full-time employment	-0.36***	-0.40***
	(0.13)	(0.13)
Net earnings (excl 0)	-790.38**	-770.57**
	(360.38)	(341.75)
Net earnings	-549.66***	-603.92***
	(200.08)	(196.81)
Observations	1323	1470

Note: 2SLS estimates of the effect of subsidiary protection status on various labor market outcomes. Donut RD estimate is based on a sample that excludes observations one month before and after the cut-off (March and April 2016). Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B5:
Robustness: RD estimates, reduced form, covariates included

	(1)	(2)	(3)	(4)	(5)
First stage estimation					
Subsidiary protection	0.19***	0.16***	0.19***	0.19***	0.16**
	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
F-statistic	24	11	21	17	6
Reduced form estimation					
Any employment	-0.07**	-0.11**	-0.09**	-0.07	-0.04
	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
Full-time employment	-0.09***	-0.11***	-0.10***	-0.11***	-0.15***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)
Monthly earnings (excl 0)	-236.62**	-196.18	-247.44**	-231.92**	-209.28
	(101.23)	(130.18)	(101.93)	(117.77)	(156.41)
Monthly earnings	-136.25***	-174.74***	-152.21***	-145.32***	-167.70**
	(43.20)	(54.82)	(46.28)	(53.59)	(71.13)
Bandwidth selection	none	none	18	12	6
Polynomial order	1	2	1	1	1
Observations	1470	1470	1399	1238	782

Note: 2SLS estimates of the effect of subsidiary protection status on various labor market outcomes. Huber-White standard errors are reported in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B6:
Robustness: Reduced form estimates, fixed effect specification

	Baseline sample			Placebo sample	
	(1)	(2)	(3)	(4)	(5)
First-stage					
Subsidiary protection	0.23***	0.23***	0.24***		
	(0.03)	(0.04)	(0.04)		
F statistic	62.40	39.58	42.31		
Reduced-form estimates					
Any employment	-0.09***	-0.08**	-0.07**	0.06	0.00
	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)
Full-time employment	-0.06***	-0.07***	-0.07**	0.01	-0.05
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Monthly earnings	-116.75***	-108.31**	-102.66**	39.71	-31.63
	(35.66)	(43.71)	(40.51)	(77.67)	(77.35)
Application FE	Yes	No	No	No	No
Arrival x application FE	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	No	Yes
Observations	1470	1470	1470	722	722

Note: Regression of subsidiary protection status (column 1) or labor market outcome on a binary variable indicating if an refugee migrant received notification of the asylum application after March 2016. Placebo sample consists of refugees who did not receive either Geneva protection status or subsidiary protection status. Cluster robust standard errors at the level of the arrival month time application month are reported in parentheses. Number of cluster: 371 (316, placebo sample).
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B7:
Robustness: OLS and IV estimates, fixed effect specification

	OLS estimate	IV estimate	Fuzzy RD estimate
Any employment	-0.03	-0.30**	-0.37**
	(0.03)	(0.13)	(0.17)
Full-time employment	-0.05**	-0.28***	-0.40***
	(0.02)	(0.10)	(0.13)
Monthly earnings	-81.30**	-427.39***	-603.92***
	(31.91)	(150.24)	(196.81)
Month of arrival FE	No	No	
Month of application FE	No	No	
Arrival x application FE	Yes	Yes	
Control variables	Yes	Yes	
Observations	1470	1470	1470

Note: OLS and IV estimates of the effect of subsidiary protection status on various labor market outcomes. Excluded instrument in the IV estimation: binary variable indicating if refugee was notified about the decision of the asylum application after March 2016. The third column reports the fuzzy RD design estimates obtained in Table 8. Cluster robust standard errors at the level of the arrival month time application month are reported in parentheses. Number of cluster: 371.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B8:
Robustness: IV estimates, fixed effect specification

	(1)	(2)	(3)	(4)
Any employment	-0.30**	-0.42	-0.38	-0.43
	(0.13)	(0.28)	(0.27)	(0.28)
Full-time employment	-0.28***	-0.52**	-0.51**	-0.52**
	(0.10)	(0.23)	(0.23)	(0.23)
Monthly earnings	-427.39***	-559.63	-510.49	-570.83*
	(150.24)	(341.78)	(332.91)	(340.60)
F statistic	42.31	11.26	11.22	11.53
Month of arrival FE	No	No	No	No
Month of application FE	No	No	No	No
Arrival x application FE	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
Application to decision (month)	No	Yes	No	Yes
Notfication to interview (month)	No	No	Yes	Yes
Observations	1470	1470	1470	1470

Note: IV estimates of the effect of subsidiary protection status on various labor market outcomes. Excluded instrument: binary variable indicating if refugee was notified about the decision of the asylum application after March 2016. Cluster robust standard errors at the level of the arrival month time application month are reported in parentheses. Number of cluster: 371.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

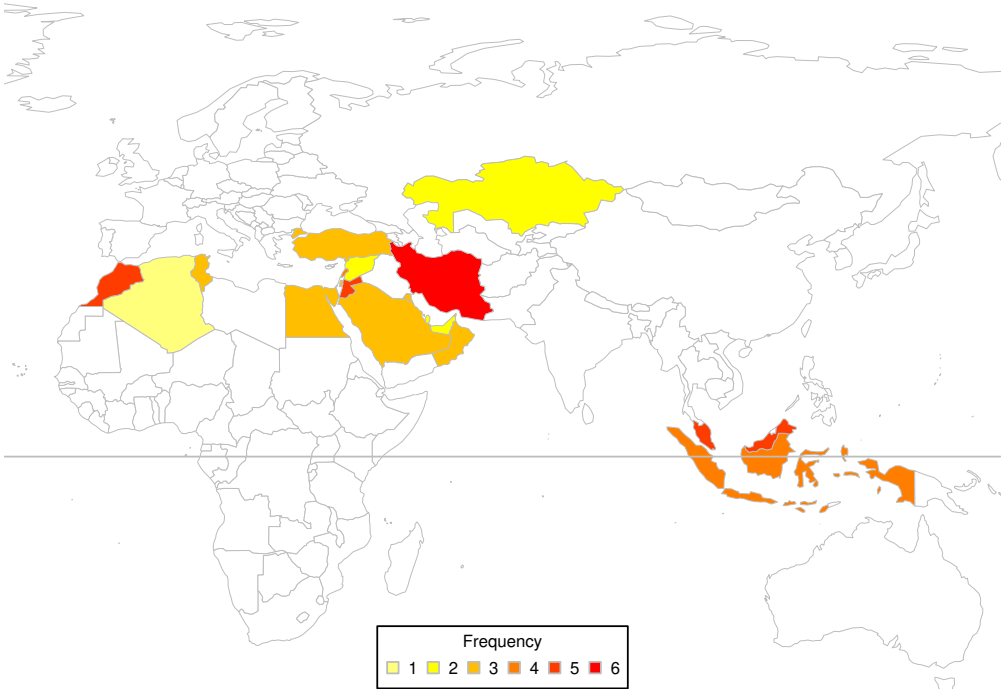
C Appendix of Essay 3

Table C1:
Countries included in study, TIMSS

Country	# Included	Muslim Share	Muslim Share Quartile
Taiwan	4	0.000	1
Czech Republic	2	0.000	1
Hong Kong	6	0.000	1
Slovakia	2	0.000	1
Chile	4	0.001	1
Japan	6	0.001	1
South Korea	6	0.001	1
Lithuania	6	0.001	1
Armenia	3	0.001	1
Hungary	6	0.002	1
Malta	2	0.002	1
Romania	4	0.003	1
Botswana	4	0.004	1
Columbia	2	0.005	2
Finland	2	0.005	2
Ireland	2	0.005	2
Moldavia	2	0.005	2
USA	5	0.008	2
New Zealand	5	0.009	2
Italy	5	0.010	2
Latvia	3	0.010	2
Norway	5	0.010	2
South Africa	4	0.015	2
Canada	3	0.020	2
Sweden	4	0.020	2
Slovenia	4	0.024	2
Philippines	2	0.051	2
Netherlands	3	0.057	3
Thailand	4	0.058	3
Georgia	3	0.099	3
Russia	5	0.117	3
Bulgaria	3	0.122	3
Singapore	5	0.149	3
Ghana	3	0.159	3
Israel	6	0.167	3
Cyprus	4	0.227	3
Macedonia	3	0.333	3
Kazakhstan	2	0.564	3
Lebanon	4	0.593	3
Malaysia	5	0.604	3
United Arab Emirates	2	0.762	3
Qatar	2	0.775	4
Bahrain	4	0.812	4
Oman	3	0.877	4
Indonesia	4	0.882	4
Syria	2	0.922	4
Egypt	3	0.946	4
Kuwait	3	0.950	4
Saudi Arabia	3	0.970	4
Palestine	3	0.980	4
Turkey	3	0.980	4
Jordan	5	0.982	4
Morocco	5	0.990	4
Iran	6	0.994	4
Tunisia	3	0.995	4

Note: All countries that participated only once in TIMSS are excluded from our analysis.

Figure C1:
Muslim majority countries included in TIMSS: 1995 to 2015



Note: The horizontal line displays the equator. We exclude Algeria from our analysis since it only participated once in TIMSS.

Table C2:
Descriptive statistics, TIMSS, 1995 to 2015, all selected variables

	Non-Muslim countries		Muslim countries	
	Mean	Std. Dev.	Mean	Std. Dev.
Test scores				
Math	494.06	107.44	415.10	92.23
Science	493.56	105.45	435.67	93.02
Fasting hours				
Year before test	11.56	2.39	12.22	1.45
Average over last 3 years	11.47	2.21	12.05	1.36
Student characteristcs				
Age	14.35	0.84	14.25	0.88
Female	0.50		0.49	
Immigration status				
Second-generation migrant	0.04		0.04	
First-generation migrant	0.07		0.13	
Parental education				
University	0.25		0.22	
Post-secondary	0.23		0.14	
Upper-secondary	0.21		0.22	
Lower-secondary	0.09		0.14	
No school	0.05		0.14	
Unknown	0.18		0.14	
Books at home				
0 to 10	0.15		0.24	
11 to 25	0.21		0.34	
26 to 100	0.27		0.24	
101 to 200	0.16		0.08	
More than 200	0.18		0.07	
Unknown	0.03		0.02	
School characteristics				
Shortage: Instructional mat.				
Not at all	0.40		0.30	
A little	0.24		0.17	
Some	0.19		0.18	
A lot	0.12		0.32	
Unknown	0.06		0.03	
Shortage: Supplies				
Not at all	0.44		0.34	
A little	0.21		0.18	
Some	0.17		0.21	
A lot	0.12		0.24	
Unknown	0.07		0.03	
Shortage: School building				
Not at all	0.35		0.23	
A little	0.23		0.20	
Some	0.21		0.26	
A lot	0.15		0.29	
Unknown	0.06		0.04	
Shortage: Heating system				
Not at all	0.43		0.27	
A little	0.23		0.22	
Some	0.17		0.22	
A lot	0.11		0.24	
Unknown	0.06		0.04	
Number of observations				
Students	805,168		364,786	
Countries	38		18	
Country-year	148		62	

Note: Mean values of all selected variables. A country is categorized as a Muslim country if more than 50 % of its population is Muslim. Applied population weights are standardized to sum up to 1 for each country-year cluster.

Table C3:
Parental country of origin by country and year, PISA

2018	
Austria	**Afghanistan**, Austria, Former Yugoslavia (Bosnia-Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Serbia, Slovenia), Germany, Missing, **Syria**, **Turkey**, another country inside of Europe (AUT), another country outside of Europe (AUT)
Belgium	A Sub-Saharan country (Africa excl. Maghreb), An Eastern European country, Another country (BEL), Belgium, France, Germany, Missing, Netherlands, **North African country (Maghreb)**, Other Western European country (BEL), **Turkey**
Switzerland	**Albania**, Another country (CHE), Austria, Former Yugoslavia (Bosnia-Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Serbia, Slovenia), France, Germany, Italy, Liechtenstein, Missing, Portugal, Spain, Switzerland, **Turkey**
Germany	Another country (DEU), Bosnia and Herzegovina, Croatia, FYR Macedonia, Germany, Greece, Italy, Missing, Montenegro, One of the former USSR republics, Poland, Serbia, Slovenia, **Turkey**
Denmark	**Afghanistan**, Denmark, Finland, Greenland, Iceland, **Iraq**, **Lebanon**, Missing, Norway, One of the former Yugoslav republics, Other European country (DNK), Other non-European country (DNK), **Pakistan**, **Somalia**, Sweden, **Syria**, The Faroe Islands, **Turkey**
Finland	**Afghanistan**, Another country (FIN), China, Estonia, Finland, Former Yugoslavia (Bosnia-Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Serbia, Slovenia), **Iraq**, Missing, Russia, **Somalia**, Sweden, **Turkey**, Vietnam
Great Britain	Another country (QSC), Another country (QUK), Ireland, Missing, Other European Union country (QSC), United Kingdom (England), United Kingdom (Scotland), United Kingdom (excl.Scotland)
Netherlands	Another country (NLD), Missing, Netherlands
2015	
Austria	Bosnia and Herzegovina, Croatia, Germany, Hungary, Poland, Romania, Russian Federation, **Turkey**, FYR Macedonia, Serbia, Montenegro and Kosovo, Another country (AUT)
Belgium	Germany, **Turkey**, France, Netherlands, A Sub-Saharan country (Africa excl. Maghreb), **North African country (Maghreb)**, Other Western European country (BEL), An Eastern European country, Another country (BEL)
Switzerland	Germany, **Turkey**, France, **Albania**, Italy, Liechtenstein, Portugal, Spain, Switzerland, One of the former Yugoslav republics, Another country (CHE)
Germany	Germany, **Turkey**, Italy, Bosnia and Herzegovina, Croatia, Poland, FYR Macedonia, Montenegro, Serbia, Slovenia, One of the former USSR republics, Another country (DEU),
Denmark	**Turkey**, One of the former Yugoslav republics, **Afghanistan**, Denmark, The Faroe Islands, Greenland, Finland, Iceland, **Iraq**, **Lebanon**, Norway, **Pakistan**, **Somalia**, Sweden, Other European country (DNK), Other non-European country (DNK)
Finland	**Turkey**, One of the former Yugoslav republics, Finland, **Iraq**, **Somalia**, Sweden, Russian Federation, China, Estonia, Thailand, Another country (FIN),
Great Britain	**Pakistan**, Germany, Poland, India, Ireland, United Kingdom, United Kingdom (excl.Scotland), United Kingdom (Scotland), Other European Union country (QSC), Another country (QUK), Another country (QSC)
Netherlands	Germany, Poland, **Turkey**, One of the former Yugoslav republics, **Iraq**, China, One of the former USSR republics, Netherlands, Romania, Belgium, Bulgaria, **Iran, Islamic Republic of**, **Morocco**, Netherlands Antilles, Suriname, United Kingdom (Great Britain), United States of America, Another country (NLD),
2012	
Austria	Austria, **Turkey**, Germany, Serbia, Montenegro and Kosovo, Another country (AUT), Missing, Croatia, Poland, Bosnia and Herzegovina, Romania, FYR Macedonia, Russian Federation, Invalid
Belgium	Belgium, Another country (BEL), **North African country (Maghreb)**, An Eastern European country, Other Western European country (BEL), **Turkey**, A Sub-Saharan country (Africa excl. Maghreb), Netherlands, Germany, France, Missing, Invalid
Switzerland	Switzerland, Another country (CHE), Italy, **Turkey**, One of the former Yugoslav republics, Austria, Germany, Spain, **Albania**, Portugal, France, Missing, Invalid, Liechtenstein
Germany	Germany, **Turkey**, One of the former USSR republics, Serbia, Another country (DEU), Poland, Italy, Missing, FYR Macedonia, Croatia, Invalid, Greece, Bosnia and Herzegovina, Slovenia, Montenegro
Denmark	Denmark, Missing, **Turkey**, **Afghanistan**, Another country (CZE), One of the former Yugoslav republics, **Lebanon**, **Somalia**, **Pakistan**, **Iraq**, Invalid
Finland	Finland, **Somalia**, Invalid, Another country (FIN), Russian Federation, Estonia, One of the former Yugoslav republics, Missing, **Iraq**, Sweden, **Turkey**, China, Thailand
Great Britain	United Kingdom, Another country (QUK), Missing, United Kingdom (excl.Scotland), United Kingdom (Scotland), **Pakistan**, Another country (QSC), Africa, India, Invalid, Other European Union country (QSC), China (incl. Hong Kong), Caribbean, **Middle Eastern country**, **Bangladesh**
Netherlands	Netherlands, Other European country (NLD), China, Another country (NLD), United States of America, **Morocco**, Missing, Belgium, Suriname, **Turkey**, Invalid, Germany, Netherlands Antilles, United Kingdom (Great Britain), Romania, **Iran, Islamic Republic of**, **Iraq**, One of the former Yugoslav republics, Poland, One of the former USSR republics, Bulgaria

2009

Austria	Austria, **Turkey**, Another country (AUT), Germany, Missing, Bosnia and Herzegovina, Serbia and Montenegro, FYR Macedonia, Russian Federation, Croatia, Romania, Invalid, Poland
Belgium	Belgium, France, Netherlands, An Eastern European country, Missing, Other Western European country (BEL), Another country (BEL), A Sub-Saharan country (Africa excl. Maghreb), Germany, **North African country (Maghreb)**, **Turkey**, Invalid
Switzerland	Germany, One of the former Yugoslav republics, Switzerland, Another country (CHE), **Turkey**, Italy, Portugal, Austria, Invalid, Missing, Spain, **Albania**, France, Liechtenstein
Germany	One of the former USSR republics, Germany, Italy, **Turkey**, Another country (DEU), Missing, Poland, Bosnia and Herzegovina, Greece, Serbia, Slovenia, Croatia, FYR Macedonia, Montenegro, Invalid
Denmark	Denmark, **Somalia**, Another country (DNK), **Turkey**, One of the former Yugoslav republics, **Iraq**, **Lebanon**, **Afghanistan**, Missing, **Pakistan**, Invalid
Finland	Finland, Another country (FIN), Missing, Sweden, Another country (QVE), Russian Federation, Estonia, Invalid
Great Britain	United Kingdom, Another country (QUK), Missing, Invalid, Another country (QSC), **Pakistan**, Other European Union country (QSC), China (incl. Hong Kong), Africa, **Middle Eastern country**, India, **Bangladesh**, Caribbean,
Netherlands	**Morocco**, Netherlands, **Turkey**, Another country (NLD), One of the former Yugoslav republics, Other European country (NLD), Netherlands Antilles, United Kingdom, Missing, Suriname, Italy, **Iraq**, Spain, Belgium, China, Germany, **Iran, Islamic Republic of**, One of the former USSR republics

2006

Austria	Austria, Hungary, Bosnia and Herzegovina, **Turkey**, Poland, Other countries (AUT), China, Missing, Romania, Serbia-Montenegro, Germany, **Albania**, Slovenia, Invalid, Slovakia, Croatia, Former Yugoslav Republic of Macedonia, Czech Republic, N/A
Belgium	**North African country (Maghreb)**, Belgium, Germany, Other countries (BEL), Missing, Netherlands, **Turkey**, Other Western European country (BEL), A Sub-Saharan country (Africa excl. Maghreb), An Eastern European country, France, Invalid, N/A
Switzerland	Switzerland, Other countries (CHE), Italy, Germany, A former Yugoslav republic, **Albania**, Missing, **Turkey**, Spain, Portugal, Austria, Invalid, France, N/A, Liechtenstein,
Germany	Germany, Missing, A former USSR republic, **Turkey**, Other countries (DEU), Poland, N/A, Montenegro, Invalid, Bosnia and Herzegovina, Italy, Croatia, Greece, Serbia, Former Yugoslav Republic of Macedonia, Slovenia
Denmark	Denmark, Other countries (DNK), **Turkey**, Missing, **Pakistan**, A former Yugoslav republic, N/A
Finland	Finland, Other countries (FIN), Invalid, Sweden, Missing, Russian Federation, Estonia, N/A
Great Britain	United Kingdom, Other countries (GBR-QUK), Missing, N/A, United Kingdom (Scotland), China (incl. HongKong), United Kingdom (excl.Scotland), **Pakistan**, India, **Bangladesh**, Other European country (QSC), Africa, **Middle Eastern country**, Other countries (GBR-QSC), Invalid, Caribbean
Netherlands	Netherlands, Other European country (NLD), Missing, Other countries (NLD), N/A, Missing

2003

Austria	Austria, **Turkey**, AUT: Other, Czech Republic, Slovakia, **Albania**, Hungary, AUT: Yugoslavia, Poland, Romania, Slowenia
Belgium	Belgium, BFL: An African country (not Maghreb), BFR: Another country of the EU, BFR: Other, **BFR: A Maghreb country**, BFR: An African country (not Maghreb), **Turkey**, **BFL: A Maghreb country**, BFL: An other country of the EU, BFL: Other, France, Netherlands, An East-European country
Switzerland	CHE: Other, CHE: Switzerland, CHE: Ex-Yugoslavia, **CHE: Albania or Kosovo**, **Turkey**, Italy, CHE: Germany or Austria, CHE: France or Belgium, Portugal, Spain
Germany	DEU: Other, Germany, DEU: Russia, Kazakhstan or another former Soviet republic, Italy, Poland, **Turkey**, Croatia, Bosnia-Herzegovina, DEU: Montenegro, Zambia, DEU: Serbia, Greece, Macedonia, Slowenia
Denmark	Denmark, DNK: Other, **Pakistan**, **Turkey**, DNK: The former Yuguslavia
Finland	Sweden, Finland, FIN: Other, Russia, Estonia
Great Britain	United Kingdom, GRB: Other, SCO: Scotland, SCO: England, Wales, N Ireland, India, **Pakistan**, SCO: Other European country, **SCO: Middle East**, SCO: Other, **Bangladesh**, SCO: China (incl Hong Kong), Africa, Caribbean
Netherlands	NLD: Other, NLD: Other European country, Netherlands

Note: Muslim majority countries in bold. Order of the countries is random.

Table C4:
Share of Muslim students per country and year, PISA study: 2003 to 2015

	Austria	Belgium	Switzerland	Germany	Denmark	Finland	Great Britain	Netherlands
2003	0.04	0.02	0.05	0.05	0.02	0.00	0.00	0.00
2006	0.04	0.05	0.03	0.04	0.02	0.00	0.00	0.00
2009	0.04	0.05	0.02	0.04	0.04	0.00	0.00	0.06
2012	0.04	0.05	0.02	0.03	0.04	0.01	0.00	0.05
2015	0.04	0.05	0.02	0.03	0.05	0.01	0.01	0.05
2018	0.05	0.05	0.02	0.03	0.04	0.01	0.00	0.00

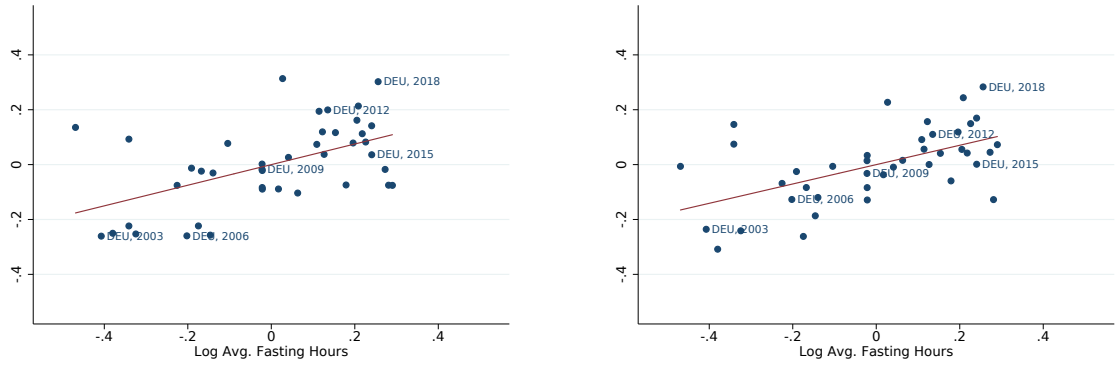
Note: The share of Muslim students is equal to 0 for the following country-year pairs: Finland 2003, 2006, 2009, Great Britain 2018, Netherlands 2003, 2006, 2018. Population weights applied.

Table C5:
Descriptive statistics, PISA, 2003 to 2018, all selected variables

	Non-Muslim students		Muslim students	
	Mean	Std. Dev.	Mean	Std. Dev.
Testscores				
Science test score	504.43	92.93	418.78	92.60
Reading test score	514.25	94.31	414.59	86.88
Math test score	515.80	88.97	432.69	81.88
Fasting hours				
Ramadan fasting	13.26	3.08	13.62	2.75
Ramadan fasting (3 year average)	12.77	3.20	13.15	2.86
Student characteristcs				
Age	15.78	0.29	15.79	0.29
Female	0.49		0.51	
Immigration status				
Native	0.85		0.00	
Second-generation migrant	0.06		0.75	
First-generation migrant	0.05		0.23	
Unknown	0.03		0.02	
Parental occupation (index)	52.18	18.34	39.46	16.91
Parental education				
None	0.01		0.06	
ISCED 1	0.01		0.09	
ISCED 2	0.05		0.19	
ISCED 3B, C	0.12		0.11	
ISCED 3A, 4	0.17		0.14	
ISCED 5B	0.22		0.13	
ISCED 5A, 6	0.30		0.17	
Unknwon	0.13		0.12	
Books at home				
0 to 10	0.12		0.32	
11 to 25	0.15		0.27	
26 to 100	0.29		0.26	
101 to 200	0.18		0.08	
201 to 500	0.15		0.03	
More than 500	0.08		0.02	
Unknwon	0.03		0.02	
School characteristics				
Avg. of Muslim in cohort	0.02	0.06	0.20	0.18
Location				
Village	0.08		0.02	
Small town	0.29		0.15	
Town	0.35		0.33	
City	0.17		0.27	
Large city	0.05		0.15	
Unknwon	0.06		0.08	
Type				
Private	0.02		0.00	
Private (gov. dep)	0.16		0.13	
Public	0.71		0.72	
Unknwon	0.11		0.14	
Country and year effects				
Number of observations	340,856		10,906	
Countries	8.00		8.00	
Country-year combination	48.00		41.00	

Note: Mean values of all selected variables. A student is defined as being Muslim if his or her mother and father are from a Muslim majority country. The countries included are Austria, Belgium, Switzerland, Germany, Denmark, Finland, Great Britain, and Netherlands. Applied population weights are standardized to sum up to 1 for each country-year cluster.

Figure C2:
Performance gaps and Ramadan fasting hours, reading and math scores, PISA



(a) Reading score **(b)** Math score

Note:: Performance gaps in reading and math scores, respectively, between students whose parents are from Muslim majority countries and all other students and logarithm of average fasting hours during Ramadan before test was taken. Both variables are adjusted by their country means.

Table C6:
Estimation results, TIMSS, 1995 to 2015, additional controls: test dates

	Math	Science	Math	Science	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline estimates						
Ramadan	2.25*	2.18**	1.40***	1.70***	1.05***	1.31***
	(1.22)	(0.95)	(0.36)	(0.41)	(0.35)	(0.40)
R squared	0.03	0.02	0.17	0.12	0.28	0.25
Test during Ramadan						
Ramadan	2.23*	2.16**	1.42***	1.72***	1.08***	1.33***
	(1.22)	(0.96)	(0.36)	(0.40)	(0.34)	(0.38)
Test during Ramadan	0.09	0.07	-0.16	-0.14	-0.22*	-0.20**
	(0.43)	(0.21)	(0.10)	(0.08)	(0.11)	(0.08)
R squared	0.03	0.02	0.17	0.12	0.28	0.25
Time between test and Ramadan						
Ramadan	2.34*	2.27**	1.19***	1.52***	0.81***	1.10***
	(1.24)	(0.96)	(0.30)	(0.38)	(0.30)	(0.37)
Time to test	-0.07	-0.08	0.09***	0.07***	0.10***	0.08***
	(0.11)	(0.05)	(0.03)	(0.02)	(0.04)	(0.02)
R squared	0.03	0.02	0.17	0.13	0.28	0.25
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No
Observations	364,786	364,786	364,786	364,786	364,786	364,786
Cluster	62	62	62	62	62	62

NOTE: Estimation of standardized achievement test scores on logarithmized average Ramadan fasting hours before the test was taken. Each panel refers to a different specification. Only Muslim majority countries included. A country is categorized as a Muslim country if more than 50 % of its population is Muslim. A list of the included control variables can be found in Table C2 in the Appendix. *Test Ramadan* is a binary variable indicating whether a test was taken during Ramadan. *Time to test* measures the days (in log) between the first day of Ramadan and the test date. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C7:
Estimation results, TIMSS, 1995 to 2015, additional control: log GDP per capita

	Only Muslim countries						All countries			
	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ramadan	1.63***	1.87***	0.96***	1.16***	0.55	0.78*	-0.42**	-0.52**	-0.78***	-0.88***
	(0.53)	(0.56)	(0.30)	(0.44)	(0.35)	(0.43)	(0.20)	(0.22)	(0.28)	(0.26)
Ramadan x % Muslim							1.34***	1.83***		
							(0.35)	(0.41)		
Ramadan x $\mathbb{1}(q_{25} < \%\text{ Muslim} \leq q_{50})$									0.56***	0.42**
									(0.21)	(0.17)
Ramadan x $\mathbb{1}(q_{50} < \%\text{ Muslim} \leq q_{75})$									0.40	0.67**
									(0.28)	(0.26)
Ramadan x $\mathbb{1}(\%\text{ Muslim} > q_{75})$									1.65***	1.93***
									(0.36)	(0.41)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Muslim-by-year FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes
R squared	0.05	0.06	0.12	0.13	0.26	0.26	0.51	0.47	0.51	0.47
Observations	364,786	364,786	364,786	364,786	364,786	364,786	1,169,954	1,169,014	1,169,954	1,169,014
Cluster	62	62	62	62	62	62	210	210	210	210

Note: Estimation of standardized achievement test scores on logarithmized average fasting hours over the last 3 years. A country is categorized as a Muslim country if more than 50 % of its population is Muslim. A list of the included control variables can be found in Table C2 in the Appendix. Additional control variable: log GDP. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C8:
Estimation results, TIMSS, 1995 to 2015, piecewise deletion: years

	Science	Math
Excluded wave		
1995	1.25***	1.00***
	(0.39)	(0.35)
1999	0.83	1.06***
	(0.55)	(0.38)
2003	1.38***	0.99**
	(0.42)	(0.41)
2007	1.41***	1.14***
	(0.39)	(0.34)
2011	0.90**	0.83*
	(0.39)	(0.42)
2015	1.81***	1.24**
	(0.45)	(0.50)

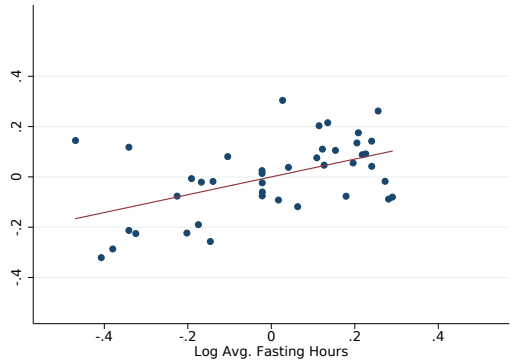
Note: Table shows result of baseline specification (Table 13 columns (5) and (6)) for various samples. Each row refers to estimation results for a sample where the respective year is excluded. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C9:
Estimation results, TIMSS, 1995 to 2015, piecewise deletion: Countries

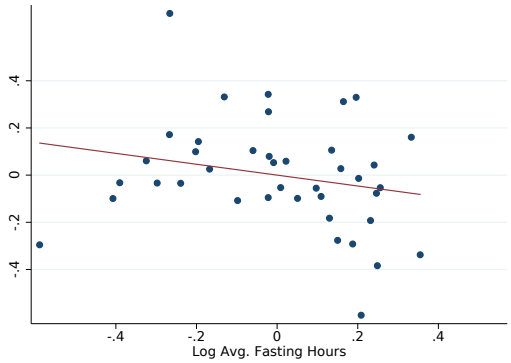
	Science	Math
Excluded country:		
Bahrain	1.34***	1.11***
	(0.41)	(0.34)
Palestine	1.34***	1.05***
	(0.40)	(0.35)
Indonesia	1.55***	1.45***
	(0.58)	(0.38)
Iran	1.35***	1.09***
	(0.40)	(0.36)
Kazhakstan	1.23***	0.99***
	(0.41)	(0.36)
Jordan	1.43***	1.18***
	(0.38)	(0.35)
Kuwait	1.26***	1.02***
	(0.40)	(0.35)
Lebanon	1.30***	1.06***
	(0.39)	(0.35)
Malaysia	1.06***	0.24
	(0.35)	(0.39)
Morocco	1.22***	0.99***
	(0.41)	(0.36)
Oman	1.37***	1.08***
	(0.41)	(0.35)
Qatar	1.33***	1.06***
	(0.40)	(0.35)
Saudi Arabia	1.33***	1.12***
	(0.40)	(0.34)
Syria	1.28***	1.02***
	(0.40)	(0.35)
United Arab Emirates	1.30***	1.04***
	(0.40)	(0.35)
Tunesia	1.27***	1.12***
	(0.41)	(0.34)
Turkey	1.07**	0.96**
	(0.44)	(0.40)
Egypt	1.36***	1.08***
	(0.38)	(0.35)

Note: Table shows result of baseline specification (Table 13 columns (5) and (6)) for various samples. Each row refers to estimation results for a sample where the respective country is excluded. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

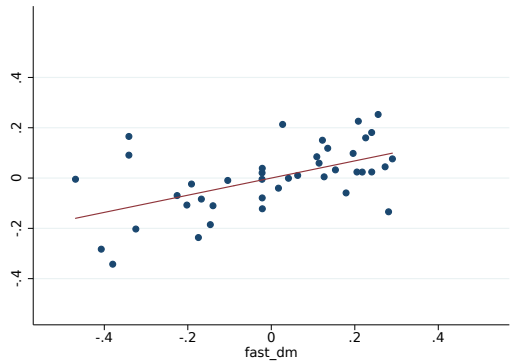
Figure C3:
Performance gaps and Ramadan fasting hours, reading and math scores, PISA



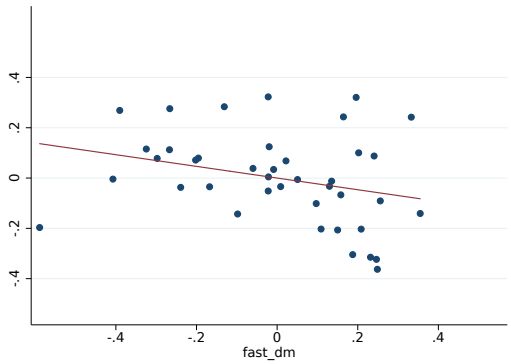
(a) Reading: Muslim immigrants vs natives



(b) Reading: Non-Muslim immigrants vs natives



(c) Math: Muslim immigrants vs natives



(d) Math: Non-Muslim immigrants vs natives

Note: Plots on the left (right) shows performance gaps in reading and math scores between students whose parents are from Muslim majority countries (non-Muslim majority countries) and natives and logarithm of average fasting hours during Ramadan before test was taken. Both variables are adjusted by their country means.

Table C10:
Non-Muslim majority countries, PISA

Austria	China, Croatia, Czech Republic, Hungary, Poland, Romania, Russia, Slovakia, Slovenia
Belgium	Germany
Switzerland	Portugal, Spain
Germany	Croatia, Greece, Poland, Slovenia
Denmark	Finland, Greenland, Iceland, Norway, Sweden
Finland	China, Russia, Thailand, Vietnam
Great Britain	Caribbean, China, Germany, India
Netherlands	Belgium, China, Germany, Italy, Poland, Romania, Spain, Suriname, United Kingdom, USA

Note: Students from a respective country (first column) are defined as being a non-Muslim immigrant if both parents are from one of the countries listed in the Table.

Table C11:
Estimation results, PISA, 2003 to 2018, non-Muslim included

	Science	Reading	Math
Muslim x Ramadan	0.33***	0.28**	0.29***
	(0.12)	(0.12)	(0.10)
Muslim	-1.14***	-0.99***	-0.98***
	(0.32)	(0.32)	(0.26)
Non-Muslim x Ramadan	-0.01	-0.07	-0.04
	(0.14)	(0.18)	(0.14)
Non-Muslim	0.10	0.25	0.21
	(0.37)	(0.48)	(0.36)
Country x Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
R squared	0.29	0.30	0.28
Observations	351,762	351,762	351,762
Cluster	48	48	48

Note: Estimation of standardized achievement test scores on logarithmized average fasting hours. Muslim and non-Muslim students are identified using the country of origin of their parents (see Table C3 and Table C10 in the Appendix). A list of the included control variables can be found in Table C5 in the Appendix. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C12:
Estimation results, PISA, 2003 to 2018, piecewise deletion: Years

	Science	Reading	Math
Excluded wave:			
2003	0.44**	0.34*	0.38**
	(0.19)	(0.18)	(0.14)
2006	0.28**	0.21	0.25**
	(0.13)	(0.14)	(0.11)
2009	0.34**	0.29**	0.30***
	(0.12)	(0.12)	(0.10)
2012	0.32**	0.25*	0.28**
	(0.13)	(0.12)	(0.11)
2015	0.33**	0.29**	0.29**
	(0.14)	(0.14)	(0.11)
2018	0.37***	0.38***	0.33***
	(0.13)	(0.13)	(0.12)

Note: Table shows result of baseline specification (Table 14 columns (2), (4), and (6)) for various samples. Each row refers to estimation results for a sample where the respective year is excluded. Reported standard errors in parentheses are cluster-robust at country times year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C13:
Estimation results, PISA, 2003 to 2018, piecewise deletion: Countries

	Science	Reading	Math
Excluded country:			
Austria	0.32**	0.22*	0.34***
	(0.13)	(0.13)	(0.10)
Belgium	0.31**	0.27*	0.26**
	(0.13)	(0.14)	(0.11)
Switzerland	0.29**	0.25*	0.24**
	(0.14)	(0.14)	(0.11)
Germany	0.31**	0.25	0.29**
	(0.15)	(0.15)	(0.13)
Denmark	0.41***	0.40***	0.35***
	(0.13)	(0.12)	(0.12)
Finland	0.40***	0.35***	0.34***
	(0.12)	(0.12)	(0.10)
Great Britain	0.32**	0.27**	0.28**
	(0.13)	(0.12)	(0.10)
Netherlands	0.28**	0.23*	0.23**
	(0.12)	(0.12)	(0.10)

Note: Table shows result of baseline specification (Table 14 columns (2), (4), and (6)) for various samples. Each row refers to estimation results for a sample where the respective country is excluded. Reported standard errors in parentheses are cluster-robust at country times year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C14:
Estimation results, PISA, 2003 to 2018, extension

	Science		Reading		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Ramadan x Share x Muslim		0.26		0.35*		0.10
		(0.17)		(0.18)		(0.15)
Ramadan x Share	0.32***	0.28***	0.33***	0.29***	0.33***	0.31***
	(0.08)	(0.08)	(0.09)	(0.09)	(0.07)	(0.08)
Ramadan x Muslim		-0.06		-0.19		0.01
		(0.13)		(0.17)		(0.13)
Muslim x Share		-0.70		-0.99**		-0.23
		(0.44)		(0.48)		(0.39)
Share	-1.13***	-1.01***	-1.13***	-1.02***	-1.16***	-1.09***
	(0.21)	(0.21)	(0.23)	(0.24)	(0.19)	(0.20)
Muslim		0.02		0.40		-0.14
		(0.32)		(0.45)		(0.32)
Country x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.30	0.30	0.30	0.30	0.29	0.29
Observations	351,762	351,762	351,762	351,762	351,762	351,762
Cluster	48	48	48	48	48	48

Note: Estimation of standardized achievement test scores on logarithmized average fasting hours. Muslim students are identified using the country of origin of their parents. *Share* is a binary variable indicating a high share of Muslim students in the school cohort of a student. A list of the included control variables can be found in Table C5 in the Appendix. In all regressions, standardized population weights are applied. Reported standard errors in parentheses are cluster-robust at country-year level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *Review of Economic Studies*, 72, 1–19.

Adetoro, N. (2010): "Internet utilisation and abuses in selected cybercafes in Ogun State, Nigeria," *African Journal of Library, Archives and Information Science*, 20, 19–28.

Adsera, A. and M. Pytlikova (2015): "The role of language in shaping international migration," *Economic Journal*, 125, 49–81.

Afifi, Z. E. M. (1997): "Daily practices, study performance and health during the Ramadan fast," *Journal of the Royal Society for the Promotion of Health*, 117, 231–235.

Aker, J. C., M. A. Clemens, and C. Ksoll (2011): "Mobiles and mobility: The effect of mobile phones on migration in Niger," *Proceedings of the German Development Economics Conference, Berlin*.

Akerlof, G. A. and R. E. Kranton (2002): "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature*, 40, 1167–1201.

Allen, T., C. d. C. Dobbin, and M. Morten (2018): "Border walls," *National Bureau of Economic Research*, No. 25267.

Almond, D., B. Mazumder, and R. Ewijk (2015): "In utero Ramadan exposure and children's academic performance," *Economic Journal*, 125, 1501–1533.

Almond, D. and B. A. Mazumder (2011): "Health capital and the prenatal environment: The effect of Ramadan observance during pregnancy," *American Economic Journal: Applied Economics*, 3, 56–85.

Angelucci, M. (2012): "US border enforcement and the net flow of Mexican illegal migration," *Economic Development and Cultural Change*, 60, 311–357.

——— (2015): "Migration and financial constraints: Evidence from Mexico," *Review of Economics and Statistics*, 97, 224–228.

Angrist, J. D. (2004): "Treatment effect heterogeneity in theory and practice," *Economic Journal*, 114, C52–C83.

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

Arendt, J. N., I. Bolvig, M. Foged, L. Hasager, and G. Peri (2020): "Integrating refugees: Language training or work-first Incentives?" *National Bureau of Economic Research*, No. 26834.

Arias-Vazquez, F. J. (2012): "A note on the effect of education on religiosity," *Economics Letters*, 117, 895–897.

Åslund, O. and D.-O. Rooth (2007): "Do when and where matter? Initial labour market conditions and immigrant earnings," *Economic Journal*, 117, 422–448.

Azlor, L., A. P. Damm, and M. L. Schultz-Nielsen (2020): "Local labour demand and immigrant employment," *Labour Economics*, 101808.

BAMF (2019): "Aktuelle Zahlen, Oktober 2019," *Bundesamt für Migration und Flüchtlinge*.

Bankston, C. L. and M. Zhou (1995): "Religious participation, ethnic identification, and adaptation of Vietnamese adolescents in an immigrant community," *Sociological Quarterly*, 36, 523–534.

Battisti, M., Y. Giesing, and N. Laurentsyeva (2019): "Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment," *Labour Economics*, 61, 101745.

Bauer, T. K., S. Braun, and M. Kvasnicka (2013): "The economic integration of forced migrants: Evidence for post-war Germany," *Economic Journal*, 123, 998–1024.

Bauernschuster, S. and M. Schlotter (2015): "Public child care and mothers' labor supply—Evidence from two quasi-experiments," *Journal of Public Economics*, 123, 1–16.

Bazzi, S. (2017): "Wealth heterogeneity and the income elasticity of migration," *American Economic Journal: Applied Economics*, 9, 219–55.

Becker, S. O. and A. Ferrara (2019): "Consequences of forced migration: A survey of recent findings," *Labour Economics*, 59, 1–16.

Becker, S. O., M. Nagler, and L. Woessmann (2017): "Education and religious participation: City-level evidence from Germany's secularization period 1890–1930," *Journal of Economic Growth*, 22, 273–311.

Becker, S. O. and L. Woessmann (2009): "Was Weber wrong? A human capital theory of Protestant economic history," *Quarterly Journal of Economics*, 124, 531–596.

Beine, M. and C. Parsons (2015): "Climatic factors as determinants of international migration," *Scandinavian Journal of Economics*, 117, 723–767.

Belot, M. and S. Ederveen (2012): "Cultural barriers in migration between OECD countries," *Journal of Population Economics*, 25, 1077–1105.

Ben-Porath, Y. (1967): "The production of human capital and the life cycle of earnings," *Journal of Political Economy*, 75, 352–365.

Bertanha, M. and G. W. Imbens (2020): "External validity in fuzzy regression discontinuity designs," *Journal of Business & Economic Statistics*, 38, 593–612.

Bertoli, S., H. Brücker, and J. Fernández-Huertas Moraga (2020): "Do processing times affect the distribution of asylum seekers across Europe?" *IZA Discussion Papers*, No. 13018.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119, 249–275.

Black, R., S. Ammassari, S. Mouillesseaux, and R. Rajkotia (2004): "Migration and Pro-Poor Policy in West Africa, University of Sussex," *Development Research Centre on Migration, Globalisation and Poverty Sussex*.

Böhme, M. H., A. Gröger, and T. Stöhr (2020): "Searching for a better life: Predicting international migration with online search keywords," *Journal of Development Economics*, 142, 102347.

Borjas, G. J. (1987): "Self-Selection and the earnings of immigrants," *American Economic Review*, 531–553.

Botticini, M. and Z. Eckstein (2005): "Jewish occupational selection: Education, restrictions, or minorities?" *Journal of Economic History*, 65, 922–948.

——— (2007): "From farmers to merchants, conversions and diaspora: Human capital and Jewish history," *Journal of the European Economic Association*, 5, 885–926.

Braga, M. (2007): "Dreaming another life. The role of foreign media in migration decisions. Evidence from Albania," *Unpublished Working Paper*.

Brake, D. (2019): *Submarine Cables: Critical Infrastructure for Global Communications*, Information Technology & Innovation Foundation.

Bratsberg, B., O. Raaum, and K. Røed (2014): "Immigrants, labour market performance and social insurance," *Economic Journal*, 124, 644–683.

Brell, C., C. Dustmann, and I. Preston (2020): "The labor market integration of refugee migrants in high-income countries," *Journal of Economic Perspectives*, 34, 94–121.

Brown, S. and K. Taylor (2007): "Religion and education: Evidence from the National Child Development Study," *Journal of Economic Behavior & Organization*, 63, 439–460.

Brücker, H., A. Hauptmann, and P. Jaschke (2020): "Beschränkungen der Wohnortwahl für anerkannte Geflüchtete: Wohnsitzauflagen reduzieren die Chancen auf Arbeitsmarktintegration," *IAB-Kurzbericht*, 03/2020.

Cai, S. (2020): "Migration under liquidity constraints: Evidence from randomized credit access in China," *Journal of Development Economics*, 142.

Campante, F. and D. Yanagizawa-Drott (2015): "Does religion affect economic growth and happiness? Evidence from Ramadan," *Quarterly Journal of Economics*, 130, 615–658.

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996): "Migration with endogenous moving costs," *American Economic Review*, 909–930.

Cattaneo, C. and G. Peri (2016): "The migration response to increasing temperatures," *Journal of Development Economics*, 122, 127–146.

Cesur, R. and N. H. Mocan (2013): "Does secular education impact religiosity, electoral participation and the propensity to vote for Islamic parties? Evidence from an education reform in a Muslim country," *National Bureau of Economic Research*, No. 19769.

Chaney, E. (2011): "Islam and human capital formation: Evidence from premodern Muslim science," in *The Oxford Handbook of the Economics of Religion*, ed. by R. M. McCleary, Oxford University Press, chap. 2.4, 81–92.

——— (2016): "Religion and the rise and fall of Islamic science," *Unpublished Working Paper, Department of Economics, Harvard University.*

Chaudhary, L. and J. Rubin (2011): "Reading, writing, and religion: Institutions and human capital formation," *Journal of Comparative Economics*, 39, 17–33.

Chavula, J., N. Feamster, A. Bagula, and H. Suleman (2014): "Quantifying the effects of circuitous routes on the latency of intra-Africa internet traffic: a study of research and education networks," in *International Conference on e-Infrastructure and e-Services for Developing Countries*, Springer, 64–73.

Chiquiar, D. and G. H. Hanson (2005): "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113, 239–281.

Chiswick, B. R. (1978): "The effect of Americanization on the earnings of foreign-born men," *Journal of Political Economy*, 86, 897–921.

Clemens, M. A. (2014): "Does development reduce migration?" *International Handbook on Migration and Economic Development*, 152, 152–176.

Clingingsmith, D., A. I. Khwaja, and M. Kremer (2009): "Estimating the impact of the Hajj: Religion and tolerance in Islam's global gathering," *Quarterly Journal of Economics*, 124, 1133–1170.

Cochran, J. K. and R. L. Akers (1989): "Beyond hellfire: An exploration of the variable effects of religiosity on adolescent marijuana and alcohol use," *Journal of Research in Crime and Delinquency*, 26, 198–225.

Cohodes, S. R. (2020): "The long-run impacts of specialized programming for high-achieving students," *American Economic Journal: Economic Policy*, 12, 127–66.

Coleman, J. S., T. Hoffer, and S. Kilgore (1982): *High school achievement: Public, Catholic, and private schools compared*, New York: Basic Books.

Connor, P. (2018): *At Least a Million Sub-Saharan Africans Moved to Europe Since 2010: Sub-Saharan Migration to the United States also Growing*, Pew Research Center.

Cortes, K. E. (2004): "Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States," *Review of Economics and Statistics*, 86, 465–480.

Czaika, M. and H. de Haas (2014): "The globalization of migration: Has the world become more migratory?" *International Migration Review*, 48, 283–323.

Dahl, G. and S. DellaVigna (2009): "Does movie violence increase violent crime?" *Quarterly Journal of Economics*, 124, 677–734.

Davezies, L. and T. Le Barbanchon (2017): "Regression discontinuity design with continuous measurement error in the running variable," *Journal of Econometrics*, 200, 260–281.

De Chaisemartin, C. and X. d'Haultfoeuille (2018): "Fuzzy differences-in-differences," *Review of Economic Studies*, 85, 999–1028.

De Haas, H. (2006): *Trans-Saharan migration to North Africa and the EU: historical roots and current trends*, Migration Policy Institute, Migration Information Source.

——— (2007): "International migration, national development and the role of governments: The case of Nigeria," in *International Migration and National Development in Sub-Saharan Africa: viewpoints and policy initiatives in the country of origin*, Brill Publishers.

Dekker, R. and G. Engbersen (2014): "How social media transform migrant networks and facilitate migration," *Global Networks*, 14, 401–418.

DellaVigna, S. and E. La Ferrara (2015): "Economic and social impacts of the media," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Strömberg, Elsevier, vol. 1, 723–768.

Deutscher Bundestag (2016): "Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke, Jan Korte, Harald Petzold (Havelland), Halina Wawzyniak und der Fraktion DIE LINKE (Drucksache 18/9657)," *Drucksache 18/9992*.

——— (2017): "Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke, Frank Tempel, Kersten Steinke, Jörn Wunderlich und der Fraktion DIE LINKE (Drucksache 18/10960)," *Drucksache 18/11473*.

Devillanova, C., F. Fasani, and T. Frattini (2018): "Employment of undocumented immigrants and the prospect of legal status: evidence from an amnesty program," *ILR Review*, 71, 853–881.

DiNardo, J. and D. S. Lee (2004): "Economic impacts of new unionization on private sector employers: 1984–2001," *Quarterly Journal of Economics*, 119, 1383–1441.

Docquier, F. and H. Rapoport (2012): "Globalization, brain drain, and development," *Journal of Economic Literature*, 50, 681–730.

Duleep, H. O. (2015): "The adjustment of immigrants in the labor market," in *Handbook of the Economics of International Migration*, ed. by B. R. Chiswick and P. W. Miller, Elsevier, vol. 1, 105–182.

Dustmann, C. (1993): "Earnings adjustment of temporary migrants," *Journal of Population Economics*, 6, 153–168.

——— (1997): "Differences in the labor market behavior between temporary and permanent migrant women," *Labour Economics*, 4, 29–46.

——— (1999): "Temporary migration, human capital, and language fluency of migrants," *Scandinavian Journal of Economics*, 101, 297–314.

——— (2000): "Temporary migration and economic assimilation," *Swedish Economic Policy Review*, 7, 213–244.

Dustmann, C., F. Fasani, T. Frattini, L. Minale, and U. Schönberg (2017): "On the economics and politics of refugee migration," *Economic Policy*, 32, 497–550.

Dustmann, C. and J.-S. Görlach (2016): "The economics of temporary migrations," *Journal of Economic Literature*, 54, 98–136.

Dustmann, C. and A. Okatenko (2014): "Out-migration, wealth constraints, and the quality of local amenities," *Journal of Development Economics*, 110, 52–63.

Evans, T. D., F. T. Cullen, R. G. Dunaway, and V. S. Burton (1995): "Religion and crime reexamined: The impact of religion, secular controls, and social ecology on adult criminality," *Criminology*, 33, 195–224.

Evans, W. N. and R. M. Schwab (1995): "Finishing high school and starting college: Do Catholic schools make a difference?" *Quarterly Journal of Economics*, 110, 941–974.

Fan, C. S. (2008): "Religious participation and children's education: A social capital approach," *Journal of Economic Behavior & Organization*, 65, 303–317.

Farré, L. and F. Fasani (2013): "Media exposure and internal migration: Evidence from Indonesia," *Journal of Development Economics*, 102, 48–61.

Fasani, F., T. Frattini, and L. Minale (2018): "(The Struggle for) Refugee Integration into the Labour Market: Evidence from Europe," *CEPR Discussion Paper*, No. 12718.

Feigenberg, B. (2020): "Fenced out: The impact of border construction on US-Mexico migration," *American Economic Journal: Applied Economics*, 12, 106–39.

Freeman, R. B. (1985): "Who escapes? The relation of church-going & other background factors to the socio-economic performance of blak male youth from inner-city poverty tracts," Working Paper 1656, National Bureau of Economic Research.

Gathmann, C. and N. Keller (2018): "Access to citizenship and the economic assimilation of immigrants," *Economic Journal*, 128, 3141–3181.

Gentzkow, M. and J. M. Shapiro (2008): "Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study," *Quarterly Journal of Economics*, 123, 279–323.

Gibson, J., D. McKenzie, and S. Stillman (2011): "The impacts of international migration on remaining household members: omnibus results from a migration lottery program," *Review of Economics and Statistics*, 93, 1297–1318.

Glaeser, E. L. and B. I. Sacerdote (2008): "Education and religion," *Journal of Human Capital*, 2, 188–215.

Greenstein, S. (2020): "The Basic Economics of Internet Infrastructure," *Journal of Economic Perspectives*, 34, 192–214.

Grembi, V., T. Nannicini, and U. Troiano (2016): "Do fiscal rules matter?" *American Economic Journal: Applied Economics*, 1–30.

Grote, J. (2018): "The changing influx of asylum seekers in 2014-2016: Responses in Germany, Focussed study by the German National Contact Point for the European Migration Network (EMN)," *Federal Office for Migration and Refugees, Working Paper*, No. 79.

Gruber, J. H. (2005): "Religious market structure, religious participation, and outcomes: Is religion good for you?" *BE Journal of Economic Analysis & Policy*, 5.

GSMA (2015): "Digital Inclusion and the role of mobile in Nigeria," *www.gsma.com*.

Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69, 201–209.

Hainmueller, J., D. Hangartner, and D. Lawrence (2016): "When lives are put on hold: Lengthy asylum processes decrease employment among refugees," *Science Advances*, 2, e1600432.

Hanson, G. H. and C. McIntosh (2010): "The great Mexican emigration," *Review of Economics and Statistics*, 92, 798–810.

Hatton, T. J. (2005): "Explaining trends in UK immigration," *Journal of Population Economics*, 18, 719–740.

——— (2015): "Asylum policy in the EU: the case for deeper integration," *CESifo Economic Studies*, 61, 605–637.

Heckman, J. J. (2013): *Giving Kids a Fair Chance*, Mit Press.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The economics and econometrics of active labor market programs," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. 3, Part A, 1865–2097.

Heckman, J. J. and Y. Rubinstein (2001): "The importance of noncognitive skills: Lessons from the GED testing program," *American Economic Review*, 91, 145–149.

Hernández-Coss, R. and C. Egwuagu Bun (2007): *The UK-Nigeria remittance corridor: challenges of embracing formal transfer systems in a dual financial environment*, The World Bank.

Hinnerich, B. T. and P. Pettersson-Lidbom (2014): "Democracy, redistribution, and political participation: Evidence from Sweden 1919–1938," *Econometrica*, 82, 961–993.

Hjort, J. and J. Poulsen (2019): "The arrival of fast internet and employment in Africa," *American Economic Review*, 109, 1032–79.

Hullegie, P. and T. J. Klein (2010): "The effect of private health insurance on medical care utilization and self-assessed health in Germany," *Health Economics*, 19, 1048–1062.

Hungerman, D. M. (2014): "The effect of education on religion: Evidence from compulsory schooling laws," *Journal of Economic Behavior & Organization*, 104, 52–63.

Iannaccone, L. R. (1998): "Introduction to the economics of religion," *Journal of Economic Literature*, 36, 1465–1495.

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.

International Organization for Migration (2018): "World migration report 2018," Tech. rep.

Iyer, S. (2016): "The new economics of religion," *Journal of Economic Literature*, 54, 395–441.

Iyer, S., C. Velu, and M. Weeks (2014): "Divine competition: Religious organisations and service provision in India," *Unpublished Working Paper*.

Jackson, C. K. (2012): "Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina," *NBER Working Paper*, No. 18624.

Jiang, Z. and P. Ding (2019): "Measurement errors in the binary instrumental variable model," *Biometrika*, 107, 238–245.

Kane, T. J., C. E. Rouse, and D. Staiger (1999): "Estimating returns to schooling when schooling is misreported," *National Bureau of Economic Research*, No. 7235.

Kende, M. and K. Rose (2015): "Promoting local content hosting to develop the internet ecosystem," *www.internetsociety.org*.

Kolesár, M. and C. Rothe (2018): "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, 108, 2277–2304.

Kosyakova, Y. and H. Brenzel (2020): "The role of length of asylum procedure and legal status in the labour market integration of refugees in Germany," *SozW Soziale Welt*, 71, 123–159.

Kroh, M., H. Brücker, S. Kühne, E. Liebau, J. Schupp, M. Siegert, and P. Trübswetter (2016): "Das Studiendesign der IAB-BAMF-SOEP-Befragung von Geflüchteten," *SOEP Survey Papers*, No. 365.

Kuran, T. (2003): "The Islamic commercial crisis: Institutional roots of economic underdevelopment in the Middle East," *Journal of Economic History*, 63, 414–446.

——— (2004): "Why the Middle East is economically underdeveloped: Historical mechanisms of institutional stagnation," *Journal of Economic Perspectives*, 18, 71–90.

——— (2018): "Islam and economic performance: Historical and contemporary links," *Journal of Economic Literature*, forthcoming.

La Ferrara, E., A. Chong, and S. Duryea (2012): "Soap operas and fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, 4, 1–31.

LaRoche, S., M. Joncas, and P. Foy (2016): "Sample design in TIMSS 2015," in *Methods and Procedures in TIMSS 2015*, ed. by M. O. Martin, I. V. Mullis, and M. Hooper, TIMSS & PIRLS International Study Center, chap. 3, 39–76.

Lee, D. S. (2008): "Randomized experiments from non-random selection in US House elections," *Journal of Econometrics*, 142, 675–697.

Lee, D. S. and D. Card (2008): "Regression discontinuity inference with specification error," *Journal of Econometrics*, 142, 655–674.

Lee, D. S. and T. Lemieux (2010): "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48, 281–355.

Lee, J. N., J. Morduch, S. Ravindran, A. Shonchoy, and H. Zaman (2020): "Poverty and migration in the digital age: Experimental evidence on mobile banking in Bangladesh," *American Economic Journal: Applied Economics (forthcomming)*.

Leiper, J. B. and A. Molla (2003): "Effects on health of fluid restriction during fasting in Ramadan," *European Journal of Clinical Nutrition*, 57, 30–38.

Lleras, C. (2008): "Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings," *Social Science Research*, 37, 888–902.

Lu, Y., H. Xie, and L. C. Xu (2016): "Telecommunication externality on migration: Evidence from Chinese villages," *China Economic Review*, 39, 77–90.

Majid, M. F. (2015): "The persistent effects of in utero nutrition shocks over the life cycle: Evidence from Ramadan fasting," *Journal of Development Economics*, 117, 48–57.

Mayda, A. M. (2010): "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23, 1249–1274.

McCleary, R. M. and R. J. Barro (2006): "Religion and economy," *The Journal of Economic Perspectives*, 20, 49–72.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

McKenzie, D. and H. Rapoport (2007): "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, 84, 1–24.

——— (2010): "Self-selection patterns in Mexico-US migration: the role of migration networks," *Review of Economics and Statistics*, 92, 811–821.

Meckel, Y., A. Ismaeel, and A. Eliakim (2008): "The effect of the Ramadan fast on physical performance and dietary habits in adolescent soccer players," *European Journal of Applied Physiology*, 102, 651–657.

Missirian, A. and W. Schlenker (2017): "Asylum applications respond to temperature fluctuations," *Science*, 358, 1610–1614.

Mocan, N. and L. Pogorelova (2017): "Compulsory schooling laws and formation of beliefs: Education, religion and superstition," *Journal of Economic Behavior & Organization*, 142, 509–539.

Munshi, K. (2003): "Networks in the modern economy: Mexican migrants in the US labor market," *Quarterly Journal of Economics*, 118, 549–599.

Mwesige, P. G. (2004): "Cyber elites: A survey of Internet café users in Uganda," *Telematics and Informatics*, 21, 83–101.

Neal, D. (1997): "The effects of Catholic secondary schooling on educational achievement," *Journal of Labor Economics*, 15, 98–123.

Oosterbeek, H. and B. van der Klaauw (2013): "Ramadan, fasting and educational outcomes," *Economics of Education Review*, 34, 219–226.

Ortega, F. and G. Peri (2013): "The effect of income and immigration policies on international migration," *Migration Studies*, 1, 47–74.

——— (2015): "Migration policies: recent advances on measurement, determinants and outcomes," *CESifo Economic Studies*, 61, 521–526.

Pei, Z. and Y. Shen (2016): "The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable," *arXiv e-prints*, arXiv:1609.01396.

Peri, G. (2016): "Immigrants, productivity, and labor markets," *Journal of Economic Perspectives*, 30, 3–30.

Phillips, R. (1998): "Religious market share and Mormon Church activity," *Sociology of Religion*, 59, 117–130.

Rapoport, H. and F. Docquier (2006): "The economics of migrants' remittances," in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. by S.-C. Kolm and J. M. Ythier, Elsevier, vol. 2, 1135–1198.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosholm, M. and R. Vejlin (2010): "Reducing income transfers to refugee immigrants: Does start-help help you start?" *Labour Economics*, 17, 258–275.

Rubin, J. (2011): "Institutions, the rise of commerce and the persistence of laws: Interest restrictions in Islam and Christianity," *Economic Journal*, 121, 1310–1339.

Sander, W. (2002): "Religion and human capital," *Economics Letters*, 75, 303–307.

Schammann, H. (2019): "Aufenthalt gegen Leistung? Der Einzug meritokratischer Elemente in die deutsche Flüchtlingspolitik," in *Flucht–Bildung–Integration? Bildungspolitische und pädagogische Herausforderungen von Fluchtverhältnissen*, ed. by M. S. Baader, T. Freytag, and D. Wirth, Springer, 43–61.

Schielke, S. (2009): "Being good in Ramadan: Ambivalence, fragmentation, and the moral self in the lives of young Egyptians," *Journal of the Royal Anthropological Institute*, 15, 24–40.

Sjaastad, L. A. (1962): "The costs and returns of human migration," *Journal of Political Economy*, 70, 80–93.

Smith, C. (2003): "Theorizing religious effects among American adolescents," *Journal for the Scientific Study of Religion*, 42, 17–30.

Smith, E. S. (1999): "The effects of investments in the social capital of youth on political and civic behavior in young adulthood: A longitudinal analysis," *Political Psychology*, 20, 553–580.

Stanley, U., N. Nkordeh, B.-M. Ibiabo, and V. M. Olu (2018): "Broadband superhighway: A detailed analysis of Nigeria Trans-Atlantic fibre optic cables: Comparing MainOne cable, Glo-1 cable, WACS and SAT3," *Proceedings of the International Conference on Industrial Engineering and Operations Management*.

Tiedemann, P. (2014): "Die Geschichte des subsidiären Flüchtlingsschutzes," in *Flüchtlingsrecht in Theorie und Praxis. Füunf Jahre Refugee Law Clinic an der Justus-Liebig-Universität Gießen*, ed. by P. Tiedemann and J. Gieseking, Nomos, 95–122.

Toda, M. and K. Morimoto (2004): "Ramadan fasting–Effect on healthy Muslims," *Social Behavior and Personality: An International Journal*, 32, 13–18.

UN DESA, Population Division (2019): *World Population Prospects 2019, Volume II: Demographic Profiles (ST/ESA/SER.A/427)*, United Nations Department of Economic and Social Affairs.

UNDP (2019): *Human Development Report 2019. Beyond income, beyond averages, beyond today: Inequalitites in human development in the 21st century*, United Nations Development Programme.

Van Ewijk, R. (2011): "Long-term health effects on the next generation of Ramadan fasting during pregnancy," *Journal of Health Economics*, 30, 1246–1260.

Wilson, J. (1978): *Religion in American Society: The Effective Presence*, Prentice Hall.

WorldPovertyClock (2020): https://worldpoverty.io/, accessed: 2020-09-09.

Wright, L. S., C. J. Frost, and S. J. Wisecarver (1993): "Church attendance, meaningfulness of religion, and depressive symptomatology among adolescents," *Journal of Youth and Adolescence*, 22, 559–568.

Yamamoto, K. and E. Kulick (2016): "TIMSS 2015 achievement scaling methodology," in *Methods and Procedures in TIMSS 2015*, ed. by M. O. Martin, I. V. Mullis, and M. Hooper, TIMSS & PIRLS International Study Center, chap. 12, 39–76.

Yang, D. (2011): "Migrant remittances," *Journal of Economic Perspectives*, 25, 129–52.

Young, A. (2013): "Inequality, the urban-rural gap, and migration," *Quarterly Journal of Economics*, 128, 1727–1785.

Zijlstra, J. and I. v. Liempt (2017): "Smart (phone) travelling: Understanding the use and impact of mobile technology on irregular migration journeys," *International Journal of Migration and Border Studies*, 3, 174–191.