To Veil or Not to Veil? Assessing the Removal of Headscarf Ban in a Muslim Country*

Wenqi Lu† Ekin Yurdakul‡

May 30, 2026

Abstract

This paper examines how removing identity-based institutional restrictions affects women's economic participation by exploiting the 2013 repeal of a longstanding headscarf ban in Turkey. We combine two nationally representative surveys and use statistical matching and machine-learning models to predict women's veiling status and identify treatment effects by comparing veiled and non-veiled women's labor market outcomes in a difference-in-differences framework. The repeal led to a significant rise in public sector employment among veiled women, driven by both higher employment rates and shifts away from self-employment and unpaid family work. In contrast, non-veiled women experienced a decline in public sector jobs, suggesting a substitution effect, with suggestive evidence of some increase in education–occupation mismatch among veiled women in the public sector after the repeal. Consistent with this pattern, effects are concentrated among more educated veiled women, the group most likely to qualify for public positions. We find no differential effects across regions with high and low veiling prevalence or local Islamist-party vote shares, suggesting that institutional access, rather than local acceptance or political favoritism, drives women's labor market response to the reform.

JEL: J16, J12, J22, K31

Keywords: headscarf ban, identity-based policy, women's economic participation

*We are particularly grateful to Luca Paolo Merlino for extensive comments and valuable guidance. We also thank Laurens Cherchye, Giuseppe Sorrenti, Sunčica Vujić, Bahadır Dursun, Thomas van Huizen and Marion Leturcq for helpful feedback and suggestions. We further thank participants at the 2024 ESPE and 2024 EALE conferences for valuable comments. All remaining errors are our own.

†Department of Economics, University of Antwerp; ECARES, Université Libre de Bruxelles; Aix-Marseille University, CNRS, AMSE (Marseille, France). E-mail: wenqi.lu@ulb.be

‡Corresponding author. Utrecht School of Economics, Utrecht University. e.yurdakul@uu.nl

1 Introduction

Economic participation is often shaped by the interaction between cultural identity and institutional rules (Akerlof and Kranton, 2000; Fernández and Fogli, 2009; Alesina et al., 2013; Alesina and Giuliano, 2015; Oh, 2023). When institutions regulate how identity can be expressed, for example, by veiling bans or dress codes, they may alter individuals' incentives and opportunities in the labor market. Recent research highlights the importance of identity-based policies in shaping behavior and welfare (Carvalho, 2013; Fouka, 2020; Saleh and Tirole, 2021; Abou Daher et al., 2025). This raises a central question: when legal barriers linked to identity are removed, do individuals adjust their economic choices, or do persistent identity-based norms continue to constrain participation? The answer matters for designing policies that aim to expand economic opportunity by changing formal institutions rather than deeply embedded cultural norms (Bursztyn et al., 2020; Jayachandran, 2021; Dhar et al., 2022).[1]

To provide evidence on whether easing identity-based restrictions can translate into greater economic inclusion, we study the repeal of a longstanding headscarf ban, which lifted a major institutional barrier to veiled women's employment in public institutions in Turkey. Until 2013, wearing headscarves in all public institutions was prohibited under the country's constitutionally secular framework. In October 2013, the government lifted this ban, allowing women to wear headscarves in public institutions, including government offices. Given that roughly 60% of Turkish women wear headscarves, this reform represents a rare and sharp nationwide elimination of a legal barrier directly linked to a widespread cultural practice, providing a distinctive setting to study how women's labor market outcomes respond to the lifting of an identity-based institutional constraint.

We use data from the 2010–2017 Turkish Household Labor Force Survey (HLFS) and employ a difference-in-differences strategy to estimate the effects of the policy change. A key empirical challenge in the literature on headscarf bans is the lack of direct information on veiling status in large-scale surveys, which has often limited identification strategies. To address this, we leverage the 2013 Turkish Demographic and Health Survey (DHS), which uniquely contains individual-level information on veiling together with detailed background characteristics. We integrate the two datasets using statistical matching techniques (Rubin, 1986) to estimate veiling probabilities in the HLFS based on observable characteristics. Unlike previous work that relies on broad proxies such as ethnicity or immigrant status to identify the effects of institutional restrictions regarding veiling, our approach uses individual-level survey data to impute veiling status as a proxy for treatment assignment, thereby improving the precision of group identification.[2] This enables a quasi-experimental comparison of labor market outcomes between veiled and non-veiled women before and after the policy change. Robustness checks using alternative prediction models and machine-learning algorithms (Mullainathan and Spiess, 2017; Athey and Imbens, 2019; Heller et al., 2024) confirm that the results are not sensitive to different model specifications.

Our findings reveal that following the removal of the headscarf ban, women predicted to be veiled experienced a 1.2-percentage-point increase in employment (about 4 percent relative to the sample mean) compared with their non-veiled counterparts. This rise is

[1] In this paper, we use the term cultural identity to refer to forms of identity grounded in religious or traditional values, such as veiling in Muslim societies. Cultural norms denote the informal behavioral expectations associated with that identity, for example, beliefs about appropriate roles for veiled women in public life and labor market.

[2] We first estimate each woman's probability of veiling using the 2013 DHS based on observable characteristics, and then apply nearest-neighbor statistical matching (Rosenbaum and Rubin, 1983) to impute veiling status in the HLFS. This two-step approach links women across surveys with comparable background characteristics and enables consistent treatment classification (see Ridder and Moffitt (2007) for related methods).

driven primarily by a substantial increase in public sector employment among veiled women, whose participation in such jobs was nearly zero before the reform. At the same time, we observe notable declines in self-employment and unpaid family work, suggesting that the repeal reshaped employment choices among women already in the workforce. Consistent with this shift, the probability of informal employment fell, while the likelihood of holding a permanent contract increased. Following the repeal, veiled women's average weekly working hours rose by about half an hour per week, reflecting higher employment rates rather than longer hours among those already working. Examining the occupational distribution within the public sector, we find the largest gains in professional and associate professional roles, such as teachers, nurses, and administrative staff, which typically require some level of educational attainment.

Interestingly, we also find that women predicted not to be veiled experienced a shift away from public sector jobs toward private-sector employment, despite overall public sector hiring remaining stable. This pattern points to a substitution effect: the repeal reallocated public employment opportunities from women predicted not to be veiled to women predicted to be veiled rather than generating new public sector jobs. We do not find evidence that this substitution reflects local political favoritism, as the post-repeal rise in public employment among women predicted to be veiled is not concentrated in politically aligned regions. At the same time, our analysis of education–occupation matching within the public sector provides suggestive evidence of some increase in underqualification and overall mismatch among women predicted to be veiled after the reform. These patterns suggest that the repeal broadened access to public employment, while also changing the composition of women employed in the public sector.

Two potential identification concerns deserve emphasis. The first is that predicted veiling status may be correlated with unobserved religiosity, which could independently influence women's labor supply decisions (Carvalho, 2013). Because religiosity is not observed in our main dataset, we proxy for it using two pre-reform regional measures of religious practice, mosque density and Quran course enrollment per capita, and find no evidence that the estimated effects are concentrated in more religious areas. Consistent with the institutional scope of the reform, the gains of the repeal are instead concentrated among relatively more educated veiled women, those more likely to access public sector jobs, while effects for less educated women are negligible. The second concern is that our identification strategy implicitly treats veiling as a stable trait. Although individual panel data on veiling are not available, we document very high regional stability in veiling rates across the 2008 and 2013 DHS waves. More importantly, we exploit two major institutional reforms, the 1997 extension of compulsory schooling and the 2002 Civil Code reform, raising the minimum marriage age to 18, and show that neither reform had any sizable effect on the probability of veiling. We also exploit interview timing within the 2013 DHS and find no evidence that lifting the ban led to a significant change in women's veiling decisions in the short run. This evidence suggests that veiling is highly persistent and unlikely to respond to short-run policy changes. To the extent that some women may have altered their veiling status in response to the policy, it would attenuate the estimates, implying that our results should be interpreted as conservative lower bounds.[3]

Given that veiling is often associated with lower employment in Western contexts and in some Muslim-majority settings where secular institutions formally restrict religious expression (Ghumman and Jackson, 2010; Abdelhadi, 2019; Dildar, 2015),[4] our findings shed

[3]We discuss exogeneity of veiling status in detail in Section 4.

[4]It is important to note that this negative association is not universal and appears instead to be context-

light on whether veiled women's lower labor market participation primarily reflects individual religious identity constraints or institutional barriers (Aksoy and Gambetta, 2016; Joslin and Nordvik, 2021; Shofia, 2022; Jacquet and Montpetit, 2023). Exploiting the repeal of a long-standing headscarf ban in a secular, yet Muslim-majority, setting, we show that removing institutional restrictions on religious expression substantially increased veiled women's employment, most visibly in the public sector, where the ban had applied directly. This finding highlights that institutional constraints can play a decisive role in shaping women's economic choices. Conceptually, our evidence suggests that reducing the cost of identity expression need not hinder economic inclusion; on the contrary, relaxing institutional barriers can facilitate the integration of women previously excluded from employment due to their religious identity.

Economic theories of veiling predict that when restrictions impose significant costs on religious expression, some women may withdraw from employment rather than compromise their religious identity, while others may adjust their practices if economic incentives outweigh these costs (Carvalho, 2013). Despite the increasing prevalence of veiling regulations worldwide, systematic evidence on their economic and social consequences remains limited, largely due to the absence of direct information on veiling status.[5] A few exceptions provide mixed evidence, reflecting this theoretical ambiguity. Abdelgadir and Fouka (2020) study the 2004 French headscarf ban in schools using country of birth to proxy for Muslim identity and find that exposure to the ban significantly reduced secondary educational attainment among female students of North African origin, with long-term negative effects on labor market outcomes and family composition. Using a similar empirical approach but focusing on a different cohort, Maurin and Navarrete H (2023) exploit nationality at birth to identify Muslim students and find that an earlier 1994 ministerial circular prohibiting veiling improved educational outcomes for female students with a Muslim background.

Unlike these studies, which analyze the effects of restrictions on wearing a headscarf, our paper provides one of the first causal estimates of the consequences of lifting such a restriction in public spaces. While the proxies used by Abdelgadir and Fouka (2020) and Maurin and Navarrete H (2023)—country and nationality at birth—are plausibly exogenous and appropriate given data limitations, they remain imperfect indicators of veiling, as only around 30% of Muslim women are veiled in France (Drouhot et al., 2023). Moreover, these proxies are not informative in Muslim-majority contexts, where national or ethnic origin does not distinguish between veiled and non-veiled women. Our approach uses prediction-based statistical matching methods in a difference-in-differences framework to identify the treatment effects of lifting veiling restrictions, offering an alternative, though still indirect, strategy that is better suited to settings where the majority is Muslim.

Given the key differences in treated populations, we further examine heterogeneity by regional veiling prevalence to provide intuition across Muslim-minority and Muslim-majority contexts. Specifically, we contrast regions where veiled women constitute a minority (below 50 percent) with those where they are the majority. We find no significant differences in

dependent. In Muslim-majority settings without such institutional restrictions, veiling may facilitate rather than hinder women's labor market entry by enabling participation in public life without violating prevailing community norms, as shown for Indonesia (Shofia, 2022) and theorized in accounts of the New Veiling Movement (Carvalho, 2013).

[5] Throughout the text, we use the term "veiling regulations" to refer to policies governing the wearing of headscarves, turbans, and chadors. These regulations exist across Europe and several Muslim-majority countries. For example, France banned all types of veiling in state-run institutions, such as schools and hospitals, in 2004. More recently, in 2023, the European Court of Justice (ECJ) ruled that EU Member States may prohibit government employees from wearing religious symbols, including headscarves. Similarly, Kazakhstan, where a majority of women practice Islam, has banned headscarves for students and teachers in schools.

labor market responses between these groups, suggesting that peer effects and local social acceptance are not the main drivers of our results. This heterogeneity analysis is, of course, specific to variation in veiling prevalence within a Muslim-majority country, and any external parallels should be viewed as suggestive rather than directly comparable. Nonetheless, the pattern we document offers indicative implications for contexts such as France, where veiled women form a minority and may face discrimination. The finding that veiled women in Turkey respond similarly regardless of local veiling prevalence suggests that easing institutional restrictions can enhance participation even in environments where veiling is less common or socially discouraged. This strengthens the external relevance of our findings by showing that institutional access matters even where social acceptance varies.

In contexts such as Turkey, our findings suggest that integration need not require assimilation into secular norms. Rather, it may involve enabling women's economic and social participation without compelling them to abandon their religious identity. From this perspective, removing identity-based restrictions can promote economic inclusion independently of cultural assimilation. Using data from the Turkish DHS, Uğur (2020) examines the impact of the 1997 ban on wearing headscarves in universities and formal education in a regression discontinuity design by exploiting birth month and year information. She finds no significant effect on university completion among veiled women, largely reflecting their already low university enrollment. Aksoy and Gambetta (2021) combine individual-level survey data on women's veiling from the Turkish Demographic and Health Surveys with regional variation from the 2004 municipal elections, where the Islamic Justice and Development Party (AKP) narrowly won or lost control in different provinces. They find no immediate effect on veiling in 2008 but a significant rise by 2013 in provinces where the AKP narrowly won, particularly among less religious and poorer women, suggesting that veiling can function strategically or instrumentally as the political and social environment becomes more permissive.[6] More closely related to our analysis, Corekcioglu (2021) studies the 2013 removal of the headscarf ban in public spaces using a difference-in-discontinuities design focused on municipal employment outcomes in closely contested mayoral elections. Consistent with our results, she finds an increase in female employment, but only in municipalities governed by Islamist mayors. Our study substantially extends this evidence by examining the entire labor market, covering both public and private sector employment, and showing that the observed gains among veiled women are not fully driven by local governance. In addition, we explore marriage market responses to the repeal and provide suggestive evidence that the removal of the headscarf ban reduced the probability of being married and increased the probability of divorce among veiled women relative to their non-veiled peers.

More broadly, our study contributes to research on how legal and institutional barriers shape women's labor market participation. Recent literature highlights the role of discriminatory laws in limiting women's economic opportunities globally, demonstrating that legal reforms promoting gender equality can significantly boost female employment rates (Roy, 2019; Hyland et al., 2020). Studies of specific institutional reforms also show that removing discriminatory constraints, such as restrictions on women's work hours or barriers in family law, leads to increased female participation in the labor market (Hallward-Driemeier and Gajigo, 2015; Gonzales et al., 2015). More recently, Abou Daher et al. (2025) conduct a field experiment evaluating the effects of lifting Saudi Arabia's ban on women driving and find that treated women are significantly more likely to be employed, though are less able to make purchases without family permission. By framing the headscarf ban as a gender-

[6] Their analysis also examines the immediate aftermath of the 2013 repeal, using veiling information from surveys conducted within two months of the reform. Our study instead focuses on women's labor market responses in the years following the repeal.

discriminatory policy, our study extends this literature, demonstrating how removing legal constraints related to religious identity affects pious women's labor market participation.

Finally, our study contributes to the literature on traditional gender norms as barriers to female employment, particularly in developing countries. Jayachandran (2021) emphasizes that restrictive norms concerning women's mobility, household responsibilities, and community expectations significantly reduce female labor force participation. Prior studies have shown limited impacts of economic empowerment policies, such as vocational training or increased control over earnings, due to deeply entrenched gender norms (Field et al., 2010; Duflo, 2012; Field et al., 2021; Gazeaud et al., 2023). Although veiled women in Turkey generally adhere to conservative gender norms, which are typically associated with lower employment rates (Dildar, 2015), we find that removing the institutional barrier to veiling significantly increased their employment, especially within the public sector. Our findings thus suggest that institutional constraints may substantially outweigh traditional norms in determining women's economic choices, challenging assumptions that conservative norms alone drive low labor force participation among veiled women.

The remainder of the paper is structured as follows: Section 2 discusses the headscarf debate and public sector employment in Turkey. Section 3 presents the conceptual framework to understand how the removal of the ban affects women's labor market outcomes. Section 4 describes the data, and Section 5 presents the empirical strategy. Section 6 presents the main findings and robustness checks. Section 7 concludes.

2 Institutional Background

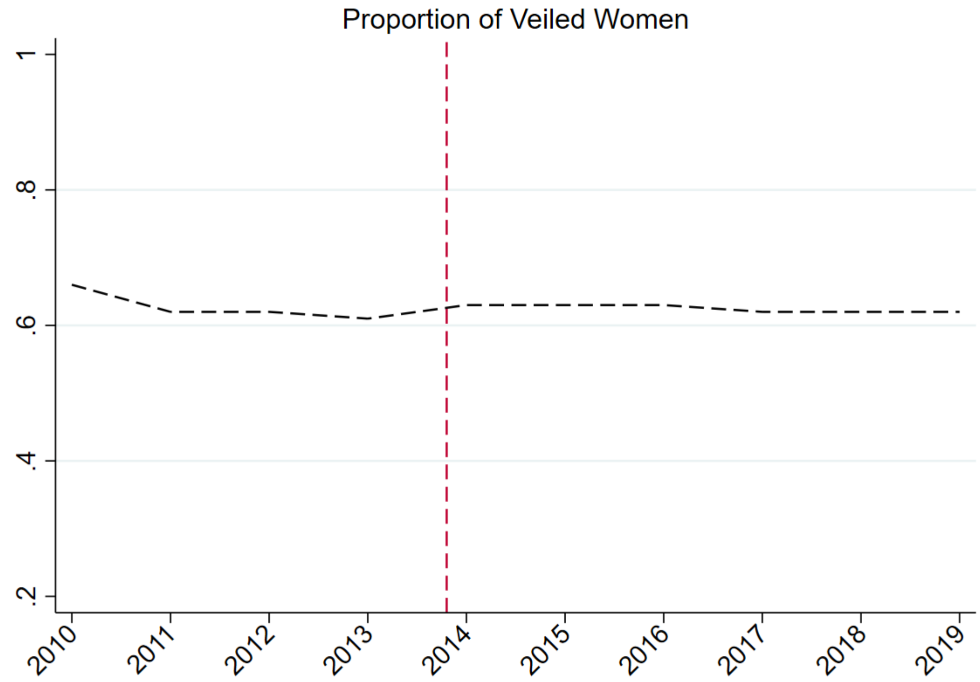
2.1 Historical and social context of veiling in Turkey

Turkey, constitutionally secular since 1924, has a predominantly Muslim population. The wearing of headscarves in public institutions has long been a contentious issue, contributing to tensions between secularism and religious practices. The historical evolution of headscarf bans in Turkey is examined in detail by Cindoglu and Zencirci (2008). While precise panel data on the prevalence of veiling are limited, multiple sources provide estimates from different time periods. According to the report by KONDA (2019), 97% of Turkey's population identifies as Muslim. Figure 1 illustrates trends in the proportion of women wearing headscarves in Turkey from 2010 to 2019, based on data from these sources. The percentage of veiled women was approximately 66% in 2010, decreasing to 62% in 2011. This figure remained stable until 2013, after which a slight increase of approximately 2 percentage points occurred. Similarly, data from the 2008 and 2013 waves of the Turkish Demographic and Health Survey (DHS) show comparable patterns: 72% of ever-married women reported wearing headscarves in 2008, while 68% of all women did so in 2013. Because the 2008 survey covers only ever-married women, the two figures are not strictly comparable; for ever-married women in the 2013 DHS, the rate is about 74.5%.

Following the 1980 military coup, the first formal headscarf ban was enacted in 1982, which prohibited headscarves in public spaces. This regulation sought to reinforce secularism, a foundational principle of the Turkish Republic. However, the ban sparked significant societal discord, especially among conservative groups who viewed it as an infringement on religious identity and emblematic of secularist authoritarianism. In 1997, the government extended the ban from universities to preparatory schools (Uğur, 2020).

In 2008, the Turkish government made a major attempt to lift the longstanding headscarf ban in universities. This initiative aimed to lift the ban in universities by amending Articles

Figure 1: Share of veiled women in Turkey, 2010–2019



Notes: Data are from KONDA survey data (KONDA, 2019).

10 and 42 of the Turkish Constitution. However, the Constitutional Court intervened, ruling on June 5, 2008, that the proposed amendments violated the preamble of the Constitution and the principle of secularism enshrined within it. The Court subsequently annulled the proposed changes, effectively blocking their implementation (Wiltse, 2008; Höjelid, 2010). Despite this setback, the Higher Education Council partially lifted the headscarf ban in 2010 by revising regulations within its jurisdiction. These regulatory changes marked a gradual easing of restrictions on veiling in Turkish universities (Bianet, 2010; BBC News, 2010).[7]

A landmark change occurred on September 30, 2013, when the government, under Prime Minister Erdoğan, announced a broader "democratization package", parts of which were subsequently implemented through legislation and executive decrees. The package introduced a set of legal and regulatory changes in several domains, including political party regulation and elections, minority-language and cultural rights, and freedom of assembly. Among other measures, it allowed electoral campaigning in languages and dialects other than Turkish, permitted political parties to adopt a co-chair system, reduced organizational requirements for party organization, lowered the vote threshold for state financial support to political parties from 7% to 3%, allowed private schools to provide education in languages and dialects traditionally used by Turkish citizens, removed criminal sanctions related to the use of the letters Q, X, and W, and revised rules governing the location, timing, and supervision of rallies and demonstrations (European Commission, 2013; Arsu and Bilefsky, 2013; Karahan and Tuğsuz, 2022).[8]

Within this broader package, the policy change most directly relevant for our setting was the repeal of the longstanding headscarf ban for most public employees. This measure came into effect on October 8, 2013, repealing the 1982 by-law that had prohibited the wearing of headscarves in public institutions.[9] Prior to this policy change, veiled women were formally

[7]We discuss in the following section how these regulatory changes may affect our empirical identification.

[8]Many of the non-headscarf components described above were later implemented through Law No. 6529, which entered into force on March 2, 2014, as recorded in the *Resmî Gazete*, the Turkish Official Gazette in which laws and regulations are formally promulgated and enter into force.

[9]The legal change entered into force through the *Resmî Gazete* under the "Regulation on Amendments to the Regulation on the Dress and Attire of Personnel Working in Public Institutions and Organizations" (No. 28789, Council of Ministers Decision No. 2013/5443), adopted on October 4, 2013 and effective on

prohibited from working in public institutions while wearing a headscarf. Although a small number of veiled women managed to circumvent the ban by wearing wigs in the workplace, this practice remained limited and was not officially recognized by authorities (Cindoğlu, 2011; Guveli, 2011). With the reform, women were legally permitted to wear headscarves in public institutions and government offices, with exceptions for personnel in the armed forces, security forces, and judiciary (Akoglu, 2015; Karahan and Tuğsuz, 2022). The ban was subsequently lifted for members of the judiciary in 2015 and for security forces in 2016.

This broader policy context is important for the interpretation of our results. While the package included several other political and civil-rights measures, these did not directly alter public sector hiring rules, recruitment procedures, or occupational access for women. By contrast, the repeal of the dress-code restriction directly changed the conditions under which veiled women could work in public institutions. Some non-headscarf elements of the package, especially those related to minority-language rights, may have influenced the broader political environment and, through that channel, labor-market behavior more generally. However, these measures did not directly relax the employment restriction faced by veiled women in public institutions. Our empirical analysis therefore focuses on the repeal of the headscarf ban as the component of the package most plausibly linked to differential changes in women's public sector employment. As a robustness check, we examine whether the results are disproportionately driven by regions in which these other components of the package may have been more salient, particularly regions with larger Kurdish populations.

While the repeal of the headscarf ban was part of the Justice and Development Party (AKP)'s broader political agenda and its general direction was therefore predictable (see Corekcioglu, 2021), the precise timing of its implementation was plausibly exogenous to women's labor market choices. Access to public sector employment typically requires passing competitive national examinations, which limits the ability of potential workers to adjust their behavior in anticipation of the policy change. Consistent with this, we show in the next section that pre-repeal trends in public sector employment were similar between treatment and control groups. We also demonstrate that our results are not driven by municipalities governed by AKP mayors, which supports our identification strategy.

2.2 Public sector employment

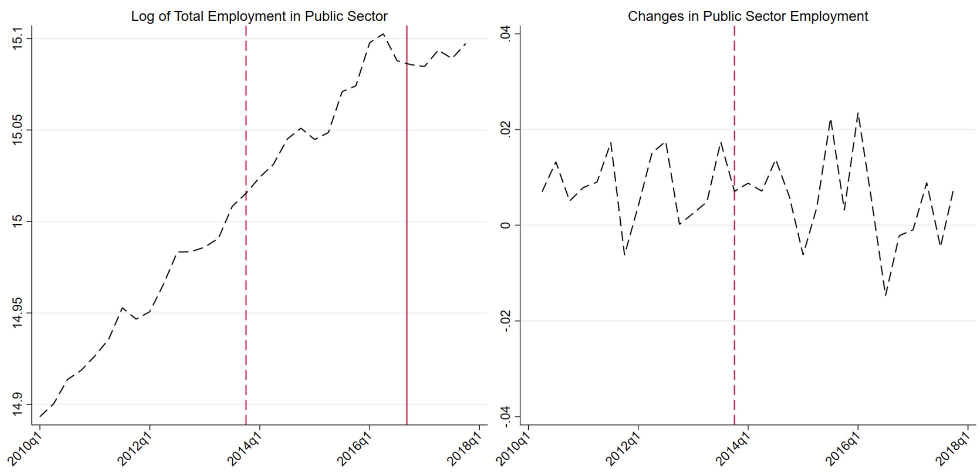
As of 2023, public sector employment in Turkey accounted for 13% of total employment, below the OECD average of 18%, indicating a relatively smaller public sector presence in the labor market (OECD, 2025). Figure 2 presents trends in total public sector employment in Turkey between 2010 and 2018, based on quarterly data from the Turkish Presidency of Strategy and Budget. The left panel of Figure 2 shows the logarithmic growth in public sector employment, which increased steadily between 2010 and 2016. The upward trend paused around mid-2016, coinciding with the July military coup attempt and a temporary hiring freeze that followed, as the government implemented extensive security reviews and purges (BBC News, 2017; Euronews, 2017). Public sector employment followed a moderate upward trajectory after this period. The right-hand graph of Figure 2 shows that the labor demand side of the public sector remained stable from 2010 to 2017, with no major disruptions apart from the temporary hiring freeze in 2016.

To understand whether the reform was anticipated, it is important to examine the centralized nature of recruitment and exam participation trends. Public sector recruitment in Turkey is centralized and highly standardized. Applicants must pass a biennial civil ser-

October 8, 2013.

vice examination administered by the Department of Measuring, Selection, and Placement (OSYM). The exam targets different candidate groups in alternating years: odd-year exams primarily recruit for positions such as teachers and district governors, while even-year exams cover a broader range of occupations, including engineers, technicians, and clerks. Figure 3 shows the share of individuals taking the exam during even years between 2008 and 2018. The absence of a noticeable pre-reform increase in participation supports the assumption that the repeal of the headscarf ban was largely unanticipated, though exam participation remains only an indirect proxy for expectations.

Figure 2: Employment in the public sector as a percentage of total employment



Notes: Data are from the Turkish Presidency of Strategy and Budget.

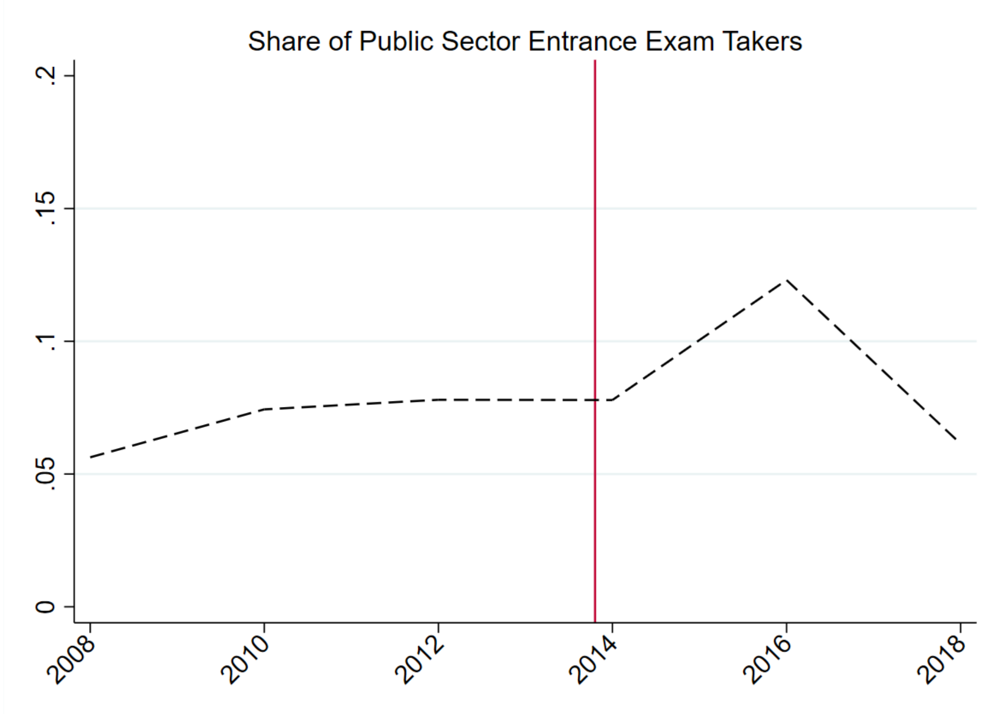
Public sector positions are predominantly professional roles, such as teaching, nursing, and engineering, that require advanced qualifications. Women hold about 57% of professional public sector positions, compared with 41% among men. Clerical positions, requiring lower educational credentials, employ 17% of women and 15% of men. Technical roles, involving specialized tasks, are the third most common for both genders. In contrast, labor-intensive occupations, including service workers, agricultural and fishery workers, craft and trade workers, and machine operators, are male-dominated, accounting for 28% of male employment versus 9% of female employment. Positions as legislators and senior officials remain uncommon, comprising 4% of male and 3% of female employment. Detailed gender breakdowns across public sector occupations are provided in Table A1 in Online Appendix.

3 Conceptual Framework

Following Carvalho (2013), we provide a conceptual framework to better understand the effects of the reform on public sector employment (see Appendix A.2 for the full setup and discussion).

Prior to the reform, veiled women were barred from public sector employment, as entry required unveiling — a condition many found prohibitively costly. Public sector jobs may provide a better working conditions but expose female employees to environments that may conflict with religious observance. Veiling serves as a commitment device that shields religious women from religiously prohibited behavior, reducing both intrinsic regret and social disapproval from the religious community, with the latter increasing in community religiosity. Removing the ban eliminates the unveiling requirement, thereby lowering the cost of public sector entry for veiled women by reducing the intrinsic regret and social disapproval they would otherwise incur.

Figure 3: Proportion of individuals taking the public sector entrance exam



Notes: Data are from the Department of Measuring, Selection, and Placement (OSYM).

Taking veiling as predetermined,[10] among veiled women who meet the educational qualification, the key distinction is between those whose labor market decisions respond to the reform and those who would not enter the public sector regardless. The former do not work in the public sector under the ban but enter upon its removal; the latter remain outside the public sector irrespective of the policy. The magnitude of the reform's effect depends on the relative size of these two groups, which varies across regions with the educational qualifications and religiosity of the local community. This framework delivers three testable predictions.

Prediction 1. Lifting the ban increases public sector employment among veiled women who meet the educational qualifications for such positions, as ban removal lowers the cost of entry for those on the margin.

Prediction 2. The effect is stronger among more educated veiled women, since less educated women are ineligible for public sector positions regardless of the ban and thus do not respond to the reform.

Prediction 3. The effect is stronger in lower-religiosity regions, where the share of educationally qualified women in the responsive window is larger, since education and religiosity are negatively correlated.

We bring these predictions to the data in the sections that follow.

4 Data

Turkish Household Labor Force Survey. The main dataset used in this study is the Turkish Household Labor Force Survey (HLFS), an annual nationally representative survey conducted by the Turkish Statistical Institute. The HLFS covers around 150,000 households per wave and provides detailed background and labor market information for each household

[10]We discuss in the full model setup the case where veiling is endogenous.

member aged 16 and above. For the analysis, we use repeated cross-sections spanning 2010 to 2017.[11]

The HLFS includes a rich set of variables such as individual characteristics (e.g., age, marital status, education, household size, and region of residence) and detailed labor market information, containing employment type and status (e.g., public or private sector, self-employment or unpaid family work), occupation, weekly hours worked, and monthly earnings. Earnings data incorporate bonuses and premiums received during the reference month.[12]

The analysis focuses on women aged 18–49, who are most likely to participate in the labor market. Because the HLFS reports earnings only for wage earners, analyses of earnings are restricted to this subgroup. In total, the analytical sample comprises 933,810 women. Column 1 of Online Appendix Table A2 presents descriptive statistics for this sample. Panel A shows that the average age is approximately 33 years, with 98% of respondents being native-born. About 35% have completed at least high school, while 15% hold a university degree. Roughly half have only a primary school degree (five years of education) or less. Around 71% are married, and the average household size is four persons. Panel B reports labor-market outcomes. About 33% of women are employed, primarily as wage earners (22%), including 5% in the public sector and 17% in the private sector. In addition, 8% work as unpaid family workers, and 3% are self-employed.[13] The remaining columns of Online Appendix Table A2 present summary statistics separately for women predicted to be veiled and those predicted not to be. In the following section, we discuss the comparability of these samples with those in the DHS, which includes direct information on headscarf use.

Turkish Demographic and Health Survey. While the HLFS provides rich labor market data, the 2013 Turkish Demographic and Health Survey (DHS) uniquely includes information on headscarf use, enabling us to distinguish between veiled and non-veiled women.[14] The 2013 DHS, conducted by Hacettepe University's Institute of Population Studies, is nationally representative and includes 9,746 women, 6,835 of whom were married at the time of the survey (administered between September and December 2013). In addition to headscarf-wearing and other religious practices (e.g., prayer and fasting), the survey contains detailed background and socioeconomic characteristics.

In the 2013 DHS sample, 68% of women aged 18–49 reported wearing a headscarf.[15] Online Appendix Table A3 compares non-veiled (column 1) and veiled (column 2) women. Non-veiled women are significantly more likely to have completed high school or university, whereas veiled women are concentrated among those with primary education or less. Among non-veiled women, 39% hold a university degree compared with only 6% of veiled women, while 65% of veiled women have primary education or less, compared with 20% of non-veiled women. These differences indicate strong educational sorting by veiling status, with veiled women being substantially less educated on average.

[11] Data from 2018 onward are excluded to avoid confounding effects from the prolonged political and economic disruptions that followed the July 2016 military coup attempt, including large-scale public sector dismissals. The choice of 2010 as a starting point is motivated by the constitutional referendum held in September 2010, which marked a shift in Turkey's political and institutional trajectory. While the referendum did not directly alter headscarf regulations, it could have shaped expectations about future reforms. By starting in 2010, we ensure that our pre-reform period is defined consistently within the new institutional environment. A remaining concern is that the referendum may have affected treatment and control groups differentially. To address this, we allow for group-specific linear trends in our estimates and show that our main results are robust to this more flexible specification.

[12] Monthly wages are deflated using the Consumer Price Index (CPI) and converted to 2017 Turkish Lira, then expressed in U.S. dollars using 2017 exchange rates.

[13] The self-employment category includes women classified as self-employed or own-account workers.

[14] To our knowledge, only the 2008 and 2013 waves of the Turkish DHS collect information on veiling. The 2008 wave covers only ever-married women, whereas the 2013 DHS includes all women aged 15–49.

[15] This figure reflects the unweighted mean; the weighted mean corresponds to 62%.

Panel B reports summary statistics for labor market outcomes. Compared with the pronounced education gap between the two groups, the difference in employment probabilities is smaller, around 16 percentage points, but still statistically and economically meaningful, with non-veiled women much more likely to be employed. In addition, only about 8% of veiled women work as wage earners, compared with a substantially higher share among non-veiled women. Public sector employment is particularly limited: only about 1% of veiled women hold public sector jobs, compared with roughly 10% of their non-veiled peers, consistent with the restrictions in place before 2013.[16]

5 Empirical strategy

We estimate the effect of lifting the headscarf ban in public institutions on women's labor market outcomes by exploiting both the timing of the policy change and variation in exposure between veiled and non-veiled women within a difference-in-differences (DiD) framework. Using out-of-sample prediction and statistical matching models based on a second dataset (described below), we compare changes in labor market outcomes for women predicted to be veiled with those predicted not to veil, before and after the repeal of the headscarf ban. Formally, we estimate the following equation:

$$y_{it} = \beta_1 + \beta_2 \left(Veiled_i \times Post_t\right) + \beta_3 Veiled_i + X_{it}'\gamma + \mu_t + \varepsilon_{it}. \tag{1}$$

where y_{it} represents the outcome of interest for woman i at time t. The indicator $Post_t$ equals one for the post-policy period ($t \geq 2014$) and zero otherwise.[17] X_{it} includes region dummies, age and age squared, household size, and an indicator for being native. The specification also controls for year and region–year fixed effects, which absorb all observable and unobservable shocks common to women within the same region and year.

In a standard DiD framework, the interaction term $Veiled_i \times Post_t$ captures the average treatment effect on the treated. In our context, however, $Veiled_i$ is a predicted rather than an observed indicator, so β_2 represents an intent-to-treat (ITT) effect, capturing the average impact of the reform on women predicted to be veiled relative to those predicted not to be. We therefore base inference on a two-stage bootstrap procedure that accounts for uncertainty in both the imputation of veiling status and the second-stage treatment-effect estimation.[18]

Defining treatment and control groups. Our empirical strategy defines treatment and comparison groups based on women's exposure to the removal of the headscarf ban. If veiling status were directly observable in the main dataset, we could estimate the average treatment effect by comparing veiled to non-veiled women. However, such information is not available in large-scale surveys. Because the HLFS does not record veiling, we impute veiling status using an out-of-sample prediction and statistical matching model based on the framework of Rosenbaum and Rubin (1983), estimated with data from the 2013 Turkish Demographic and Health Survey (DHS), which provides nationally representative information

[16] The presence of veiled women in public sector employment before the reform likely reflects cases where women circumvented the ban by wearing wigs in the workplace, as discussed in Section 2.

[17] The headscarf ban in public institutions was lifted in October 2013. Because our data are annual and the survey month is not reported, we designate 2014 as the start of the post-treatment period in the main analysis. As a robustness check, we also re-estimate the specifications treating 2013 as the first post-policy year.

[18] We use 100 bootstrap replications. In each replication, we resample the DHS data, re-estimate the first-stage prediction model, re-impute veiling status in the HLFS using the corresponding prediction and matching procedure, and re-estimate the second-stage difference-in-differences specification. Confidence intervals are constructed from the empirical distribution of the bootstrap estimates. This procedure is applied analogously for the baseline Probit-based imputation and for the alternative machine-learning-based imputations reported in the robustness checks.

on veiling.

The estimation proceeds in two stages. In the first stage, using data from the 2013 DHS,[19] we estimate the probability of wearing a headscarf by fitting a Probit model. The dependent variable equals one if a woman wears a headscarf and zero otherwise. The regressors include education-level indicators, an indicator for being native, age, household size, and region dummies at the NUTS-2 level. Table 1 reports the Probit estimates and marginal effects of each covariate on the probability of veiling. Using these coefficients, we predict headscarf-wearing probabilities for both the DHS and HLFS samples. Although model fit could be improved with additional covariates, the HLFS lacks relevant information beyond these controls.[20]

Table 1: *Probit estimates of the probability of wearing a headscarf*

Variables	Coef. (1)	Std. Err. (2)	Marginal Eff. (3)
Age	-0.010***	0.003	-0.002***
Native	0.626***	0.135	0.158***
Household size	0.168***	0.017	0.042***
Education level:			
Primary school degree	-0.337***	0.097	-0.068***
Junior high school degree	-0.771***	0.111	-0.186***
High school degree	-1.253***	0.106	-0.347***
University degree	-2.010***	0.114	-0.600***
Obs.	7,172		
Pseudo R^2	0.279		

Notes: Data are from the 2013 Turkish Demographic and Health Survey. The sample includes all women aged 18-49. The estimates include dummies for each region at the NUTS2 level. The omitted category for education is having no formal education. ***, significant at the 1 percent level; **, significant at the 5 percent level; *, significant at the 10 percent level.

Figure 4 illustrates the distribution of predicted probabilities of veiling. The left panel displays the distribution based on the DHS sample, and the right panel shows the corresponding distribution for women in the HLFS. The substantial overlap between the two distributions indicates that each woman in the HLFS has comparable counterparts in the DHS based on observable characteristics.

In the second stage, we impute a binary veiling status, equal to one for veiled and zero for non-veiled women in the HLFS, by matching each HLFS observation to its nearest neighbors in the DHS sample based on predicted probabilities.[21] The matched HLFS sample closely mirrors the DHS in population-weighted averages: approximately 61% of women in the HLFS are predicted to be veiled (s.d. 0.48), nearly identical to the DHS figure of 62%.[22] Columns 2 and 3 of Online Appendix Table A2 present descriptive statistics by predicted

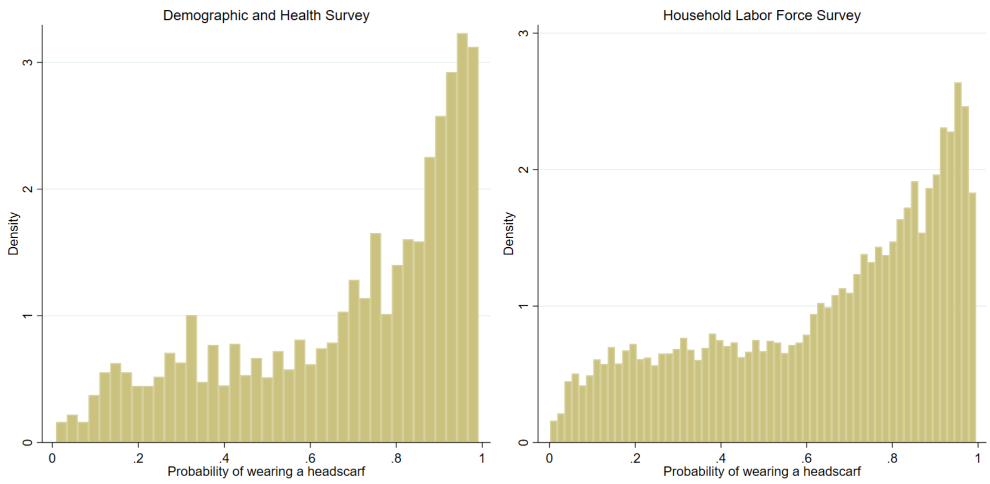
[19] In our main analysis, we estimate veiling probabilities using the 2013 DHS. Although we do not expect major changes in veiling patterns after the policy change, relying on the 2013 data could, in principle, introduce bias if the reform affected veiling choices. To address this concern, we re-estimate veiling probabilities using data from the 2008 DHS wave as a robustness check.

[20] Marital status is excluded to avoid potential endogeneity, as it is analyzed later as an outcome variable. Estimates that include marital status yield results consistent with our baseline.

[21] Matching is performed with replacement. Each DHS observation can be used as a match for multiple HLFS observations. We set the maximum allowed distance in predicted probabilities to 0.05; observations exceeding this threshold are dropped. The excluded observations total 1,416 (less than 1% of the estimation sample).

[22] These figures are based on survey sampling weights, which ensure national representativeness. In unweighted terms, the corresponding shares are 68 percent in the DHS and 65 % in the HLFS, reflecting differences in sampling composition rather than population structure.

Figure 4: Distribution of predicted probabilities of wearing a headscarf



Notes: The left-hand graph shows the distribution of predicted probabilities of veiling based on data from the DHS, while the right-hand graph presents the distribution using data from the HLFS.

veiling status. For comparison, Online Appendix Table A3 reports the same statistics for the DHS, where veiling is directly observed. The relative gaps between predicted veiled and non-veiled women in the HLFS closely resemble those between observed groups in the DHS, both in background characteristics and in labor market outcomes. This similarity indicates that the predicted veiling status in the HLFS successfully captures the same background differences observed between veiled and non-veiled women in the DHS, suggesting that our imputation model provides a reliable classification for the analysis.

We further examine whether our findings are sensitive to the specific prediction and imputation models used to predict veiling status. To this end, we implement a set of alternative out-of-sample prediction models based on machine-learning methods. As in the baseline approach, these models are trained in the DHS, where veiling status is observed, using the set of background characteristics common to the DHS and the HLFS, and are then applied to the HLFS to generate alternative predicted veiling measures. Our main machine-learning specification relies on an ensemble classifier, while we also consider individual Support Vector Machine, Random Forest, K-Nearest Neighbor, and Neural Network models as robustness checks. These methods allow for more flexible, nonlinear relationships between observable characteristics and veiling status than the baseline Probit specification.[23]

A limitation of both the baseline Probit-based matching procedure and the alternative machine-learning imputations is that they rely on predicted rather than observed veiling status. As a result, the constructed treatment indicator is subject to misclassification error: some truly veiled women may be assigned to the predicted non-veiled group, while some truly non-veiled women may be assigned to the predicted veiled group. When these errors simply mix truly veiled and truly non-veiled women across the two predicted groups, the resulting bias works toward attenuation. A truly non-veiled woman classified as predicted veiled enters the treatment group even though she was not directly constrained by the ban (false positive), reducing the average post-reform gain among women classified as veiled. Conversely, a truly veiled woman classified as predicted non-veiled enters the comparison group (false negative) and may increase its post-reform average after the repeal. Both types of error therefore pull the post-reform averages of the two predicted groups toward each other, which would tend to attenuate the estimated treatment-control difference.

[23]Details on the algorithms and parameter choices are provided in Appendix A.3, and the corresponding treatment-effect estimates are reported in the robustness section.

To evaluate the extent of such misclassification, we reserve 20% of the DHS, where veiling status is directly observed, as a hold-out sample. We first estimate each prediction model on the remaining 80% of the DHS sample and then evaluate classification accuracy in the hold-out sample. The baseline Probit model performs well, with a ROC area (AUC) of 0.847 (s.e. 0.010; 95% CI [0.828, 0.867]), meaning that a randomly chosen veiled woman is correctly ranked above a randomly chosen non-veiled woman about 85% of the time. We choose the probability cutoff by maximizing Youden's J statistic, which balances correct classification of truly veiled and truly non-veiled women.[24] At this cutoff, the model correctly classifies 83% of truly veiled women and 71% of truly non-veiled women in the hold-out sample. Table 2 reports the corresponding false-negative, false-positive, and total accuracy rates for the baseline Probit model and the alternative machine-learning classifiers.

Table 2: *Misclassification rates of imputation methods*

	Probit	Ensemble	Support Vector Machine	Random Forest	K-Nearest Neighbor	Neural Network
False negative rate	16.98%	11.40%	18.07%	12.50%	13.43%	13.01%
False positive rate	29.06%	43.58%	37.33%	39.76%	40.97%	41.49%
Total accuracy rate	79%	78%	76%	79%	78%	78%

Notes: The table reports misclassification rates of each imputation method evaluated on the DHS test set, which comprises 20% of the DHS sample held out from training. The false negative rate is the share of truly veiled women predicted as unveiled. The false positive rate is the share of truly non-veiled women predicted as veiled. Total accuracy rate is the share of correctly classified observations.

The validation exercise shows that all imputation methods generate some classification error, but the predicted measures contain substantial information about true veiling status. For the baseline Probit specification, the false-negative rate is 16.98% and the false-positive rate is 29.06%, implying the model distinguishes true veiling status substantially better than chance. The machine-learning methods generate somewhat different error profiles: false-negative rates range from 11.40% to 18.07%, while false-positive rates range from 37.33% to 43.58%, with total accuracy rates between 76% and 79%. This variation across methods provides a useful robustness check: if results are sensitive to which specific women are misclassified, estimates should differ substantially across imputation approaches.

While the simple misclassification described above would tend to attenuate the estimated effect, this attenuation logic need not hold if classification errors are systematically related to unobserved determinants of labor-market outcomes. Misclassification in DiD settings may generate bias beyond simple attenuation when the misclassified group differs in counterfactual outcome trends or treatment-effect heterogeneity; in such cases, the DiD estimand may even have the wrong sign (Denteh and Kédagni, 2022; Negi and Negi, 2025). In our context, classification errors could be differential because the prediction model relies on observable demographics but cannot capture all unobserved determinants of both veiling behavior and labor-market participation. Truly veiled women who are harder to classify as veiled based on observables may differ in unobserved religiosity or gender norms. If these women are less constrained by religiosity or traditional norms, they may also be more responsive to the repeal. More generally, any unobserved characteristic that simultaneously predicts veiling and labor-market behavior could generate differential misclassification through the same mechanism.

However, for our DiD estimates to reflect a sign reversal, the true average treatment effect (ATT) would have to be negative, meaning that the repeal of the headscarf ban reduced employment among truly veiled women. This is difficult to reconcile with the institutional setting, as discussed in Section A.2. The reform expanded the set of legal employment options

[24] The resulting threshold is 0.63.

available to veiled women by removing institutional barriers to public sector employment. Women who preferred their existing employment arrangements could remain in them, while those previously excluded from public sector jobs could now enter them. A negative true ATT would therefore require a mechanism through which removing a legal barrier actively harmed the employment prospects of the women it directly affected.

To further characterize this concern, we examine the observed characteristics of misclassified women in the DHS hold-out sample. The DHS contains direct information on veiling status, as well as measures of religiosity and gender-role attitudes that are not available in the HLFS, allowing us to assess whether prediction errors are systematically related to the dimensions through which differential misclassification could affect our estimates. Appendix Table A4 compares false negatives with correctly classified veiled women and false positives with correctly classified non-veiled women across all imputation methods. For the baseline Probit model, false negatives among truly veiled women are substantially more educated than correctly classified veiled women: while 19.4% of correctly classified veiled women have no formal education, none of the false negatives fall in this category, and 79.3% of false negatives have completed at least high school compared with only 5.7% of correctly classified veiled women. Notably, false negatives do not differ significantly in religious practice, the share praying regularly is 89.4% among false negatives versus 87.5% among correctly classified veiled women ($p = 0.455$), but are significantly less traditional in gender-role attitudes (53.8% versus 61.6%, $p = 0.043$). This pattern suggests that misclassified truly veiled women face lower norm-based constraints on labor market participation rather than lower religiosity per se, and are therefore more likely to respond positively to the reform. False positives, truly non-veiled women incorrectly classified as veiled, are less educated and hold more traditional gender attitudes than correctly classified non-veiled women, suggesting they face stronger norm-based constraints on labor market participation and are therefore unlikely to generate strong post-reform employment gains within the predicted-veiled group. Their presence in the treatment group therefore would tend to attenuate rather than inflate our estimates. This pattern is consistent across all imputation methods in Online Table A4, suggesting it reflects a systematic feature of veiling prediction from observables rather than an artifact of any particular method.

Taken together, these validation exercises suggest that misclassification is unlikely to mechanically generate the positive public sector employment effects we estimate. The predicted veiling measures are informative, the main misclassification patterns are consistent with attenuation, and the characteristics of misclassified women suggest that the most plausible differential-misclassification channel places more labor-market-oriented truly veiled women in the comparison group, compressing rather than amplifying the treatment-control gap. We therefore interpret $\hat{\beta}_2$ as an intent-to-treat effect for women predicted to be veiled, while recognizing treatment misclassification as an important limitation of the empirical design.

Two additional checks further support this interpretation. First, we restrict the estimation sample to women with high-confidence predictions, defined as those with predicted veiling probabilities above 0.80 or below 0.20, based on the fitted values from the first-stage Probit imputation model used to assign veiling status. In this subset, classification uncertainty is lowest because observations far from the classification threshold are less likely to be incorrectly assigned to the predicted veiled or predicted non-veiled group. If the estimated effects were primarily driven by women with ambiguous predicted veiling status, excluding these observations should substantially alter the estimates. Online Appendix Table A11 reports this sensitivity exercise.[25] Second, if differential misclassification through unobserved

[25] Online Appendix Table A11 reports estimates for the full matched sample (Panel A) and for increasingly

factors such as religiosity or gender norms were generating a spurious positive effect, we would expect effects to be concentrated in high-religiosity regions, where differences between more and less religious veiled women are likely to be larger. Our heterogeneity analysis, discussed in the following section, does not support this pattern.

5.1 Identifying assumptions

The credibility of our empirical strategy rests on two key conditions: first, that our out-of-sample prediction and matching procedure reliably captures true differences between veiled and non-veiled women; and second, that, absent the reform, labor market outcomes for these groups would have followed parallel trends. We assessed the validity of the first condition in the preceding subsection, showing that predicted veiling status provides a close approximation of actual veiling and enables meaningful comparisons between treatment and control groups. The second condition, the standard identifying assumption in difference-in-differences designs, requires that no other time-varying shocks differentially affected the two groups before the policy change. We examine this assumption using an event-study specification that interacts the veiling indicator with relative-year dummies, taking the last full pre-reform year (2012) as the reference period and controlling for covariates.

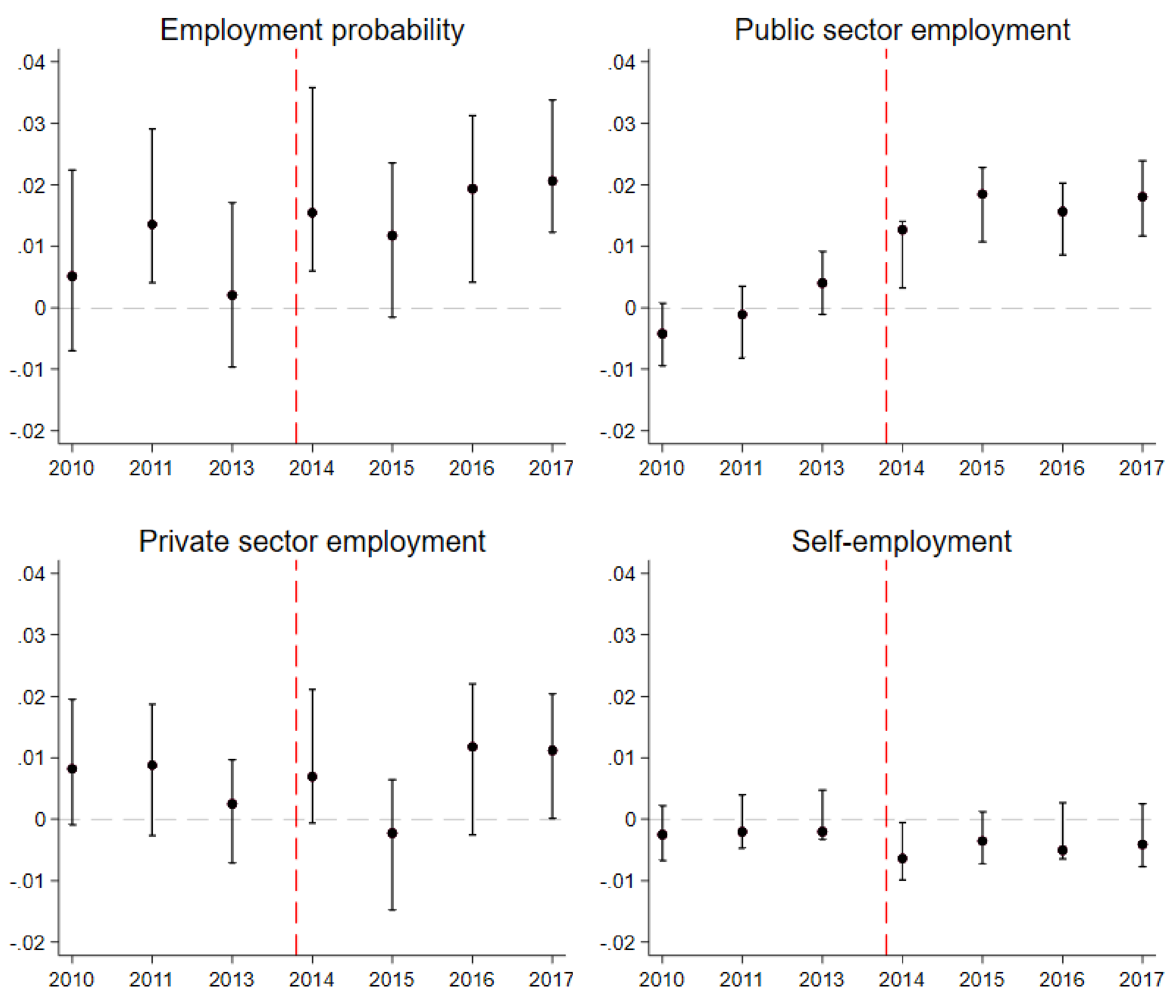
The coefficients plotted in Figure 5 show the annual differences between women predicted to be veiled and those predicted not to be veiled in the probability of being employed and of working in the public sector, private sector, or self-employment, each measured relative to the 2012 gap, with 95 percent confidence intervals. For public and private sector employment, as well as self-employment, the estimates display no statistically significant differences in the pre-reform years before the transition year, supporting the parallel-trends assumption. Consistently, bootstrap Wald tests of the 2010 and 2011 interaction terms yield $p = 0.102$, $p = 0.277$, $p = 0.253$, and $p = 0.483$ for overall employment, public sector employment, private sector employment, and self-employment, respectively. The test for overall employment is marginally significant, indicating a modest upward movement between 2011 and 2012, while the remaining outcomes show no statistically significant pre-reform differences, supporting the parallel-trends assumption. Given that the reform was implemented in October 2013 and our data are annual, this pattern is not unexpected and likely reflects that 2013 is a transition year rather than a clean pre-treatment year. Some women predicted to be veiled may have begun adjusting their employment behavior soon after the policy change, even though 2013 is coded as pre-treatment in our baseline specification.

We address these issues in three ways. First, to account for the transition-year concern directly, we extend Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, thereby allowing for a differential change in outcomes for women predicted to be veiled in the transition year. This specification accounts for the possibility that 2013 captures differential transition-year adjustments for women predicted to be veiled, rather than treating this variation as part of the pre-reform trend. Throughout the analysis, we report this transition-year-adjusted specification alongside the baseline estimates. Second, in the following section, we further examine the sensitivity of our findings to deviations from the parallel-trends assumption using the framework proposed by Rambachan and Roth (2023). Third, for the main labor market outcomes, we also report a trend-adjusted version of the baseline specification that allows the treatment group to follow its own linear time trend

strict high-confidence subsamples (Panels B-D). The estimates are generally similar across thresholds. Under the strictest threshold, based on predicted probabilities above 0.80 or below 0.20, some estimates become smaller and less precise, which is expected because the sample size falls substantially when observations with more ambiguous predicted veiling status are excluded. Overall, the table provides no evidence that the main results are driven primarily by women near the classification boundary.

by interacting the veiled indicator with a continuous year variable. In this specification, the post-reform coefficient β_2 captures the change in the veiled–non-veiled gap in the post period net of the veiled group's own linear trend. Because adjustment to the reform may unfold gradually over time, especially for labor market outcomes such as public sector employment, this trend-adjusted specification may absorb part of the treatment effect. We therefore interpret it as a stringent sensitivity check rather than our preferred specification.

Figure 5: Trends in labor market outcomes between treatment and control groups



Notes: Data are from the 2010–2017 HLFS. The figure reports event-study coefficient estimates from the matched HLFS sample, with 2012 omitted as the reference year. The controls include whether the woman is native, dummies for each education level, age, age squared, household size, and region and year fixed effects, along with region-specific year effects. Confidence intervals are 95% bootstrap confidence intervals obtained from the full two-step bootstrap procedure, which resamples both the DHS imputation sample and the HLFS analysis sample and re-estimates the matching and second-stage regression in each replication. Data are weighted using the cross-sectional weights for the wave in which the outcome is measured.

In addition to examining pre-reform trends, it is important to account for other policy changes that may have independently influenced women's labor market outcomes. As discussed in Section 2, the gradual lifting of the headscarf ban in universities beginning in 2010 may have affected women's educational trajectories, particularly by shaping decisions to pursue higher education. Moreover, Turkey implemented two major compulsory schooling reforms known to influence women's education and labor market outcomes. The 1997 reform required individuals born after 1986 to complete junior high school (eight years of schooling), implying that the oldest affected individuals were 30 years old in our sample. A later reform in 2012 extended compulsory education to high school (12 years), but the oldest affected individuals were only 19 by 2017, so only a small share of our sample was exposed. To mitigate confounding from these reforms, we include cohort-based exposure indicators that flag (i) eligibility for the gradual lifting of the university headscarf ban beginning in 2010 and (ii) exposure to the 1997 compulsory-schooling reform as robustness checks.

Our identification also assumes that veiling is a relatively stable individual trait and exogenous to the repeal of the headscarf ban over our study period. Sociological and anthropological research emphasizes that veiling is not a transient behavioral choice but a deeply embedded moral and social practice, expressing enduring values of religious identity (e.g., Göle, 1996; Alvi, 2013). These values are cultivated through long-term socialization within families and communities and tend to change only slowly, if at all, in response to short-term political or institutional shifts (e.g., Fleischmann and Phalet, 2012; Drouhot, 2021). While recent work highlights that veiling may also serve instrumental or political purposes (e.g., Shofia, 2022; Aksoy and Gambetta, 2021), such adaptations typically emerge gradually as new social or political equilibria form. We further discuss this assumption in the following section, providing evidence on regional stability in veiling and on how other policy changes that might influence veiling decisions affect our results. Even if the 2013 repeal encouraged some women to begin veiling, whether as an expression of identity or for strategic reasons, such switching would blur treatment and comparison groups and attenuate the DiD contrast, making our estimates conservative lower bounds.

A remaining concern is potential selection on unobservables, particularly religiosity. Because religiosity is positively correlated with veiling and negatively correlated with labor market participation, any omitted-variable bias is likely to attenuate our estimates toward zero. We discuss this issue in detail in the following section, where we present additional analyses designed to account for differences in unobserved religiosity among treated women.

6 Main Results

6.1 Effects on Labor Market Outcomes of Women

We begin by presenting the intent-to-treat (ITT) estimates of the effect of lifting the headscarf ban on women's labor market outcomes. Table 3 reports estimates for three specifications. Panel A presents the baseline difference-in-differences estimates from Eq. (1). Panel B augments the baseline model by including an indicator for the year 2013 and its interaction with the predicted veiling indicator, allowing women predicted to be veiled to experience a differential change in the transition year.[26] Panel C reports the trend-adjusted specification, which allows for a group-specific linear trend by interacting the predicted veiling indicator with a centered year trend, as discussed in Section 5.1. Our focus is on the interaction term $Veiled \times Post$, which captures the relative change in outcomes for women predicted to be veiled compared to those predicted not to veil after the reform.

Panel A shows that the removal of the headscarf ban increased the employment probability of women predicted to be veiled relative to their non-veiled peers. The coefficient in column 1 implies a 1.2 percentage-point increase, with a bootstrap percentile 95 percent confidence interval of [0.005, 0.019]. This corresponds to an increase of roughly 4–5 percent relative to the pre-policy mean for veiled women. Columns 2 and 3 decompose wage employment into public and private sector components. The strongest response is observed in public sector employment, where the ban applied most directly: the probability that women

[26] This specification directly addresses the concern that 2013 may not be a clean pre-treatment year, given that the reform was implemented in October 2013.

predicted to be veiled worked in the public sector increased by 1.7 percentage points relative to their non-veiled peers. This estimate is precisely estimated, with a bootstrap percentile 95 percent confidence interval of [0.011, 0.018]. Given that the pre-reform mean among veiled women was only about 1.7 percent, this effect is substantial, roughly doubling their pre-reform rate, as reported in the last row of Table 3. By contrast, the coefficient on private sector employment is small and statistically insignificant, suggesting that the reform's effects were concentrated in the public sector rather than the private sector. This significant increase in public sector employment is consistent with Prediction 3 of the conceptual framework: lifting the ban increases public sector employment among women predicted to be veiled because it removes the institutional constraint that had directly excluded them from public sector jobs.

Columns 4 and 5 show that the increase in public sector employment reflected not only higher employment probability but was also accompanied by declines in self-employment and unpaid family work. In Panel A, self-employment falls by about 0.3 percentage points, marginally significant at the 10 percent level, though the bootstrap percentile 95 percent confidence interval includes zero. Unpaid family work declines by about 1.0 percentage point and is precisely estimated, with a bootstrap percentile 95 percent confidence interval of [-0.012, -0.007]. These patterns suggest that the reform not only increased employment among women predicted to be veiled but also changed the allocation of work among those already employed, with especially clear evidence of a shift away from unpaid family work and toward public sector jobs.

In Panel B, we explicitly account for the ambiguity of 2013 as a transition year. The estimates remain very similar to those in Panel A in both magnitude and statistical significance: the employment and public sector employment effects are nearly unchanged, the coefficient on private sector employment remains small and statistically insignificant, and unpaid family work continues to decline. Thus, allowing for differential transition-year adjustments leaves the main conclusions unchanged. Panel C reports the trend-adjusted specification, which allows women predicted to be veiled to follow their own linear time trend. The overall employment effect remains positive, although it is less precisely estimated than in Panels A and B. The public sector employment coefficient also remains positive but declines in magnitude and is no longer statistically distinguishable from zero under the bootstrap confidence interval, though the interval extends further into positive territory $[-0.004, 0.011]$, leaving the direction of the effect uncertain but leaning positive. This attenuation is expected, as a group-specific linear trend may absorb part of the treatment effect when labor market adjustment unfolds gradually after the reform. The coefficients for self-employment and unpaid family work remain negative, although the unpaid family work estimate becomes smaller and less precise.

Overall, the findings indicate that the repeal of the headscarf ban not only increased employment among women predicted to be veiled but also changed the sectoral allocation of their work. The sharp rise in public sector employment, coupled with declines in self-employment and unpaid family work, suggests that some veiled women who were already in the labor market reallocated from informal or unpaid jobs toward formal public positions once these became accessible.

To better understand how the removal of the headscarf ban affected women's labor market outcomes, we next examine weekly working hours, where non-employed women are coded as working zero hours. Column 1 of Table 4 shows that lifting the headscarf ban increased weekly working hours among women predicted to be veiled by about 0.5 hours in Panel A, corresponding to roughly 4–5 percent rise relative to the pre-policy mean for veiled women.

Table 3: *Effects of the headscarf ban removal on women's labor market outcomes*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A					
Veiled × Post	0.012***	0.017***	0.002	-0.003*	-0.010***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.005, 0.019]	[0.011, 0.018]	[-0.004, 0.009]	[-0.005, 0.001]	[-0.012, -0.007]
Veiled	-0.010***	-0.006***	-0.003	0.001	0.001
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel B					
Veiled × Post	0.011***	0.018***	0.001	-0.003	-0.011***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.003, 0.019]	[0.013, 0.020]	[-0.006, 0.009]	[-0.005, 0.001]	[-0.016, -0.008]
Veiled	-0.009**	-0.007**	-0.002	0.001	0.003*
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel C					
Veiled × Post	0.011**	0.009	0.002	-0.005**	-0.003
	(0.007)	(0.004)	(0.006)	(0.003)	(0.002)
	[0.001, 0.029]	[-0.004, 0.011]	[-0.004, 0.021]	[-0.009, -0.001]	[-0.007, 0.005]
Veiled	-0.010**	-0.009***	-0.003	-0.001	0.004*
	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)
Observations	933,810	933,810	933,810	933,810	933,810
Mean Dep. Var.	0.275	0.017	0.116	0.039	0.110

Notes: Data are from the 2010-2017 HLFS. The sample includes all women aged 18-49. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\}\times t$, where t is a centered year variable ($t = \text{year} - 2010$). Each estimate includes a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Last row reports the pre-policy mean of the dependent variable among women predicted to be veiled. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

The estimate remains positive and statistically significant in Panel B, where we allow for a differential change in 2013 for women predicted to be veiled. The trend-adjusted estimate in Panel C is also positive and statistically significant, though less precisely estimated. This increase primarily reflects higher employment among veiled women rather than longer hours among those already employed. If the policy had affected hours at the intensive margin, we would expect a larger increase in average weekly hours. Instead, the magnitude aligns closely with the rise in employment probability reported in Table 3, indicating that the reform operated mainly through new employment opportunities rather than increased working hours for veiled women who were already employed before the repeal of the ban.

Columns 2 and 3 turn to indicators of job quality. Column 2 shows that the reform reduced the probability that women predicted to be veiled worked without social security coverage by about 0.9 percentage points in Panel A, consistent with a shift from informal to formal public sector jobs. Column 3 shows a corresponding improvement in contract stability: the probability of holding a permanent contract increased by 1.8 percentage points, equivalent to roughly a 16 percent rise relative to the pre-policy mean. These estimates are nearly unchanged in Panel B, indicating that they are not sensitive to allowing for differential trends in 2013. In Panel C, the informal-employment coefficient remains negative but becomes less precise, while the effect on permanent employment remains positive and statistically significant. These findings are consistent with the observed rise in public sector employment, as such positions in Turkey typically provide formal contracts and social security coverage.

Columns 4–6 of Table 4 explore the occupational distribution of veiled women within the public sector. Each outcome is a binary indicator equal to one if a woman is employed in a specific public occupation and zero otherwise. Column 4 captures professional occupations requiring higher education, such as teachers, doctors, and nurses; Column 5 covers technicians and associate professionals; and Column 6 includes clerical occupations such as clinical assistants, midwives, administrative secretaries, and office clerks. The results indicate that women predicted to be veiled gained access to public sector occupations, especially technical and clerical positions. In Panel A, the probability of working as a technician rose by about 0.9 percentage points, while the probability of working in clerical occupations increased by about 0.3 percentage points. The coefficient for professional occupations is positive but less precisely estimated. Panel B yields similar results. Under the trend-adjusted specification in Panel C, the occupational coefficients remain positive but become smaller and less precise, consistent with the more demanding nature of this specification. Overall, these results suggest that the expansion in public sector employment among veiled women was concentrated in technical and clerical occupations, reflecting improved access to formal, higher-quality jobs once the ban was lifted.

The last column of Table 4 examines monthly labor earnings, measured only for employed wage earners.[27] The estimates show no robust effect: Panel A suggests a modest positive coefficient, roughly 6 percent of the pre-policy mean, but it is imprecisely estimated. The coefficient remains positive but statistically insignificant in Panels B and C. This result should be interpreted cautiously, as post-reform selection into employment likely biases the estimates, limiting causal interpretation.

Finally, to complement our main results, Online Appendix Table A5 presents analogous estimates for women predicted not to be veiled. We re-estimate equation (1) in a simplified form by replacing the year fixed effects with a single post-period indicator. This coefficient captures aggregate differences in outcomes between the pre- and post-policy periods for non-veiled women and should be interpreted as descriptive, reflecting compositional adjustments rather than causal effects of the reform. The estimates show that, after 2013, non-veiled women became less likely to work in the public sector, while their likelihood of private sector employment and self-employment increased. Given that the overall size of public sector employment remained stable during this period (Figure 2), these patterns are consistent with a reallocation in the composition of public sector jobs, with veiled women gaining access once the ban was lifted and non-veiled women correspondingly shifting toward private- or self-employment.

All in all, the evidence paints a clear picture of how the repeal of the headscarf ban reshaped women's labor market opportunities. Women predicted to be veiled gained access to public sector jobs that had previously been closed to them, leading to higher overall employment. The increase in public sector employment was accompanied not only by higher

[27]Earnings are available only for wage earners, so the analysis excludes the self-employed and unpaid family workers.

Table 4: *Effects of the removal of the headscarf ban on working hours, job quality, and occupations*

Outcome	Working hours (weekly) (1)	Informal employment (2)	Permanent job (3)	Professionals (4)	Technicians (5)	Clerks (6)	Earnings (monthly) (7)
Panel A							
Veiled × Post	0.491***	-0.009***	0.018***	0.003	0.009***	0.003***	10.695
	(0.170)	(0.002)	(0.003)	(0.001)	(0.008)	(0.001)	(7.268)
	[0.194, 0.793]	[-0.012, -0.003]	[0.012, 0.025]	[-0.001, 0.004]	[0.007, 0.010]	[0.002, 0.005]	[-10.01, 16.48]
Veiled	-0.388**	-0.001	-0.009***	-0.001	-0.004***	-0.001***	-4.336
	(0.157)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(6.362)
Panel B							
Veiled × Post	0.419**	-0.009***	0.018***	0.001	0.012***	0.004***	11.451
	(0.195)	(0.002)	(0.004)	(0.001)	(0.001)	(0.001)	(7.936)
	[0.057, 0.804]	[-0.013, -0.003]	[0.011, 0.026]	[-0.001, 0.004]	[0.007, 0.010]	[0.002, 0.005]	[-10.01, 16.48]
Veiled	-0.318*	-0.001	-0.009***	0.002	-0.006***	-0.002***	-5.082
	(0.173)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(6.665)
Panel C							
Veiled × Post	0.685**	-0.007	0.011**	0.006	0.002	0.001	10.750
	(0.348)	(0.004)	(0.007)	(0.003)	(0.001)	(0.001)	(13.24)
	[0.025, 1.272]	[-0.014, 0.001]	[0.001, 0.027]	[-0.003, 0.007]	[-0.001, 0.004]	[-0.002, 0.003]	[-17.67, 31.14]
Veiled	-0.316	-0.001	-0.011***	0.001	-0.006***	-0.002***	-4.316
	(0.215)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(8.092)
Observations	915,873	933,810	933,810	933,810	933,810	933,810	790,783
Mean Dep. Var.	10.95	0.181	0.114	0.008	0.004	0.003	180.8

Notes: Data are from the 2010-2017 HLFS. The sample includes all women aged 18-49. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable ($t = \text{year} - 2010$). Each estimate includes a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. The working hours and labor market earnings variables take zero for those who are unemployed. The earnings variable is only available for those who are employed as wage earners in the survey. Last row reports the pre-policy mean of the dependent variable among women predicted to be veiled. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

employment rates but also by a decline in unpaid family work, suggesting a transition from unpaid or more marginal forms of employment toward formal public sector positions. These new opportunities were concentrated in technical and clerical occupations, roles that typically require some educational qualifications, suggesting that better-educated veiled women were among the main beneficiaries. Meanwhile, women predicted not to be veiled became less likely to hold public sector jobs compared with the pre-policy period, consistent with a compositional shift in which veiled women partly substituted for non-veiled women once the ban was lifted. Overall, the evidence underscores how institutional barriers to identity-based expression can shape not only labor market participation but also the distribution of high-quality jobs across groups.

Exogeneity of veiling status. A potential concern for interpreting our results arises if the policy change induced some women to adopt veiling. For instance, a woman who previously refrained from veiling to pursue a career in the public sector might begin veiling and subsequently enter public employment after the ban was lifted. Because our empirical design classifies women into treatment and comparison groups based on pre-policy predictions of veiling status, which we assume to be time-invariant, such switchers would be misclassified.

These women, who experience a positive labor market response to the reform, would remain in the comparison group, thereby contaminating its post-policy outcomes. Conversely, some women with low predicted veiling probabilities might begin veiling strategically after the repeal, adopting the headscarf not out of religious conviction but to access newly opened public sector opportunities. Since their pre-policy characteristics place them in the non-veiled (comparison) group, post-policy changes in their labor market outcomes would again raise outcomes in the control group, reducing the estimated treatment effect.

While we cannot directly observe such switching in our main data due to the lack of individual panel information,[28] we present several pieces of evidence suggesting that veiling is highly stable over time and unlikely to respond to short-term policy changes.

We first assess the persistence of veiling at the aggregate level. Figure 6 plots the regional share of ever-married women aged 18 and above who report wearing a headscarf in 2008 and 2013, using DHS data.[29] The two top panels show the regional shares in 2008 and 2013, while the bottom panel presents the change in percentage points between the two surveys. The comparison reveals a striking degree of regional stability: the Pearson correlation in regional veiling rates between 2008 and 2013 is 0.93, and the Spearman rank correlation is 0.89 (both $p < 0.001$). This implies that both absolute levels and the relative ranking of regions remained almost unchanged over the five-year period.

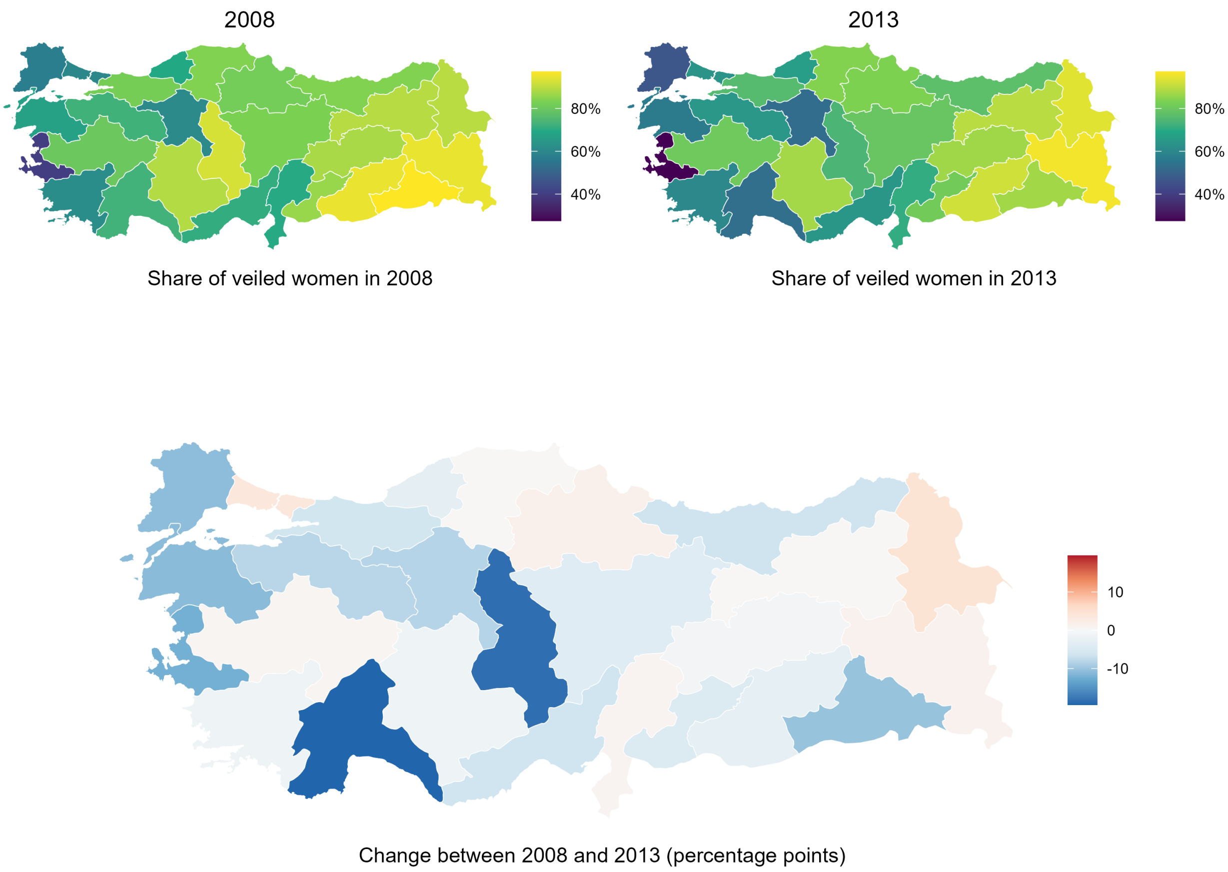
Second, we examine whether earlier institutional reforms in Turkey affected veiling behavior. The first is the 1997 extension of compulsory schooling from five to eight years, which is well documented to have increased women's educational attainment and labor market participation (Erten and Keskin, 2018; Güneş, 2016; Merlino and Yurdakul, 2026). The second is the 2002 reform of the Civil Code, which eliminated several legal provisions favoring men (e.g., polygyny, unilateral divorce, inheritance rights) and raised the minimum marriage age to 18 for both genders (Anıl, 2002; Kırdar et al., 2018). Using the 2013 DHS and exploiting variation in birth month and year in a regression discontinuity design, we assess both reforms (see Online Appendix Section A and Tables A6–A7). Across specifications, we find no statistically or economically significant effects of either reform on the probability of veiling. We also find no impact on regular prayer, and only a small negative effect of the 2002 reform on fasting during Ramadan.

Third, similar to Aksoy and Gambetta (2021), we exploit interview timing within the 2013 DHS to test whether veiling changed immediately after the October 2013 repeal. Since the survey was fielded from September to December 2013 and the reform occurred in October, September is clearly pre-reform and November–December are clearly post-reform, while October is ambiguous. We therefore report two specifications in Online Appendix Table A8: one that treats October as part of the pre-reform period, comparing September–October with November–December, and another that treats October as part of the post-reform period, comparing September with October–December. Neither specification shows evidence of an overall increase in veiling during the fieldwork period. Allowing the post-reform change to vary by AKP-governed province, where any immediate policy-induced increase in veiling might be expected to be strongest, likewise yields no evidence of a differential increase in politically aligned areas. Under both ways of classifying October 2013, the implied total post-reform change in AKP-governed provinces is close to zero and statistically insignificant. This suggests that the removal of the headscarf ban did not cause an immediate change in women's veiling decisions.

[28] Only the 2008 and 2013 Turkish DHS waves collect information on veiling. To our knowledge, KONDA is the only institution maintaining a panel dataset on veiling, but access is proprietary and requires purchase.

[29] We restrict the sample to ever-married women because the 2008 DHS collects veiling information only for this group.

Figure 6: Regional stability in veiling, 2008–2013



Notes: Data are from the 2008 and 2013 Turkish Demographic and Health Survey. This figure plots the regional share of ever-married women aged 18 and above who report wearing a headscarf since the 2008 DHS covers data only for the sample of ever married women. Regions correspond to the 26 NUTS-2 areas of Turkey. The two panels in the top row show the proportion in 2008 and 2013, while the bottom panel shows the change in percentage points.

Taken together, the high degree of regional stability in veiling between 2008 and 2013, the null effects of earlier institutional reforms, and the absence of any immediate increase in veiling during the 2013 DHS fieldwork period suggest that veiling is a relatively persistent practice, shaped more by enduring cultural and religious norms than by short-run policy variation. Nevertheless, to the extent that some switching in response to the removal of the headscarf ban may have occurred, such misclassification would attenuate our estimated effects by introducing treated individuals into the comparison group. As a result, our estimates should be interpreted as conservative lower bounds of the true causal impact of the reform.

Unobservable heterogeneity. Another potential concern is that veiling status may be correlated with unobserved religiosity, which can influence labor supply decisions, as discussed by Carvalho (2013). This correlation could bias our estimates in either direction. For instance, if the most religious women, those for whom veiling is non-negotiable, were also the most constrained by the ban, they might respond most strongly to its repeal. In this case, our estimates would capture both the policy effect and a differential response among highly religious women (an upward bias). However, in the Turkish context specifically, religiosity is negatively associated with female employment (Dildar, 2015; Kızılca, 2016; Gevrek and Gevrek, 2023), suggesting that substantial upward bias from unobserved religiosity is less likely.

Because religiosity is unobserved, we proxy for it using two regional measures of religious practice, both constructed from pre-reform data. The first is mosque density, the number of mosques per capita at the NUTS2 level, drawn from administrative records compiled by Livny (2020) and originally collected by the Turkish Statistical Institute. The second is Quran course enrollment per capita, the number of women enrolled in voluntary Quran courses per 10,000 women, sourced from the Presidency of Religious Affairs and likewise compiled by Livny (2020).[30] We construct above-median binary indicators for each proxy and estimate a triple-difference model interacting each with our main $Veiled \times Post$ term.[31] If unobserved religiosity were driving the results, we would expect larger effects in regions with higher religious intensity. Panels A and B of Table 5 show no such pattern. Across both proxies, the triple interaction term is either statistically insignificant or significantly negative, indicating that the reform effect is, if anything, smaller, not larger, in more religiously intense regions. For overall employment probability, neither the mosque density indicator (Panel A: -0.007) nor the Quran enrollment indicator (Panel B: -0.013) yields a statistically significant triple interaction. For public sector employment, the baseline $Veiled \times Post$ effect is around 2 percentage points in both panels, and the triple interaction is negative in both panels, though statistically significant only for mosque density, meaning women predicted to be veiled in high-religiosity regions gained somewhat less, not more, from the reform. This pattern is inconsistent with upward bias from unobserved religiosity. If anything, it suggests a potential downward bias: less religious women predicted to be veiled, who were more likely to be primarily labor-market-oriented and constrained by the ban, appear to have responded more strongly to the new economic opportunities opened by the reform. Combined with the measurement error in predicted veiling status, explained above, this suggests that our estimates are likely conservative.

The stronger effects in lower-religiosity regions are also consistent with Prediction 3 of the conceptual framework, although this heterogeneity exercise should be interpreted as suggestive rather than as a direct test. The mosque-density and Quran-course measures are regional proxies for religious intensity and do not capture individual religiosity. Still, the pattern is consistent with the mechanism in the framework: because public sector jobs require educational qualifications that are less prevalent in higher-religiosity regions, lower-religiosity regions may contain a larger share of qualified women predicted to be veiled who are able to respond to the reform.

Although we assume religiosity to be time-invariant, it may still evolve differently across groups. To strengthen our causal interpretation, we examine heterogeneity by education. To do so, we compare veiled women with no formal education or only primary school degree (5 years of education) to those with at least a junior high school degree ($HigherEduc$). Since the headscarf ban applied only to public institutions, its repeal should primarily benefit relatively more educated veiled women, who are more likely to be eligible for such jobs rather than those with only a primary school degree or no formal education. Panel C of Table 5 confirms this pattern, consistent with the second prediction of the conceptual framework that the effect of ban removal is stronger among more educated veiled women. The triple-interaction coefficients are positive and highly significant for both the probability

[30] To calculate per capita rates for both proxies, we use female population counts by NUTS2 region from the Address-Based Population Registration System (ADNKS) for 2013. Both measures are constructed from 2013 data, prior to the reform, and are therefore predetermined with respect to post-reform outcomes. Both datasets are available online at `https://www.alivny.com/data`.

[31] Thus, identification comes from comparing the post-reform change in the outcome gap between women predicted to be veiled and those predicted not to be veiled within a region–year cell, and testing whether this change is larger in above-median mosque regions than in below-median mosque regions. Region–year fixed effects absorb all shocks common to women in the same region and year, such as local labor demand conditions or regional policy changes, ensuring the estimate is not confounded by such factors.

Table 5: *Heterogeneous effects*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Number of mosques per capita					
Veiled × Post × HighMosque	-0.007	-0.007**	0.006	-0.008***	-0.001
	(0.007)	(0.004)	(0.006)	(0.002)	(0.003)
	[-0.023, 0.005]	[-0.015, -0.002]	[-0.009, 0.017]	[-0.011, -0.002]	[-0.011, 0.003]
Veiled × Post	0.015***	0.020***	-0.001	0.001	-0.010***
	(0.005)	(0.002)	(0.006)	(0.001)	(0.002)
	[0.005, 0.026]	[0.013, 0.023]	[-0.010, 0.010]	[-0.002, 0.004]	[-0.010, -0.004]
Panel B: Number of Quran course attendants per capita					
Veiled × Post × HighQuran	-0.013	-0.004	-0.002	-0.007***	0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.004)
	[-0.024, 0.001]	[-0.011, 0.001]	[-0.015, 0.011]	[-0.018, -0.002]	[-0.008, 0.005]
Veiled × Post	0.018***	0.018***	0.003	0.001	-0.010***
	(0.006)	(0.002)	(0.006)	(0.002)	(0.002)
	[0.006, 0.028]	[0.012, 0.021]	[-0.007, 0.014]	[-0.002, 0.004]	[-0.012, -0.004]
Panel C: Education level					
Veiled × Post × HigherEduc	0.025***	0.011***	0.004	0.006	0.007
	(0.008)	(0.003)	(0.007)	(0.003)	(0.004)
	[0.001, 0.030]	[0.006, 0.017]	[-0.012, 0.020]	[-0.001, 0.010]	[-0.008, 0.009]
Veiled × Post	-0.014	0.003	-0.006	-0.005	-0.009
	(0.006)	(0.001)	(0.006)	(0.003)	(0.004)
	[-0.021, 0.004]	[-0.001, 0.002]	[-0.009, 0.017]	[-0.008, 0.002]	[-0.011, 0.003]
Panel D: Regional share of Kurdish population					
Veiled × Post × HighShareKurd	0.034***	0.012*	0.002	-0.002	0.007***
	(0.012)	(0.007)	(0.008)	(0.003)	(0.007)
	[0.022, 0.073]	[-0.002, 0.027]	[-0.017, 0.016]	[-0.009, 0.005]	[0.004, 0.031]
Veiled × Post	0.010**	0.016***	0.002	-0.003**	-0.010***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.002, 0.015]	[0.011, 0.017]	[-0.005, 0.009]	[-0.004, -0.001]	[-0.014, -0.008]
Panel E: Political alignment (AKP vote share)					
Veiled × Post × HighVoteShare	0.018	0.007	0.003	0.005	-0.001
	(0.007)	(0.003)	(0.006)	(0.003)	(0.003)
	[-0.009, 0.019]	[-0.005, 0.008]	[-0.009, 0.014]	[-0.002, 0.009]	[-0.006, 0.007]
Veiled × Post	0.004**	0.013***	0.001	-0.006**	-0.009***
	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)
	[0.001, 0.019]	[0.009, 0.018]	[-0.004, 0.009]	[-0.007, -0.001]	[-0.013, -0.004]
Observations	933,810	933,810	933,810	933,810	933,810

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports heterogeneity estimates based on Eq. (1). $HighMosque$ is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median number of mosques per capita. $HighQuran$ is a binary variable equal to one if the woman resides in a region with an above-median number of female Quran course enrollees per 10,000 women. $HigherEduc$ is a binary variable equal to one if the woman holds at least a junior high school degree, and zero if she has no formal education or only a primary school degree. $HighShareKurd$ is a binary variable equal to one if the woman resides in a region where more than 50 percent of women report Kurdish as their mother tongue. $HighVoteShare$ is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median vote share of the Justice and Development Party (AKP) in the 2014 municipal elections. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

of employment and public sector employment, indicating that the observed effects of the reform are concentrated mainly among higher-educated veiled women. The reform increases the probability of employment primarily for the relatively more educated group (about 1.1 percentage points), whereas the effect for less-educated veiled women is small and statistically insignificant. Similarly, the probability of public sector employment rises for both groups, but the effect is roughly five times larger for more educated peers (1.4 percentage points), than the low-educated veiled women who experience a 0.3 percentage point increase. These results suggest that the removal of the ban provided public sector opportunities primarily to women with some educational qualifications required for such positions. The declines in self-employment and unpaid family work observed in the aggregate are driven mainly by low-educated veiled women. In contrast, higher-educated veiled women exhibit net employment gains concentrated in the public sector. If religiosity alone explained labor market entry, we would expect broad increases across education groups, including among low-educated veiled women. Instead, the effects are concentrated exactly where they should be if the ban constrained access to public employment among educated veiled women. This pattern is consistent with the institutional setting: public sector positions in Turkey typically require some educational degree, meaning the ban disproportionately constrained veiled women who already possessed the qualifications needed to compete for such jobs. The concentration of effects among higher-educated veiled women therefore reflects the structure of the policy itself rather than selection on unobserved religiosity.

A further concern for interpretation is that the 2013 democratization package included measures beyond the repeal of the headscarf ban, most notably expanded minority-language rights, which may have been especially salient in Kurdish-majority regions. If these other components of the package independently affected women's labor market behavior, our estimates could partly reflect broader political and social changes rather than the headscarf reform itself. To assess this possibility, Panel D of Table 5 interacts our main $Veiled \times Post$ term with an indicator for Kurdish-majority regions, defined as NUTS2 regions in which more than 50 percent of women report Kurdish as their mother tongue in the 2013 DHS.[32] The triple interaction is positive for overall employment (0.034) and public sector employment (0.012), indicating that the estimated effects are, if anything, larger in Kurdish-majority regions. However, the key point for identification is that the baseline $Veiled \times Post$ coefficient, which captures the effect in non-Kurdish-majority regions, remains positive and statistically significant for both overall employment (0.010) and public sector employment (0.016). Thus, the main employment effects are not driven solely by regions where the minority-language components of the democratization package were likely to be most salient. Taken together, these results support our interpretation that the estimated effects are more consistent with the repeal of the headscarf ban than with other contemporaneous components of the broader reform package.[33]

While the results above support that our findings are primarily attributable to the repeal of the headscarf ban, a remaining concern in the Turkish context is whether the observed gains are partly driven by political favoritism rather than the policy itself. If municipalities governed by Islamist parties were more inclined to hire veiled women after the reform, the estimated effects might reflect preferential hiring in politically aligned areas rather than the removal of an institutional barrier (Corekcioglu, 2021). To explore this mechanism, we construct a measure of Islamist political influence based on the vote share of the Justice and Development Party (AKP) in the 2014 municipal elections. Regions with an above-median AKP vote share (46.8 percent) are classified as high political influence, and this indicator is interacted with our difference-in-differences terms in Eq. (1). It is important to note, however, that this analysis should not be interpreted as causal heterogeneity. If the

[32] This classification identifies four NUTS2 regions as Kurdish-majority. Using ethnic self-identification data from the KONDA Barometer yields the same regional classification, so we do not report those results separately.

[33] Corresponding estimates including an indicator for the year 2013 and its interaction with the veiled indicator and de-trended estimates reported in Online Appendix Tables A9 and A10, are similar to those in Table 5.

reform itself affected subsequent political support for the AKP, political alignment may be endogenous to the policy change. We therefore view this analysis as a descriptive mediation test, assessing whether the labor-market effects in Table 3 are concentrated in AKP-governed regions. The results in Panel E of Table 5 show that the effects are not confined to politically aligned regions. The triple interaction term is statistically insignificant across all outcomes, indicating that the labor market effects do not differ systematically between high- and low-AKP regions. The baseline $Veiled \times Post$ coefficient, capturing effects in below-median AKP regions, is positive and significant for both overall employment (0.004, $p < 0.05$) and public sector employment (0.013, $p < 0.01$). The modest overall employment effect in these regions, despite a larger public sector effect, reflects simultaneous declines in self-employment and unpaid family work, suggesting that in below-median AKP regions the reform primarily reshuffled veiled women from informal work into public sector employment, with smaller net positive labor market effects than in high-AKP regions. No differential effects are found for private sector employment. These findings suggest that while Islamist-leaning municipalities may have amplified access for veiled women, the primary driver of the positive labor market effects is the nationwide removal of the institutional barrier that had previously excluded them from public employment.[34]

6.2 Robustness checks

We assess the robustness of our main results in Table 3 to remaining pre-trend differences using the falsification-robust procedure of Rambachan and Roth (2023). The method places a relative-magnitude bound M on any post-treatment deviation from parallel trends, scaling it to the largest adjacent pre-period deviation. For each outcome we report a breakdown value $\bar{M}$, defined as the smallest M at which the falsification-robust confidence interval (FLCI) first contains zero. Online Appendix Figure A1 plots the 95% FLCIs together with the baseline point estimates for $M \in \{0, 0.1, \ldots, 0.5\}$. As M increases, the parallel-trends assumption is relaxed and the FLCIs widen. An effect is robust up to $M^\star$ if the FLCI remains strictly above (or below) zero for all $M \leq M^\star$; the breakdown value $\bar{M}$ marks the first relaxation at which the confidence interval touches zero.

The results show a clear pattern of robustness across outcomes. For the employment probability, the effect remains statistically positive up to $\bar{M} \approx 0.4$, meaning it survives any post-policy drift up to 40% as large as the worst pre-policy deviation; beyond this threshold (e.g. $M = 0.5$) the effect is no longer distinguishable from zero. Public sector employment is more robust, with $\bar{M} > 0.5$ (no breakdown on our grid), indicating that the positive effect persists even under sizeable deviations from parallel trends. By contrast, private sector employment exhibits $\bar{M} \approx 0$, consistent with no effect. Finally, the negative effects on self-employment and unpaid family work are highly robust, as the FLCIs remain strictly below zero for all $M \leq 0.5$.

[34] We also assess whether other policy changes could explain our results. First, as discussed in Section 2, universities in Turkey gradually began allowing female students and employees to wear headscarves from 2010, with institutional variation in implementation. Since this may have affected women born in or after January 1992, who were more likely to enter university after these changes, we interact an indicator for this cohort with *Post*. The results in Panel A of Online Appendix Table A12 remain consistent with our main estimates, suggesting that earlier university-level relaxations of the ban do not drive our findings. Second, we account for the 1997 compulsory education reform, which extended compulsory schooling from five to eight years for cohorts born after 1986 and is known to have affected women's education and labor market outcomes (Erten and Keskin, 2018; Merlino and Yurdakul, 2026; Güneş, 2016). We interact an indicator for being born after 1986 with the post-policy period; as shown in Panel B of Table A12, our results remain robust. Finally, Panels A and B of Table A13 reproduce these estimates while also including an interaction between $Veiled$ and the 2013 wave indicator to account for differential trends in the transition year, yielding consistent estimates.

Moreover, Figure 6 shows that although regional veiling rates are highly stable overall, two regions experienced relatively larger changes between the 2008 and 2013 DHS waves.[35] To assess that our results are not driven by these regions, Panel C of Online Appendix Table A12 re-estimates Eq. (1) after excluding observations from both regions. The estimates are virtually unchanged relative to the baseline: the $Veiled \times Post$ coefficient remains positive and statistically significant for overall employment (0.012) and public sector employment (0.016), with the 99% bootstrap percentile confidence intervals excluding zero.[36] These results suggest that our estimates are not sensitive to excluding the regions with the largest changes in veiling rates between the two DHS waves.

Our main analysis imputes veiling status using a model trained on the 2013 wave of the Turkish Demographic and Health Survey (DHS). To assess whether our results depend on this particular training sample, we re-estimate the main specification using veiling probabilities predicted from the 2008 DHS, which reports veiling information for a representative sample of ever-married women. The estimated treatment effects, reported in Panel D of Online Appendix Table A12, are very similar to our baseline results in both magnitude and significance. A potential limitation of this exercise is that the 2008 DHS includes only ever-married women, who may be more selected and whose veiling behavior may be less responsive to labor-market incentives. To address this concern, we repeat the full prediction and estimation procedure for women aged 25 and above, among whom the share ever married is substantially higher. Specifically, we estimate veiling probabilities using the 2008 DHS restricted to women aged 25 and above and apply the same imputation procedure to the corresponding HLFS sample. The results, reported in Panel E of Online Appendix Table A12, remain similar to the baseline estimates. Taken together, these exercises support the stability of the underlying relationship between individual characteristics and veiling behavior, and indicate that our findings are not sensitive to the choice of DHS wave used for prediction and imputation.

To further validate our results, we explore alternative treatment and control groups. In the first exercise, we redefine our treatment group as women who hold traditional attitudes toward gender roles, those who agree with the statement that *"the husband should work while the wife stays home"* (i.e., traditional). If the estimated treatment effects for women who are predicted to be veiled closely resemble those for traditional women, it would be difficult to determine whether the effect is specific to veiled women or the sample of women with traditional attitudes. In principle, given the low correlation between being traditional towards gender role attitudes and being veiled (0.20 in the DHS sample), we should not expect similar effects of the reform on these samples. We predict traditional attitudes following the same methodology as our main analysis and re-estimate Eq. (1) using the variable $Traditional$ instead of $Veiled$. Results presented in Panel F indicate smaller and statistically distinct effects for traditionally-minded women. The Wald test of equality of coefficients confirms significant differences in the estimated effects for employment (p-value = 0.000) and public sector employment (p-value = 0.000). These findings support that the effects of the reform are indeed specific to veiled women rather than women with traditional attitudes more generally.

In the second exercise, we use the frequency of praying daily to validate our main results. Given the strong correlation between regular prayer and veiling status (0.42 in the DHS

[35]These two NUTS2 regions are Konya–Karaman and Kırıkkale–Aksaray–Niğde–Nevşehir–Kırşehir, which exhibit the largest declines in the share of veiled women between the 2008 and 2013 DHS waves.

[36]Panel C of Online Appendix Table A13 reproduces these estimates while additionally including an interaction between $Veiled$ and the 2013 wave indicator to account for differential trends in the transition year; the estimated effects are similar in magnitude and precision.

sample), this provides a more behaviorally anchored proxy for veiling. If the estimated effects are similar for women who pray daily, this would further validate our out-of-sample predictions. The results reported in Panel G indicate significant increases in employment probability and public sector employment for women who pray regularly, along with declines in unpaid family work. Importantly, the Wald test for employment (p-value = 0.734) suggests no significant difference between effects estimated for veiled women and those who pray daily, further supporting the validity of our prediction model. However, the effect on public sector employment is larger for veiled women (p-value = 0.000), which is expected given that the reform directly addressed institutional barriers specific to veiling.

To enhance the validity of the veiling imputation, we employ machine learning (ML) techniques to predict headscarf wearing. We then use the ML-predicted veiling information to estimate the policy effects. The ML models accommodate high-dimensional and nonlinear relationships between observables and veiling, thereby strengthening the credibility of the imputation. Overall, the results remain broadly consistent with those obtained using baseline Probit imputation. The procedure involves two steps. First, the DHS sample is divided into a training set (80%) and a test set (20%). The training set is used to estimate a predictive model that links women's background characteristics to veiling status, and the test set is used to evaluate the model's predictive performance. Second, the trained model is applied to the HLFS to generate out-of-sample predictions of veiling.

Our main approach relies on ensemble learning, which combines four commonly used machine learning models — Support Vector Machines, Random Forest, K-Nearest Neighbors, and Neural Networks. Each model is first trained separately to predict the probability that a respondent is veiled, and a grid search is performed over a predefined parameter space to identify the configuration that maximizes training accuracy. The four models are then combined through soft voting, which aggregates the predicted probabilities from each model and assigns the final veiling status based on the average probability across models. A detailed discussion of each model and its parameter configuration is provided in Appendix A.3. The ensemble model achieves approximately 80% accuracy on the DHS training set, and when applied to the HLFS, predicts that about 66% of women wear a headscarf, closely matching the DHS average and providing validation for our prediction approach.[37]

Tables A15–A17 report the estimated policy effects on women's labor market outcomes using imputation of five ML models. Table A15 presents results for specification A, corresponding to Panel A in Table 3, Table A16 for specification B corresponding to Panel B, and Table A17 for specification C corresponding to Panel C. All tables report both point estimates and bootstrap percentile 95 percent confidence intervals. Consistent with the main analysis, the results show statistically significant increases in employment probability and public sector employment among veiled women following the policy change, accompanied by declines in self-employment and unpaid family work. While the magnitudes and significance levels vary slightly across models, the overall patterns remain stable across ML algorithms and econometric specification, supporting the robustness of our main findings.

Tables A18–A20 examine the policy effects on working hours, earnings, job quality, and occupational composition using ML imputation methods, serving as a robustness check to Table 4. The three tables correspond to specifications A-C in Table 4 respectively. The results consistently show positive effects on women's working hours and the likelihood of holding a permanent job, alongside a corresponding decline in informal employment. The estimates also indicate that veiled women are more likely to work in occupations requiring

[37]The Neural Network produces continuous predicted probabilities rather than discrete classifications. As a result, the reported coefficient reflects the interaction between the predicted probability of veiling and the post-policy indicator.

lower educational attainment, such as technician and clerical positions, following the policy change. In contrast, all models show insignificant effects on monthly earnings. Overall, these robustness checks confirm the stability of our main findings across different imputation methods and econometric specifications.

In sum, these robustness exercises indicate that our results are not driven by arbitrary classification rules, omitted variable bias, or concurrent policy shocks. More importantly, the ML-based out-of-sample prediction approaches corroborate our main findings across a range of methodologies, thereby providing a rigorous framework for evaluating policy effects in settings where direct treatment assignment is unobservable.

6.3 Discussion

Our findings indicate that lifting the headscarf ban significantly increased public sector employment among women predicted to be veiled, with the gains concentrated among relatively more educated women who are most likely to meet the formal requirements for public sector jobs. This suggests that the ban had previously restricted access to public employment for women who were otherwise qualified to enter these positions.

These results carry important implications for other Muslim-majority contexts where veiling restrictions remain in place. For example, Kazakhstan introduced a ban on wearing hijabs in schools in 2023.[38] While such policies are often justified on secularist or uniformity grounds, our evidence suggests that institutional restrictions may impose substantial productivity costs by limiting the labor market participation of relatively more productive veiled women.

Although our analysis focuses on a Muslim-majority setting, the findings also raise questions for Muslim-minority contexts, such as those in Western Europe. Within Turkey, the effects of the reform do not significantly differ across regions with above- versus below-median veiling rates (Panel A of Table 5). In Online Appendix Table A21, we further compare women who are a local minority (in regions with veiling rates below 50%) to those in regions where veiling is more common, and again find no significant differential effects. These results suggest that the impact of removing institutional barriers is not confined to contexts where veiling is prevalent. That said, extrapolation to Europe must be made with caution: veiling rates in Turkey remain much higher than in most European countries, and veiled women in Europe may face different challenges, including labor market discrimination and social exclusion, beyond the formal institutional barriers we study here. Thus, our evidence is best viewed as suggestive for Muslim-minority contexts.

Moreover, in the previous section, our results show a substitution effect of the reform: women predicted to be veiled entered public employment while women predicted not to be veiled shifted toward private-sector jobs, implying a reallocation of employment opportunities rather than the creation of new ones. This raises a related question about the composition of the public sector workforce after the repeal: did the expanded access to public employment change the education–occupation match quality of women employed in the public sector?

To examine this question, we restrict the sample to women employed in the public sector and estimate Eq. (1) for three education–occupation match indicators based on ISCO skill levels: (i) overqualification, defined as having education above the occupation's skill requirement; (ii) underqualification, defined as having education below the requirement; and (iii) overall mismatch, defined as either overqualification or underqualification. These indicators provide suggestive evidence on the match between workers' education and occupational

[38] In countries such as Egypt, Tunisia, Azerbaijan, Kazakhstan, Kosovo, and Kyrgyzstan, various forms of veiling bans persist in schools or public spaces.

requirements, although they should not be interpreted as direct measures of productivity.

The results are reported in Online Appendix Table A22. In Panel A, the baseline specification shows no statistically significant change in overqualification, but indicates a 1.9 percentage-point increase in underqualification and a 1.3 percentage-point increase in overall mismatch among women predicted to be veiled relative to women predicted not to be veiled. Panel B, which allows for a differential transition-year effect in 2013, yields a similar pattern: underqualification increases by 2.6 percentage points and overall mismatch increases by 3.2 percentage points. In Panel C, which allows for differential linear trends, the estimated effects on underqualification and mismatch are smaller and less precise. In columns 4 and 5, we find no statistically significant effect on age or hourly earnings across the three specifications. The negative sign of the earnings coefficient is consistent with the seniority-based wage structure of the Turkish public sector, where pay progression depends primarily on tenure and formal pay scales rather than short-run productivity differences.

Overall, these estimates suggest that the expansion of access to public employment may have been accompanied by some increase in underqualification and education–occupation mismatch among women predicted to be veiled. One interpretation is that, after the reform, some women predicted to be veiled entered public sector occupations for which their formal educational attainment was lower than the occupation-specific skill classification, replacing or offsetting employment among women predicted not to be veiled with relatively higher formal education for the same occupational category. This points to a change in the composition of women employed in the public sector, but it should not be interpreted as direct evidence of lower productivity. Future work linking changes in public sector employment composition to sectoral outcomes, such as student achievement or patient health indicators, could provide a more direct assessment of the reform's longer-term implications for public sector productivity.

Beyond the labor market, a natural question is whether the repeal of the headscarf ban also affected family outcomes. In Online Appendix Table A23, we examine marriage and divorce probabilities before and after the repeal of the headscarf ban. Panel A shows no effects, but when we adjust for group-specific trends in Panel B, the estimates suggest a decline in marriage and a small increase in divorce among veiled women. Robustness checks following Rambachan and Roth (2023) support this pattern (see Online Appendix Figure A2), but the divergence between Panels A and B indicates that the parallel-trends assumption may be violated for marriage outcomes, making these results more suggestive than definitive. Even so, the pattern is intriguing: as access to public sector jobs expands, veiled women may delay or forgo marriage due to rising opportunity costs, while the increase in divorce could reflect stronger bargaining power within the household. These findings point to potential empowerment effects of lifting institutional restrictions within the family, extending beyond women's labor market outcomes. Examining these channels with richer longitudinal data on marriage timing, household decision-making, and veiling behavior remains a promising direction for future research.

A further avenue for research concerns the dynamics of veiling itself. Our empirical design treats veiling as time-invariant, but the broader question of how labor market opportunities affect women's veiling decisions remains largely unexplored. Individual-level panel data from sources such as KONDA could allow researchers to study whether the repeal of the ban altered women's veiling choices directly, and more generally, how changes in labor market opportunities interact with religious expression.[39]

[39]We attempted a preliminary analysis using the 2008 and 2013 DHS, but the lack of variation in women's employment probabilities across these two points prevented us from drawing meaningful conclusions. More frequent survey waves capturing a longer time period would be needed to examine this question systematically.

7 Conclusion

This paper examines how the repeal of the headscarf ban in public institutions in Turkey affects women's labor market outcomes. Using two complementary datasets, the Turkish Household Labor Force Survey and the Turkish Demographic and Health Survey, we predict and impute women's veiling status from observable characteristics through statistical matching and machine learning methods. We then compare the labor market outcomes of veiled and non-veiled women before and after the 2013 reform within a difference-in-differences framework.

The results reveal three main patterns. First, the repeal substantially increases public sector employment among women predicted to be veiled. This increase reflects both a higher probability of employment and a shift away from self-employment and unpaid family work, suggesting that the reform reshaped women's occupational choices rather than simply expanding jobs. Second, most of the gains for veiled women occur in professional and associate-professional roles, such as teachers, nurses, and administrative staff, which typically require higher levels of education. Consistent with this pattern, the effects are much stronger among more-educated veiled women. Third, non-veiled women experience a decline in public sector employment and a corresponding rise in private sector work, indicating a substitution of job opportunities from non-veiled to veiled women rather than an overall increase in public hiring. Analysis of education–occupation matching shows no systematic signs of notable efficiency losses in the public sector.

The heterogeneity analysis across regions that differ in veiling prevalence and political alignment provides deeper insight into the reform's impacts. Veiled women living in low-veiling regions, who are on average less religious, respond more strongly to the reform. This pattern indicates that the observed policy effects are not driven by religiosity, and that the estimated impact is likely a lower bound on the true causal effect of lifting the ban. Veiled women in regions under stronger Islamist political influence also experience larger gains, particularly in overall and public sector employment. However, the effect on public sector employment remains positive and statistically significant even in less Islamist regions. These results suggest that while Islamist-leaning municipalities may have facilitated more access for veiled women, the main driver of the observed gains is the nationwide removal of the institutional barrier that had previously excluded them from public employment.

Future research can build on this study in several directions. One avenue is to examine more closely the reform's implications for public sector efficiency and output quality by collecting detailed data to measure productivity at the institutional or sectoral level. Another promising direction is to investigate marriage market outcomes in greater depth. The current analysis provides suggestive evidence that the reform increased women's empowerment, with veiled women delaying marriage and experiencing a modest rise in divorce. Access to richer longitudinal data would allow more precise estimates of effects on marriage timing and intra-household decision-making. Finally, future work could explore how the reform influenced the educational choices of younger religious women and whether subsequent changes in educational attainment feed back into veiling practices over time.

Taken together, these findings highlight that easing institutional restrictions on religious expression can significantly improve women's economic inclusion. This suggests that policy interventions expanding institutional access, rather than attempts to reshape individual identities or cultural norms, may be more effective in improving economic opportunities for underrepresented groups. While the prediction of veiling status and the relatively short post-reform period impose some limitations, the results are robust across specifications and offer credible short-run evidence.

References

Abdelgadir, A. and V. Fouka (2020). Political secularism and Muslim integration in the West: Assessing the effects of the French headscarf ban. *American Political Science Review 114*(3), 707–723.

Abdelhadi, E. (2019). The Hijab and Muslim women's employment in the United States. *Research in Social Stratification and Mobility 61*, 26–37.

Abou Daher, C., E. Field, K. Swanson, and K. Vyborny (2025). Drivers of change: Employment responses to the lifting of the Saudi female driving ban. *American Economic Review 115*(9), 3248–3271.

Aha, D. W., D. Kibler, and M. K. Albert (1991). Instance-based learning algorithms. *Machine learning 6*, 37–66.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Akoglu, K. S. (2015). Piecemeal freedom: Why the headscarf ban remains in place in Turkey. *Boston College International and Comparative Law Review 38*, 277.

Aksoy, O. and D. Gambetta (2016). Behind the veil: The strategic use of religious garb. *European Sociological Review 32*(6), 792–806.

Aksoy, O. and D. Gambetta (2021). The politics behind the veil. *European Sociological Review 37*(1), 67–88.

Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of economic literature 53*(4), 898–944.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alvi, A. (2013). Concealment and revealment: The Muslim veil in context. *Current Anthropology 54*(2), 177–199.

Anıl, E. (2002). *The new legal status of women in Turkey.* Women for Women's Human Rights-New Ways.

Arsu, S. and D. Bilefsky (2013, October). Turkey lifts longtime ban on head scarves in state offices. *The New York Times*. Accessed: 24 April 2026.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*(1), 685–725.

BBC News (2010). Turkey drops Islamic headscarf ban at universities. Accessed: 2025-01-06.

BBC News (2017). Turkey's Erdogan backs removing headscarf ban in military. Accessed: 2025-01-03.

Bianet (2010). Headscarf ban at universities softened. Accessed: 2025-01-06.

Breiman, L. (1996). Bagging predictors. *Machine learning 24*, 123–140.

Breiman, L. (2001). Random forests. *Machine learning 45*, 5–32.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review 110*(10), 2997–3029.

Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics 128*(1), 337–370.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Chen, A. Y. and J. McCoy (2024). Missing values handling for machine learning portfolios. *Journal of Financial Economics 155*, 103815.

Cindoğlu, D. (2011). *Headscarf ban and discrimination: Professional headscarved women in the labor market*. Tesev Publications İstanbul, Turkey.

Cindoglu, D. and G. Zencirci (2008). The headscarf in Turkey in the public and state spheres. *Middle Eastern Studies 44*(5), 791–806.

Corekcioglu, G. (2021). Unveiling the effects of a headscarf ban: Evidence from municipal jobs in Turkey. *Journal of Comparative Economics 49*(2), 382–404.

Cortes, C. (1995). Support-vector networks. *Machine Learning*.

Denteh, A. and D. Kédagni (2022). Misclassification in difference-in-differences models. *arXiv preprint arXiv:2207.11890*.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review 112*(3), 899–927.

Dildar, Y. (2015). Patriarchal norms, religion, and female labor supply: Evidence from Turkey. *World Development 76*, 40–61.

Drouhot, L., P. Simon, and V. Tiberj (2023). La diversité religieuse en france: transmissions intergénérationnelles et pratiques selon les origines. *Références-INSEE*, 39–48.

Drouhot, L. G. (2021). Cracks in the melting pot? Religiosity and assimilation among the diverse Muslim population in France. *American Journal of Sociology 126*(4), 795–851.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic literature 50*(4), 1051–1079.

Erten, B. and P. Keskin (2018). For better or for worse? Education and the prevalence of domestic violence in Turkey. *American Economic Journal: Applied Economics 10*(1), 64–105.

Euronews (2017). Turkey purge continues with 7000 public sector workers dismissed. Accessed: 2025-01-03.

European Commission (2013). Commission Staff Working Document: Turkey 2013 Progress Report. Staff Working Document, European Commission. Accompanying the document "Communication from the Commission to the European Parliament and the Council: Enlargement Strategy and Main Challenges 2013–2014", COM(2013) 700 final.

Fernández, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics 1*(1), 146–177.

Field, E., S. Jayachandran, and R. Pande (2010). Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India. *American Economic Review 100*(2), 125–129.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *American Economic Review 111*(7), 2342–2375.

Fleischmann, F. and K. Phalet (2012). Integration and religiosity among the Turkish second generation in Europe: A comparative analysis across four capital cities. *Ethnic and Racial Studies 35*(2), 320–341.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in US schools after World War I. *The Review of Economic Studies 87*(1), 204–239.

Gazeaud, J., N. Khan, E. Mvukiyehe, and O. Sterck (2023). With or without him? Experimental evidence on cash grants and gender-sensitive trainings in Tunisia. *Journal of Development Economics 165*, 103169.

Gevrek, Z. E. and D. Gevrek (2023). Social norms and the gender gap in labor force participation: Evidence from turkey. *Applied Economics Letters 30*(15), 2102–2107.

Ghumman, S. and L. Jackson (2010). The downside of religious attire: The Muslim headscarf and expectations of obtaining employment. *Journal of organizational behavior 31*(1), 4–23.

Gogas, P. and T. Papadimitriou (2021). Machine learning in Economics and Finance. *Computational Economics 57*, 1–4.

Goldstein, I., C. S. Spatt, and M. Ye (2021). Big data in Finance. *The Review of Financial Studies 34*(7), 3213–3225.

Göle, N. (1996). *The forbidden modern: Civilization and veiling.* University of Michigan Press.

Gonzales, M. C., M. S. Jain-Chandra, M. K. Kochhar, and M. M. Newiak (2015). *Fair play: More equal laws boost female labor force participation.* International Monetary Fund.

Güneş, P. M. (2016). The impact of female education on teenage fertility: Evidence from Turkey. *The BE Journal of Economic Analysis & Policy 16*(1), 259–288.

Guveli, A. (2011). Social and economic impact of the headscarf ban on women in Turkey. *European Societies 13*(2), 171–189.

Hallward-Driemeier, M. and O. Gajigo (2015). Strengthening economic rights and women's occupational choice: The impact of reforming Ethiopia's family law. *World Development 70*, 260–273.

Heller, S. B., B. Jakubowski, Z. Jelveh, and M. Kapustin (2024). Machine learning can predict shooting victimization well enough to help prevent it. *Review of Economics and Statistics*, 1–45.

Höjelid, S. (2010). Headscarves, judicial activism, and democracy: The 2007–8 constitutional crisis in Turkey. *The European Legacy 15*(4), 467–482.

Hyland, M., S. Djankov, and P. K. Goldberg (2020). Gendered laws and women in the workforce. *American Economic Review: Insights 2*(4), 475–490.

Jacquet, A. and S. Montpetit (2023). Veiling and the Economic Integration of Muslim women in France.

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. *IMF Economic Review 69*(3), 576–595.

Joslin, K.-E. and F. M. Nordvik (2021). Does religion curtail women during booms? Evidence from resource discoveries. *Journal of Economic Behavior & Organization 187*, 205–224.

Karahan, H. and N. Tuğsuz (2022). Political climate and the headscarf issue in Turkey. *Insight Turkey 24*(1), 211–228.

Kırdar, M. G., M. Dayıoğlu, and İ. Koç (2018). The effects of compulsory-schooling laws on teenage marriage and births in Turkey. *Journal of Human Capital 12*(4), 640–668.

Kızılca, F. K. (2016). Breaking with dogma: Unorthodox consumption patterns and women's labor market outcomes in turkey. *Feminist Economics 22*(4), 1–30.

KONDA (2019). July 2019 Barometer: Public opinion research report. Accessed: 2025-01-06.

Livny, A. (2020). *Trust and the Islamic advantage: religious-based movements in Turkey and the Muslim world*. Cambridge University Press.

Maurin, E. and N. Navarrete H (2023). Behind the veil: The effect of banning the Islamic veil in schools. *Economic Policy 38*(113), 63–98.

Merlino, L. P. and E. Yurdakul (2026). Gender Education Gaps and Household Labor Supply: Causal evidence from Turkey.

Meyer, B. D., W. K. Mok, and J. X. Sullivan (2015). Household surveys in crisis. *Journal of Economic Perspectives 29*(4), 199–226.

Mullainathan, S. and Z. Obermeyer (2022). Diagnosing physician error: A machine learning approach to low-value health care. *The Quarterly Journal of Economics 137*(2), 679–727.

Mullainathan, S. and J. Spiess (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Negi, A. and D. S. Negi (2025). Difference-in-differences with a misclassified treatment. *Journal of Applied Econometrics 40*(4), 411–423.

Nguimkeu, P., A. Denteh, and R. Tchernis (2019). On the estimation of treatment effects with endogenous misreporting. *Journal of econometrics 208*(2), 487–506.

OECD (2025). Government at a glance 2025.

Oh, S. (2023). Does identity affect labor supply? *The American Economic Review 113*(8), 2055–2083.

Rambachan, A. and J. Roth (2023). A more credible approach to parallel trends. *Review of Economic Studies 90*(5), 2555–2591.

Ridder, G. and R. Moffitt (2007). The econometrics of data combination. *Handbook of econometrics 6*, 5469–5547.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Roy, S. (2019). Discriminatory laws against women: A survey of the literature. *World Bank Policy Research Working Paper* (8719).

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics 4*(1), 87–94.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning representations by back-propagating errors. *Nature 323*(6088), 533–536.

Saleh, M. and J. Tirole (2021). Taxing identity: Theory and evidence from early Islam. *Econometrica 89*(4), 1881–1919.

Shofia, N. (2022). Why veil? Religious headscarves and the public role of women. *Religious Headscarves and The Public Role of Women (July 21, 2022)*.

Tommasi, D. and L. Zhang (2024). Identifying program benefits when participation is misreported. *Journal of Applied Econometrics 39*(6), 1123–1148.

Uğur, Z. B. (2020). Unveiled: The effect of the headscarf ban on women's tertiary education in Turkey. *Feminist Economics 26*(2), 187–217.

Wiltse, E. Ç. (2008). The gordian knot of Turkish politics: Regulating headscarf use in public. *South European Society and Politics 13*(2), 195–215.

A Online Appendix

A.1 Effects of the 1997 Education Reform and The 2002 Civil Code on Veiling Decision

To identify the causal effect of the 1997 education reform, we assume that aside from the increase in education induced by the 1997 reform, women born just before and just after January 1987 are otherwise comparable in observable and unobservable characteristics. Based on this assumption, we exploit the exogenous variation in the birth months and years generated by the change in the compulsory schooling law in Turkey. Our empirical strategy is based on a regression discontinuity (RD) design, comparing the probability of being veiled of women born in or after January 1987 (who must complete eight years of education) to those women was born before January 1987 (who could drop out after five years of compulsory education). Formally, we estimate the following equation:

$$y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \alpha_3 (D_i \times X_i) + \gamma' Z_i + \varepsilon_i, \tag{2}$$

where y_i denotes the probability of veiling for woman i, D_i is a binary indicator equal to one if the woman was born in or after January 1, 1987. The term X_i is the normalized running variable around January 1987, defined as the number of months from the cutoff date, flexibly capturing smooth trends in outcomes as a function of birth timing.[40] The interaction term $X_i \times D_i$ allows the slope of these trends to differ on either side of the cutoff. Our coefficient of interest, α_1, captures the average discontinuity at the cutoff and is interpreted as the causal effect of exposure to the 1997 reform for women born just after January 1987, relative to those born just before.

Estimation of model (2) involves specifying the functional form of the running variable, X_i. We model this function as a linear polynomial, determining an optimal bandwidth using the algorithm proposed by Cattaneo et al. (2019). To address potential fuzziness in the treatment status, we also estimate Eq. (2) by excluding women born within three months before or after January 1, 1987. The rationale for using a ± 3 month window is that children born closest to the cutoff date are the most likely, if any, to have started school a year earlier or later than their assigned cohort, due to possible deviations from standard enrollment age.[41] By omitting births within this window, we ensure cleaner separation between cohorts exposed to the different schooling regimes with minimal sample loss.

We control for a vector of variables Z_i comprising dummies for childhood place of residence, mother tongue, whether the woman's mother is literate, dummies for father's education level, the woman's month-of-birth fixed effects, and region-of-residence fixed effects. Standard errors are clustered at the month-year of birth level.

To identify the causal effect of the 2002 Civil Code reform, we assume that aside from the restrictions induced by the increase in the minimum legal marriage age, women born just before and just after January 1985 are otherwise comparable in observable and unobservable characteristics. Based on this assumption, we exploit exogenous variation in birth timing around the January 1985 cutoff, which determines the duration of exposure to the new Civil Code in Turkey.

[40] The running variable is centered at the January 1987 cutoff; positive values indicate birth months after the cutoff, and negative values indicate months before.

[41] For example, a child born in February 1987 would typically start school in September 1993, the year they turn six. However, due to enrollment practices at the time, such a child could have started a year early—in September 1992, at age five and a half. In contrast, a child born in August 1986 would be too old to delay entry by a full year, making early or late enrollment less likely. This justifies our focus on the ± 3-month window, where deviations from standard enrollment age are more plausible.

Our empirical strategy is based on a regression discontinuity (RD) design with dose intensity, comparing veiling probabilities of women born between 1983 and 1986. Women born before January 1985 were not affected by the reform, while those born between January 1985 and December 1986 were affected to varying degrees: a woman born in January 1985 was exposed for one month, whereas a woman born in December 1986 was exposed for 24 months.

$$y_i = \beta_0 + \beta_1 \, \text{ExposureMonths}_i + \beta_2 X_i + \gamma' Z_i + \varepsilon_i \tag{3}$$

where y_i denotes the veiling probability for woman i, ExposureMonths_i is the number of months for which she was legally restricted by the reform (equal to zero for those born before January 1985, and between 1 and 24 for those born between January 1985 and December 1986), and X_i is the normalized running variable defined as the number of months from the January 1985 cutoff.[42] Our coefficient of interest, β_1, captures the causal effect of an additional month of exposure to the higher marriage age requirement on the probability of veiling. In practice, we report this effect both per month and scaled to a 12-month increase in exposure. We restrict the sample around a $\pm$24-month window around the cutoff, and also report robustness checks with alternative cohorts.

A.2 Conception framework

We draw on Carvalho (2013) to provide a conceptual framework for our empirical findings.

Setup. Consider a community of women indexed by type $i \in \{r, s\}$, where r denotes religious and s denotes secular. The share of religious types in the community is $q \in (0, 1)$. Each woman chooses whether to veil $v \in \{0, 1\}$ and whether to work in public sector $\ell \in \{0, 1\}$. Public sector work requires minimum education $\bar{e}$, and exposes women to temptation to engage in religiously prohibited behavior $p_1 > p_0 = 0$ and yields return $b_1 > b_0 = 0$, representing higher wages and better working conditions. Women with $e_i < \bar{e}$ cannot work in the public sector regardless of the ban.[43] [44]

Preferences. A woman's expected utility is:

$$U_i = p_\ell(1 - v)\lambda_i + p_\ell q(1 - v)\lambda_r + (1 - q)p_\ell(1 - v)\lambda_s + b_\ell - c(v) \tag{4}$$

where $p_\ell(1 - v)$ is the probability of engaging in religiously prohibited behavior. The first term captures intrinsic payoff, where $\lambda_r < 0$ reflects regret for religious types and $\lambda_s > 0$ reflects enjoyment for secular types. The second and third terms capture social payoff from community opinion, weighted by type shares q. The term b_ℓ is the return to public sector work and $c(v)$ is the cost of veiling reflecting discomfort and social discrimination.

Veiling and public sector work. A woman veils ($v = 1$) if and only if $q > \bar{q}_i$, where $\bar{q}_s < \bar{q}_r$.

[42] The running variable is centered at January 1985; positive values indicate birth months after the cutoff, and negative values indicate months before.

[43] Women also work in private or informal sectors, which are not subject to the veiling ban removal. We therefore focus on public sector employment, the margin directly affected by the reform.

[44] We introduce several modifications to the theoretical framework of Carvalho (2013). First, veiling choice is modeled as binary rather than continuous, to align with our empirical setting. Second, we explicitly model "integration" as employment in the public sector choices and assume symmetric effects of the introduction and removal of the veiling ban. Third, we allow the share of qualified women, $N(q)$, to vary with community religiosity q, whereas it is held constant in the original framework. This extension captures the qualification requirements associated with public sector employment.

Without a ban, a woman works in the public sector if $b > b_i(q)$, where:

$$b_i(q) = -p_1(1-v)[\lambda_i + q\lambda_r + (1-q)\lambda_s] + c(v) \tag{5}$$

evaluated at optimal $v \in \{0,1\}$. Under a ban forcing $v = 0$ when working in public sector, the threshold rises to:

$$\beta_i(q) = -p_1[\lambda_i + q\lambda_r + (1-q)\lambda_s] + c(0) \tag{6}$$

Since veiling reduces cost, $\beta_i(q) > b_i(q)$ for regions with $q > \bar{q}_i$; in regions below $\bar{q}_i$, women do not veil so the ban is irrelevant. [45]

Among women with $e_i \geq \bar{e}$, four groups emerge based on community religiosity q and return to public sector work b; women with $e_i < \bar{e}$ never work regardless and are absorbed into the "Never work" group:

- **Low religiosity regions** ($q \leq \bar{q}_i$): women do not veil ($v = 0$) and work decisions are unchanged across regimes; women work if $b > b_i(q)$.
- **Always work** ($q > \bar{q}_i$, $b > \beta_i(q)$): women work in both regimes; veil without ban ($v = 1$), unveil under ban ($v = 0$).
- **Compliers** ($q > \bar{q}_i$, $b_i(q) < b < \beta_i(q)$): veil but do not work under ban ($v = 1, \ell = 0$); veil and work after ban removal ($v = 1, \ell = 1$).
- **Never work** ($q > \bar{q}_i$, $b < b_i(q)$): veil and never work in public sectors ($v = 1, \ell = 0$) throughout.

The aggregate effects of the reform depend on the relative population shares across the four groups, providing a test of which group is most relevant across regions.

Compliers are the only group whose work decision responds to the ban removal while veiling remains unchanged. Ban removal therefore shifts their public sector work choice from $\ell = 0$ to $\ell = 1$ holding $v = 1$ fixed. Compliers are women living in regions above the religiosity threshold, where greater social disapproval of working unveiled leads them to always veil regardless of their work decision. They face moderate returns to public sector work, high enough to work when veiling is permitted but insufficient to offset the social cost of working without the veil under the ban. Hence, the reform shifts only their work decision while leaving their veiling status unchanged, **providing theoretical grounding for our assumption that veiling is a stable characteristic that identifies the treated population.**

The size of the complier pool in a region depends on two forces. First, only educationally qualified women ($e_i \geq \bar{e}$) face the work decision, so regions with more qualified women mechanically have more women in the responsive window $b_i(q) < b < \beta_i(q)$. Since education and religiosity are negatively correlated across regions, let $N(q)$ denote the share of qualified women ($e_i \geq \bar{e}$) in a region with religiosity q, so that $\frac{dN(q)}{dq} < 0$. Low q regions therefore have more qualified women in the responsive window $b_i(q) < b < \beta_i(q)$. Figure 7 provides empirical evidence that the share of women with at least primary school education declines with community religiosity. Second, the width of the responsive window $\beta_i(q) - b_i(q)$ itself varies with q. In lower religiosity regions, the social cost of working without the veil is smaller, reducing $\beta_i(q)$ and narrowing the window. These two forces work in opposite directions, with

[45] The veiling decision follows Proposition 1 of Carvalho (2013). The public sector participation thresholds, both in the absence of a ban and under a ban, are derived from Proposition 7.

low q regions having more qualified women but a narrower responsive window, and high q regions having fewer qualified women but a wider window. Whether the education force dominates, and the reform generate stronger employment effects is a prediction we can address.

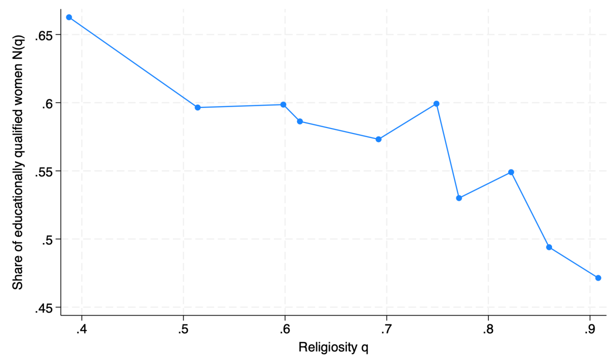


Figure 7: Share of educationally qualified women by religiosity

Testable predictions. The framework delivers three predictions:

1. **Veiled women increase public sector employment after ban removal:** Ban removal lowers work threshold from $\beta_i(q)$ to $b_i(q)$ for qualified veiled women, enabling those with $b_i(q) < b < \beta_i(q)$ to enter the public sector.

2. **Stronger effects among higher educated women:** Only women with $e_i \geq \bar{e}$ face the decision of working in public sectors. Ban removal is irrelevant for unqualified women, generating stronger effects among higher educated women.

3. **Stronger effects in lower q regions:** Since $\frac{dN(q)}{dq} < 0$, low q regions have more qualified women in the responsive window $b_i(q) < b < \beta_i(q)$, producing larger employment responses to ban removal.

Prediction (1) is consistent with our DiD result in Table 3, while predictions (2) and (3) are consistent with our triple DiD findings.

A.3 Machine learning techniques

Machine learning (ML) techniques are increasingly applied to address missing data problems (Chen and McCoy, 2024) and to improve out-of-sample predictions (Athey and Imbens, 2019). Compared with other statistical methods, ML approaches offer advantages such as flexible functional forms, greater computational efficiency and higher accuracy (Mullainathan and Spiess, 2017). Given the common challenges of poor data quality in various fields, a variety of ML methods have been proposed to tackle missing value issues in different contexts (Gogas and Papadimitriou, 2021). Applications in economics include healthcare data (Mullainathan and Obermeyer, 2022), asset pricing (Goldstein et al., 2021), and other areas.

The ML approaches are typically categorized into supervised, unsupervised, and semi-supervised learning. The key distinction lies in whether labeled datasets are available—that is, whether clear relationships between input features and corresponding output are established. In this paper, we use supervised learning, as we have observed demographic characteristics as inputs and veiling status as outputs derived from the DHS dataset. Supervised learning achieves high accuracy when the training dataset is abundant and well-labeled. Since the DHS is a high-quality, nationally representative, and randomized dataset, we can reasonably expect strong performance from the applied ML techniques.

The procedure is divided into two main steps: training and prediction. In the training phase, we analyze the DHS dataset to identify patterns between observed demographic characteristics and the headscarf-wearing behavior of female respondents. Using a selected ML algorithm, we develop a model capable of predicting headscarf-wearing outcomes for unseen inputs. To train the model, we split the DHS dataset into training and test sets. The algorithm uses the training set to fit the model, and its predictions are compared against the true outputs in the test set. The model's parameters are adjusted iteratively to minimize the difference between its predictions and the actual outcomes. In the prediction phase, the optimized model is applied to the HLFS dataset to predict unknown veiling outcomes based on observed demographic inputs. Several key algorithms have demonstrated strong performance in recent studies.

Ensemble learning combines multiple models to improve accuracy, robustness, and flexibility by aggregating their predictions (Breiman, 1996). The approach is particularly effective when individual algorithms capture different data patterns or when a single model lacks sufficient accuracy. The method also helps to mitigate overfitting in small or noisy datasets. A variety of models can be incorporated into ensemble learning.

For our robustness checks in scarf prediction, we implement several machine learning approaches. First, we apply Support Vector Machines, Random Forests, K-Nearest Neighbors and Neural Networks individually. We then combine these models through ensemble learning to leverage their complementary strengths. A detailed discussion of each model and its parameter setup is provided in the following sections.

A.3.1 Support Vector Machines

Support Vector Machines (SVM) is a powerful algorithm for data classification (Cortes, 1995). The core principle is to find an optimal decision boundary, known as a hyperplane, that separates two classes of data points. The objective is to maximize the margin—the distance between the hyperplane and the nearest data points from each class, called support vectors—while ensuring accurate classification.[46]

Consider a two-class classification problem with training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i$ are feature vectors and $y_i \in \{-1, 1\}$ are class labels. In the primal form of SVM, the goal is to find a hyperplane that optimally separates the two classes. The decision function is a linear function defined as:

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b$$

where $\mathbf{w} \in \mathbb{R}^n$ is the weight vector (orthogonal to the decision hyperplane), $\mathbf{x} \in \mathbb{R}^n$ is the input feature vector, and $b \in \mathbb{R}$ is the bias term. The dot product $\mathbf{w} \cdot \mathbf{x}$ represents the projection of the data point x onto vector w, which plays a key role in defining the decision boundary.

For classification, the predicted class is determined by the sign of $f(x)$. If $f(x) > 0$, predict class is $+1$; and if $f(x) < 0$, predict class is -1.

The primal optimization problem aims to find the weight vector w and bias term b that define the optimal separating hyperplane. This is formulated as:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

[46] Maximizing the margin helps create a more robust decision boundary, making it less sensitive to noise and outliers.

subject to the classification constraints:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \text{for all } i$$

The objective function $\frac{1}{2}\|\mathbf{w}\|^2$ is minimized to maximize the margin between the two classes, while the constraints ensures that each data point is classified correctly.

When the data is non-linearly separable, the approach is to map the data into a higher-dimensional space using a feature mapping function $\phi(x)$. In this higher-dimensional space, the data becomes linearly separable, and the decision function is given by:

$$f(x) = \mathbf{w} \cdot \phi(\mathbf{x}) + b$$

Instead of explicitly computing $\phi(x)$, the kernel trick is used to replace the dot product $\mathbf{w} \cdot \mathbf{x}_i$ with a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$, which directly computes the similarity between data points in the higher-dimensional space.

The dual form of the SVM optimization problem is then written in terms of the Lagrange multipliers α_i:

$$\max_{\alpha_i} \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right)$$

subject to the constraints:

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C$$

Here the C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors. $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function. We have several commonly used kernel functions include: linear, polynomial and radial basis function Kernel. A key advantage of SVM is its flexibility in handling complex data structures by employing different kernel functions, making it well-suited for nonlinear classification problems. After performing a grid search to optimize model parameters for maximum training accuracy, we configure the SVM model as follows:

```
svm_model = SVC(C=5, kernel='poly', gamma='auto', degree=3, coef0=0.0,
                class_weight='balanced', probability=True, random_state=77)
```

We use a polynomial kernel of degree $d = 3$ to capture nonlinear decision boundaries in the higher-dimensional feature space, defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \cdot \mathbf{x}_j + r)^d$$

where γ is set to `auto`, computed as $1/n$ where n is the dimension of feature vector $\mathbf{x}_i$, scaling the dot product in the kernel function. The intercept term r is set to 0.0. The regularization parameter $C = 5$ controls the trade-off between maximizing the margin and minimizing classification errors, as defined in the dual optimization problem above. Finally, the balanced class weight accounts for class imbalance by automatically adjusting weights inversely proportional to class frequencies, preventing the model from being biased toward the majority class.

A.3.2 Random Forest

Random Forest is a powerful machine learning algorithm known for its ability to identify important features and its robustness to noise and outliers, making it highly resistant to overfitting (Breiman, 2001). It is a reliable choice for a variety of tasks, especially classification. The implementation involves several key steps:

The first step is "Bootstrap Sampling". We generate multiple bootstrap samples from the original dataset. Each sample $D_{\text{bootstrap}}$ is a random subset created by sampling with replacement. [47]

$$D_{\text{bootstrap}} = \{\text{Sample from } D \text{ with replacement}\}$$

where D is the original dataset.

In the second step "Training Decision Trees", a decision tree is independently trained on each bootstrap sample $D_{\text{bootstrap}}$. Here's how the training process works:

At each node of the tree, the algorithm evaluates which feature and corresponding split threshold best separates the data into subsets. This is done by assessing potential splits using impurity measures such as Gini impurity or entropy. The goal is to find the feature and threshold that minimize the impurity, creating more homogeneous subsets.

$$Gini(D) = 1 - \sum_{i=1}^{C} p_i^2$$

where p_i is the proportion of class i in the node, and C is the number of classes. A lower Gini impurity means the node is more pure (i.e., most of the samples in the node belong to the same class), while a higher Gini impurity indicates that the samples are more mixed across different classes.

Once the best feature and threshold are identified for the current node, the dataset is split into two child nodes based on the chosen feature and value. This process is repeated recursively for each child node in the same way, each time selecting the best feature to further split the data, until a stopping criterion is met. Some common stopping criteria include: maximum depth of the tree, minimum samples per leaf, minimum impurity split and etc.

The third step is "Prediction". Once all trees are trained, predictions are made by aggregating the outputs of all trees. A "majority vote" is taken to determine the final class label. The class that receives the most votes from the individual trees is selected as the final prediction:

$$\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T)$$

where $\hat{y}_i$ is the predicted class from tree i, and T is the total number of trees.

After performing a grid search to find the optimal parameters for the best accuracy, we configure the RF model as follows:

```
rf_model = RandomForestClassifier(n_estimators=100, max_depth=10,
                min_samples_leaf=4, min_samples_split=10, random_state=77)
```

- `n_estimators` = 100: This specifies the number of trees in the random forest. A higher number improves the model's robustness but also increases computation time.

The following three parameters are stopping criteria that control how the individual decision trees are trained:

- `max_depth` = 10: This sets the maximum depth of each tree. Limiting depth prevents the trees from becoming too complex and overfitting to the training data.

- `min_samples_leaf` = 4: This controls the minimum number of samples required in each leaf node after a split. It ensures each leaf contains at least 4 samples, which prevents overly specific splits and helps avoid overfitting.

[47] Some data points may be repeated, and some may be left out.

- `min_samples_split` = 10: This defines the minimum number of samples required to split an internal node. Nodes with fewer than 10 samples cannot be split further. This helps in reducing overfitting by ensuring that splits are based on sufficiently large subsets of data.

A.3.3 K-Nearest Neighbors

K-Nearest Neighbor is a simple and powerful instance-based algorithm (Aha et al., 1991). Unlike other models, K-NN doesn't require an explicit training phase. Instead, it stores the training dataset and uses it directly for prediction. To predict the output for a new data point, K-NN calculates the distance between the new point and all existing points in the training set, typically using Euclidean distance or another distance measure. It then identifies the K-nearest neighbors and classifies or regresses based on the labels or values of the neighbors. We select the optimal parameters for the model as follows:

```
knn_model = KNeighborsClassifier(n_neighbors=15, algorithm='ball_tree',
    leaf_size=50, metric='minkowski', p=1, weights='uniform')
```

The parameters are as follows:

- `n_neighbors` = 15: This sets the number of nearest neighbors considered when making a prediction. A larger value produces smoother decision boundaries and reduces sensitivity to noise.

- `algorithm=ball_tree`: This specifies the algorithm used to compute nearest neighbors. The Ball Tree algorithm organizes data points into a tree structure of nested hyperspheres, enabling efficient nearest neighbor searches in higher-dimensional spaces.

- `leaf_size` = 50: This controls the size of the leaf nodes in the Ball Tree structure. A larger leaf size of 50 reduces the depth of the tree, speeding up construction and query time at the cost of slightly slower nearest neighbor searches within each leaf.

`metric=minkowski, p=1`: This defines the Minkowski distance metric to compute distances between data points. With p=1, the Minkowski metric reduces to the Manhattan distance, which measures the sum of absolute differences across all features.

`weights=uniform`: This sets all neighbors to contribute equally to the prediction regardless of their distance to the query point, treating each of the 15 nearest neighbors with equal importance.

A.3.4 Neural Networks

Neural networks are powerful machine learning models inspired by the structure of the human brain, consisting of interconnected neurons organized into layers that perform mathematical operations to recognize complex patterns in data (Rumelhart et al., 1986). Unlike the other classifiers, the neural network retains the continuous predicted probability directly as the treatment variable in the DiD regression, while in the ensemble learning model a threshold of 0.5 is applied to the neural network's predicted probabilities when combining it with the other classifiers to produce a binary veiling assignment. After performing a grid search to find the optimal parameters, we establish the following architecture:

```
nn_model = Sequential([
        Input(shape=(X_train_scaled.shape[1],)),
        Dense(64, activation='relu'),
        BatchNormalization(),
        Dropout(0.2),
        Dense(64),
```

```
        LeakyReLU(negative_slope=0.1),
        Dropout(0.2),
        Dense(32),
        LeakyReLU(negative_slope=0.1),
        Dense(1, activation='sigmoid')])
    nn_model.compile(optimizer=Adam(learning_rate=0.001),
        loss='binary_crossentropy',metrics=['accuracy'])
    nn_model.fit(X_train_scaled, y_train,
        epochs=30,batch_size=32,validation_data=(X_test_scaled, y_test),verbose=0)
```

- `Input Layer`: Contains $X\_train.shape[1]$ neurons, equal to the number of input features.

- `Hidden Layer1`: A dense layer with 64 neurons using the ReLU activation function $f(x) = \max(0, x)$ to introduce non-linearity, followed by Batch Normalization to stabilize and accelerate training by normalizing layer outputs, and a Dropout rate of 0.2 to randomly deactivate 20% of neurons during training to prevent overfitting.

- `Hidden Layer 2`: A dense layer with 64 neurons using the Leaky ReLU activation function $f(x) = \max(0.1x, x)$ with a negative slope of 0.1, allowing small gradients for negative inputs to prevent the dying ReLU problem, followed by a Dropout rate of 0.2.

- `Hidden Layer 3`: A dense layer with 32 neurons using the Leaky ReLU activation function with a negative slope of 0.1.

- `Output Layer`: A single neuron with a sigmoid activation function $f(x) = \frac{1}{1+e^{-x}}$ to produce binary classification probabilities in the range [0,1].

The model is compiled using the Adam optimizer with a learning rate of 0.001, which adaptively adjusts parameter updates during training. Binary cross-entropy is used as the loss function, appropriate for binary classification tasks. The model is trained for 30 epochs with a batch size of 32, and accuracy is monitored on the validation set during training to assess performance.

A.4 Misclassification bias

Both the PSM and the alternative machine-learning imputations rely on predicted rather than observed veiling status, making the constructed treatment indicator subject to misclassification error (Meyer et al., 2015; Tommasi and Zhang, 2024). In this section, we discuss the direction of bias and argue that sign reversal is unlikely in our setting.

In the imputation, some truly veiled women may be assigned to the predicted non-veiled group (false negatives), while some truly non-veiled women may be assigned to the predicted veiled group (false positives). Our imputation relies on observable demographics such as education and region, but cannot capture all unobserved determinants of veiling status and labor market participation, such as religiosity or gender norms. These unobserved characteristics may simultaneously correlated with veiling and labor market outcomes, making our misclassification differential. [48] While formal correction methods for differential misclassification exist (Nguimkeu et al., 2019), they are not applicable to our setting due to data limitations.[49]

Instead, we draw on Theorem 2 of Negi and Negi (2025) to argue that sign reversal is implausible in our setting, and further discuss about the direction of our misclassification bias. [50] The naive DID estimator converges to the true estimator ATT plus bias terms:

$$\text{DiD} = \text{ATT} + \underbrace{A\delta}_{\text{covariate mean bias}} + \underbrace{B\theta}_{\text{misclassification bias}} + \underbrace{C}_{\text{differential bias}}$$

The covariate mean bias ($A\delta$) arises because false negatives are excluded from the observed treatment group, causing the regression to use the wrong covariate reference point and misspecifying the treatment effect. The misclassification bias ($B\theta$) arises because truly veiled false negatives appear in the control group rather than the treatment group, where their positive employment gains from the ban removal inflate the control group trend and compress the estimated treatment-control difference. Under one-sided misclassification, both $A\delta$ and $B\theta$ attenuate the DiD estimate, though their signs depend on the covariate and outcome structure of the false negatives relative to correctly classified treated women.

For our positive DiD estimates to reflect a spurious positive after sign reversal, the necessary condition is that the true ATT must be negative, meaning the ban removal reduced employment of truly veiled women. This is economically unlikely as the ban removal expanded the set of jobs available to veiled women by lifting legal restrictions on their employment in the public sector. An expanded choice set is unlikely to reduce veiled women's employment.

Furthermore, even if the necessary condition is met, the differential bias term would additionally need to be positive and large. This would require false negatives to have smaller employment gains, or larger employment declines, than correctly classified veiled women. [51]

[48] There are two types of misclassification error: nondifferential and differential. Nondifferential misclassification means the classification error is independent of outcomes conditional on true treatment status, producing an attenuation bias but preserves the sign of the true effect. Differential misclassification means the error correlates with outcomes, making the bias direction harder to characterize.

[49] The correction methods commonly require an instrument that affects misclassification but not the true treatment or outcomes, which is not available in our data.

[50] Negi and Negi (2025) exclusively address one-sided misclassification with false negatives only, while our setting involves two-sided errors. False positives — truly unveiled women incorrectly imputed as veiled — dilute the treatment group with women who experience no treatment effect, attenuating the DiD estimate toward zero without reversing its sign. False negatives — truly veiled women incorrectly imputed as unveiled — are misassigned to the control group and can in principle generate sign reversal under differential misclassification when the false negative rate is sufficiently high. Hence, we draw on this paper to characterize the false negative bias channel, as it provides the most explicit formal framework for examining when sign reversal can occur.

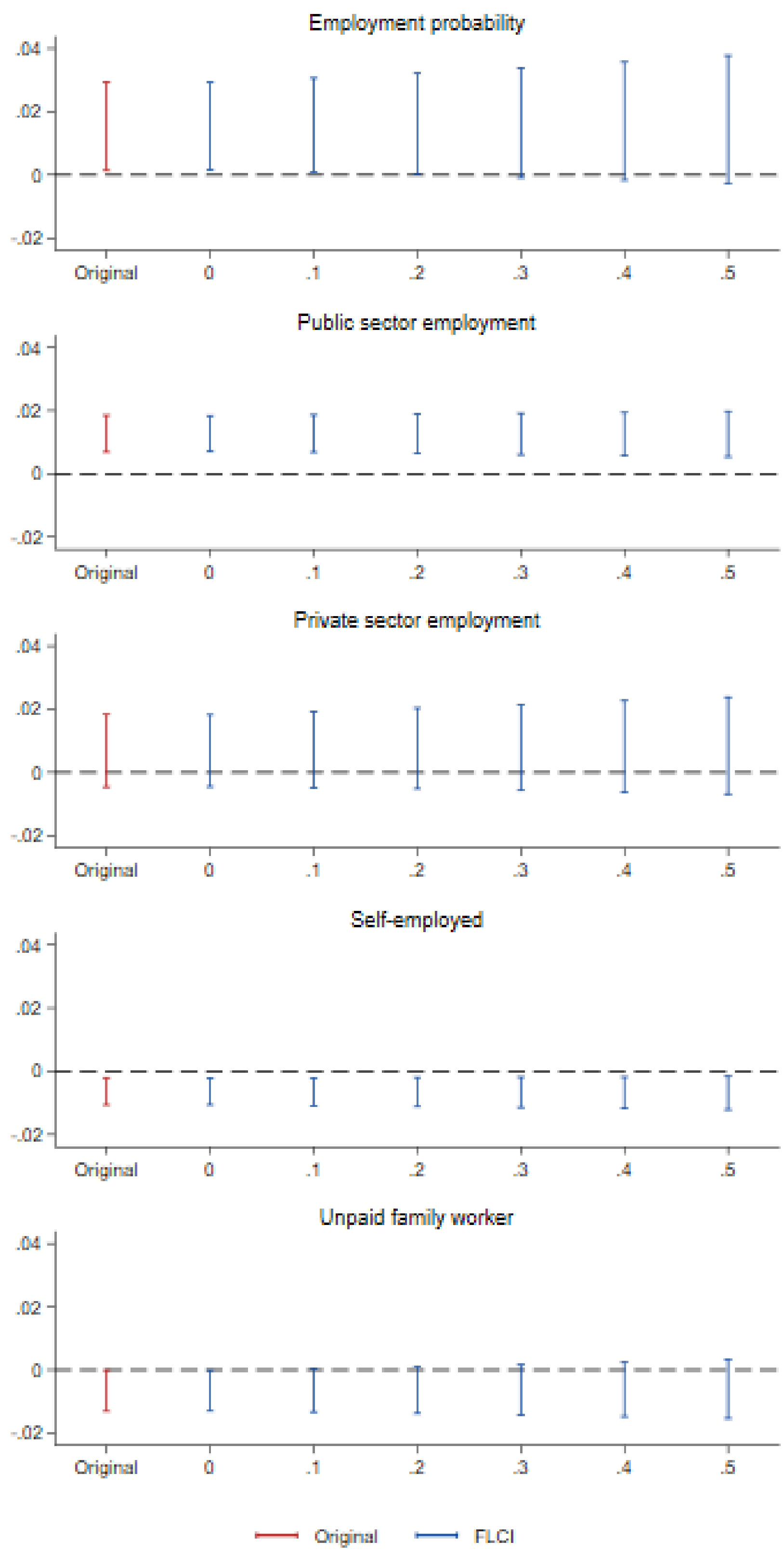
[51] See proof of Theorem 2. This argument is consistent with the Roy selection condition shown in Denteh and Kédagni (2022).

This condition contradicts the economic mechanism of the policy change and is not supported by the data. Table A4 shows that false negatives tend to be more highly educated, suggesting they stand to gain more from ban removal rather than experience more severe declines, which would make C negative and further attenuate rather than reverse the DiD estimate.

Taken together, our imputation procedure inevitably introduces misclassification errors that bias our DiD estimates. Since the necessary condition for sign reversal is economically unlikely and the additional condition is implausible given our data, we argue that our positive DiD estimates reflect a genuine positive ATT rather than a spurious reversal of a negative effect. However, we cannot determine the direction of bias of our DiD estimates relative to the true ATT, that is, whether our estimates attenuate or exaggerate the true effect. Correcting for this bias would require either a valid instrument that affects misclassification but not outcomes, or direct observation of true veiling status in THLFS, neither of which is available in our data, and we therefore leave this as a direction for future research.

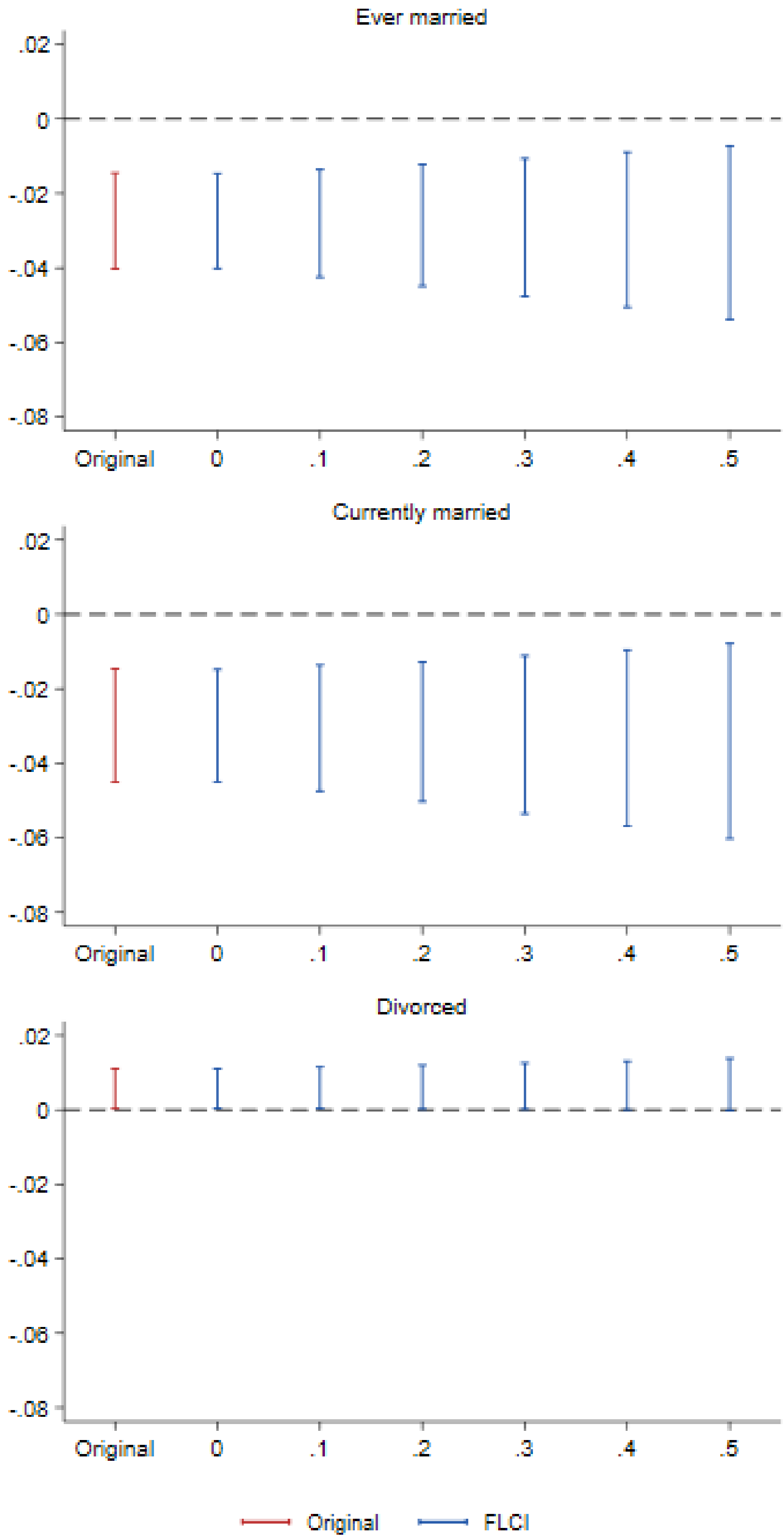
A.5 Figures

Figure A1: Effects on labor market outcomes: sensitivity analysis



Notes: Figures display sensitivity analysis of estimated treatment effects on labor market outcomes of women to potential violations of the parallel trends assumptions outlined in Rambachan and Roth (2023). The red dashed bar in each panel represents the 95% confidence interval of our baseline estimates. The black solid bars represent corresponding 95% confidence intervals when allowing for pre-reform period violations of parallel trends of up to M, indicating the largest allowable change in the slope of an underlying linear trend between two consecutive years. The baseline estimate is obtained from our two-step estimation procedure, and inference is based on a covariance matrix constructed from the full two-step bootstrap, which resamples both the imputation stage and the second-stage estimation.

Figure A2: Effects on marriage probability: sensitivity analysis



Notes: Figures display sensitivity analysis of estimated treatment effects on the marriage probability of women to potential violations of the parallel trends assumptions outlined in Rambachan and Roth (2023). The red dashed bar in each panel represents the 95% confidence interval of our baseline estimates. The black solid bars represent corresponding 95% confidence intervals when allowing for pre-reform period violations of parallel trends of up to M, indicating the largest allowable change in the slope of an underlying linear trend between two consecutive years. The baseline estimate is obtained from our two-step estimation procedure, and inference is based on a covariance matrix constructed from the full two-step bootstrap, which resamples both the imputation stage and the second-stage estimation.

A.6 Tables

Table A1: *Occupations in the public sector (%)*

	Men	Women
Legislators and Senior Officials	4.06	3.02
Professionals	41.09	57.81
Technicians and Associate Professionals	12.49	12.71
Clerks	14.69	17.34
Service Workers and Shop and Market Sales Workers	15.14	6.03
Skilled Agricultural and Fishery Workers	0.29	0.11
Craft and Related Trades Workers	2.59	0.37
Plant and Machine Operators and Assemblers	2.44	0.08
Elementary Occupations	7.19	2.54
Total	100.00	100.00

Notes: The share of men and women in each occupation is calculated using employed samples from the Turkish Household Labor Force Survey. We follow the SIC-92 classifications.

Table A2: *Summary statistics using the HLFS dataset*

	All sample	Non-veiled	Veiled
	Mean (S.D) (1)	Mean (S.D) (2)	Mean (S.D) (3)
Panel A: Background characteristics			
Age	32.88	31.46	33.79
	(8.906)	(8.564)	(9.000)
Native	0.978	0.960	0.989
	(0.148)	(0.196)	(0.105)
University degree	0.151	0.322	0.042
	(0.358)	(0.467)	(0.200)
High school degree and above	0.352	0.640	0.169
	(0.478)	(0.480)	(0.374)
Primary school degree and below	0.494	0.228	0.664
	(0.499)	(0.420)	(0.472)
Married	0.713	0.582	0.797
	(0.452)	(0.493)	(0.402)
Household size	3.925	3.235	4.271
	(1.914)	(1.310)	(2.070)
Panel B: Labor market outcomes			
Employed	0.330	0.395	0.289
	(0.470)	(0.489)	(0.453)
Employed as wage earner	0.222	0.339	0.147
	(0.415)	(0.474)	(0.354)
Public sector employment	0.051	0.102	0.019
	(0.221)	(0.302)	(0.138)
Self-employed	0.034	0.0325	0.036
	(0.184)	(0.177)	(0.187)
Unpaid family worker	0.083	0.044	0.108
	(0.276)	(0.204)	(0.311)
Obs.	933,810	328,862	604,948

Notes: Data are from the Turkish HLFS 2010-17. The sample covers all women aged 18-49. The table presents the means, and standard deviations (in parenthesis) of selected variables for the sample of all women in column 1, for the sample of women who are predicted to be non-veiled women in column 2, and for those predicted to be veiled in column 3, respectively. Data are weighted using the cross-sectional weights for the wave at which the outcome was measured.

Table A3: *Summary statistics for veiled and non-veiled women from the 2013 DHS*

	Non-veiled	Veiled	
	(1)	(2)	(3)
	Mean	Mean	Difference
	(S.D)	(S.D)	(1)-(2)
Panel A: Background characteristics			
Age	31.17	33.61	-2.446***
	(8.631)	(8.799)	(0.234)
Native	0.951	0.989	-0.037***
	(0.215)	(0.106)	(0.005)
University degree	0.387	0.067	0.320***
	(0.487)	(0.250)	(0.011)
High school degree and above	0.468	0.109	0.359***
	(0.499)	(0.311)	(0.012)
Primary school degree and below	0.204	0.645	-0.441***
	(0.008)	(0.007)	(0.011)
Married	0.616	0.841	-0.225***
	(0.487)	(0.366)	(0.012)
Household size	3.752	5.067	-1.314***
	(1.335)	(2.296)	(0.047)
Panel B: Labor market outcomes			
Employed	0.427	0.268	0.159***
	(0.495)	(0.443)	(0.013)
Employed as wage earner	0.334	0.0839	0.250***
	(0.472)	(0.277)	(0.011)
Employed in public sector	0.095	0.012	0.083***
	(0.293)	(0.108)	(0.006)
Self-employed	0.038	0.054	-0.016***
	(0.190)	(0.227)	(0.005)
Unpaid family worker	0.022	0.091	-0.068***
	(0.147)	(0.287)	(0.005)
Obs.	2,793	6,005	

Notes: The data are from the 2013 Turkish Demographic and Health Survey. Sample covers all women aged 18-49. The table presents the means, and standard deviations (in parenthesis) of selected variables for non-veiled women in column 1, and for veiled women in column 2, and the difference between non-veiled and veiled women in column 3. ***, significant at the 1 percent level; **, significant at the 5 percent level; *, significant at the 10 percent level.

Table A4: *Characteristics of misclassified and correctly classified women in the DHS hold-out sample*

Group	Education			Religious	Traditional
	No educ.	Junior high+	High school+		
Panel A: Probit					
False negatives	0.000	0.939	0.793	0.894	0.538
Correctly classified veiled	0.194	0.205	0.057	0.875	0.616
Difference	-0.194***	0.735***	0.736***	0.019	-0.078**
False positives	0.067	0.374	0.147	0.497	0.519
Correctly classified non-veiled	0.000	0.970	0.907	0.580	0.409
Difference	0.067***	-0.596***	-0.760***	-0.083*	0.110**
Panel B: Ensemble					
False negatives	0.000	0.965	0.922	0.864	0.597
Correctly classified veiled	0.198	0.244	0.093	0.858	0.598
Difference	-0.198***	0.720***	0.829***	0.006	-0.001
False positives	0.000	0.994	0.969	0.506	0.308
Correctly classified non-veiled	0.064	0.522	0.279	0.466	0.429
Difference	-0.064	0.472***	0.690***	0.040	-0.121***
Panel C: Support Vector Machine					
False negatives	0.000	0.907	0.776	0.840	0.559
Correctly classified veiled	0.213	0.203	0.063	0.862	0.607
Difference	-0.213***	0.703***	0.713***	-0.022	-0.047
False positives	0.000	0.992	0.939	0.497	0.321
Correctly classified non-veiled	0.074	0.447	0.214	0.474	0.429
Difference	-0.074***	0.545***	0.725***	0.023	-0.107**
Panel D: Random Forest					
False negatives	0.000	0.946	0.932	0.857	0.614
Correctly classified veiled	0.200	0.242	0.086	0.859	0.596
Difference	-0.200***	0.704***	0.847***	-0.002	0.018
False positives	0.000	0.988	0.953	0.503	0.324
Correctly classified non-veiled	0.070	0.485	0.236	0.467	0.418
Difference	-0.070***	0.504***	0.718***	0.036	-0.094**
Panel E: K-Nearest Neighbor					
False negatives	0.000	0.950	0.868	0.848	0.596
Correctly classified veiled	0.202	0.234	0.087	0.860	0.598
Difference	-0.202***	0.716***	0.781***	-0.012	-0.002
False positives	0.000	0.988	0.944	0.506	0.333
Correctly classified non-veiled	0.067	0.506	0.280	0.464	0.400
Difference	-0.067***	0.482***	0.663***	0.042	-0.067
Panel F: Neural Network					
False negatives	0.000	0.994	0.942	0.864	0.591
Correctly classified veiled	0.201	0.230	0.079	0.858	0.599
Difference	-0.201***	0.763***	0.863***	0.006	-0.008
False positives	0.000	0.997	0.958	0.511	0.302
Correctly classified non-veiled	0.065	0.508	0.280	0.459	0.440
Difference	-0.065	0.489***	0.677***	0.051	-0.137***

Notes: The table uses the DHS hold-out sample. False negatives are truly veiled women predicted as non-veiled. Correctly classified veiled women are truly veiled women predicted as veiled. False positives are truly non-veiled women predicted as veiled. Correctly classified non-veiled women are truly non-veiled women predicted as non-veiled. Entries report group means. Difference rows report the mean for the misclassified group minus the mean for the corresponding correctly classified group. p-values are from two-sided tests of equality of means. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table A5: *Effects on labor market outcomes of non-veiled women*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A					
Veiled × Post	0.012***	0.016***	0.002	-0.003**	-0.009***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.004, 0.019]	[0.011, 0.018]	[-0.004, 0.008]	[-0.005, -0.001]	[-0.012, -0.006]
Post	0.034***	-0.027***	0.048***	0.013***	0.009***
	(0.003)	(0.002)	(0.004)	(0.001)	(0.001)
	[0.028, 0.041]	[-0.029, -0.023]	[0.041, 0.056]	[0.011, 0.016]	[0.007, 0.012]
Panel B					
Veiled × Post	0.011**	0.018***	0.001	-0.004*	-0.010***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.003, 0.019]	[0.013, 0.020]	[-0.007,0.008]	[-0.005, 0.001]	[-0.015, -0.007]
Post	0.037***	-0.028***	0.055***	0.014***	0.006***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.001)
	[0.031, 0.044]	[-0.030, -0.024]	[0.048, 0.063]	[0.011, 0.016]	[0.004, 0.009]
Panel C					
Veiled × Post	0.013**	0.007	0.003	-0.003	-0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.003)
	[0.002, 0.026]	[-0.004, 0.009]	[-0.002 ,0.020]	[-0.009, 0.001]	[-0.004, 0.007]
Post	0.022***	-0.025***	0.020***	0.006***	0.026***
	(0.005)	(0.003)	(0.004)	(0.006)	(0.002)
	[0.011, 0.030]	[-0.027, -0.016]	[0.009, 0.026]	[0.004, 0.011]	[0.020, 0.029]
Year fixed effects	No	No	No	No	No
Observations	933,810	933,810	933,810	933,810	933,810

Notes: Data are from the 2010-2017 HLFS. The sample includes all women aged 18-49. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable ($t = \text{year} - 2010$). Each estimate includes a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region fixed effects, with interaction terms between regions and *Post*. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A6: *Effects of the 1997 education reform on women's religious practices*

Outcome	(1) Veiled	(2) Regularly pray	(3) Fast during Ramadan
Panel A: Static bandwidth			
Born after 1986	-0.058	0.046	-0.013
	(0.045)	(0.038)	(0.035)
Bandwidth	60	60	60
Covariates	No	No	No
Number of obs.	2,875	2,873	2,860
Mean	0.60	0.71	0.80
Panel B: Static bandwidth			
Born after 1986	-0.026	0.053	-0.002
	(0.044)	(0.036)	(0.034)
Bandwidth	60	60	60
Covariates	Yes	Yes	Yes
Number of obs.	2,786	2,784	2,772
Mean	0.58	0.70	0.83
Panel C: Donut-hole			
Born after 1986	-0.022	0.043	-0.001
	(0.047)	(0.037)	(0.034)
Bandwidth	60	60	60
Covariates	Yes	Yes	Yes
Number of obs.	2,633	2,631	2,620
Mean	0.58	0.70	0.83
Panel D: Optimal bandwidth CCT			
Born after 1986	-0.007	0.037	-0.024
	(0.048)	(0.031)	(0.030)
Bandwidth	49	82	74
Covariates	Yes	Yes	Yes
Number of obs.	2,277	3,840	3,440
Mean	0.58	0.70	0.83

Notes: The sample includes all women drawn from the 2013 DHS. Columns 1–3 report local linear RDD estimates for the probability of veiling, regular prayer, and fasting during Ramadan, respectively. All specifications include a linear function of month–year of birth and its interaction with the treatment indicator (born after January 1987). Panel A reports point estimates using a static bandwidth of 60 months, obtained with the Calonico-Cattaneo-Titiunik (CCT) algorithm proposed by Cattaneo et al. (2019), around the cutoff (January 1987). Panel B reports the same estimates while including background covariates: dummies for childhood place of residence, mother tongue, whether the woman's mother is literate, father's education level, the woman's month-of-birth fixed effects, and region-of-residence fixed effects. Panel C presents donut-hole RDD estimates, excluding women born within three months of the cutoff, and Panel D uses outcome-specific CCT optimal bandwidths. Standard errors are clustered at the month–year-of-birth level. ***, significant at the 1 percent level; **, significant at the 5 percent level; *, significant at the 10 percent level.

Table A7: *Effects of the 2002 civil code on women's religious practices*

Outcome	(1) Veiled	(2) Regularly pray	(3) Fast during Ramadan
Panel A: 1983–86 birth cohorts			
MonthExposure	-0.006	-0.003	-0.008***
	(0.004)	(0.004)	(0.003)
Number of obs.	1,144	1,143	1,138
Mean	0.60	0.71	0.84
Panel B: 1982–86 birth cohorts			
MonthExposure	-0.004	-0.004	-0.006**
	(0.003)	(0.003)	(0.003)
Number of obs.	1,452	1,451	1,445
Mean	0.60	0.72	0.84
Panel C: 1981–86 birth cohorts			
MonthExposure	-0.003	-0.003	-0.005**
	(0.003)	(0.003)	(0.002)
Number of obs.	1,754	1,753	1,746
Mean	0.60	0.72	0.84

Notes: The sample includes all women born between 1983 and 1986 in the 2013 DHS. Columns 1–3 report local linear RD estimates of the probability of veiling, regular prayer, and fasting during Ramadan, respectively, based on the dose–response design around the January 1985 cutoff. The running variable is the number of months from the cutoff, centered at January 1985, and interacted with an indicator for post-reform cohorts. Panel A reports point estimates using the 1983–1986 cohorts, Panel B expands the sample to 1982–1986 cohorts, and Panel C to 1981–1986 cohorts. All specifications include dummies for childhood place of residence, mother tongue, whether the woman's mother is literate, father's education level, the woman's month-of-birth fixed effects, and region-of-residence fixed effects. Standard errors are clustered at the month–year-of-birth level. ***, significant at the 1 percent level; **, at the 5 percent level; *, at the 10 percent level.

Table A8: *Short–run effects on veiling probability using 2013 DHS*

	(1)	(2)	(3)	(4)
Panel A. Overall change in veiling				
Post	-0.009	-0.036*		
	(0.018)	(0.019)		
Panel B. Differential change by AKP municipality				
Post			-0.031	-0.059***
			(0.041)	(0.019)
AKP province			0.053	0.039
			(0.037)	(0.033)
Post × AKP province			0.033	0.037
			(0.049)	(0.031)
Observations	8,796	8,796	8,796	8,796
R-squared	0.330	0.330	0.331	0.332

Notes: Data are from the 2013 Turkish DHS. The dependent variable is an indicator equal to one if the respondent wears a headscarf. Columns (1) and (2) report overall before-after differences in veiling, while columns (3) and (4) additionally allow these differences to vary by AKP-governed province. Columns (1) and (3) define the post period as November–December 2013 and the pre period as September–October 2013, so October 2013 is treated as part of the pre period. Columns (2) and (4) define the post period as October–December 2013 and the pre period as September 2013, so October 2013 is treated as part of the post period. AKP-governed province is an indicator equal to one if the respondent lives in a province governed by an AKP mayor. All regressions control for age, age squared, household size, native status, education, and NUTS2 fixed effects. In columns (3) and (4), the coefficient on Post captures the post-period change in non-AKP provinces, while the interaction term captures the additional post-period change in AKP provinces. The implied total post-period change in AKP provinces, computed as the sum of the coefficients on Post and Post × AKP-governed province, is 0.002 ($p = 0.914$) in column (3) and -0.023 ($p = 0.382$) in column (4). Standard errors clustered at the province level are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A9: *Heterogeneous effects - Specification B*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Number of mosques per capita					
Veiled × Post × HighMosque	-0.007	-0.007**	0.006	-0.008***	-0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.003)
	[-0.023, 0.005]	[-0.015, -0.002]	[-0.009, 0.017]	[-0.011, -0.002]	[-0.011, 0.003]
Veiled × Post	0.014**	0.021***	-0.001	0.001	-0.011***
	(0.006)	(0.003)	(0.006)	(0.002)	(0.002)
	[0.004, 0.026]	[0.015, 0.025]	[-0.012, 0.010]	[-0.002, 0.004]	[-0.013, -0.004]
Panel B: Number of Quran course attendants per capita					
Veiled × Post × HighQuran	-0.013	-0.004	-0.003	-0.007***	0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.004)
	[-0.024, 0.001]	[-0.011, 0.001]	[-0.015, 0.011]	[-0.012, -0.002]	[-0.008, 0.005]
Veiled × Post	0.017***	0.020***	0.003	0.001	-0.011***
	(0.006)	(0.003)	(0.006)	(0.002)	(0.002)
	[0.004, 0.028]	[0.013, 0.023]	[-0.009, 0.013]	[-0.002, 0.004]	[-0.014, -0.005]
Panel C: Education level					
Veiled × Post × HigherEduc	0.025***	0.011***	0.004	0.006	0.007
	(0.008)	(0.003)	(0.007)	(0.003)	(0.004)
	[0.001, 0.030]	[0.006, 0.017]	[-0.012, 0.020]	[-0.001, 0.010]	[-0.008, 0.009]
Veiled × Post	-0.014	0.004	-0.007	-0.005	-0.010
	(0.007)	(0.001)	(0.006)	(0.003)	(0.004)
	[-0.021, 0.006]	[-0.001, 0.004]	[-0.015, 0.006]	[-0.008, 0.002]	[-0.013, 0.002]
Panel D: Regional share of Kurdish population					
Veiled × Post × HighShareKurd	0.034***	0.012*	0.002	-0.002	0.007***
	(0.013)	(0.007)	(0.008)	(0.003)	(0.007)
	[0.022, 0.073]	[-0.002, 0.027]	[-0.017, 0.016]	[-0.009, 0.005]	[0.004, 0.031]
Veiled × Post	0.009**	0.017***	0.001	-0.003	-0.012***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.001, 0.016]	[0.012, 0.019]	[-0.007, 0.009]	[-0.004, 0.001]	[-0.017, -0.009]
Panel E: Political alignment (AKP vote share)					
Veiled × Post × HighVoteShare	0.018	0.007	0.003	0.005	-0.001
	(0.007)	(0.003)	(0.006)	(0.003)	(0.003)
	[-0.009, 0.019]	[-0.005, 0.008]	[-0.009, 0.014]	[-0.002, 0.009]	[-0.006, 0.007]
Veiled × Post	0.003	0.014***	0.001	-0.006**	-0.011***
	(0.005)	(0.002)	(0.004)	(0.002)	(0.002)
	[-0.001, 0.017]	[0.010, 0.020]	[-0.005, 0.009]	[-0.007, -0.001]	[-0.015, -0.006]
Observations	933,810	933,810	933,810	933,810	933,810

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports heterogeneity estimates based on Eq. (1), augmented by an additional interaction term between $Veiled$ and an indicator for the 2013 survey wave to account for differential trends in the transition year. $HighMosque$ is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median number of mosques per capita. $HighQuran$ is a binary variable equal to one if the woman resides in a region with an above-median number of female Quran course enrollees per 10,000 women. $HigherEduc$ is a binary variable equal to one if the woman holds at least a junior high school degree, and zero if she has no formal education or only a primary school degree. $HighShareKurd$ is a binary variable equal to one if the woman resides in a region where more than 50 percent of women report Kurdish as their mother tongue. $HighVoteShare$ is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median vote share of the Justice and Development Party (AKP) in the 2014 municipal elections. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A10: *Heterogeneous effects - Specification C*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Number of mosques per capita					
Veiled × Post × HighMosque	-0.007	-0.007**	0.006	-0.008***	-0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.003)
	[-0.023, 0.005]	[-0.015, -0.002]	[-0.008, 0.017]	[-0.011, -0.002]	[-0.011, 0.003]
Veiled × Post	0.015*	0.012*	-0.001	-0.001	-0.002
	(0.008)	(0.004)	(0.007)	(0.003)	(0.003)
	[-0.001, 0.031]	[-0.001, 0.014]	[-0.009, 0.020]	[-0.008, 0.001]	[-0.007, 0.007]
Panel B: Number of Quran course attendants per capita					
Veiled × Post × HighQuran	-0.013	-0.004	-0.002	-0.007**	0.001
	(0.007)	(0.003)	(0.006)	(0.002)	(0.004)
	[-0.024, 0.001]	[-0.011, 0.001]	[-0.015, 0.011]	[-0.012, -0.002]	[-0.008, 0.006]
Veiled × Post	0.017**	0.011*	0.003	-0.002	-0.003
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[0.001, 0.030]	[-0.001, 0.016]	[-0.007, 0.021]	[-0.008, 0.003]	[-0.009, 0.006]
Panel C: Education level					
Veiled × Post × HigherEduc	0.025***	0.011***	0.004	0.006	0.007
	(0.008)	(0.003)	(0.007)	(0.003)	(0.004)
	[0.001, 0.030]	[0.006, 0.017]	[-0.012, 0.020]	[-0.001, 0.009]	[-0.008, 0.009]
Veiled × Post	-0.015	-0.004***	-0.007	-0.007**	-0.003
	(0.008)	(0.003)	(0.008)	(0.003)	(0.005)
	[-0.023, 0.008]	[-0.017, -0.002]	[-0.014, 0.015]	[-0.013, -0.001]	[-0.007, 0.011]
Panel D: Regional share of Kurdish population					
Veiled × Post × HighShareKurd	0.034***	0.012*	0.002	-0.002	0.007***
	(0.013)	(0.007)	(0.008)	(0.003)	(0.007)
	[0.022, 0.073]	[-0.002, 0.027]	[-0.017, 0.016]	[-0.009, 0.004]	[0.005, 0.031]
Veiled × Post	0.009	0.008	0.002	-0.005**	-0.003
	(0.007)	(0.004)	(0.006)	(0.002)	(0.003)
	[-0.003, 0.022]	[-0.004, 0.010]	[-0.005, 0.019]	[-0.011, -0.002]	[-0.009, 0.004]
Panel E: Political alignment (AKP vote share)					
Veiled × Post × HighVoteShare	0.018	0.007	0.003	0.005	-0.001
	(0.007)	(0.003)	(0.006)	(0.003)	(0.003)
	[-0.009, 0.019]	[-0.005, 0.008]	[-0.009, 0.014]	[-0.002, 0.009]	[-0.006, 0.007]
Veiled × Post	0.003	0.005	0.001	-0.007**	-0.003
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.008, 0.024]	[-0.010, 0.006]	[-0.006, 0.020]	[-0.013, -0.002]	[-0.006, 0.008]
Observations	933,810	933,810	933,810	933,810	933,810

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports robustness estimates of the baseline specification in Panel C of Table 3 that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable ($t = \text{year} - 2010$). *HighMosque* is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median number of mosques per capita. *HighQuran* is a binary variable equal to one if the woman resides in a region with an above-median number of female Quran course enrollees per 10,000 women. *HigherEduc* is a binary variable equal to one if the woman holds at least a junior high school degree, and zero if she has no formal education or only a primary school degree. *HighShareKurd* is a binary variable equal to one if the woman resides in a region where more than 50 percent of women report Kurdish as their mother tongue. *HighVoteShare* is a binary variable equal to one if the woman resides in a NUTS2 region with an above-median vote share of the Justice and Development Party (AKP) in the 2014 municipal elections. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A11: *Sensitivity to high-confidence veiling predictions*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Full sample					
Veiled × Post	0.012***	0.017***	0.002	-0.003*	-0.010***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.005, 0.019]	[0.011, 0.018]	[-0.004, 0.009]	[-0.005, 0.001]	[-0.012, -0.007]
Veiled	-0.010***	-0.006***	-0.003	0.001	0.001
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Observations	933,810	933,810	933,810	933,810	933,810
Panel B: *pscore* > 0.70 **or** *pscore* < 0.30					
Veiled × Post	0.014***	0.019***	0.002	-0.001	-0.012***
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
	[0.006, 0.025]	[0.013, 0.023]	[-0.006, 0.011]	[-0.003, 0.003]	[-0.016, -0.008]
Veiled	-0.017***	-0.010***	-0.003	-0.000	-0.001
	(0.005)	(0.003)	(0.004)	(0.001)	(0.003)
Observations	648,740	648,740	648,740	648,740	648,740
Panel C: *pscore* > 0.75 **or** *pscore* < 0.25					
Veiled × Post	0.016***	0.020***	0.004	-0.002	-0.013***
	(0.006)	(0.003)	(0.004)	(0.002)	(0.003)
	[0.005, 0.030]	[0.013, 0.025]	[-0.006, 0.012]	[-0.004, 0.004]	[-0.017, -0.008]
Veiled	-0.022***	-0.010***	-0.007	-0.001	-0.001
	(0.005)	(0.004)	(0.004)	(0.002)	(0.003)
Observations	551,579	551,579	551,579	551,579	551,579
Panel D: *pscore* > 0.80 **or** *pscore* < 0.20					
Veiled × Post	0.006	0.015***	-0.000	-0.001	-0.014***
	(0.007)	(0.005)	(0.005)	(0.002)	(0.003)
	[-0.004, 0.023]	[0.010, 0.027]	[-0.011, 0.011]	[-0.005, 0.004]	[-0.020, -0.006]
Veiled	-0.017**	-0.009***	-0.004	-0.001	-0.000
	(0.006)	(0.005)	(0.006)	(0.002)	(0.004)
Observations	455,742	455,742	455,742	455,742	455,742

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports estimates from the baseline specification in Panel A of Table 3. Panel A uses the full matched sample. Panels B–D restrict the sample to observations with high-confidence predicted veiling status, defined using increasingly strict thresholds based on the predicted probabilities generated by the first-stage Probit imputation model used to assign veiling status. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A12: *Robustness tests for labor market outcomes - Specification A*

Outcome	Employment probability	Public sector employment	Private sector employment	Self-employed	Unpaid family worker
	(1)	(2)	(3)	(4)	(5)
Panel A: Controlling for removing the ban in universities					
Veiled × Post	0.012***	0.017***	0.002	-0.003*	-0.010***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.004, 0.019]	[0.011, 0.018]	[-0.004, 0.008]	[-0.004, 0.001]	[-0.013, -0.006]
After1991 × Post	-0.001	-0.021***	0.011***	-0.001	0.006**
	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)
	[-0.008, 0.004]	[-0.023, -0.019]	[0.006, 0.016]	[-0.003, 0.001]	[0.002, 0.010]
Panel B: Controlling for the 1997 education reform					
Veiled × Post	0.010***	0.015***	0.001	-0.003*	-0.009***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.002, 0.017]	[0.009, 0.016]	[-0.005, 0.008]	[-0.004, 0.000]	[-0.012, -0.006]
After1986 × Post	-0.018***	-0.003***	-0.018***	0.001	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[-0.022, -0.013]	[-0.005, -0.001]	[-0.021, -0.014]	[-0.001, 0.002]	[0.001, 0.007]
Panel C: Excluding regions with changes in veiling rates					
Veiled × Post	0.012***	0.016***	0.003	-0.003*	-0.010***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.004, 0.018]	[0.010, 0.017]	[-0.004, 0.009]	[-0.004, 0.001]	[-0.013, -0.006]
Veiled	-0.010**	-0.006***	-0.003	0.001	0.002
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel D: Using 2008 DHS for prediction and imputation					
Veiled × Post	0.010***	0.014***	-0.001	-0.002	-0.008***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.001, 0.015]	[0.010, 0.018]	[-0.009, 0.006]	[-0.005, 0.001]	[-0.011, -0.004]
Veiled	-0.016***	-0.007***	-0.005	0.001	-0.001***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel E: Using 2008 DHS for prediction and imputation (age $\geq$ 25)					
Veiled × Post	0.007**	0.019**	-0.008	-0.002	-0.010***
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
	[0.003, 0.022]	[0.013, 0.024]	[-0.009, 0.007]	[-0.005, 0.001]	[-0.013, -0.005]
Veiled	-0.019***	-0.012***	-0.003	0.001	-0.001
	(0.005)	(0.003)	(0.004)	(0.001)	(0.003)
Panel F: Alternative sample using traditional gender role attitudes					
Traditional × Post	-0.003	0.004***	-0.003	-0.002	-0.004
	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)
	[-0.006, 0.008]	[0.001, 0.007]	[-0.008, 0.004]	[-0.003, 0.001]	[-0.006, 0.001]
Traditional	0.002	-0.001*	0.003	0.001	0.002
	(0.003)	(0.001)	(0.003)	(0.001)	(0.002)
Wald test (p value)	0.087	0.000	0.410	0.667	0.003
Panel G: Alternative sample using information on praying					
Praying × Post	0.014	0.003*	0.011	-0.001	-0.003
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[-0.001, 0.013]	[-0.001, 0.007]	[-0.001, 0.013]	[-0.004, 0.001]	[-0.007, 0.001]
Praying	-0.010	-0.003	-0.004	0.001	-0.002
	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)
Wald test (p value)	0.419	0.000	0.063	0.339	0.000

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports robustness estimates of the baseline specification in Panel A of Table 3. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Reported Wald test p-values in the last row of Panels F and G rely on conventional robust inference conditional on the matched sample. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A13: *Robustness tests for labor market outcomes - Specification B*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Controlling for removing the ban in universities					
Veiled × Post	0.011***	0.018***	0.001	-0.003	-0.011***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.004, 0.019]	[0.013, 0.019]	[-0.007, 0.008]	[-0.004, 0.001]	[-0.015, -0.008]
After1991 × Post	-0.001	-0.021***	0.011***	-0.001	0.006**
	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)
	[-0.008, 0.004]	[-0.023, -0.019]	[0.006, 0.016]	[-0.003, 0.001]	[0.002, 0.010]
Panel B: Controlling for the 1997 education reform					
Veiled × Post	0.009**	0.016***	0.001	-0.003	-0.011***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
	[0.001, 0.017]	[0.011, 0.018]	[-0.007, 0.007]	[-0.004, 0.001]	[-0.015, -0.007]
After1986 × Post	-0.018***	-0.003***	-0.018***	0.001	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[-0.022, -0.012]	[-0.005, -0.001]	[-0.021, -0.014]	[-0.001, 0.002]	[0.001, 0.007]
Panel C: Excluding regions with changes in veiling rates					
Veiled × Post	0.010***	0.017***	0.002	-0.003	-0.012***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[0.003, 0.019]	[0.012, 0.019]	[-0.006, 0.008]	[-0.005, 0.001]	[-0.016, -0.008]
Veiled	-0.008**	-0.007***	-0.002	0.001	0.003*
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel D: Using 2008 DHS for prediction and imputation					
Veiled × Post	0.006	0.016***	-0.004	-0.002	-0.009***
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
	[-0.002, 0.016]	[0.010, 0.019]	[-0.011, 0.005]	[-0.004, 0.001]	[-0.013, -0.005]
Veiled	-0.012***	-0.008***	-0.001	0.001	0.001
	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)
Panel E: Using 2008 DHS for prediction and imputation (age $\geq$ 25)					
Veiled × Post	0.004**	0.021***	-0.010	-0.003	-0.012***
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
	[0.001, 0.021]	[0.015, 0.026]	[-0.013, 0.006]	[-0.006, 0.002]	[-0.015, -0.007]
Veiled	-0.017***	-0.014***	-0.002	0.002	0.001
	(0.005)	(0.003)	(0.004)	(0.002)	(0.003)
Panel F: Alternative sample using traditional gender role attitudes					
Traditional × Post	-0.003	0.004***	-0.003	-0.003	-0.004
	(0.004)	(0.002)	(0.003)	(0.002)	(0.006)
	[-0.006, 0.009]	[0.001, 0.008]	[-0.010, 0.003]	[-0.003, 0.001]	[-0.006, 0.002]
Traditional	0.003	-0.001*	0.003	0.001	0.001
	(0.003)	(0.001)	(0.003)	(0.001)	(0.002)
Panel G: Alternative sample using information on praying					
Praying × Post	0.015	0.003**	0.013	-0.001	-0.003
	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)
	[-0.001, 0.015]	[0.001, 0.008]	[-0.003, 0.014]	[-0.004, 0.002]	[-0.008, 0.001]
Praying	-0.011	-0.003	-0.006	-0.001	-0.002
	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports robustness estimates of the specification in Panel B of Table 3, which augments Eq. (1) with an additional interaction term between $Veiled$ and an indicator for the 2013 survey wave to account for differential trends in the transition year. All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A14: *Robustness tests for labor market outcomes - Specification C*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Controlling for removing the ban in universities					
Veiled × Post	0.011*	0.008	0.002	-0.005**	-0.002
	(0.007)	(0.004)	(0.006)	(0.002)	(0.003)
	[-0.001, 0.025]	[-0.004, 0.010]	[-0.005, 0.020]	[-0.011, -0.002]	[-0.008, 0.005]
After1991 × Post	-0.001	-0.021***	0.011***	-0.001	0.006**
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
	[-0.008, 0.004]	[-0.023, -0.019]	[0.006, 0.016]	[-0.003, 0.001]	[0.002, 0.010]
Panel B: Controlling for the 1997 education reform					
Veiled × Post	0.010	0.008	-0.001	-0.005**	-0.002
	(0.007)	(0.004)	(0.006)	(0.002)	(0.003)
	[-0.002, 0.023]	[-0.003, 0.011]	[-0.007, 0.018]	[-0.011, -0.002]	[-0.008, 0.005]
After1986 × Post	-0.018***	-0.003***	-0.018***	0.001	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[-0.022, -0.012]	[-0.005, -0.001]	[-0.021, -0.014]	[-0.001, 0.002]	[0.001, 0.007]
Panel C: Excluding regions with changes in veiling rates					
Veiled × Post	0.011*	0.009	0.001	-0.004**	-0.003
	(0.008)	(0.004)	(0.006)	(0.003)	(0.003)
	[-0.002, 0.025]	[-0.004, 0.011]	[-0.005, 0.020]	[-0.011, -0.001]	[-0.008, 0.005]
Veiled	-0.010**	-0.008***	-0.004	-0.001	0.004*
	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)
Panel D: Using 2008 DHS for prediction and imputation					
Veiled × Post	0.027*	0.006	0.009	-0.001	0.004
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.001, 0.031]	[-0.002, 0.011]	[-0.005, 0.020]	[-0.008, 0.002]	[-0.007, 0.007]
Veiled	-0.009*	-0.010***	-0.001	0.001	0.003
	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)
Panel E: Using 2008 DHS for prediction and imputation (age $\geq$ 25)					
Veiled × Post	0.021*	0.007	0.005	-0.003	0.001
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.001, 0.033]	[-0.004, 0.014]	[-0.005, 0.024]	[-0.010, 0.003]	[-0.009, 0.007]
Veiled	-0.014***	-0.016***	0.001	0.001	0.003
	(0.006)	(0.003)	(0.005)	(0.002)	(0.003)
Panel F: Alternative sample using traditional gender role attitudes					
Traditional × Post	-0.002	0.001	-0.005	-0.001	-0.001
	(0.010)	(0.004)	(0.008)	(0.003)	(0.004)
	[-0.014, 0.018]	[-0.006, 0.008]	[-0.018, 0.014]	[-0.006, 0.006]	[-0.009, 0.009]
Traditional	0.003	-0.002**	0.003	0.001	0.002
	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)
Panel G: Alternative sample using information on praying					
Praying × Post	0.001	0.001	0.010	-0.006	-0.005
	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)
	[-0.017, 0.025]	[-0.010, 0.009]	[-0.018, 0.030]	[-0.010, 0.003]	[-0.011, 0.007]
Praying	-0.015	-0.004	-0.007	-0.001	-0.002
	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. Each panel reports robustness estimates of the baseline specification in Panel C of Table 3 that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable (t = year − 2010). All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A15: *Effects on labor market outcomes using ML imputations- Specification A*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Ensemble Learning					
Veiled × Post	0.015***	0.025***	-0.003	-0.002**	-0.015***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.012, 0.026]	[0.022, 0.027]	[-0.005, 0.008]	[-0.005, -0.001]	[-0.017, -0.012]
Veiled	-0.041***	-0.014***	-0.007	0.001	-0.014***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel B: Support Vector Machines					
Veiled × Post	0.011***	0.021***	-0.001	-0.004**	-0.016***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.005, 0.023]	[0.017, 0.025]	[-0.006, 0.007]	[-0.004, -0.001]	[-0.018, -0.012]
Veiled	-0.001	-0.029***	0.017	0.003	0.014
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel C: Random Forest					
Veiled × Post	0.020***	0.024***	0.004	-0.002**	-0.014***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.009, 0.025]	[0.021, 0.028]	[-0.006, 0.007]	[-0.005, -0.001]	[-0.016, -0.012]
Veiled	-0.045***	-0.007*	-0.015	-0.001	-0.015**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel D: K-Nearest Neighbor					
Veiled × Post	0.013***	0.023***	-0.001	-0.003**	-0.015***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.008, 0.024]	[0.019, 0.026]	[-0.005, 0.008]	[-0.005, -0.001]	[-0.017, -0.011]
Veiled	-0.020***	-0.013***	-0.003	0.001	0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel E: Neural Network					
PrVeiled × Post	0.020***	0.059***	-0.015	-0.008***	-0.041***
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
	[0.013, 0.042]	[0.042, 0.063]	[-0.017, 0.008]	[-0.012, -0.004]	[-0.042, -0.028]
PrVeiled	-0.154***	-0.124**	-0.011	-0.020**	0.013
	(0.007)	(0.003)	(0.006)	(0.003)	(0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A16: *Effects on labor market outcomes using ML imputations- Specification B*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Ensemble Learning					
Veiled × Post	0.009***	0.027***	-0.007	-0.002	-0.018***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.005, 0.023]	[0.024, 0.030]	[-0.011, 0.006]	[-0.004, -0.000]	[-0.020, -0.014]
Veiled	-0.036***	-0.016***	-0.003	0.001	-0.011**
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Panel B: Support Vector Machines					
Veiled × Post	0.004	0.023***	-0.006	-0.003	-0.020***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
	[-0.002, 0.020]	[0.018, 0.027]	[-0.011, 0.006]	[-0.004, -0.000]	[-0.022, -0.015]
Veiled	0.005	-0.030***	0.021*	0.002	0.018**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel C: Random Forest					
Veiled × Post	0.018**	0.026***	0.002	-0.002**	-0.017***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.003, 0.023]	[0.023, 0.031]	[-0.012, 0.005]	[-0.005, -0.000]	[-0.020, -0.014]
Veiled	-0.043***	-0.009**	-0.013	-0.001	-0.013
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Panel D: K-Nearest Neighbor					
Veiled × Post	0.010***	0.025***	-0.005	-0.002**	-0.018***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
	[0.002, 0.022]	[0.020, 0.029]	[-0.011, 0.006]	[-0.004, -0.000]	[-0.019, -0.013]
Veiled	-0.017*	-0.016***	0.000	0.001	0.004
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Panel E: Neural Network					
PrVeiled × Post	0.006*	0.068***	-0.032	-0.007***	-0.048***
	(0.005)	(0.003)	(0.005)	(0.002)	(0.002)
	[-0.001, 0.037]	[0.047, 0.071]	[-0.031, 0.002]	[-0.012, -0.003]	[-0.049, -0.032]
PrVeiled	-0.142***	-0.132**	0.004	-0.021**	0.019
	(0.008)	(0.003)	(0.006)	(0.003)	(0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero. The Neural Network's private sector employment estimate falls marginally outside the bootstrap percentile CI, because the interval is centered around the bootstrap mean rather than the original estimate to capture full resampling variation. Given the negligible magnitude of the discrepancy and the low probability of this occurring across all estimates, this does not affect our conclusions.

Table A17: *Effects on labor market outcomes using ML imputations- Specification C*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A: Ensemble Learning					
Veiled × Post	0.042***	0.013***	0.017**	-0.006***	0.004
	(0.005)	(0.003)	(0.005)	(0.002)	(0.002)
	[0.019, 0.056]	[0.007, 0.019]	[0.001, 0.037]	[-0.011, -0.003]	[-0.004, 0.007]
Veiled	-0.032**	-0.018***	-0.000	-0.001	-0.007*
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Panel B: Support Vector Machines					
Veiled × Post	0.051***	0.012***	0.029***	-0.010***	0.007
	(0.005)	(0.002)	(0.004)	(0.002)	(0.002)
	[0.030, 0.060]	[0.006, 0.019]	[0.006, 0.039]	[-0.012, -0.004]	[-0.002, 0.010]
Veiled	0.013	-0.032***	0.027*	0.000	0.022***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Panel C: Random Forest					
Veiled × Post	0.026***	0.012***	0.007	-0.007***	0.002
	(0.005)	(0.003)	(0.005)	(0.002)	(0.002)
	[0.010, 0.042]	[0.006, 0.018]	[-0.006, 0.024]	[-0.011, -0.001]	[-0.006, 0.004]
Veiled	-0.043***	-0.011***	-0.014	-0.003	-0.009
	(0.003)	(0.001)	(0.003)	(0.001)	(0.001)
Panel D: K-Nearest Neighbor					
Veiled × Post	0.031***	0.009***	0.018***	-0.008***	-0.000
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
	[0.014, 0.045]	[0.003, 0.016]	[0.004, 0.035]	[-0.012, -0.003]	[-0.008, 0.004]
Veiled	-0.014	-0.018***	0.003	-0.001	0.007
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Panel E: Neural Network					
PrVeiled × Post	0.104***	0.009**	0.077***	-0.016***	-0.001
	(0.010)	(0.006)	(0.009)	(0.004)	(0.005)
	[0.045, 0.111]	[0.001, 0.029]	[0.024, 0.079]	[-0.022, -0.008]	[-0.016, 0.009]
PrVeiled	-0.124***	-0.142**	0.022	-0.023**	0.027
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A18: *Effects of the removal of the headscarf ban on working hours, job quality, and occupations using ML imputations - Specification A*

Outcome	Working hours (weekly)	Informal employment	Permanent job	Professionals	Technicians	Clerks	Earnings (monthly)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Ensemble Learning							
Veiled × Post	0.552***	-0.013***	0.025***	-0.000	0.016***	0.006***	-1.465
	(0.118)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(5.174)
	[0.381, 1.036]	[-0.016, -0.009]	[0.022, 0.035]	[-0.003, 0.002]	[0.014, 0.017]	[0.005, 0.008]	[-8.767, 15.804]
Veiled	-1.319**	-0.013***	-0.024***	0.000	-0.008***	-0.005***	-3.153
	(0.114)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(4.008)
Panel B: Support Vector Machines							
Veiled × Post	0.382***	-0.013***	0.024***	0.000	0.013***	0.006***	7.006
	(0.111)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(4.356)
	[0.084, 0.934]	[-0.016, -0.008]	[0.017, 0.033]	[-0.003, 0.002]	[0.012, 0.015]	[0.004, 0.006]	[-7.026, 18.037]
Veiled	0.098	0.013	-0.014**	-0.013***	-0.008***	-0.006***	11.550
	(0.102)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(3.430)
Panel C: Random Forest							
Veiled × Post	0.798***	-0.010***	0.029***	-0.000	0.015***	0.006***	5.205
	(0.116)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(5.044)
	[0.320, 0.990]	[-0.015, -0.007]	[0.018, 0.033]	[-0.002, 0.003]	[0.013, 0.017]	[0.005, 0.008]	[-14.041, 14.376]
Veiled	-1.419**	-0.015*	-0.025***	0.003	-0.007***	-0.003***	-2.551
	(0.118)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(4.136)
Panel D: K-Nearest Neighbor							
Veiled × Post	0.536***	-0.013***	0.023***	-0.000	0.015***	0.006***	-2.699
	(0.115)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.882)
	[0.260, 0.976]	[-0.016, -0.007]	[0.020, 0.033]	[-0.003, 0.003]	[0.013, 0.015]	[0.004, 0.007]	[-11.236, 15.461]
Veiled	-0.618*	0.002	-0.019***	-0.000	-0.007***	-0.004***	1.564
	(0.102)	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)	(3.564)
Panel E: Neural Network							
PrVeiled × Post	0.522**	-0.034***	0.050***	-0.000	0.037***	0.015***	-34.123
	(0.229)	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)	(11.439)
	[0.155, 1.642]	[-0.038, -0.021]	[0.034, 0.065]	[-0.009, 0.004]	[0.029, 0.039]	[0.010, 0.017]	[-55.868, 8.229]
PrVeiled	-5.242***	-0.013	-0.130***	-0.051*	-0.027***	-0.032***	-383.990***
	(0.347)	(0.005)	(0.006)	(0.002)	(0.001)	(0.001)	(12.612)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A19: *Effects of the removal of the headscarf ban on working hours, job quality, and occupations using ML imputations - Specification B*

Outcome	Working hours (weekly)	Informal employment	Permanent job	Professionals	Technicians	Clerks	Earnings (monthly)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Ensemble Learning							
Veiled × Post	0.284**	-0.014***	0.023***	-0.004***	0.021***	0.008***	-6.138
	(0.126)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(5.415)
	[0.048, 0.909]	[-0.017, -0.010]	[0.018, 0.035]	[-0.007, -0.001]	[0.019, 0.022]	[0.007, 0.009]	[-16.558, 13.317]
Veiled	-1.066*	-0.011**	-0.022**	0.004	-0.012***	-0.006***	1.266
	(0.121)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.262)
Panel B: Support Vector Machines							
Veiled × Post	0.065	-0.015***	0.021***	-0.004***	0.017***	0.007***	1.879
	(0.118)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.530)
	[-0.266, 0.825]	[-0.018, -0.009]	[0.013, 0.033]	[-0.007, -0.002]	[0.015, 0.020]	[0.005, 0.008]	[-12.648, 15.318]
Veiled	0.395	0.015	-0.011	-0.010*	-0.011***	-0.007***	16.377
	(0.108)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(3.592)
Panel C: Random Forest							
Veiled × Post	0.681**	-0.010***	0.029***	-0.004***	0.019***	0.007***	4.546
	(0.125)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(5.316)
	[0.002, 0.904]	[-0.018, -0.007]	[0.015, 0.033]	[-0.007, -0.001]	[0.018, 0.022]	[0.006, 0.009]	[-20.106, 14.181]
Veiled	-1.304**	-0.014	-0.025**	0.007**	-0.011***	-0.005***	-1.907
	(0.126)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.448)
Panel D: K-Nearest Neighbor							
Veiled × Post	0.358*	-0.014***	0.022***	-0.004**	0.019***	0.007***	-4.758
	(0.123)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(5.105)
	[-0.063, 0.892]	[-0.017, -0.007]	[0.016, 0.033]	[-0.007, -0.000]	[0.016, 0.020]	[0.005, 0.008]	[-16.372, 13.616]
Veiled	-0.449	0.003	-0.018***	0.003	-0.011***	-0.005***	3.524
	(0.109)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(3.820)
Panel E: Neural Network							
PrVeiled × Post	-0.279	-0.037***	0.043***	-0.007***	0.049***	0.019***	-49.828
	(0.247)	(0.003)	(0.005)	(0.003)	(0.002)	(0.001)	(12.195)
	[-0.627, 1.294]	[-0.039, -0.021]	[0.028, 0.064]	[-0.018, -0.004]	[0.038, 0.051]	[0.013, 0.021]	[-79.252, 2.300]
PrVeiled	-4.500**	-0.011	-0.124***	-0.044	-0.038***	-0.035***	-369.377***
	(0.357)	(0.005)	(0.006)	(0.003)	(0.002)	(0.002)	(13.205)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A20: *Effects of the removal of the headscarf ban on working hours, job quality, and occupations using ML imputations - Specification C*

Outcome	Working hours (weekly) (1)	Informal employment (2)	Permanent job (3)	Professionals (4)	Technicians (5)	Clerks (6)	Earnings (monthly) (7)
Panel A: Ensemble Learning							
Veiled × Post	1.931***	-0.005	0.032***	0.007***	0.004***	0.002	32.457*
	(0.243)	(0.003)	(0.005)	(0.002)	(0.001)	(0.001)	(10.587)
	[0.909, 2.843]	[-0.014, 0.001]	[0.016, 0.050]	[0.002, 0.013]	[0.001, 0.006]	[-0.001, 0.004]	[-1.314, 61.141]
Veiled	-0.834	-0.010	-0.021*	0.003	-0.012***	-0.007***	8.866
	(0.134)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.942)
Panel B: Support Vector Machines							
Veiled × Post	2.380***	-0.001	0.041***	0.007***	0.003***	0.002	52.125***
	(0.229)	(0.003)	(0.004)	(0.002)	(0.001)	(0.001)	(8.902)
	[1.332, 2.858]	[-0.010, 0.004]	[0.021, 0.052]	[0.003, 0.011]	[0.002, 0.005]	[-0.001, 0.004]	[14.934, 72.138]
Veiled	0.806	0.017	-0.008	-0.011*	-0.011***	-0.008***	27.824*
	(0.121)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(4.143)
Panel C: Random Forest							
Veiled × Post	1.097***	-0.006**	0.019***	0.007**	0.004**	0.001	10.875
	(0.239)	(0.003)	(0.005)	(0.002)	(0.001)	(0.001)	(10.290)
	[0.344, 1.977]	[-0.016, -0.000]	[0.007, 0.035]	[0.001, 0.012]	[0.001, 0.005]	[-0.001, 0.004]	[-12.052, 42.485]
Veiled	-1.308*	-0.014	-0.028**	0.006	-0.011***	-0.005***	-0.445
	(0.139)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(5.178)
Panel D: K-Nearest Neighbor							
Veiled × Post	1.553***	-0.007**	0.026***	0.005*	0.003***	0.001	22.583*
	(0.237)	(0.003)	(0.004)	(0.002)	(0.001)	(0.001)	(9.971)
	[0.754, 2.364]	[-0.017, -0.001]	[0.015, 0.043]	[-0.001, 0.010]	[0.001, 0.005]	[-0.001, 0.004]	[-0.326, 50.752]
Veiled	-0.258	0.004	-0.018***	0.001	-0.011***	-0.006***	10.595
	(0.123)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(4.510)
Panel E: Neural Network							
PrVeiled × Post	5.343***	-0.012**	0.085***	-0.000	0.006***	0.003	90.154***
	(0.474)	(0.006)	(0.009)	(0.006)	(0.002)	(0.003)	(23.397)
	[2.400, 5.364]	[-0.034, -0.001]	[0.041, 0.093]	[-0.006, 0.016]	[0.002, 0.012]	[-0.001, 0.007]	[5.779, 126.625]
PrVeiled	-3.485*	-0.005	-0.117***	-0.051*	-0.038***	-0.037***	-338.460**
	(0.376)	(0.006)	(0.007)	(0.003)	(0.002)	(0.002)	(14.498)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 2010–2017 HLFS. The sample includes all women aged 18–49. Panel A reports difference-in-differences estimates using scarf predictions from ensemble learning methods; Panel B reports estimates using Support Vector Machines; Panel C reports estimates using Random Forests; Panel D reports estimates using K-Nearest Neighbor; and Panel E reports estimates using Neural Networks, where the continuous predicted probability of veiling is used as the treatment variable to capture veiling intensity. The estimation follows Equation 1. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A21: *Heterogeneous effects by regions with a high share of veiled women ($\geq$ 50%)*

Outcome	Employment probability (1)	Public sector employment (2)	Private sector employment (3)	Self-employed (4)	Unpaid family worker (5)
Panel A:					
Veiled × Post × MeanShare	0.007	0.005	0.002	-0.006	0.003
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.016, 0.017]	[-0.007, 0.008]	[-0.009, 0.019]	[-0.010, 0.001]	[-0.008, 0.007]
Veiled × Post	0.006	0.013***	0.001	0.002	-0.013***
	(0.007)	(0.004)	(0.006)	(0.003)	(0.004)
	[-0.006, 0.024]	[0.007, 0.021]	[-0.013, 0.012]	[-0.003, 0.006]	[-0.016, -0.002]
Panel B:					
Veiled × Post × MeanShare	0.007	0.005	0.002	-0.006	0.003
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.016, 0.017]	[-0.007, 0.008]	[-0.009, 0.019]	[-0.010, 0.001]	[-0.008, 0.007]
Veiled × Post	0.005	0.014***	0.000	0.002	-0.014***
	(0.007)	(0.004)	(0.006)	(0.003)	(0.004)
	[-0.006, 0.022]	[0.009, 0.023]	[-0.014, 0.011]	[-0.003, 0.006]	[-0.018, -0.004]
Panel C:					
Veiled × Post × MeanShare	0.007	0.005	0.002	-0.006	0.004
	(0.008)	(0.004)	(0.007)	(0.003)	(0.004)
	[-0.016, 0.017]	[-0.007, 0.008]	[-0.009, 0.019]	[-0.010, 0.001]	[-0.008, 0.007]
Veiled × Post	0.006	0.005	0.000	-0.000	-0.006
	(0.010)	(0.005)	(0.009)	(0.003)	(0.005)
	[-0.015, 0.033]	[-0.007, 0.012]	[-0.016, 0.018]	[-0.008, 0.004]	[-0.011, 0.007]
Observations	933,810	933,810	933,810	933,810	933,810

Notes: Data are drawn from the 2010–2017 HLFS and the sample includes all women aged 18–49. $MeanShare$ is a binary indicator equal to one if the woman resides in a region where at least 50 percent of women are veiled (based on the 2013 DHS), and zero otherwise. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable (t = year − 2010). All specifications include a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A22: *Worker efficiency in public sector before and after the headscarf removal*

Outcome	Match quality: Over qualified (1)	Match quality: Under qualified (2)	Match quality: Mismatch (3)	Age (4)	Earnings per hour (5)
Panel A:					
Veiled × Post	-0.006	0.019***	0.013**	-0.382	-0.022
	(0.016)	(0.006)	(0.017)	(0.280)	(0.015)
	[-0.019, 0.043]	[0.011, 0.034]	[0.004, 0.067]	[-0.732, 0.388]	[-0.058, 0.004]
Veiled	0.001	-0.011***	-0.011	0.651	0.002
	(0.012)	(0.005)	(0.013)	(0.494)	(0.012)
Panel B:					
Veiled × Post	0.005	0.026***	0.032***	-0.269	-0.020
	(0.018)	(0.006)	(0.019)	(0.311)	(0.015)
	[-0.015, 0.056]	[0.012, 0.038]	[0.010, 0.085]	[-0.639, 0.442]	[-0.056, 0.009]
Veiled	-0.011	-0.019***	-0.030**	0.537	-0.001
	(0.014)	(0.005)	(0.015)	(0.524)	(0.013)
Panel C:					
Veiled × Post	-0.057	0.003	-0.055	-0.137	-0.055
	(0.028)	(0.012)	(0.029)	(0.571)	(0.031)
	[-0.092, 0.007]	[-0.014, 0.031]	[-0.084, 0.018]	[-1.168, 1.082]	[-0.096, 0.029]
Veiled	-0.021**	-0.018**	-0.039**	0.752	-0.012
	(0.016)	(0.007)	(0.017)	(0.601)	(0.016)
Observations	51,149	51,149	51,149	51,149	43,752
Mean Dep. Var.	0.281	0.024	0.306	34.23	2.835

Notes: Data are from the 2010-2017 HLFS. The sample includes all women aged 18-49. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable ($t = \text{year} - 2010$). Each estimate includes a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared (except column 4), household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.

Table A23: *Effects of removing the headscarf ban on marriage probability*

Outcome	Ever married (1)	Currently married (2)	Divorced (3)
Panel A:			
Veiled × Post	0.003	0.003	0.001
	(0.003)	(0.003)	(0.001)
	[-0.005, 0.007]	[-0.006, 0.007]	[-0.001, 0.003]
Veiled	-0.002	0.001	-0.004*
	(0.004)	(0.004)	(0.001)
Panel B:			
Veiled × Post	0.007	0.008	0.001
	(0.004)	(0.004)	(0.001)
	[-0.002, 0.012]	[-0.002, 0.012]	[-0.002, 0.004]
Veiled	-0.006	-0.003	-0.004*
	(0.005)	(0.005)	(0.001)
Panel C:			
Veiled × Post	-0.030***	-0.034***	0.006
	(0.007)	(0.008)	(0.003)
	[-0.044, -0.016]	[-0.050, -0.018]	[-0.001, 0.010]
Veiled	-0.014***	-0.013**	-0.002
	(0.005)	(0.006)	(0.002)
Observations	933,810	933,810	933,810
Mean Dep. Var.	0.797	0.756	0.025

Notes: Data are from the 2010-2017 HLFS. The sample includes all women aged 18-49. Panel A reports baseline difference-in-differences estimates from Eq. (1), comparing labor market outcomes between women predicted to be veiled and those predicted not to be veiled. Panel B augments Eq. (1) by including an indicator for the year 2013 and its interaction with the veiled indicator, allowing for a differential effect in the transition year. Panel C reports trend-adjusted estimates that include $1\{\text{Veiled}\} \times t$, where t is a centered year variable ($t = \text{year} - 2010$). Each estimate includes a dummy variable indicating whether the woman is native, dummies for the woman's education level, age, age squared, household size, and region and year fixed effects, along with their interactions. Parentheses report bootstrap standard errors obtained from a two-step resampling procedure that re-estimates both the veiling imputation stage and the second-stage regression. Brackets report bootstrap percentile 95% confidence intervals. All estimates are weighted using cross-sectional survey weights from the wave in which the outcome was measured. * indicates that the 90% bootstrap percentile confidence interval excludes zero, ** indicates that the 95% bootstrap percentile confidence interval excludes zero, and *** indicates that the 99% bootstrap percentile confidence interval excludes zero.