



Returns to schooling in Kazakhstan: an update using a pseudo-panel approach

Saule Kemelbayeva¹

Received: 22 June 2019 / Revised: 18 February 2020 / Accepted: 21 February 2020 /
Published online: 3 April 2020
© The Author(s) 2020

Abstract

This paper represents the first attempt to employ a long series of repeated cross-sectional data from the Kazakhstani National Statistics as the pseudo-panel for estimating returns to schooling. We found the returns to be relatively high and internationally comparable. The cohort effect turned out to be negative, suggesting the interpretation of the business cycle's impact. The gender gap in returns has additionally been revealed: while females tend to earn less, the returns are higher for them, which can likely be explained by gender differences in labor allocation across sectors and industries and, in turn, explains the higher levels of education amongst women.

Keywords Returns to schooling · Pseudo-panel methodology · Kazakhstan

JEL Classification I26 · J24

Abbreviations

2SLS	Two-stage least squares
CPI	Consumer price index
FE	Fixed effects
GDP	Gross domestic product
HEI	Higher education institution
ILO	International Labour Organization
IV	Instrumental variables
OLS	Ordinary least squares
PhD	Doctor of Philosophy
PSU	Primary sampling unit
RE	Random effects
USD	United States Dollar

✉ Saule Kemelbayeva
S.Kemelbayeva2@newcastle.ac.uk

¹ Newcastle University Business School, Newcastle upon Tyne, UK

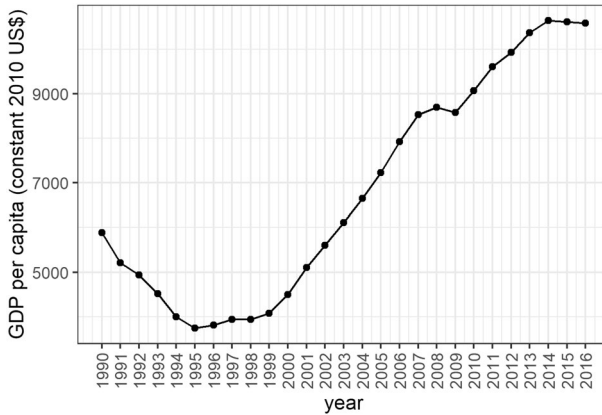


Fig. 1 GDP per capita, (constant 2010 USD). Source: World Bank Data

1 Introduction

The ‘returns to schooling (education)’ concept, as developed by Mincer (1974), was subsequently theoretically enriched and empirically tested in various contexts, and contributed to the evaluation of the economic role of education, labor market conditions and human capital productivity. Although the model in its basic form has certain conceptual flaws—in particular, it ignores bias potentially caused by unobservable factors that can influence both schooling and earnings—methods have been proposed to overcome them such as a use of panel data allowing to explicitly control for unobserved heterogeneity, assuming it is time-invariant. The aim of this study is to estimate the returns to schooling in Kazakhstan with the use of massive repeated cross-sectional data collected by the Household Budget Survey in 2002–2016, as a synthetic or pseudo-panel. The approach proposed by Deaton (1985) suggests adopting a pseudo-panel of cohort means, where a ‘cohort’ we consider to be a group of people of the same gender born in the same year who are assumed to share some common, unobserved characteristics.

There are no assessments available for the returns to education in Kazakhstan during the Soviet period, but they are believed to be low due to wage levelling, wage ‘grids’, and the centralized allocation of the labor force. However, according to a few post-Soviet examinations, they soared with the transition. In Kazakhstan, whose independence can be roughly divided into two sub-periods—the severe crisis of the 1990s and the oil boom of the 2000s (Fig. 1)—the later economic growth might additionally have contributed to the increase in returns via several channels. Demand for education consistently grew during the period of independence, with the number

of university students increasing from 287,367 in 1990/91 to 542,458 in 2018/19, and the number of college students (technical and vocational education and training) from 247,650 to 489,818 for the same years, respectively (with a corresponding net increase in population of around 1.5 million).¹ Kazakhstan's educational institutions expanded accordingly: from 55 HEIs in 1990/91 to 124 in 2018/19, and from 247 colleges to 769 for the corresponding years.² On the other hand, with this nearly twofold increase in the inflow of educated people, it might be reasonable to predict a decrease in returns to schooling over time. Additionally, the period under consideration represents the oil boom decade when GDP per capita grew from 1658.00 USD in 2002 to 13,890.80 USD in 2013,³ but with signs of recession starting in 2014 when world commodity prices plummeted, dragging down an economy that was (and still is) highly dependent on oil and gas exports, which might also have intriguing effects.

With the pseudo-panel approach, we found the returns to schooling to be relatively high (7–13% with the fixed effects and 8–11% with Mundlak random effects, depending on a set of additional control variables) and essentially identical to simple OLS estimates obtained from individual data (8–12% for men and 10–13% for women), which are in turn very similar to the only previous examination that used the instrumental variables approach (Arabsheibani and Mussurov, 2007). Though the results for schooling are robust across models regardless of controlling for cohort heterogeneity, with the Mundlak model the cohort effect (between-estimator) turned out to be highly significant and negative: while an increase in cohorts' average schooling over time increases their wages, less educated cohorts earn more than more educated ones. More educated cohorts in the sample are the younger individuals whose school-leaving age fell roughly within the recession of the 1990s, suggesting the business cycle impact interpretation: cohorts entering the labor market during a recession and facing a lack of jobs apparently end up getting more education and lower lifetime wages.

The study additionally uncovers other curious results. First, though real wages rocketed during the observed period (by about 500–600% for men and 300–400% for women), the returns to schooling dropped (by about 4–5% and 2–3% for men and women, respectively). Second, the rapid growth in real wages over the period could only partially be explained by the changes in the working population's observed characteristics, including education, by about 30% for men and 40% for women, leaving the remaining part likely due to the oil boom growth. Third, despite females earning less, their returns to schooling were consistently higher for all models. The latter could probably be explained by gender differences in the labor force allocation between industries and sectors, with men mainly employed in market-oriented, riskier, but better paid industries with predominantly private ownership that probably value education less than the public sector and those industries absorbing the female labor force, where a certain level of schooling is often formally required and, indeed, rewarded. This, in turn, complies with the higher level of education amongst

¹ The Committee on Statistics of the Republic of Kazakhstan, www.stat.gov.kz.

² The Committee on Statistics of the Republic of Kazakhstan, www.stat.gov.kz.

³ The Committee on Statistics of the Republic of Kazakhstan, www.stat.gov.kz.

women compared to men due to their rational decisions under the prevailing labor market conditions.

The paper is organized as follows. The following section discusses the theoretical framework, the pseudo-panel methodology and its possible drawbacks, and briefly reviews its previous applications worldwide. It also details some of the very few research efforts to examine the returns to schooling in Kazakhstan and the region. Section 3 depicts the sampling methodology and the questionnaire, stating data limitations and caveats with regard to the interpretation of the results so arising. It familiarizes the reader with descriptive statistics and visualizes the most important individual-level variables disaggregated by gender, as well as the cohort-level data. The following section discusses the main findings from the estimated models and their possible interpretations in the context of the Kazakhstani labor market, as summarized by the conclusions.

2 Theoretical framework and previous examinations

A definition of returns to schooling was given by Mincer in his seminal work as “a full quantitative accounting of the effects of the distribution of investment in human capital on observed earnings inequality” (Mincer 1974, p. 43). Mincer’s earning function, in its attempt to explain the extent to which earnings depend on schooling or education, is still widely used in many empirical studies as a key concept for the analysis of private returns. As Heckman et al. (2003, p. 1) note, “Mincer’s model of earnings... is the framework used to estimate returns to schooling, returns to schooling quality, and to measure the impact of work experience on male-female wage gaps”. Comprehensive reviews of existing empirical applications are given by Harmon and Walker (2001) and Card (1999, 2001).

In its basic form, Mincer’s model suggests the log of earnings (or wages) to be linearly dependent on either years of schooling or a level of education attained by an individual and other relevant control variables, such as their experience (practically, often substituted by its proxy, age, normally both in linear and quadratic terms to allow for diminishing returns to experience), gender, region and others. Depending on what is used as an explanatory variable—schooling or level of education attained—the model estimates the returns to either schooling or credentials. The debate in Labor Economics with regard to this topic has given rise to a number of hypotheses, among which the ‘sheepskin effect’ might be considered as potentially promising for testing in Kazakhstan, where the current education system has been widely criticized by society as adding little value in terms of human capital productivity due to overall low-profile staff, outdated content and learning facilities, and poor links to industry. The concept suggests that completing a degree provides better returns than the same years of schooling with no degree awarded (Hungerford and Solon 1987) and echoes human capital signalling theory, indicating education’s filtering and signalling role: in a market of asymmetric information with employers having limited access to information on potential employees and no opportunity to conduct formal tests for productivity, they can only rely on information regarding

their level of education as a signal of potential productivity, which is the main channel leading from education to the labor market returns rather than the value added by education (Spence 1978; Arrow 1973; Stiglitz 1975).

The biggest challenge with Mincer's specification, as discussed in the academic literature, is that it treats schooling (or education) as exogenous, ignoring any possible endogeneity caused by potential correlation of unobservable factors influencing wages (such as inner ability, motivation or family background) with schooling (education). This strong assumption generates omitted variable bias—so-called 'ability bias' (Griliches 1977)—and methods to deal with it have been proposed and empirically tested. One such is the fixed effects model, the implementation of which requires panel data. Generally, the whole idea behind the use of panel data is motivated by the possibility of being able to solve the omitted variable problem (Wooldridge 2010).

Using panel data — repeated observation of the same individuals overtime — allows unobservable variable(s) influencing wages to be held constant while obtaining the partial effects of the observable explanatory variables. With the wage equation:

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, t = 1, \dots, T \quad (1)$$

where

y_{it} —individual i 's wage observed in a period t

\mathbf{x}_{it} —observable variables reflecting various factors influencing i 's wage in a period t

c_i —time-constant unobserved component

u_{it} —an idiosyncratic error term,

this might be achieved through either differencing or transformation, with both eliminating the unobserved component (Wooldridge 2010):

$$\Delta y = \Delta \mathbf{x} \boldsymbol{\beta} + \Delta u \quad (2)$$

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \boldsymbol{\beta} + u_{it} - \bar{u}_i \quad (3)$$

Both are known as the fixed effects model.

In reality, especially with regard to developing countries, genuine micro-level panel data is rarely available. Deaton (1985) proposed the use of a time-series of independent repeated cross-sections as a synthetic or *pseudo*-panel. In particular, he “considers the possibility of tracking ‘cohorts’”, “with a ‘cohort’ defined as a group with fixed membership” assuming that they share some common characteristics, whilst the use of intra-cohort means represents an alternative to that of individual data (Deaton 1985, p. 109).

This approach has been employed in a number of empirical research efforts (Dickerson et al. 2001; Brunello and Comi 2004; Warunsiri and McNown 2010; Himaz and Aturupane 2016; Bhattacharya and Sato 2017), with the most common

treatment of cohorts being those of age and gender groups, as initially proposed by Deaton (1985) and as replicated in this study. We argue that although unobserved heterogeneity might not be fully eliminated (since ability or parental background are determined at an individual, not group, level), it will at least be in part. By this, we assume the economic and social conditions witnessed and shared by people of the same generation can potentially have a similar effect once we account for gender, or the so-called ‘cohort effect’. In our case, this includes labor market conditions (demand and supply, institutional framework including labor market policies, and so on), content and quality of education, educational policies and reforms, and possible external shocks, which might be particularly pronounced in the country during the transition from the communist regime to the market economy. Interestingly, synthetic panel data might even hold some advantages over the genuine panel data, particularly while estimating the returns to schooling. The schooling variable in genuine panel data usually varies only incrementally, where one normally observes an increase in schooling only once for a particular individual. By contrast, this could be rather variable in a pseudo-panel.

Some issues arise with the pseudo-panel methodology, the potentially most serious of which is the error-in-variables caused by averaging observations at the cohort level, which in turn might create attenuation bias and additional noise. However, Verbeek and Nijman (1992) argue that with a large enough cohort (where by ‘large enough’ they assume 100 or more individual observations per cohort), the sampling error can be disregarded, and estimates may thus be considered to be unbiased. On the other hand, increasing cohort size results in a decrease in the number of cohorts (which is the number of observations in a pseudo-panel); this, in turn, reduces precision. Thus, empirically, there is always a trade-off between the number of cohorts and their size.

Another problem is heteroscedasticity, which arises with variations in cohort size. The efficient estimator is achieved by weighting each observation by the square root of the cohort size (or any other appropriate weight), as validated by Deaton (1985).

With Deaton’s (1985) synthetic panel approach, one can adopt any method allowed with the genuine panel data, such as the fixed effects or the random effects methods, with the latter being more efficient since it utilizes both the within- and between-group variations; however, it implies a strong assumption of no correlation between explanatory and unobserved variables. Mundlak (1978) suggested a technique justifying the use of the random effects model in situations when one might expect endogeneity. Mundlak’s (1978) correlated random effects model is essentially the random effects model with added group (cohort) means of the variable(s) which are believed to be endogenous, varying within the group and over time. This ‘within-between’ estimator is based on the decomposition of the unobserved component from the model (1) as:

$$c_i = \psi + \bar{x}_i \xi + a_i, E(a_i | \bar{x}_i) = 0, \quad (4)$$

which includes correlated (with explanatory variable(s)) and uncorrelated components. Further, substituting Eq. (4) into the wage equation (1) allows one to reach strict exogeneity:

$$E(y_{it}|\mathbf{x}_i) = E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\xi, t = 1, \dots, T \quad (5)$$

In addition, the degree of statistical significance of the group-mean estimates serves as a test for endogeneity (Mundlak test).

Over the last few decades, a variety of empirical studies have appeared that attempt to establish the causal effect of schooling on earnings. In his famous review of the studies evaluating the returns to schooling in a number of developed countries' databases, Card (2001) implied the returns to schooling found from these studies to be around 7% for OLS estimations and around 9% for instrumental variables (IV) estimations. Overall, the studies employing quasi-experimental designs tend to find higher returns compared with OLS estimations: “average returns to schooling from simple regression methods are around 6% internationally but over 9% from these alternative methods” (Harmon and Walker 2001, p. 6). This seems not to be the case for pseudo-panel estimations, where the empirical results worldwide are mixed with pseudo-panel models providing both higher and lower outcomes than OLS. However, overall, examinations in developing countries generally show somewhat higher returns coefficients, probably reflecting diminishing returns to education due to the accumulation of human capital in advanced economies as the average level of schooling grows (Psacharopoulos and Patrinos 2004).

In the Soviet and post-Soviet economies, research studies on the returns to schooling have been limited. As Fleisher et al. (2005) remark, prior to the late 1980s reforms, the returns in the USSR were less than 5%, which is explained by the “wage compression imposed by the [wage] grids” (Fleisher et al. 2005, p. 352), as compatible with the communist ideology of equality and the favoritism of the working class. Kapelyushnikov (2008) reports even lower estimates—at least by the end of the Soviet era, returns to schooling were not more than 1–2%. However, this changed in the post-reform period. According to Fleisher et al. (2005, p. 352), returns in transition economies tended “to rise almost immediately following reform, albeit at different speeds”. There is very little empirical evidence pertaining to Kazakhstan in this regard. Barro and Lee (2010) estimated the rate of returns for an additional year of schooling worldwide, finding it to be a little more than 8% for ‘Europe and Central Asia’. Arabsheibani and Mussurov (2007)—having used OLS and IV methodologies (with spouse education and smoking habits as instruments)—indicated that the returns to schooling in Kazakhstan have increased with the transition, with OLS estimations of 8% for men and 11.5% for women and 2SLS estimations of 11% for married men and 13.7% for married women.

This study attempts to partially fill the gap in the empirical analysis of the returns to schooling in Kazakhstan by making use of the pseudo-panel technique in conjunction with national statistics data.

3 Data and methodology

The study analyzes the Household Budget Survey data from the Kazakhstan Committee on Statistics for 2002–2016. The methodology of the survey, first introduced in 2002, has changed several times. Before 2011, monthly data were recorded, whilst

after 2011 the survey has been conducted on a quarterly basis. To achieve comparability, the data for 2002–2010 have been aggregated to a quarterly level. According to the current methodology (MNE 2015), the survey is designed in the form of rotating repeated cross-sections with one-third of the 12,000 participating households being replaced at the end of each year.⁴

A two-stage stratified random sample design has been adopted for sampling. In the first stage, the population is stratified into 30 strata representing the country's 14 provinces (*oblasts*) with urban and rural places of residence, and the two biggest cities (the current capital and the previous capital, the latter of which still remains the main financial and business centre in the country) considered separately. 400 territorial units are selected as the primary sampling units (PSUs) with a probability proportional to the stratum size (number of households per stratum). In the second stage, 30 households per PSU are randomly selected for interview from a register of dwellings; the distribution of the PSU by strata is found in Appendix 1).

In some years, the sample consists of fewer observations (2006, 2007, 2008 and 2010). The final dataset used for estimations is the pooled quarterly data comprising of 588,100 employee-individuals. Unemployed, economically inactive, self-employed and employee-respondents having reached the official state retirement age (63 for men and 58 for women) were filtering off in order to ensure accuracy and better comparability.

The survey consists of questions about employment, household spending and savings, and individual incomes. The question regarding employment changes in 2015: before 2015, the respondents were asked if they had worked for at least 1 h during the past 7 days and received monetary payment or payment in kind, which allowed them to be considered as employed in accordance with the ILO approach; from 2015 onwards, they have been asked if they have worked at least 1 h in the past 30 days. Whether or not both questions may cause inaccuracy is open to question, as no data on hours worked by the individual are recorded. By using wage data aggregated quarterly (not hourly) and with no information on full- or part-time employment, we violate two conditions set by Griliches (1977), thus requiring additional caution when interpreting schooling and experience estimates.

The dependent variable in all models is the natural logarithm of the real wage from employment; thus, other earnings (income from self-employment, benefits, property income and other incomes) are excluded from the analysis. Wages for 2003–2016 are adjusted by the CPI officially reported by the Committee on Statistics, with 2002 as the base year.

Schooling is a derivative variable transformed from the levels of education attained that are recorded in the survey⁵:

⁴ There are fewer observations appearing in the sample in some years.

⁵ The coding for level of education attained in the original dataset changed in 2011, where the coding introduced in 2011 is presented in the text. Before 2011, there was no Master's degree recorded separately, while TVET was classified into two groups: 'initial vocational training' requiring a minimum of 11 years of schooling, and 'secondary vocational education', requiring a minimum of 12 years of schooling.

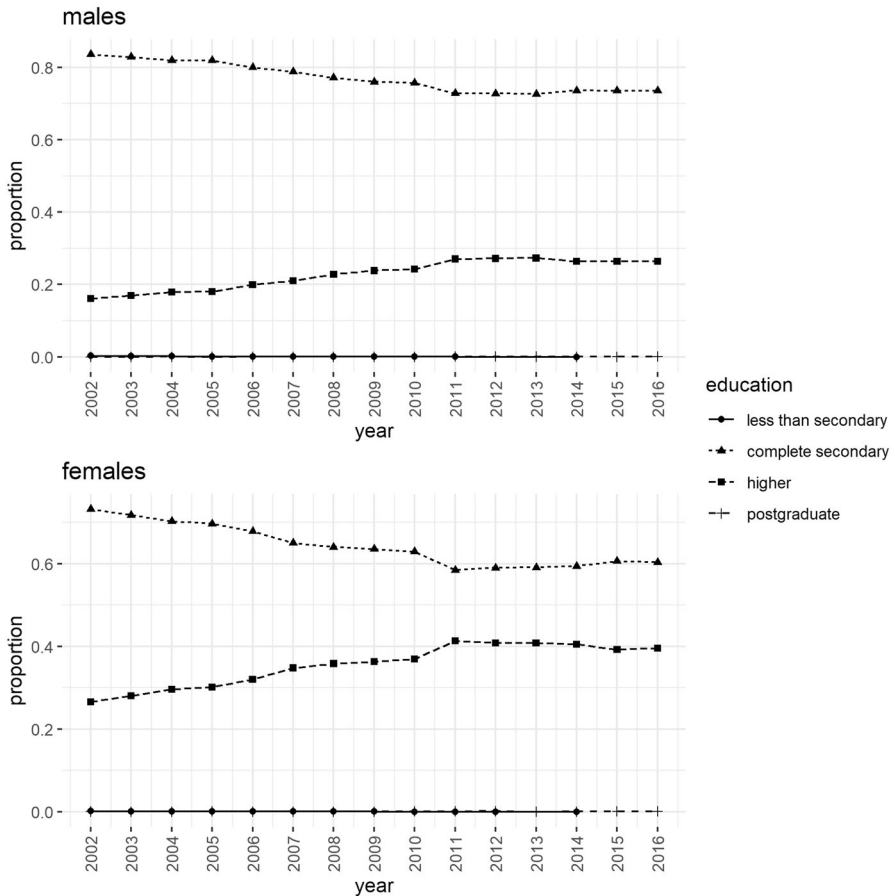


Fig. 2 Proportions of the respondents by attained level of education in the corresponding year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

- no education: 0 years of schooling;
- primary education: 4 years of compulsory schooling (from the age of 6–7);
- basic secondary education: 9 years of compulsory schooling;
- general secondary education and TVET: minimum of 11 years of schooling (9 years of compulsory schooling plus either 2 years of university-preparatory secondary school or 2–3 years of specialized technical and vocational training);
- higher education: minimum of 15 years of schooling (11 years of secondary schooling plus a Bachelor’s or ‘specialist’ degree requiring a minimum duration of 4 years;⁶)
- postgraduate education (Master’s degree): minimum of 16 years of schooling;⁷

⁶ In some fields, such as Medicine, 5 or 6 years.

⁷ There are two types of Master’s degree in Kazakhstan: a professional 1-year programme and a research 2-year programme.

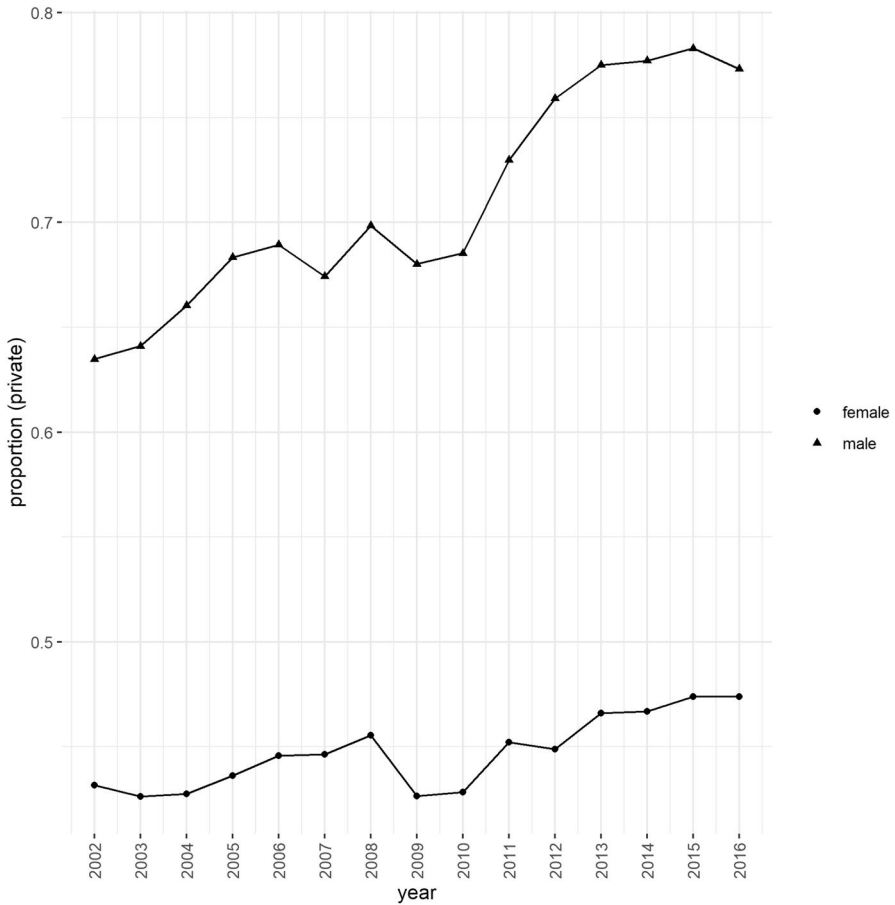


Fig. 3 Proportions of the respondents employed by the private sector in the corresponding year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

- academic degree (*uchenaya stepen*): (1) Ph.D. and (2) “Candidate of Science” and ‘Doctor of Science’ degrees from the previously existing Soviet system—minimum of 18 years of schooling.⁸

Therefore, with the given variable, we estimate the returns to credentials rather than the returns to schooling, albeit with its average rate, accepting that, as Harmon and Walker note from comparison between returns to schooling (linear specification) and returns to credentials (with nonlinearities between completion of different qualification assumed) computed for the same sample of individuals, “a linear form seems

⁸ ‘Candidate of Science’ degree was a research degree accessible upon completion of the analogue of the Bachelor’s degree (*specialist*), requiring a minimum of 3 years of training and research; the ‘Doctor of Science’ degree requires an additional 3 years of research work.

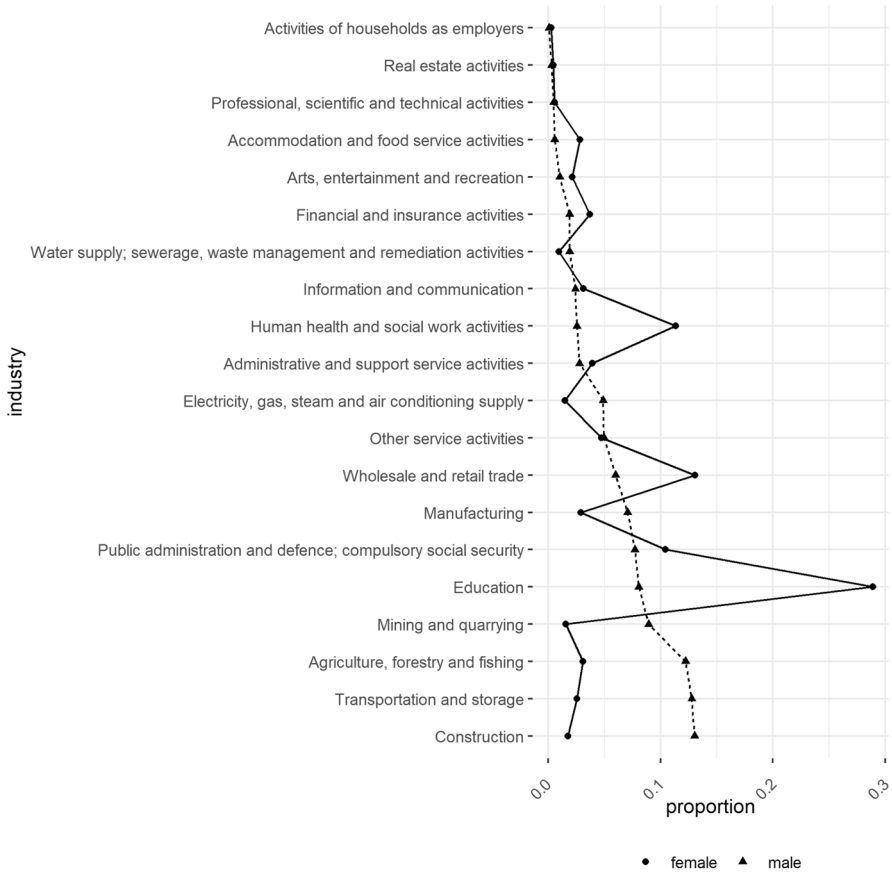


Fig. 4 Distribution of 2011–2016 respondents by industry. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

to be a reasonable approximation” (Harmon and Walker 2001, p. 31). However, we cannot directly test the sheepskin effect with no direct data on years of schooling.

Descriptive characteristics of the sample, as divided by gender, are given below (descriptive statistics can be found in Appendix 2). The following figures demonstrate summary statistics on gender subsamples and shed some light on the character of employment dissimilarities between the genders.

Figure 2 describes the distribution of the respondents’ highest attained levels of education combined in four wider groups by year of observation. For both genders, the majority of respondents had attained a general secondary education or TVET. The share of respondents with a degree in higher education grew until 2011 and was consistently higher for females in each year (mean of schooling in the pooled data is 0.62 years higher for females than males). This corresponds to the official aggregated statistics reporting the share of people having attained at least a higher

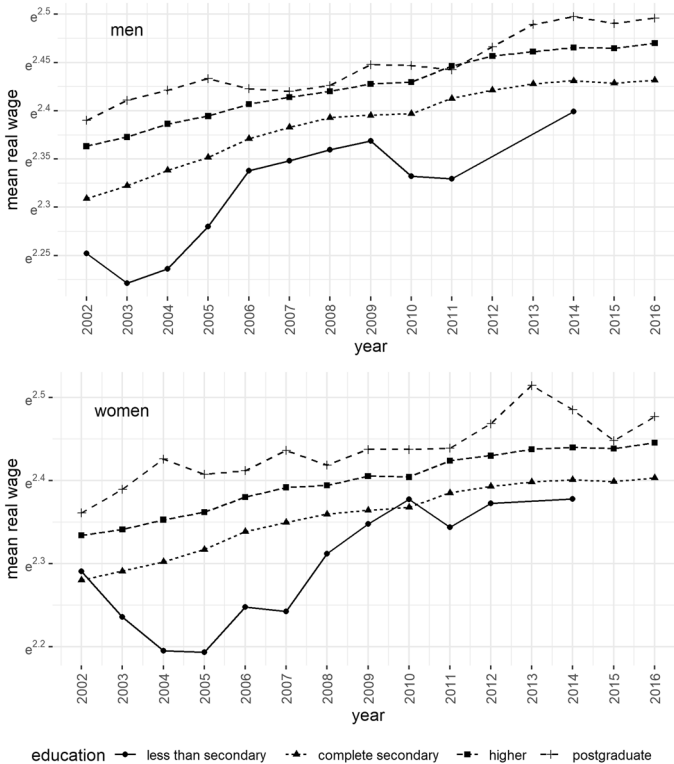


Fig. 5 Mean real wages by attained level of education and year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

education as comprising 35.7% of the employed population in 2016 (31.7% and 40.3% for men and women, respectively) (MNE 2017). The number of people with no schooling and with postgraduate degrees (including academic degrees) is very low for both subsamples, respectively.

As can be seen from Fig. 3, men are mostly employed by the private sector and the share of such increases over time, while women are approximately evenly distributed between public and private sector employment. It is noticeable that females' employment by sector is almost static. This is also reflected in the industry of employment (more precisely, the 'type of economic activity', which we further refer to as an 'industry'), with plot 4 showing the number of employees in different industries built for respondents for 2011–2016 only, since the earlier data does not record industry. Leading industries for male workers are those with primarily private ownership (construction, transportation, mining and quarrying, agriculture), while nearly 30% of working females in the sample are employed in education (with the majority in public secondary education) (Fig. 4).

Figures 5, 6, 7, 8 and 9 document average real wages by the respondents' selected characteristics used as explanatory variables in different specifications, as separately

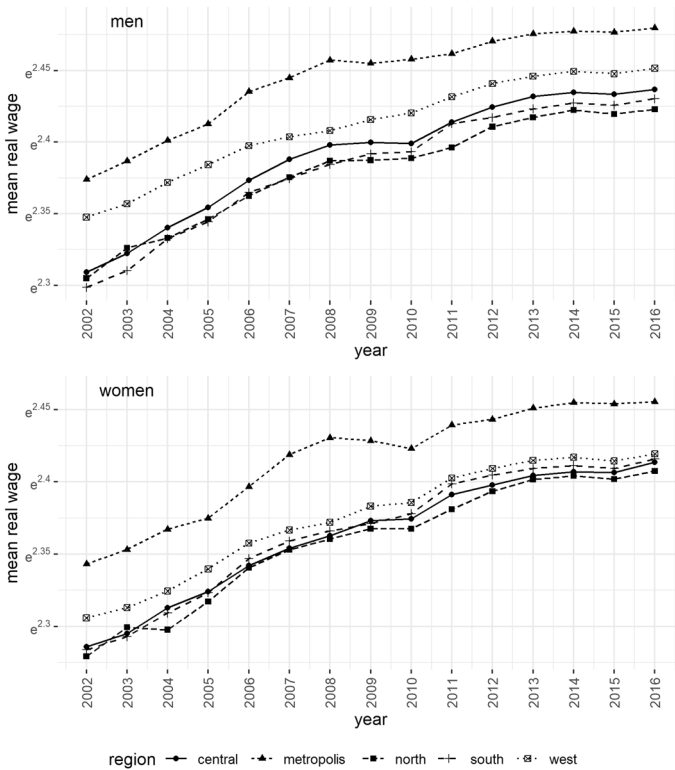


Fig. 6 Mean real wages by region and year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

computed from the pooled data for each gender.⁹ Region is aggregated into four geographical groups and the ‘metropolis’ category, which includes the two largest cities having the highest wages, followed by the western regions specializing on oil and gas exports. Sector of employment is derived from the categories listed in the questionnaire: ‘private company employee’, ‘farm-worker’ and ‘those employed by individuals’, the latter group mainly consisting of shadow sector wage-earners—combined into the ‘private sector’, and ‘public company employee’ comprising the ‘public sector’. As seen from Fig. 5, higher educational attainments consistently provide higher wages, on average. Excluding the highest and lowest levels of education (which both have very few observations), the log transformed average real wages demonstrate a somewhat parallel pattern for the two largest groups in each of the genders. There is a gender gap in almost every category for every variable observed.

The pseudo-panel was designed from individual data based on the respondents’ recorded years of birth and gender. The youngest and oldest cohorts are dropped due

⁹ For consistency with the Mincerian specification and pseudo-panel technique, we first log transform real wages, and then average them, and therefore the figures show exponentiated real wages rather than being in units of the national currency.

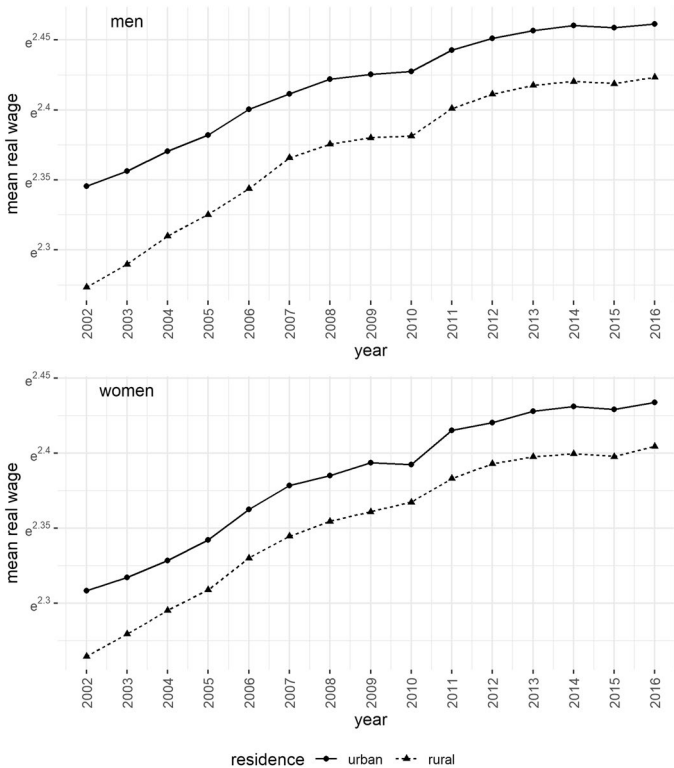


Fig. 7 Mean real wages by residence and year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

to either having only a few or no observations for the particular year(s). The final cohorts are:

- Male cohorts: 1954–1986, one per each year of birth;
- Female cohorts: 1959–1986, one per each year of birth.

The final number of cohorts is 61 (33 male cohorts and 28 female cohorts). The cohort size is sufficiently large, with mean numbers of observations of 9272 and 10,075 per cohort for the male and female cohorts, respectively, and smallest cohort sizes of 6352 (male) and 7156 (female)—see Appendices 3 and 4. However, the size of the cohort varies substantially over the years, therefore, in accordance with Deaton’s approach, observations are weighted in the pseudo-panel data by the square root of the corresponding cohort size in any given year of observation.

Since the data used for this analysis has a rather large time-series dimension observed over the period of the rapid economic growth the country experienced with the oil boom, one might be concerned about a possible non-stationarity issue potentially leading to spurious regression (Kao 1999; Phillips and Moon 2000). Indeed,

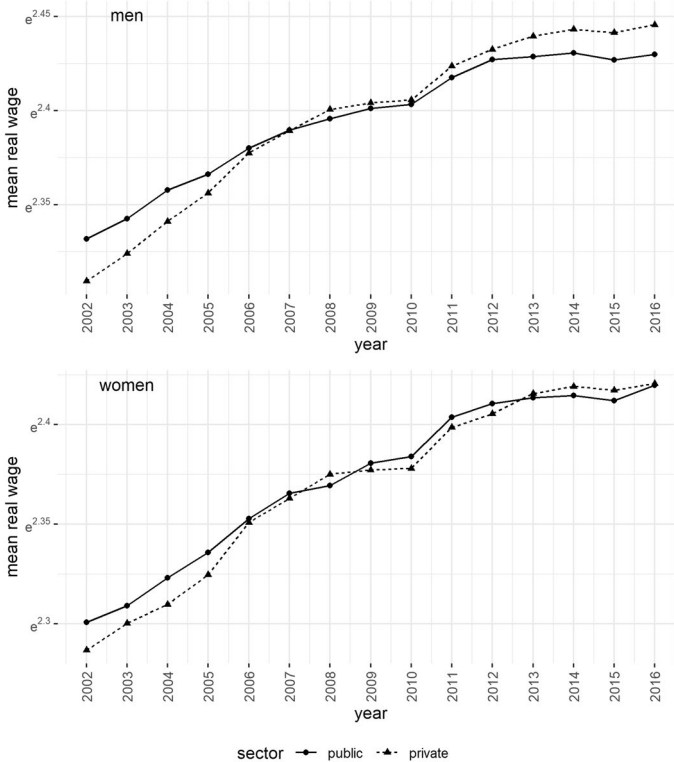


Fig. 8 Mean real wages by sector of employment and year. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

although it is unlikely this can be tested properly with 15 time series, the growth of real wages over the years is easily visually detected (Figs. 5, 6, 7, 8). We cannot disregard this possibility, however:

- (1) we rely on the fact that the number of cross-section units (n) is still larger than the number of times series periods (T), allowing one to interpret them as “a set of cross-sections... [rather than] a set of time series” (Smith and Fuertes 2010, p. 8), and it is believed that the non-stationarity problem normally arises in the panel data context when $n \leq T$ (Phillips and Moon 2000; Baltagi and Kao 2001; Smith and Fuertes 2010);
- (2) according to Kao (1999, p. 2) even with non-stationary data in the panel regression, the “LSDV estimator... is consistent for its true value”, though the standard errors could be biased. Thus, in this study we employ ‘classical’ micro-panel methods (Chamberlain 1984; Wooldridge 2010), additionally detrending them by explicitly controlling for the year dummy variables.

Descriptive statistics on cohort data are given in Appendices 5 and 6. Figure 10 demonstrates the decomposition of real wages by cohort. Each line represents the

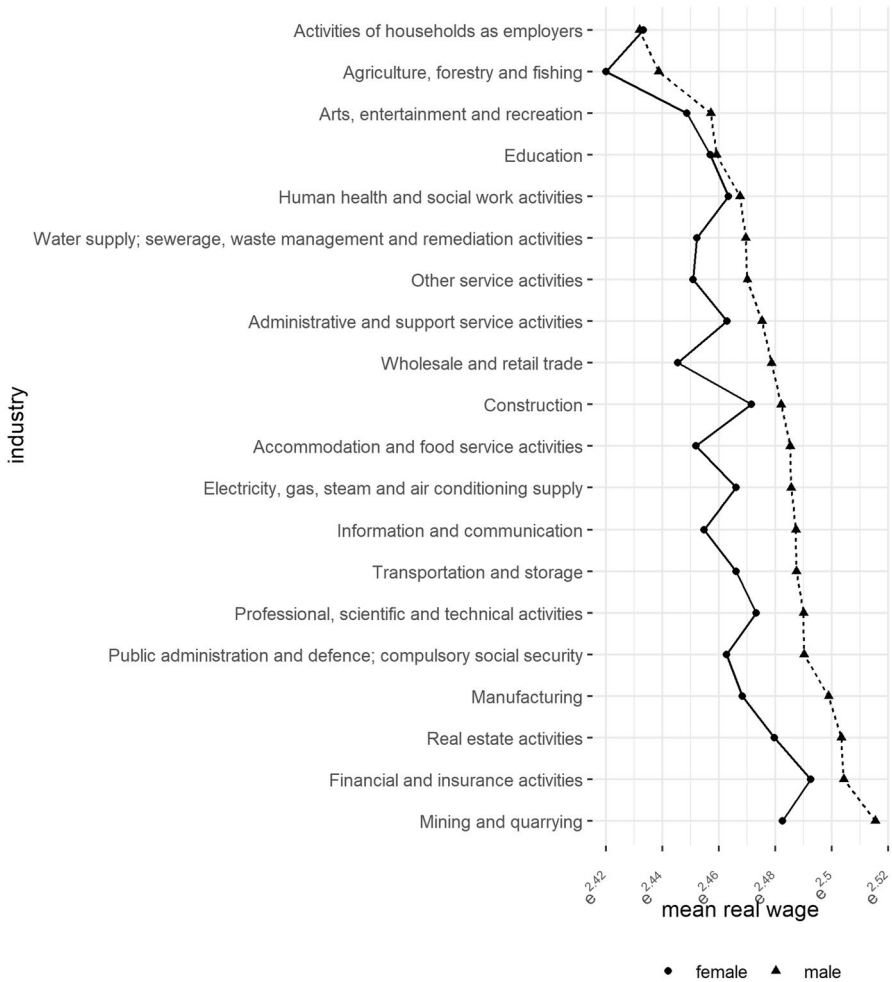


Fig. 9 Mean real wages by industry, pooled data, 2011–2016. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

evolution of the particular cohort's mean real wages over time; every third cohort is plotted to keep the plot less 'busy'.

As surmised, the youngest cohorts earn the lowest wages, and the age-wage profile has a somewhat inverted U-shape for both males and females. The picture additionally reflects the recessions of 2008 and 2014 due to the world market commodity bust when all cohorts' real wages dropped slightly. This is also evident from Fig. 11, which shows the year effect, where each dot represents the cohort's mean real wage observed for each year and the line represents the mean of each year's means.

In this analysis, we start with the 'classical' Mincerian specification, with the age in linear and quadratic terms as the only control variables, further augmenting with additional controls: region, residence and sector of employment (private vs. public):

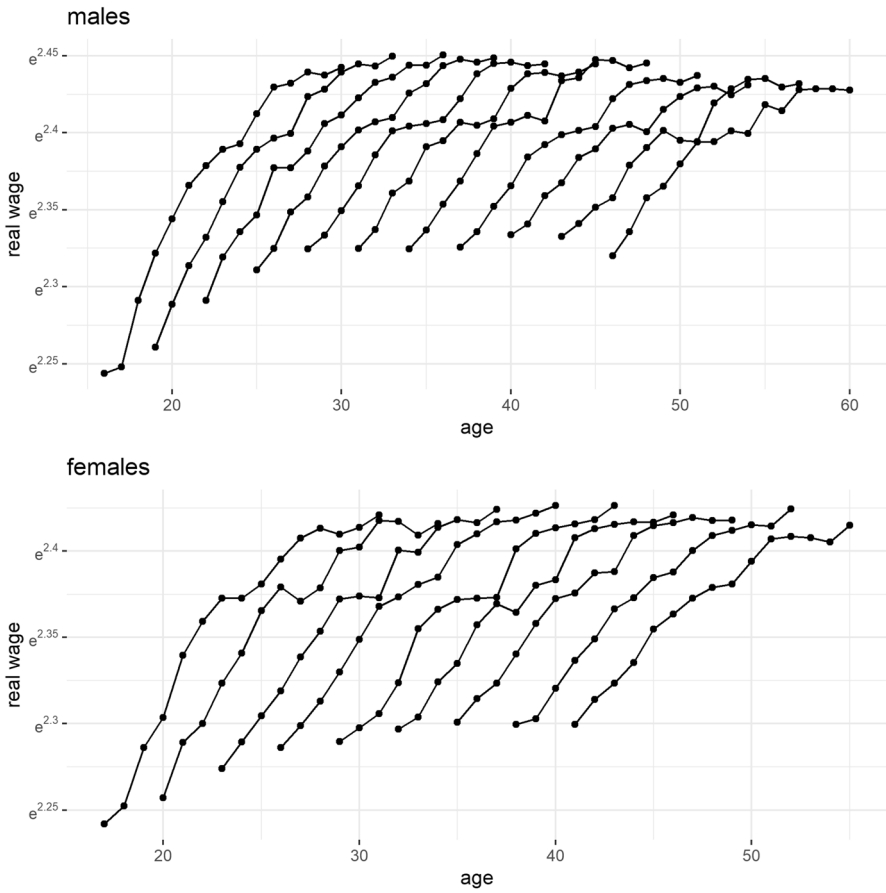


Fig. 10 Decomposition of real wages by cohort and age effect for every third cohort. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

$$\log w = \beta_0 + \beta_1 S + \beta_2 A + \beta_3 A^2 + X\beta + \epsilon \tag{6}$$

where

- w —real wage
- S —years of schooling
- A —age
- X —additional control variables
- ϵ —composite error term.

For pseudo-panel models, we run log and squared transformations before taking means.

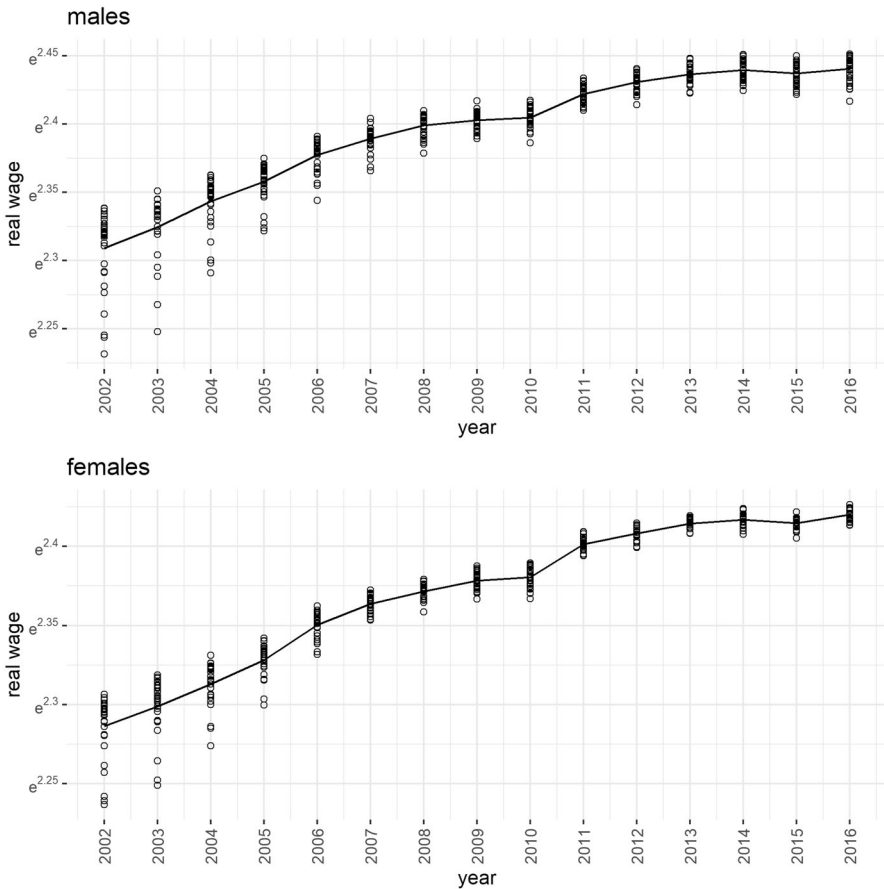


Fig. 11 Decomposition of real wages by cohort, year effect. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

When estimating models, we follow three slightly different approaches. First, we start with the basic OLS models which do not control for the cohort effect and which, therefore, likely suffer from omitted variable bias. Considering the long series of the repeated cross sections, we further concentrate on the first and the final years and elaborate on the changes that took place over the period in question by decomposing the wage equation using the Blinder–Oaxaca technique. This allows us to observe the year effects and, further, we additionally focus on the gender differences in the returns to schooling computed from the individual data. Second, we estimate the fixed effects and the Mundlak random effects models on the pseudo-panel data with cohort means treated as individual observations. Finally, to grasp the returns' variations across cohorts, we estimate the OLS model, separately controlling for cohort dummies and their interaction with schooling for individual gender subsamples.

Table 1 Returns to schooling estimated from pooled individual data by OLS, men. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.109*** (0.001)	0.091*** (0.001)	0.075*** (0.001)	0.075*** (0.001)
Age	0.063*** (0.001)	0.064*** (0.001)	0.065*** (0.001)	0.065*** (0.001)
Age squared	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)
Central		− 0.461*** (0.004)	− 0.307*** (0.004)	− 0.307*** (0.004)
North		− 0.577*** (0.004)	− 0.387*** (0.004)	− 0.387*** (0.004)
South		− 0.575*** (0.003)	− 0.335*** (0.004)	− 0.335*** (0.004)
West		− 0.261*** (0.004)	− 0.079*** (0.004)	− 0.079*** (0.004)
Rural			− 0.395*** (0.002)	− 0.395*** (0.002)
Private				0.003 (0.002)
Constant	7.680*** (0.020)	8.299*** (0.019)	8.452*** (0.019)	8.447*** (0.019)
F statistic	9849.2***	10472***	11590***	11178***
N	305,990	305,990	305,990	305,990
Adj. R2	0.385	0.437	0.486	0.486
Year dummies	Yes	Yes	Yes	Yes
Joint significance of year dummies				
Wald chi2	132253.4***	128566.2***	153383.3***	150551.9***
F test	9446.7***	9183.3***	10956***	10754***

(1) OLS computed in R. HC1 robust standard errors computed with ‘sandwich’ package (Zeileis 2004) in parentheses. Wald statistics computed with ‘aod’ package (Lesnoff and Lancelot 2012). (2) Reference categories: metropolis (region); urban (residence); public (sector)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4 Outcomes and discussion

4.1 Year effects

Tables 1 and 2 report the outcomes of the basic Mincer model computed from the pooled individual data with additional control variables introduced step-by-step, separately, for each gender. The returns to additional year of schooling vary from 7.75 to 11.50% for men and from 10.12 to 12.73% for women, depending on specification.

To examine the year dynamic, the same models are estimated for each year’s subsamples independently. Tables 3 and 4 document detailed outcomes for the first and the final years for men and women, respectively, while Appendix 7 details the descriptive statistics for these 2 years’ subsamples.

For both genders, regional and residency disparities have mitigated over the analyzed period, and employment in the private sector has become more lucrative,

Table 2 Returns to schooling estimated from pooled individual data by OLS, women. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.120*** (0.001)	0.106*** (0.001)	0.099*** (0.001)	0.096*** (0.001)
Age	0.034*** (0.001)	0.037*** (0.001)	0.037*** (0.001)	0.037*** (0.001)
Age squared	− 0.0004*** (0.00001)	− 0.0004*** (0.00001)	− 0.0004*** (0.00001)	− 0.0004*** (0.00001)
Central		− 0.442*** (0.004)	− 0.374*** (0.004)	− 0.379*** (0.004)
North		− 0.483*** (0.004)	− 0.394*** (0.005)	− 0.400*** (0.005)
South		− 0.440*** (0.004)	− 0.330*** (0.004)	− 0.340*** (0.004)
West		− 0.332*** (0.004)	− 0.248*** (0.004)	− 0.256*** (0.004)
Rural			− 0.198*** (0.002)	− 0.205*** (0.002)
Private				− 0.034*** (0.002)
Constant	7.727*** (0.022)	8.228*** (0.022)	8.288*** (0.022)	8.351*** (0.022)
F statistic	10906***	10384***	10579***	10139***
N	282,110	282,110	282,110	282,110
Adj. R2	0.417	0.453	0.466	0.466
Year dummies	Yes	Yes	Yes	Yes
Joint significance of year dummies				
Wald chi2	121185.1***	121269.3***	129101.9***	129086.0***
F test	8656.1***	8662.1***	9221.6***	9220.4***

(1) OLS computed in R. HC1 robust standard errors computed with ‘sandwich’ package (Zeileis 2004) in parentheses. Wald statistics computed with ‘aod’ package (Lesnoff and Lancelot 2012). (2) Reference categories: metropolis (region); urban (residence); public (sector)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

especially for men. The returns to schooling decreased over the analyzed period for both genders, though the difference in the returns’ estimates in 2016 versus 2002 is slightly larger for men. This is also seen from Fig. 12, which plots the schooling coefficients with their confidence intervals extracted from the each year’s subsamples’ models. To test if the difference is statistically significant, we run the models for the pooled data with an interaction between schooling and year (reported in Appendices 8, 9) separately for each gender, and found them to vary between -4.23 and -5.28% for men and between -2.43 and -2.71% for women. For comparison, the differences in real wages in 2016 compared to 2002, as computed from the same models, were 516 – 610% and 365 – 396% for men and women, respectively. The wages increased dramatically with the oil boom, while the returns to schooling dropped. This might be related to the corresponding increasing trend in years of schooling observed with the descriptive statistics.

To identify the extent to which the observed characteristics contributed to the change in the real wages of working employees between 2002 and 2016, following Lassibille and Gomez (1998) we decompose the wage equation with the

Table 3 Returns to schooling estimated from individual data for 2002 and 2016, men. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Dependent variable	(2)		(3)		(4)	
	2002	2016	2002	2016	2002	2016
Log real wage						
Schooling	0.147*** (0.004)	0.108*** (0.004)	0.129*** (0.002)	0.090*** (0.004)	0.106*** (0.003)	0.077*** (0.002)
Age	0.112*** (0.006)	0.045*** (0.004)	0.116*** (0.006)	0.040*** (0.006)	0.115*** (0.005)	0.044*** (0.004)
Age squared	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)
Central			-0.566*** (0.019)	-0.359*** (0.011)	-0.396*** (0.018)	-0.225*** (0.012)
North			-0.625*** (0.020)	-0.504*** (0.012)	-0.384*** (0.020)	-0.345*** (0.012)
South			-0.681*** (0.017)	-0.472*** (0.010)	-0.351*** (0.018)	-0.298*** (0.011)
West			-0.212*** (0.023)	-0.230*** (0.011)	-0.014 (0.022)	-0.082*** (0.012)
Rural					-0.603*** (0.012)	-0.293*** (0.007)
Private						-0.599*** (0.012)
Constant	6.439*** (0.100)	9.365*** (0.091)	7.101*** (0.097)	10.018*** (0.089)	7.379*** (0.094)	10.076*** (0.087)
F statistic	830.6***	1372.1***	683.59***	1064.5***	918.72***	1142.6***
N	18,999	25,225	18,999	25,225	18,999	25,225
Adj. R2	0.118	0.138	0.185	0.211	0.283	0.262

(1) OLS computed in R. HC1 robust standard errors computed with 'sandwich' package (Zeileis 2004) in parentheses. (2) Reference categories: metropolis (region); urban (residence); public (sector)
 *p<0.1; **p<0.05; ***p<0.01

Table 4 Returns to schooling estimated from individual data for 2002 and 2016, women. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Dependent variable	(2)		(3)		(4)	
	2002	2016	2002	2016	2002	2016
Log real wage						
Schooling	0.144*** (0.004)	0.120*** (0.004)	0.127*** (0.004)	0.108*** (0.001)	0.120*** (0.004)	0.102*** (0.004)
Age	0.032*** (0.008)	0.040*** (0.008)	0.042*** (0.008)	0.036*** (0.004)	0.039*** (0.008)	0.039*** (0.008)
Age squared	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Central						
North						
South						
West						
Rural						
Private						
Constant	7.461*** (0.127)	8.842*** (0.102)	7.917*** (0.122)	9.370*** (0.098)	8.056*** (0.121)	9.388*** (0.097)
F statistic	568.85***	2183.9***	394.96***	1127.6***	429.56***	1142.6***
N	14,714	24,972	14,714	24,972	14,714	24,972
Adj. R2	0.122	0.204	0.172	0.254	0.200	0.277

(1) OLS computed in R. HCl robust standard errors computed with 'sandwich' package (Zeileis 2004) in parentheses. (2) Reference categories: metropolis (region); urban (residence); public (sector)

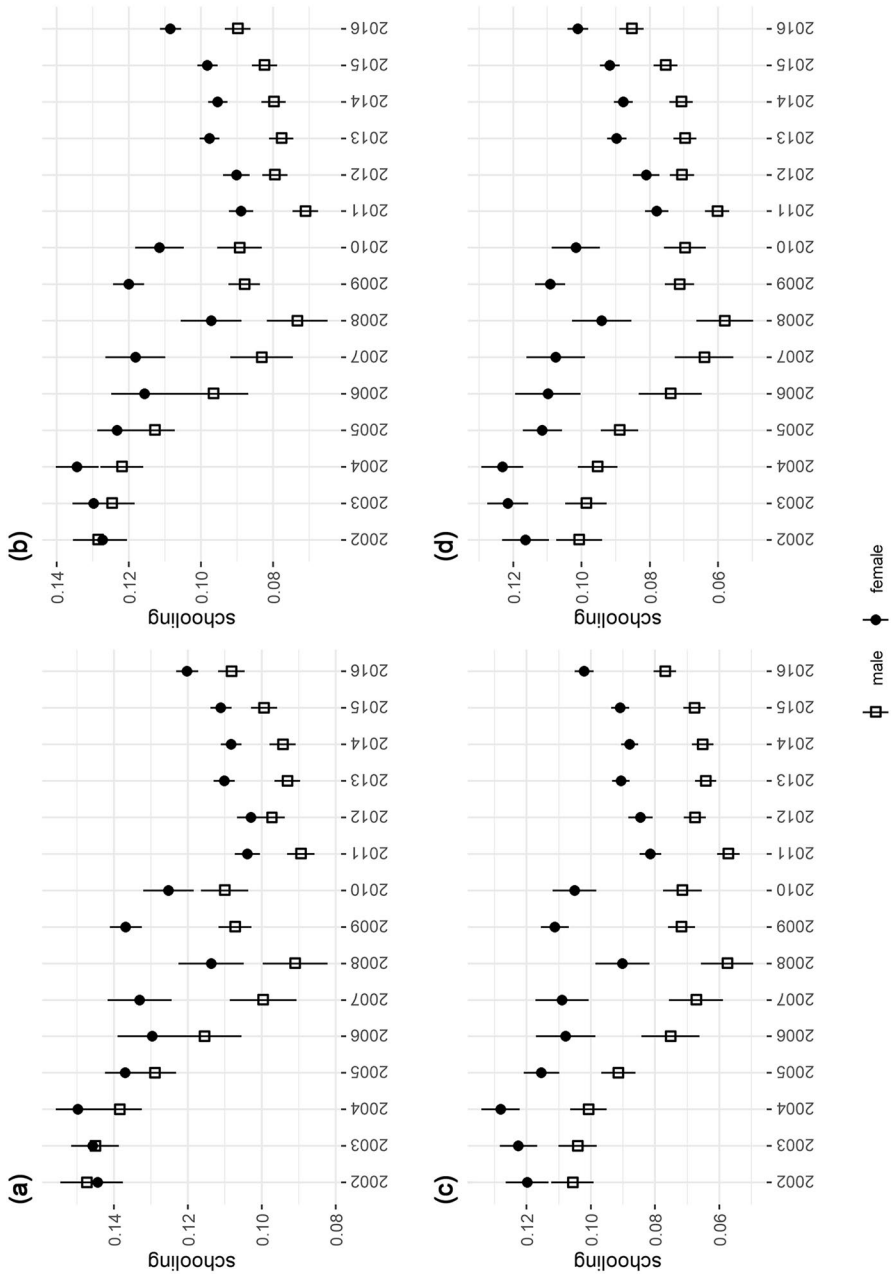
*p<0.1; **p<0.05; ***p<0.01

Blinder–Oaxaca decomposition for these 2 years for each of the gender subsamples (Blinder 1973; Oaxaca 1973). The results presented in the Tables 5 and 6 allow us to examine which part of the wage difference between the 2 years can be explained by the changes in covariates and which is unexplained and should be attributed to the “differences in the pay structure” (Lassibille and Gomez 1998, p. 7) or other structural changes across the period under consideration. 2002 is set as the reference year.

The mean real wage of male employees in 2002 is 10.15 log points, and in 2016 is 11.50 log points. Within the difference of approximately 1.35 log points, only 0.43 log points (or about 32%) is explained by the change in the observed characteristics and 0.91 log points (or 68%) remains unexplained. Though the log real wages are lower for females for both years (9.92 in 2002 and 11.25 in 2016), the cross year difference is nearly identical at 1.33 log points, but the model turned out to be slightly more powerful in terms of explaining the coefficients dynamic, as almost 40% of the difference is explained. The main contributor to the unexplained part for both genders is the constant term, which likely picks up the oil boom-driven differences in economic conditions over time. Among the observed employees’ characteristics, the age in both linear and quadratic terms contributes the most, which is also shown by the variable-by-variable decomposition of the unexplained differentials plotted in Figs. 13 and 14, which might be partially attributed to the observed difference in the age between year subsamples for each gender (average age in 2002 is 34 versus 45 in 2016 for males and 33 versus 44 for females—see Appendix 7). However, a more interesting and meaningful pattern appears with the schooling coefficient’s unexplained part. Since the 2016 sample is more educated (mean of schooling is 11.96 for males and 12.53 for females versus 11.74 and 12.37 in 2002, respectively), the returns to schooling contribute to the observed wage differentials. At the same time, taking 2002 as a baseline year, schooling is significantly ‘underestimated’ in 2016 for both genders. Other coefficients (except for west region residency) have different patterns. For example, employment in the private sector provided a higher wage premium for men in 2016 than it was otherwise ‘expected’ to provide, accounting for the difference across years which is likely to be explained by structural changes in the economy during the same period. On the other hand, the interpretation of the nominal variables (region, residence or sector) coefficients’ unexplained portion might be biased and meaningless (Jann 2008), unlike those for schooling and age having a natural zero point.

4.2 Gender effects

If anything, the results presented above reveal systematically higher returns to schooling for females, which are additionally shown in the models run using the pooled data with the schooling-gender interaction term included, as performed as a robustness exercise (Appendix 10). Male respondents earn about 0.9–1.3% less premium for each additional year of schooling than the female respondents, depending on specification. This is often the case in developing countries, and is usually attributed to “lower base levels of education of females compared to males” (Warunsiri and McNown 2010). However, this interpretation is hardly appropriate



◀**Fig. 12** Returns to schooling with 95% confidence intervals independently computed for each year of observation from models with schooling, age, age squared and: **a** with no additional control variables; **b** region; **c** region and residence; **d** region, residence and sector of employment. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

for Kazakhstan with its observed low (or absence of) educational gender disparities during the Soviet period, which turned into the current higher overall level of education amongst females, as may be noted from the analyzed database. Along with the overall lower female earnings, this suggests that the observed returns gap might be attributed to the sector-industry allocation of the labor force: as mentioned above, females are mostly employed by (less risky but worse paid) industries are primarily in public ownership, which might 'value' education more than the others in the sense that the jobs concentrated in these industries require a degree and provide better degree premiums. This observation might be in line with frequently observed worldwide "gender segregation in occupations, industries, firms, and jobs" (Garcia-Aracil 2007, p. 431; Meng 2004; Bielby and Baron 1986).

To test this, we run OLS on individual data gender and year subsamples with the basic Mincerian specification and the interaction term of the years of schooling with the sector of employment—the interaction term estimates are shown in Fig. 15. Despite the fact that for both genders, employees in the public sector are better educated than those in the private sector (the average schooling is 12.50 for men and 12.88 for women in the public versus 11.74 for men and 12.19 for women in the private sector,¹⁰) the returns to schooling in the private sector are significantly higher for women for almost all of the years observed, and higher for men until 2011 though with a clearly decreasing trend over time. The latter, along with the observed shift of the male respondents towards the private sector, likely contributes to the decreasing returns to schooling in males. However, the observed trend does not explain the higher returns to schooling amongst females.

Hypothesizing that the schooling returns premium gap could be better explained by industries rather than private versus public ownership, we compute the same specification models for the industry (instead of the sector) of employment, which appears in the survey from 2011; the associated results are presented in Fig. 16,¹¹ additionally plotting the number of employees in each industry. We drop the least populated industry ("Activities of households as employers") and assign "Education" as the reference category. Additionally, we show the average years of schooling for each industry and gender—see Fig. 17.

Indeed, in the majority of industries, females' returns to schooling are greater than or equal to those of males, although this is not the case for the main female industries of employment. Along with this, for both genders the main industries of employment (according to Fig. 4) are among those with both (gender-specific)

¹⁰ Computed for the pooled data, the pattern persists in each year.

¹¹ The results presented are computed for the pooled dataset; we do not report the yearly estimations for simplicity, as the outcomes do not vary significantly across years.

Table 5 Blinder–Oaxaca decomposition of log real wage differential between 2002 and 2016, men. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Parameter	Contributions of parameters to log real wage differential		Contribution of parameters as % of total differential	
	Endowments	Pay structure	Endowments	Pay structure
Schooling	0.0260	– 0.1883	1.93	– 14.00
Age	1.1147	– 3.0141	82.89	– 224.14
Age squared	– 0.7696	1.5119	– 57.23	112.43
Central	0.0275	0.0354	2.05	2.63
North	– 0.0001	0.0118	0.00	0.88
South	0.0303	0.0198	2.25	1.48
West	0.0003	– 0.0144	0.02	– 1.07
Rural	– 0.0330	0.1305	– 2.45	9.70
Private	0.0299	0.1489	2.22	11.07
Constant	0.0000	2.2772	0.00	169.34
Total	0.4260	0.9187	31.68	68.32
	1.3447		100.00	

Blinder–Oaxaca decomposition is computed in R with ‘oaxaca’ package (Hlavac 2018)

Table 6 Blinder–Oaxaca decomposition of log real wage differential between 2002 and 2016, women. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Parameter	Contributions of parameters to log real wage differential		Contribution of parameters as % of total differential	
	Endowments	Pay structure	Endowments	Pay structure
Schooling	0.0222	– 0.1944	1.68	– 14.66
Age	1.0810	– 0.5957	81.52	– 44.92
Age squared	– 0.6024	0.2190	– 45.43	16.52
Central	0.0203	0.0353	1.53	2.66
North	0.0041	0.0251	0.31	1.89
South	0.0143	0.0238	1.07	1.80
West	– 0.0146	– 0.0012	– 1.10	– 0.09
Rural	– 0.0088	0.0184	– 0.66	1.39
Private	0.0035	0.0092	0.27	0.69
Constant	0.0000	1.2670	0.00	95.54
Total	0.5196	0.8064	39.19	60.81
	1.3261		100.00	

Blinder–Oaxaca decomposition is computed in R with ‘oaxaca’ package (Hlavac 2018)

lowest returns to schooling and the lowest levels of education attained. The only exemption is “Education” for females, which provides relatively high returns to schooling and demands more schooling for both genders. At the same time, the industries with the highest wages for each of the genders—particularly, “Mining

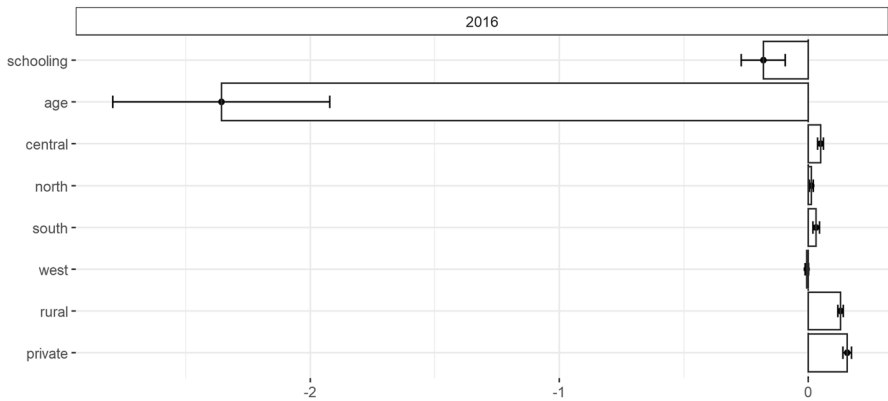


Fig. 13 Blinder–Oaxaca decomposition—unexplained portion in wage differential between 2002 and 2016, men (with 95% confidence intervals, 2002 is the reference year). Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

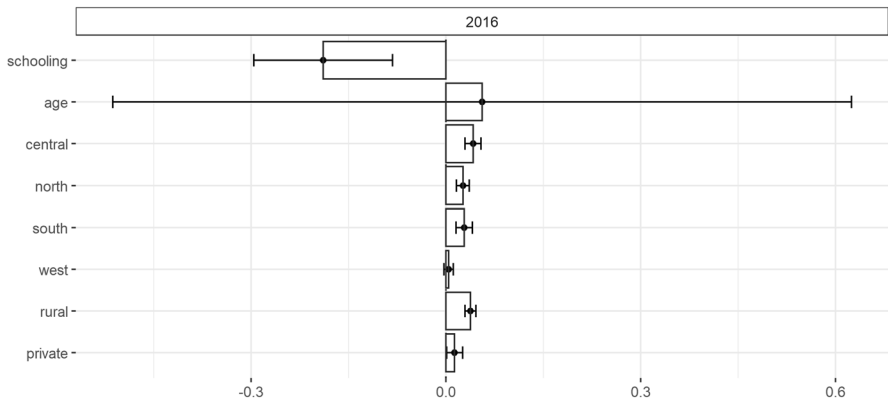


Fig. 14 Blinder–Oaxaca decomposition—unexplained portion in wage differential between 2002 and 2016, women (with 95% confidence intervals, 2002 is the reference year). Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

and quarrying” and “Manufacturing”—are predominantly occupied by men. The exception is “Financial and insurance activities”, where there are more females than males observed. Unlike “Mining and quarrying” and “Manufacturing”, this sector ‘requires’ more schooling (the highest average schooling in the sample for both genders). This suggests that, possibly for females, a higher attained level of education serves as a pass to at least some of the best-paid industries within the mining-oriented economy. On the other hand, they could additionally choose less demanding (in terms of schooling and probably competition for jobs) but worse-paid economic activities, such as “Education”, “Human health and social work” or “Wholesale and retail trade”. The former two might additionally provide relatively improved levels of social security.

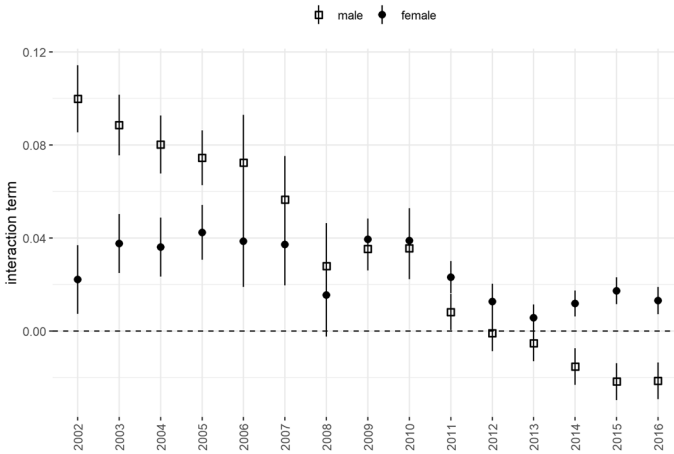


Fig. 15 Returns to schooling in the private sector with 95% confidence intervals computed for each year of observation and gender independently from models with schooling, age, age squared, year dummy and schooling with sector of employment interaction term (reference sector: public). Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Interestingly, experience coefficients (as reported in Tables 1, 2) are nearly twice as small for females than males; as a result, the age turning point differs with gender: about 41 for males and 47 for females. This might again reflect differences in the nature of employment and social security across genders. The gender returns gap additionally explains higher females' schooling that may be considered as their rational decision based on their expectations about future employment opportunities.

4.3 Cohort effects: pseudo-panel models

The results of the pseudo-panel models are introduced in Tables 7 and 8. Observations are weighted by cohort size, varying with both cohort and year.¹² Both models additionally include year dummies to capture year fixed effects, and other proportional control variables computed as proportions of respondents belonging to a certain group (for example, with rural versus urban residence) within a cohort in each observed year.

Interestingly, the fixed effects and the Mundlak random effects models produce results that are nearly identical to the individual data models, which are also very similar to the previous examinations that used the IV approach (Arabsheibani and Mussurov 2007). Attempts to control for cohort effects do not noticeably change

¹² Since there is no option to use weights varying in both—cohort and period—dimensions with the `xtreg` command, to compute the FE we used the `'regress'` command in Stata with cohort dummies and the `aweight` option, whilst to compute the RE we used the user-written `xtregre2` with `aweight` option (Merryman 2005).

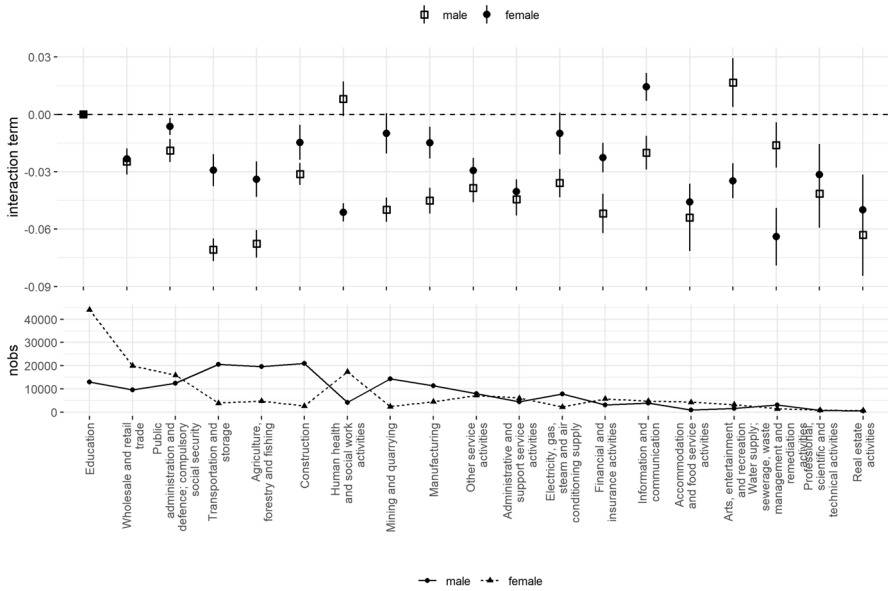


Fig. 16 Returns to schooling in industry with 95% confidence intervals computed for each gender independently from models with schooling, age, age squared, year dummy and schooling with industry of employment interaction term (reference industry: Education), pooled data, 2011–2016. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

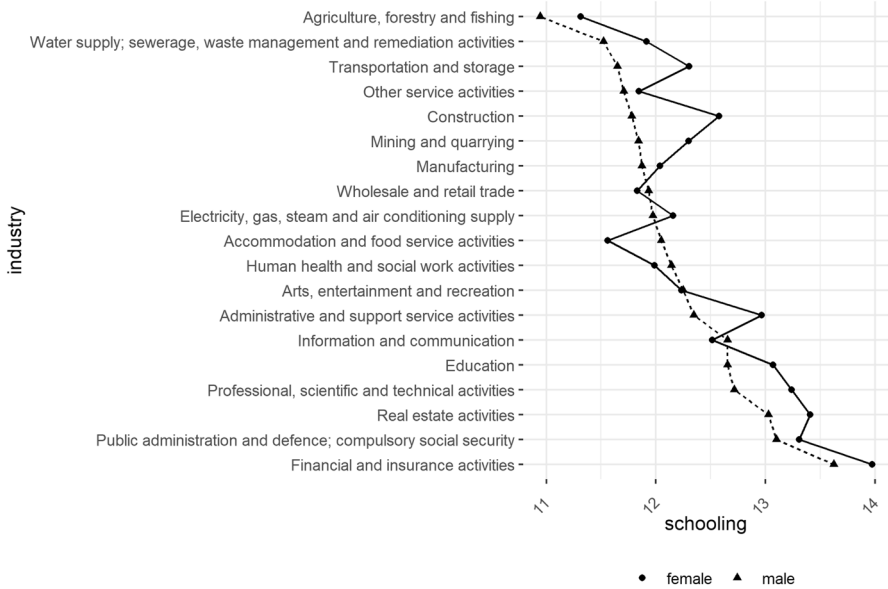


Fig. 17 Average years of schooling, pooled data, 2011–2016. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

estimations for the returns to schooling, only slightly decreasing the Mundlak model's estimates. At the same time, the mean of schooling estimate in the Mundlak model is highly significant, suggesting an unobserved cohort effect (from comparison of FE and RE). It is also negative, where this should be interpreted as follows: though the increase in the cohorts' average level of schooling over time resulted in an increase in the average level of their earnings, the cohorts with higher average levels of schooling earned less in comparison to those with lower levels.

The figures in Appendices 11 and 12 show that the more educated cohorts are the younger cohorts in the sample,¹³ and thus the negative relation could imply an age (experience) effect. It might also indicate a business cycle impact; the cohorts who entered the labor market during the 1990s recession and faced a lack of jobs apparently ended up getting more education and—because of worse economic conditions—lower lifetime wages, and vice versa (Betts and McFarland 1995; Dellas et al. 1996; Kahn 2010; Clark 2011; Oreopoulos et al. 2012; Liu et al. 2016). The other possible explanation could be a widely perceived difference in the quality of education and its link to the labor market between the Soviet and post-Soviet eras; however, this could not be further tested with the data at hand.

While OLS estimates reported in the previous sections have the expected signs and magnitudes, some of the additional control variables' estimates from the pseudo-models look rather controversial. This might reflect an error-in-variables bias generated by errors in the survey data, as noted by Griliches (1977, p. 12): “in cross-sectional household interview data all of the variables are subject to some error. Even if errors are small, their effect will be magnified as more variables are added to the equation in an attempt to control for “other possible sources of bias”. Regardless, the schooling coefficients are fairly robust and consistent in all models.

4.4 Cohort effects: individual level models

Finally, to detail how the returns to schooling vary across cohorts, we set up individual-level OLS models controlling for the set of cohort dummies and an interaction term between schooling and cohort dummies, as well as the other control variables, separately for each gender. The coefficients for the interaction terms with their confidence intervals for four specifications are shown in Figs. 18 and 19. The baseline category is the first (the oldest) cohort: the 1954 cohort for men and the 1959 cohort for women.

For men, with some few exemptions and for the youngest cohorts (born after 1983), the difference in the returns across cohorts seems to be statistically insignificant. Minor fluctuations observed in the returns look somewhat similar to the business cycle fluctuations and might reflect the external backgrounds (economic conditions, unemployment rate, labor market conditions and policies, skills mismatch, external shocks) cohorts are faced with when entering the labor market and over their life-time cycle accordingly impacting their returns. However, for females this

¹³ With the exception of the very youngest who had not completed their education on the date of the survey.

Table 7 Returns to schooling estimated from pseudo-panel, fixed effects. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.126*** (0.0080)	0.114*** (0.0080)	0.0938*** (0.0083)	0.0679*** (0.0090)
Age squared	– 0.0006*** (0.0000)	– 0.0006*** (0.0000)	– 0.0006*** (0.0000)	– 0.0006*** (0.0000)
Central		– 0.304*** (0.0819)	– 0.247** (0.0799)	– 0.335*** (0.0788)
North		– 0.274** (0.0898)	– 0.187* (0.0879)	– 0.261** (0.0863)
South		– 0.437*** (0.0825)	– 0.262** (0.0836)	– 0.386*** (0.0834)
West		0.0049 (0.0905)	0.0938 (0.0887)	0.0453 (0.0866)
Rural			– 0.369*** (0.0507)	– 0.419*** (0.0499)
Private				– 0.298*** (0.0437)
Constant	10.05*** (0.147)	10.46*** (0.172)	10.75*** (0.172)	11.33*** (0.187)
Year dummies	Yes	Yes	Yes	Yes
Cohort dummies	Yes	Yes	Yes	Yes
F statistic	1242.19***	1270.72***	1333.91***	1390.15***
N	915	915	915	915
Adj. R2:	0.990	0.991	0.992	0.992

(1) Fixed effects computed in Stata with regress command with analytical weights. Standard errors in parentheses. (2) Reference categories: metropolis (region); urban (residence); public (sector)

(3) Age in linear term is omitted to avoid collinearity

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

variation appears to be somewhat more systematic, with higher rates for the older cohorts and a downward trend towards the younger ones, namely those who entered the labor market after the collapse of the Soviet Union. This might mirror the different nature of employment for the genders, as referred to above: particular industries could be more sensitive to over-education or to perceived differences in Soviet-type and post-Soviet-type education. On the other hand, this could reflect (or be contaminated with) the age effect, and disentangling them is hardly feasible.

5 Conclusion

This paper represents the first attempt to employ a long series of repeated cross-sectional data from the Kazakhstani National Statistics to estimate the returns to schooling. The few such previous examinations found that the returns immediately increased from the very low rates typical of the Soviet era to internationally comparable rates with the transition.

Table 8 Returns to schooling estimated from pseudo-panel, Mundlak random effects. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.108*** (0.0081)	0.0883*** (0.0083)	0.0709*** (0.0087)	0.0764*** (0.0096)
Age	0.0484*** (0.0021)	0.0511*** (0.0021)	0.0518*** (0.0021)	0.0552*** (0.0021)
Age squared	− 0.0006*** (0.0000)	− 0.0007*** (0.0000)	− 0.0007*** (0.0000)	− 0.0007*** (0.0000)
Central		− 0.309*** (0.0889)	− 0.252** (0.0872)	− 0.215* (0.0909)
North		− 0.282** (0.0979)	− 0.198* (0.0965)	− 0.171 (0.101)
South		− 0.395*** (0.0899)	− 0.233* (0.0918)	− 0.162 (0.0967)
West		0.0514 (0.0978)	0.133 (0.0964)	0.163 (0.0998)
Rural			− 0.333*** (0.0553)	− 0.300*** (0.0575)
Private				0.160*** (0.0426)
Mean (schooling)	− 0.484*** (0.0168)	− 0.473*** (0.0149)	− 0.480*** (0.0149)	− 0.419*** (0.0193)
Constant	13.77*** (0.196)	14.06*** (0.193)	14.36*** (0.198)	13.33*** (0.318)
Year dummies	Yes	Yes	Yes	Yes
Wald chi2	76465.61 ***	75566.87***	79688.86***	72545.22***
<i>N</i>	915	915	915	915
R2				
Within	0.9904	0.9911	0.9916	0.9909
Between	0.7517	0.7610	0.7548	0.8031
Overall	0.9733	0.9747	0.9747	0.9779

(1) Random effects computed in Stata with xtregre2 with analytical weights (Merryman 2005). Standard errors in parentheses. (2) Reference categories: metropolis (region); urban (residence); public (sector)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Using the Household Budget Survey data for 2002–2016, we found three to six-fold growth in the real wages observed over this period. However, this dramatic growth in wages in real terms is only partially explained by the improvement in the observed labor force characteristics (education, residency and the sector of employment), by around 30% for working men and 40% for working women. The unexplained part can probably be attributed to the external economic background—fast economic growth due to the oil boom with the GDP growing by 6% on average annually.¹⁴ This growth in the GDP could have resulted in the increase in demand

¹⁴ The Committee on Statistics of the Republic of Kazakhstan, www.stat.gov.kz.

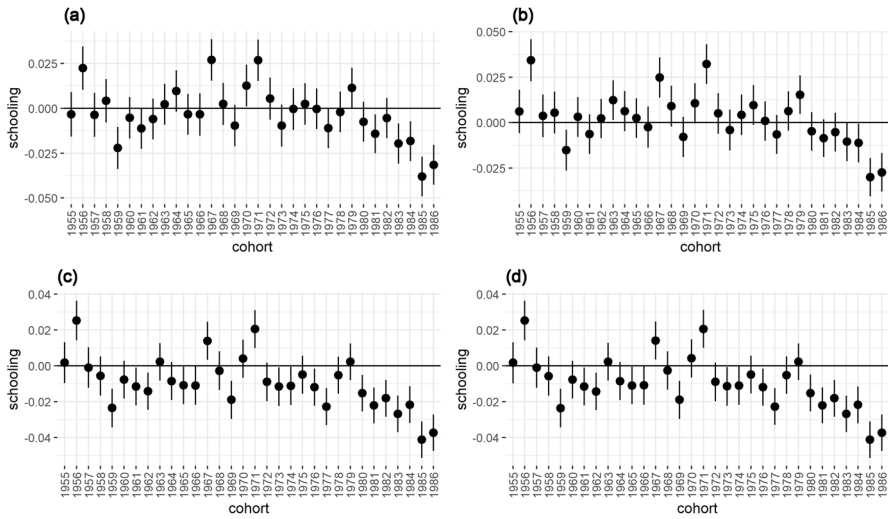


Fig. 18 Male cohorts: Returns to schooling for each cohort with 95% confidence intervals from model with schooling, age, age squared, cohort, schooling*cohort and: **a** with no additional control variables; **b** region; **c** region and residence; **d** region, residence and sector of employment. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

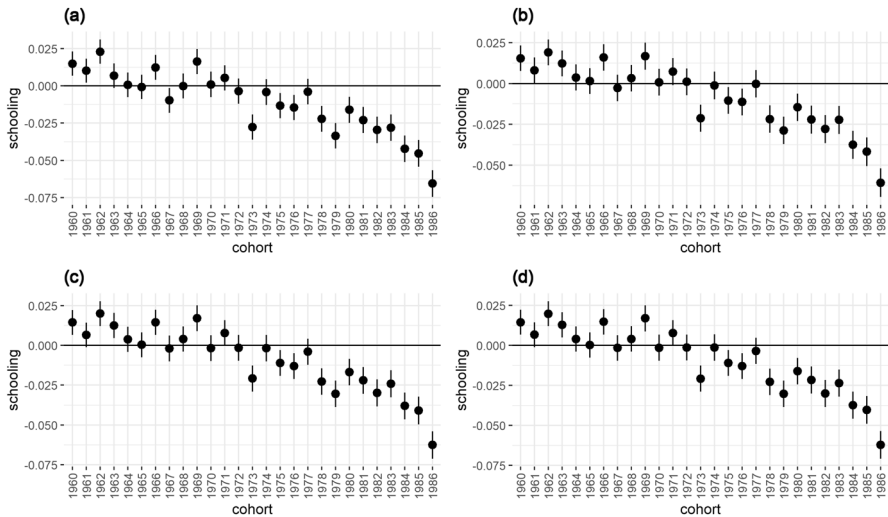


Fig. 19 Female cohorts: returns to schooling for each cohort with 95% confidence intervals from model with schooling, age, age squared, cohort, schooling*cohort and: **a** with no additional control variables; **b** region; **c** region and residence; **d** region, residence and sector of employment. Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

for a more educated labor force; on the other hand, with the associated dramatic increase in schooling, one might expect a corresponding decrease in the returns to schooling over the years.

With the various models developed, our results reveal the returns to be robust across models and specifications, statistically significant and relatively high, though decreasing over the analyzed years despite the economic boom and with the magnitude of this decrease being larger for men. Albeit that the estimates do not particularly change with the pseudo-panel data, the comparison between the fixed effects and the Mundlak random effects models suggests the presence of the cohort effect. We tend to consider the negative sign of the mean of schooling estimate as an effect of the business cycle phase which cohorts are faced with at their school-leaving age: a downward swing with a lack of jobs leaves them with a little choice but to get more schooling which, however, in turn provides them with lower returns (compared to older cohorts).

We find—at least, with the data at hand—gender differences in the nature of employment: while men prefer employment in the private sector with higher wages and apparently more uncertainty, women seem to be inclined towards worse-paid, but more secure industries and the public sector. This trait represents a likely explanation for the higher rates of the returns to schooling observed for females. We argue that some particular jobs, mainly in the public sector, formally require some level of schooling and, accordingly, value schooling to a greater degree than the private sector, possibly due to the wage grid still being persistent to some extent. These jobs tend to be occupied by females, which, in turn, reasonably explains their higher levels of education. On the other extreme, the females employed by the best-paid industries, as heavily dominated by male employees, have systematically higher levels of schooling (than males), which seems to serve as a pass to at least some of the best-paid industries for women within the mining-oriented economy.

Finally, we uncovered a downward trend in the returns pattern across cohorts, with younger cohorts demonstrating lower rates of returns (confirming the Mundlak model outcomes), which might reflect a decrease in returns to schooling due to a labor market glut but that might also reflect an age effect, and untangling these two effects is hardly feasible. This feature is more pronounced in females, which might again point to gender differences in the nature of employment. At the same time, for males, variations in the returns rates across cohorts can likely be explained by external economic conditions, such as labor market oscillations and the business cycle. Overall, the mechanism(s) determining the returns to schooling in Kazakhstan is (are) probably more complicated and require more sophisticated and detailed data to fully establish.

Acknowledgements I am grateful to my supervisors, Professor John Wildman and Professor Nils Braakmann, for their guidance and advice. I thank my colleagues Altay Mussurov, Natalia Yemelina, Aliya Abdikarimova and Olzhas Zhorayev, the participants of the Eurasia Business and Economics Society's (EBES) 25th Conference in Berlin, the University of Warwick Global Research Priorities on International Development Postgraduate Conference—2018 and Newcastle University Business School Seminar for PhD students in Economics, and two anonymous reviewers for very helpful comments and suggestions. Errors remain my own.

Compliance with ethical standards

Conflict of interest The author states that there are no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Distribution of primary sampling units by strata

Region/province	Number of households			Number of PSUs		
	Urban	Rural	Total	Urban	Rural	Total
Metropolis						
City of Astana	148,587	–	148,587	22	–	22
City of Almaty	386,251	–	386,251	30	–	30
Central						
Akmola	115,888	79,089	194,977	12	16	28
Karaganda	378,012	66,854	444,866	20	12	32
East-Kazakhstan	299,061	171,035	470,096	14	16	30
North						
Kostanai	179,666	127,047	306,713	12	15	27
Pavlodar	190,793	63,953	254,746	12	16	28
North-Kazakhstan	97,757	114,127	211,884	9	13	22
South						
Almaty	110,045	260,502	370,547	8	16	24
Zhambyl	123,593	117,878	241,471	9	14	23
South-Kazakhstan	232,170	260,099	492,269	10	16	26
Kyzyl-Orda	55,226	69,545	124,771	8	12	20
West						
Aktobe	133,540	32,803	166,343	12	16	28
Atyrau	56,823	31,931	88,754	10	8	18
West-Kazakhstan	100,630	76,727	177,357	8	14	22
Mangustau	73,270	16,828	90,098	12	8	20
Total	2,681,312	1,488,418	4,169,730	208	192	400

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan
 'Region' is a dummy variable combining several provinces geographically. East Kazakhstan province is combined with 'central' for simplicity

Appendix 2: Descriptive statistics, pooled individual data

Variable	Male subsample, N = 305,990	Female subsample, N = 282,110
Schooling		
Min.	0.00	0.00
Median	11.00	12.00
Mean	11.95	12.57
Max.	18.00	18.00
Age		
Min.	16.0	16.00
Median	40.0	39.00
Mean	39.6	38.98
Max.	62.0	57.00
Log nominal quarterly wage		
Min.	6.11	5.70
Median	11.70	11.42
Mean	11.52	11.30
Max.	14.96	14.41
Log real quarterly wage (adjusted by CPI, base year—2002)		
Min.	5.98	5.55
Median	11.13	10.90
Mean	11.05	10.81
Max.	14.22	13.80
Number of observations		
Region		
Metropolis	34,817	36,845
Central	70,566	68,526
North	52,539	50,817
South	84,764	69,909
West	63,304	56,013
Residence		
Urban	161,414	164,922
Rural	144,576	117,188
Sector of employment		
Public	85,930	155,401
Private	220,060	126,709

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 3: Cohort size in each year, male cohorts

Cohort	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Size
male1954	542	564	593	555	150	152	157	518	253	560	565	450	422	440	431	6352
male1955	581	664	603	591	161	166	158	579	301	606	594	497	501	421	380	6803
male1956	587	653	660	649	199	177	180	708	342	643	627	525	506	504	467	7427
male1957	656	715	786	769	206	223	199	712	343	719	705	624	570	557	515	8299
male1958	670	717	791	798	204	224	230	813	404	792	754	752	746	668	632	9195
male1959	751	793	772	741	243	236	231	787	400	756	796	763	723	742	631	9365
male1960	840	937	951	955	278	274	286	972	481	880	791	772	726	698	712	10553
male1961	894	998	1014	1044	243	251	248	1093	510	915	902	880	814	776	761	11343
male1962	839	899	948	930	220	212	216	980	493	874	976	914	921	876	824	11122
male1963	776	898	900	926	218	216	234	938	472	906	820	856	797	812	797	10566
male1964	639	742	771	812	249	262	256	969	494	842	861	898	783	845	779	10202
male1965	785	871	859	905	209	211	207	943	459	898	929	916	826	836	769	10623
male1966	650	741	794	777	193	191	199	850	432	748	785	809	827	761	774	9531
male1967	620	717	723	795	196	193	198	879	423	784	798	806	781	860	834	9607
male1968	718	718	714	786	189	180	204	752	372	739	837	775	780	806	856	9426
male1969	582	651	682	742	169	197	203	771	411	880	869	881	885	844	781	9548
male1970	644	689	677	715	190	184	167	864	452	857	825	787	814	871	869	9605
male1971	621	699	675	664	169	188	210	605	331	864	933	879	789	783	761	9171
male1972	627	693	699	737	192	199	182	771	383	820	807	814	838	860	746	9368
male1973	607	646	640	652	167	176	174	644	362	786	874	817	805	790	741	8881
male1974	588	618	668	675	182	164	167	680	336	741	839	837	759	727	752	8733
male1975	579	635	671	682	169	175	162	728	350	776	766	808	862	950	867	9180
male1976	560	674	652	675	169	182	174	693	382	852	840	861	873	839	845	9271

Cohort	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Size
male1977	585	607	594	602	191	198	187	770	405	914	893	890	928	862	821	9447
male1978	512	574	593	639	178	191	195	829	422	799	830	878	870	846	804	9160
male1979	546	582	625	647	163	143	164	680	369	940	887	866	799	820	824	9055
male1980	402	521	544	647	167	170	181	635	302	916	1002	927	878	786	785	8863
male1981	475	521	639	698	196	203	199	724	370	944	974	957	872	825	904	9501
male1982	394	545	635	683	164	147	153	727	350	900	796	822	859	962	842	8979
male1983	299	439	532	590	174	195	218	717	390	898	949	937	910	939	813	9000
male1984	241	341	471	590	216	225	233	797	410	1034	952	971	1002	910	876	9269
male1985	145	265	431	598	163	182	206	867	443	979	1039	1048	1066	1080	1053	9565
male1986	44	113	263	432	122	175	227	794	416	1112	1126	1155	1076	946	979	8980

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 4: Cohort size in each year, female cohorts

Cohort	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Size
fem1959	776	925	924	999	278	294	285	1140	599	1027	1034	1025	958	829	835	11928
fem1960	919	996	1099	1132	320	335	353	1173	611	1055	1033	1049	989	927	885	12876
fem1961	906	1035	1152	1153	340	354	346	1438	714	961	1009	1066	1059	1071	1037	13641
fem1962	884	1076	1154	1221	310	308	319	1435	694	1045	1037	1023	969	906	867	13248
fem1963	813	992	1083	1066	241	243	251	1147	618	1018	913	890	883	861	885	11904
fem1964	687	795	928	959	294	288	293	1083	548	1065	1062	1131	1003	945	936	12017
fem1965	765	935	1011	1000	234	240	264	1134	569	1009	1105	1108	1125	1132	1115	12746
fem1966	632	764	863	909	265	278	275	1144	568	1050	1100	1144	1181	1006	1002	12181
fem1967	612	760	805	915	241	253	253	1111	561	910	933	999	935	932	885	11105
fem1968	619	753	838	916	235	236	255	1136	565	1051	974	987	974	949	833	11321
fem1969	537	659	765	879	207	228	238	1031	502	813	859	955	1026	935	983	10617
fem1970	538	653	672	719	200	204	213	887	483	878	953	971	964	1008	984	10327
fem1971	578	721	737	699	185	190	221	917	458	999	958	933	925	949	975	10445
fem1972	539	634	688	693	165	191	203	928	470	889	977	884	914	986	1005	10166
fem1973	481	617	664	649	161	181	189	824	414	918	958	835	957	947	957	9752
fem1974	546	598	585	641	174	176	193	731	407	773	795	836	886	934	972	9247
fem1975	514	599	659	663	195	194	184	794	449	846	908	896	879	841	866	9487
fem1976	565	544	582	657	171	184	188	841	399	835	893	878	850	877	854	9318
fem1977	486	543	591	620	159	149	182	755	374	773	827	871	855	811	799	8795
fem1978	419	469	489	559	162	179	192	774	401	801	801	807	856	889	835	8633
fem1979	431	479	567	502	153	153	152	723	384	747	809	849	880	912	890	8631
fem1980	377	489	523	510	152	156	146	656	332	792	860	895	823	804	808	8323
fem1981	388	487	533	536	136	143	154	606	331	770	757	745	808	755	770	7919

Cohort	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Size
fem1982	272	385	476	505	161	148	168	652	344	700	796	774	799	796	765	7741
fem1983	211	370	484	527	137	163	155	628	370	796	719	726	808	835	871	7800
fem1984	102	217	297	428	144	159	166	618	334	845	859	849	892	843	855	7608
fem1985	102	178	241	378	103	153	185	716	314	802	857	854	763	778	732	7156
fem1986	15	70	209	308	98	114	148	690	389	946	835	814	946	825	771	7178

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 5: Descriptive statistics, pseudo-panel, male cohorts

Cohort	Log real wage mean	Log real wage SD	Schooling mean	Schooling SD	Age mean	Age SD	Central mean	North mean	South mean	West mean	Rural mean	Private mean
Male1954	10.92	0.38	11.88	0.15	55	4.47	0.26	0.19	0.26	0.18	0.45	0.69
Male1955	10.89	0.39	11.77	0.19	54	4.47	0.28	0.18	0.25	0.2	0.47	0.67
Male1956	10.95	0.38	11.85	0.28	53	4.47	0.28	0.18	0.25	0.18	0.44	0.72
Male1957	10.96	0.39	11.85	0.32	52	4.47	0.27	0.18	0.28	0.17	0.5	0.72
Male1958	10.95	0.41	11.84	0.14	51	4.47	0.28	0.18	0.26	0.17	0.49	0.7
Male1959	10.97	0.39	11.83	0.23	50	4.47	0.27	0.18	0.28	0.19	0.49	0.7
Male1960	10.97	0.38	11.9	0.18	49	4.47	0.28	0.18	0.26	0.18	0.5	0.7
Male1961	11	0.36	11.83	0.25	48	4.47	0.25	0.19	0.27	0.19	0.49	0.71
Male1962	10.98	0.35	11.86	0.14	47	4.47	0.28	0.17	0.29	0.18	0.5	0.69
Male1963	11	0.36	11.99	0.16	46	4.47	0.25	0.19	0.29	0.18	0.49	0.7
Male1964	10.98	0.43	11.88	0.11	45	4.47	0.24	0.21	0.29	0.18	0.51	0.7
Male1965	11	0.41	11.95	0.14	44	4.47	0.25	0.16	0.29	0.2	0.52	0.7
Male1966	11.02	0.45	11.85	0.17	43	4.47	0.24	0.17	0.31	0.18	0.51	0.7
Male1967	11.04	0.42	11.95	0.15	42	4.47	0.24	0.16	0.29	0.19	0.47	0.68
Male1968	11.07	0.45	11.92	0.11	41	4.47	0.25	0.17	0.27	0.2	0.46	0.7
Male1969	11.05	0.44	11.87	0.17	40	4.47	0.26	0.12	0.3	0.19	0.48	0.68
Male1970	11.09	0.4	11.97	0.16	39	4.47	0.24	0.17	0.28	0.18	0.45	0.7
Male1971	11.05	0.42	11.98	0.14	38	4.47	0.22	0.19	0.28	0.2	0.46	0.7
Male1972	11.05	0.39	11.87	0.18	37	4.47	0.23	0.19	0.29	0.18	0.45	0.69
Male1973	11.04	0.45	11.9	0.1	36	4.47	0.21	0.17	0.33	0.17	0.48	0.7
Male1974	11.05	0.45	12	0.22	35	4.47	0.24	0.17	0.3	0.19	0.47	0.7

Cohort	Log real wage mean	Log real wage SD	Schooling mean	Schooling SD	Age mean	Age SD	Central mean	North mean	South mean	West mean	Rural mean	Private mean
Male1975	11.04	0.49	12	0.16	34	4.47	0.22	0.17	0.3	0.18	0.45	0.69
Male1976	11.02	0.49	12.05	0.16	33	4.47	0.23	0.17	0.3	0.18	0.43	0.72
Male1977	11.01	0.49	11.97	0.16	32	4.47	0.2	0.17	0.31	0.21	0.42	0.7
Male1978	10.98	0.5	12	0.16	31	4.47	0.19	0.19	0.3	0.2	0.46	0.72
Male1979	10.97	0.55	12.14	0.21	30	4.47	0.24	0.15	0.27	0.22	0.44	0.72
Male1980	10.95	0.54	12.15	0.26	29	4.47	0.23	0.15	0.3	0.2	0.44	0.73
Male1981	10.95	0.58	12.14	0.35	28	4.47	0.23	0.16	0.29	0.2	0.45	0.74
Male1982	10.92	0.61	12.13	0.48	27	4.47	0.22	0.17	0.28	0.2	0.44	0.75
Male1983	10.85	0.64	12	0.59	26	4.47	0.23	0.16	0.28	0.22	0.47	0.75
Male1984	10.82	0.7	12.03	0.62	25	4.47	0.23	0.15	0.29	0.22	0.46	0.77
Male1985	10.8	0.72	11.89	0.82	24	4.47	0.21	0.13	0.33	0.21	0.46	0.8
Male1986	10.73	0.71	11.52	1.27	23	4.47	0.21	0.17	0.3	0.2	0.47	0.8

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 6: Descriptive statistics, pseudo-panel, female cohorts

Cohort	Log real wage mean	Log real wage SD	Schooling mean	Schooling SD	Age mean	Age SD	Central mean	North mean	South mean	West mean	Rural mean	Private mean
Fem1959	10.74	0.41	12.45	0.21	50	4.47	0.31	0.17	0.23	0.18	0.4	0.44
Fem1960	10.75	0.4	12.38	0.2	49	4.47	0.26	0.19	0.24	0.18	0.42	0.4
Fem1961	10.71	0.4	12.37	0.26	48	4.47	0.26	0.21	0.24	0.16	0.44	0.4
Fem1962	10.74	0.4	12.38	0.14	47	4.47	0.26	0.18	0.26	0.16	0.44	0.41
Fem1963	10.76	0.41	12.48	0.26	46	4.47	0.27	0.17	0.27	0.18	0.43	0.4
Fem1964	10.74	0.45	12.42	0.17	45	4.47	0.27	0.16	0.29	0.17	0.42	0.42
Fem1965	10.76	0.46	12.51	0.1	44	4.47	0.25	0.19	0.28	0.18	0.42	0.41
Fem1966	10.76	0.44	12.43	0.14	43	4.47	0.26	0.19	0.27	0.18	0.46	0.41
Fem1967	10.78	0.44	12.55	0.13	42	4.47	0.27	0.17	0.26	0.19	0.43	0.4
Fem1968	10.73	0.45	12.55	0.19	41	4.47	0.24	0.19	0.27	0.17	0.44	0.41
Fem1969	10.74	0.44	12.54	0.21	40	4.47	0.27	0.19	0.25	0.16	0.44	0.44
Fem1970	10.75	0.46	12.55	0.12	39	4.47	0.26	0.19	0.23	0.17	0.39	0.42
Fem1971	10.73	0.46	12.41	0.11	38	4.47	0.22	0.19	0.27	0.19	0.4	0.44
Fem1972	10.72	0.46	12.58	0.25	37	4.47	0.24	0.18	0.27	0.18	0.42	0.43
Fem1973	10.7	0.49	12.58	0.18	36	4.47	0.26	0.19	0.26	0.16	0.4	0.43
Fem1974	10.7	0.45	12.51	0.12	35	4.47	0.26	0.2	0.25	0.15	0.37	0.48
Fem1975	10.72	0.47	12.53	0.1	34	4.47	0.25	0.19	0.25	0.17	0.39	0.47
Fem1976	10.73	0.5	12.72	0.18	33	4.47	0.24	0.17	0.28	0.18	0.38	0.48
Fem1977	10.7	0.47	12.65	0.16	32	4.47	0.25	0.18	0.25	0.18	0.4	0.48
Fem1978	10.73	0.48	12.74	0.18	31	4.47	0.24	0.16	0.25	0.2	0.37	0.46
Fem1979	10.65	0.52	12.72	0.17	30	4.47	0.25	0.18	0.23	0.22	0.37	0.5

Cohort	Log real wage mean	Log real wage SD	Schooling mean	Schooling SD	Age mean	Age SD	Central mean	North mean	South mean	West mean	Rural mean	Private mean
Fem1980	10.71	0.51	12.91	0.21	29	4.47	0.23	0.19	0.21	0.23	0.37	0.53
Fem1981	10.68	0.5	12.87	0.38	28	4.47	0.23	0.17	0.25	0.22	0.4	0.49
Fem1982	10.65	0.54	12.86	0.56	27	4.47	0.22	0.16	0.27	0.21	0.36	0.52
Fem1983	10.63	0.53	12.79	0.72	26	4.47	0.25	0.17	0.24	0.22	0.38	0.54
Fem1984	10.59	0.61	12.6	0.85	25	4.47	0.25	0.15	0.27	0.19	0.38	0.55
Fem1985	10.59	0.62	12.56	1.07	24	4.47	0.24	0.15	0.25	0.21	0.39	0.59
Fem1986	10.56	0.66	12.23	1.36	23	4.47	0.21	0.14	0.28	0.23	0.4	0.61

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 7: Descriptive statistics, the first and the final year subsample

Variable	Male subsample, N = 44,224		Female subsample, N = 39,686	
	2002 N = 18,999	2016 N = 25,225	2002 N = 14,714	2016 N = 24,972
Schooling				
Mean	11.74	11.96	12.37	12.53
SD	1.65	1.88	1.73	2.03
Age				
Mean	34.3	44.59	32.92	43.97
SD	8.42	9.18	7.12	7.92
Log real wage				
Mean	10.15	11.5	9.92	11.25
SD	0.87	0.59	0.77	0.53
Number of observations				
Region				
Metropolis	1949	3122	1948	3395
Central	5062	5125	3971	5482
North	3275	4352	2781	4471
South	5966	6174	3950	5775
West	2747	6452	2064	5849
Residence				
Urban	11,509	12,267	10,087	13,261
Rural	7490	12,958	4627	11,711
Sector of employment				
Public	6938	5720	8362	13,140
Private	12,061	19,505	6352	11,832

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 8: Returns to schooling estimated by OLS with schooling*year interaction term, men

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.151*** (0.004)	0.135*** (0.004)	0.121*** (0.003)	0.122*** (0.003)
Age	0.061*** (0.001)	0.062*** (0.001)	0.063*** (0.001)	0.063*** (0.001)
Age squared	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)
Schooling*year 2003	− 0.003 (0.005)	− 0.006 (0.005)	− 0.005 (0.004)	− 0.005 (0.004)

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling*year 2004	– 0.011** (0.005)	– 0.010** (0.005)	– 0.011*** (0.004)	– 0.011*** (0.004)
Schooling*year 2005	– 0.022*** (0.005)	– 0.021*** (0.004)	– 0.023*** (0.004)	– 0.023*** (0.004)
Schooling*year 2006	– 0.036*** (0.006)	– 0.035*** (0.006)	– 0.037*** (0.006)	– 0.037*** (0.006)
Schooling*year 2007	– 0.052*** (0.006)	– 0.050*** (0.006)	– 0.052*** (0.006)	– 0.052*** (0.006)
Schooling*year 2008	– 0.060*** (0.006)	– 0.058*** (0.005)	– 0.060*** (0.005)	– 0.060*** (0.005)
Schooling*year 2009	– 0.043*** (0.004)	– 0.045*** (0.004)	– 0.047*** (0.004)	– 0.047*** (0.004)
Schooling*year 2010	– 0.040*** (0.005)	– 0.042*** (0.005)	– 0.045*** (0.004)	– 0.045*** (0.004)
Schooling*year 2011	– 0.062*** (0.004)	– 0.064*** (0.004)	– 0.066*** (0.004)	– 0.066*** (0.004)
Schooling*year 2012	– 0.054*** (0.004)	– 0.057*** (0.004)	– 0.059*** (0.004)	– 0.059*** (0.004)
Schooling*year 2013	– 0.058*** (0.004)	– 0.059*** (0.004)	– 0.063*** (0.004)	– 0.063*** (0.004)
Schooling*year 2014	– 0.056*** (0.004)	– 0.057*** (0.004)	– 0.062*** (0.004)	– 0.062*** (0.004)
Schooling*year 2015	– 0.051*** (0.004)	– 0.055*** (0.004)	– 0.060*** (0.004)	– 0.060*** (0.004)
Schooling*year 2016	– 0.043*** (0.004)	– 0.049*** (0.004)	– 0.054*** (0.004)	– 0.054*** (0.004)
Additional control variables	Year	Year region	Year region resi- dence	Year region resi- dence sector
F statistic	5518.2***	6444.9***	7235.3***	7088***
N	305,990	305,990	305,990	305,990
Adj. R2	0.386	0.439	0.488	0.488

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 9: Returns to schooling estimated by OLS with schooling*year interaction term, women

	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.144*** (0.004)	0.130*** (0.004)	0.125*** (0.004)	0.123*** (0.004)
Age	0.033*** (0.001)	0.036*** (0.001)	0.036*** (0.001)	0.036*** (0.001)
Age squared	- 0.0004*** (0.00001)	- 0.0004*** (0.00001)	- 0.0004*** (0.00001)	- 0.0004*** (0.00001)
Schooling*year 2003	0.0004 (0.005)	- 0.0001 (0.005)	- 0.001 (0.005)	- 0.001 (0.005)
Schooling*year 2004	0.005 (0.005)	0.006 (0.005)	0.004 (0.005)	0.004 (0.005)
Schooling*year 2005	- 0.009* (0.005)	- 0.008* (0.005)	- 0.011** (0.005)	- 0.011** (0.005)
Schooling*year 2006	- 0.015** (0.006)	- 0.014** (0.006)	- 0.016*** (0.006)	- 0.016*** (0.006)
Schooling*year 2007	- 0.011* (0.006)	- 0.008 (0.006)	- 0.012** (0.006)	- 0.011** (0.006)
Schooling*year 2008	- 0.029*** (0.006)	- 0.026*** (0.006)	- 0.029*** (0.006)	- 0.029*** (0.006)
Schooling*year 2009	- 0.009* (0.005)	- 0.009** (0.004)	- 0.013*** (0.004)	- 0.013*** (0.004)
Schooling*year 2010	- 0.021*** (0.005)	- 0.021*** (0.005)	- 0.025*** (0.005)	- 0.024*** (0.005)
Schooling*year 2011	- 0.040*** (0.004)	- 0.040*** (0.004)	- 0.043*** (0.004)	- 0.042*** (0.004)
Schooling*year 2012	- 0.042*** (0.004)	- 0.042*** (0.004)	- 0.044*** (0.004)	- 0.044*** (0.004)
Schooling*year 2013	- 0.034*** (0.004)	- 0.033*** (0.004)	- 0.036*** (0.004)	- 0.036*** (0.004)
Schooling*year 2014	- 0.036*** (0.004)	- 0.035*** (0.004)	- 0.038*** (0.004)	- 0.037*** (0.004)
Schooling*year 2015	- 0.033*** (0.004)	- 0.032*** (0.004)	- 0.035*** (0.004)	- 0.034*** (0.004)
Schooling*year 2016	- 0.025*** (0.004)	- 0.025*** (0.004)	- 0.027*** (0.004)	- 0.027*** (0.004)
Additional control variables	Year	Year region	Year region resi- dence	Year region resi- dence sector
F statistic	6181.2***	6466.7***	6671.1***	6498.9***
N	282,110	282,110	282,110	282,110
Adj. R2	0.419	0.454	0.467	0.468

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 10: Returns to schooling estimated from pooled individual data by OLS with schooling*gender interaction term

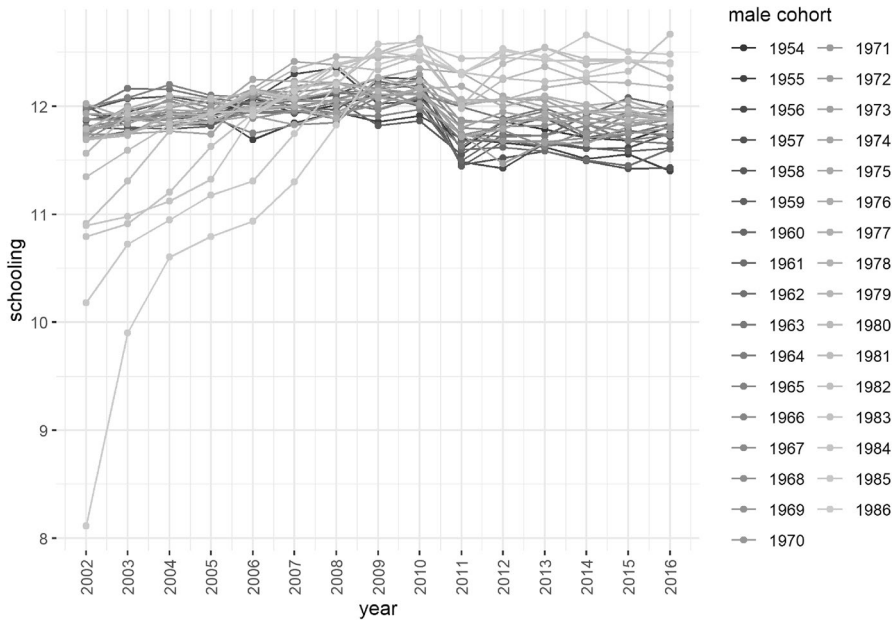
	Dependent variable			
	Log real wage			
	(1)	(2)	(3)	(4)
Schooling	0.119*** (0.001)	0.104*** (0.001)	0.093*** (0.001)	0.092*** (0.001)
Age	0.053*** (0.001)	0.055*** (0.001)	0.056*** (0.001)	0.055*** (0.001)
Age squared	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)	− 0.001*** (0.00001)
Male	0.452*** (0.011)	0.478*** (0.010)	0.509*** (0.010)	0.516*** (0.010)
Schooling*male	− 0.009*** (0.001)	− 0.011*** (0.001)	− 0.013*** (0.001)	− 0.013*** (0.001)
Constant	7.401*** (0.016)	7.493*** (0.015)	7.390*** (0.015)	7.398*** (0.015)
Additional control variables	Year	Year region	Year region residence	Year region residence sector
F statistic	19634***	19985***	21188***	20404***
N	588,100	588,100	588,100	588,100
Adjusted R2	0.412	0.454	0.482	0.482

Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

(1) OLS computed in R. HC1 robust standard errors computed with ‘sandwich’ package (Zeileis 2004) in parentheses

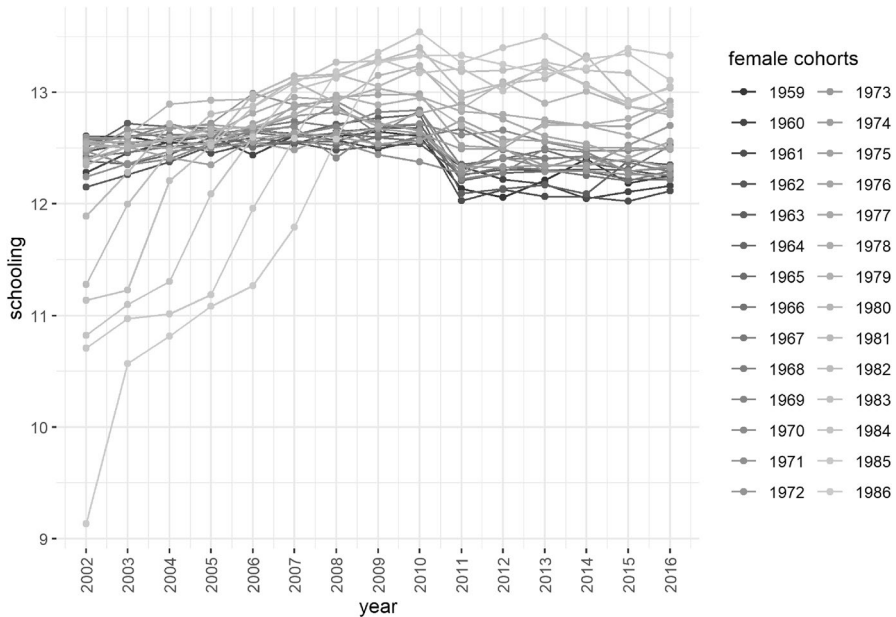
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix 11: Mean of schooling across cohorts and years, male cohorts



Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

Appendix 12: Mean of schooling across cohorts and years, female cohorts



Source: Household Budget Survey, the Committee on Statistics of the Republic of Kazakhstan

References

- Arabsheibani, G. R., & Mussurov, A. (2007). Returns to schooling in Kazakhstan: OLS and instrumental variables approach. *The Economics of Transition*, 15(2), 341–364.
- Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, 2(3), 193–216.
- Baltagi, B. H., & Kao, C. (2000). *Nonstationary panels, cointegration in panels and dynamic panels: A survey*, vol. 136. Center for Policy Research. <https://surface.syr.edu/cpr/136>. Accessed 8 Mar 2020.
- Barro, R., Caselli, F., & Lee, J. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104, 184–198.
- Betts, J. R., & McFarland, L. L. (1995). Safe port in a storm: The impact of labor market conditions on community college enrollments. *Journal of Human Resources*, 30(4), 741–765.
- Bhattacharya, P., & Sato, T. (2017). Estimating regional returns to education in India: A fresh look with pseudo-panel data. *Progress in Development Studies*, 17(4), 282–290.
- Bielby, W., & Baron, J. (1986). Sex segregation within occupations. *The American Economic Review*, 76(2), 43–47.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.
- Brunello, G., & Comi, S. (2004). Education and earnings growth: evidence from 11 European countries. *Economics of Education Review*, 23(1), 75–83.
- Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, vol. 3 (pp 1801–1863). Elsevier.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127–1160.
- Chamberlain, G., 1984. Panel data. In *Handbook of Econometrics*, vol. 2 (pp. 247–1318). Elsevier.
- Clark, D. (2011). Do recessions keep students in school? The impact of youth unemployment on enrolment in post-compulsory education in England. *Economica*, 78(311), 523–545.
- CS MNE. (2015). Metodika postroeniya vyborki domashnih hozyaystv po obsledovaniyu urovnya zhizni. <http://stat.gov.kz>. Accessed 5 July 2019.
- CS MNE. (2017). O situacii na rynke truda v 4 kvartale 2016 goda. <http://stat.gov.kz>. Accessed 5 July 2019.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1), 109–126.
- Dellas, H., & Sakellaris, P. (2003). On the cyclicity of schooling: theory and evidence. *Oxford Economic Papers*, 55(1), 148–172.
- Dickerson, A., Green, F., Arbache, J.S., 2001. Trade liberalization and the returns to education: A pseudo-panel approach. In *Department of Economics discussion paper*, No. 1, 14. Canterbury: Department of Economics, University of Kent.
- Fleisher, B. M., Sabirianova, K., & Wang, X. (2005). Returns to skills and the speed of reforms: Evidence from Central and Eastern Europe, China, and Russia. *Journal of Comparative Economics*, 33(2), 351–370.
- García-Aracil, A. (2007). Gender earnings gap among young European higher education graduates. *Higher Education*, 53(4), 431–455.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1), 1–22.
- Harmon, C., & Walker, I. (2001). *The returns to education: A review of evidence, issues and deficiencies in the literature*. Research Report No. 254. DfEE. <https://dera.ioe.ac.uk/4663/1/RR254.pdf>. Accessed 8 Mar 2020.
- Heckman, J. J., Lochner, L., & Todd, P. E. (2003). *Fifty years of mincer earnings regressions*. NBER Working Paper No. 9732. <https://www.nber.org/papers/w9732.pdf>. Accessed 8 Mar 2020.
- Himaz, R., & Aturupane, H. (2016). Returns to education in Sri Lanka: a pseudo-panel approach. *Educational Economics*, 24(3), 300–311.
- Hlavac, M., 2018. *oaxaca: Blinder-Oaxaca decomposition in R*. R package version 0.1.4. Bratislava: Central European Labour Studies Institute (CELSI). <https://CRAN.R-project.org/package=oaxaca>. Accessed 20 Nov 2019.
- Hungerford, T., & Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 69(1), 175–177.

- Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal: Promoting Communications on Statistics and Stata*, 8(4), 453–479.
- Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2), 303–316.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, 90(1), 1–44.
- Kapelyushnikov, R. (2008). *Zapiska ob otechestvennom chelovecheskom kapitale*. Moscow: National Research University Higher School of Economics. https://www.hse.ru/data/2010/05/04/1216406890/WP3_2008_01.pdf. Accessed 08 Mar 2020.
- Lassibille, G., & Gómez, L. N. (1998). The evolution of returns to education in Spain 1980–1991. *Education Economics*, 6(1), 3–9.
- Lesnoff, M., & Lancelot, R. (2012). aod: Analysis of overdispersed data. R package version 1.3.1. <https://cran.r-project.org/package=aod>. Accessed 14 Mar 2019.
- Liu, K., Salvanes, K. G., & Sørensen, E. Ø. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84(C), 3–17.
- Meng, X. (2004). Gender earnings gap: The role of firm specific effects. *Labour Economics*, 11(5), 555–573.
- Merryman, S. (2005). *XTREGRE2: Stata module to estimate random effects model with weights*. Boston: Boston College Department of Economics.
- Mincer, J. (1974). *Schooling, experience, and earnings*. Cambridge: National Bureau of Economic Research.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1), 69–85.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693–709.
- Oreopoulos, P., von Wachter, T., & Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1), 1–29.
- Phillips, P. C. B., & Moon, H. R. (2000). Nonstationary panel data analysis: An overview of some recent developments. *Econometric Reviews*, 19(3), 263–286.
- Psacharopoulos, G., & Patrinos, H. A. P. (2004). Returns to investment in education: A further update. *Education Economics*, 12(2), 11–134.
- Smith, R. P., & Fuertes, A.-M. (2010). *Panel time-series*. Princeton: Citeseer.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Stiglitz, J. E. (1975). The theory of ‘screening’, education, and the distribution of income. *The American Economic Review*, 65(3), 283–300.
- Verbeek, M., & Nijman, T. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics*, 17(1), 9–23.
- Warunsiri, S., & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development*, 38(11), 1616–1625.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.
- Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10), 1–17.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.