The Journal of Comparative Economic Studies, Vol.14, 2019, pp. 19–51.

The Contradictory Nature of Lifelong Learning in the Post-transition Russia[†]

Vasiliy A. ANIKIN *

* *National Research University Higher School of Economics, Russian Federation; vanikin@hse.ru*

Abstract: The literature tends to neglect the role of individuals in formal skills training in Russia during the period of economic growth between 2001 and 2014. The present study addresses this oversight. Although to a certain extent, studies have associated the prosperous years of recent economic growth in Russia with training, they have not considered that the Russian population insisted on better qualifications. The present study shows that such insistence came primarily from skilled non-manual workers who resided in cities, worked more than eight hours per day, had second jobs, and were in great demand by organizations. Drawing on the panel data of the Russian Longitudinal Monitoring Survey—Higher School of Economics, we argue that individual heterogeneity over the period of economic growth between 2001 and 2014 in Russia significantly contributed to the unobserved variation in training and cannot be ignored in applied socioeconomic studies of human capital. After accounting for important within- and between-person characteristics, we find that 26% of the variation in training during the studied years is attributable to the unobserved characteristics of individuals and 7% to jobs, whereas only 0.2% is accounted for by the time trend. We use multilevel longitudinal probit models with cross-classifications to partition the variation in skills training into individual and job-specific levels in the context of Russia's recent economic growth. Our results show that in some knowledge-based societies, lifetime learning may slightly be a function of the years of economic prosperity and more likely based on the unobserved individual traits.

Keywords: skills training, human capital, post-Soviet Russia, heterogeneity

JEL Classification Numbers: J24 M53 P36 C33 C51

1. Introduction

Observers agree that the period between late autumn 2014 and winter 2016 was a time of crisis for the Russian economy. It was caused by large-scale Western sanctions against Russia in relation to the Ukraine Crisis in July 2014 and by a fall in oil prices. These external events rendered the Russian stock market and rouble more volatile and consequently less predictable (Obizhaeva, 2016; Schmidbauer, Rösch, Uluceviz, & Erkol, 2016). This recent intense difficulty for the Russian economy poses questions about the role of the internal drivers of economic prosperity during 2001–2014, with a particular focus on the contribution of human capital to such prosperity and the importance of skills development for Russian employees.

The recent literature on knowledge-based societies reconsiders the role of education and examines the inflated optimism about the expansion of education and its social value (Alvesson & Benner, 2016). In this regard, one study in particular expects 'to see higher levels of workplace skill formation to generate both the work skills that cannot be learned during school and college education' (Green, Felstead, Gallie, Inanc, & Jewson, 2016, p. 424). The literature on knowledge-based societies suggests that training and skills acquisition increases during economic growth if such growth is based on human capital and thereby marks a transition towards a knowledge economy.

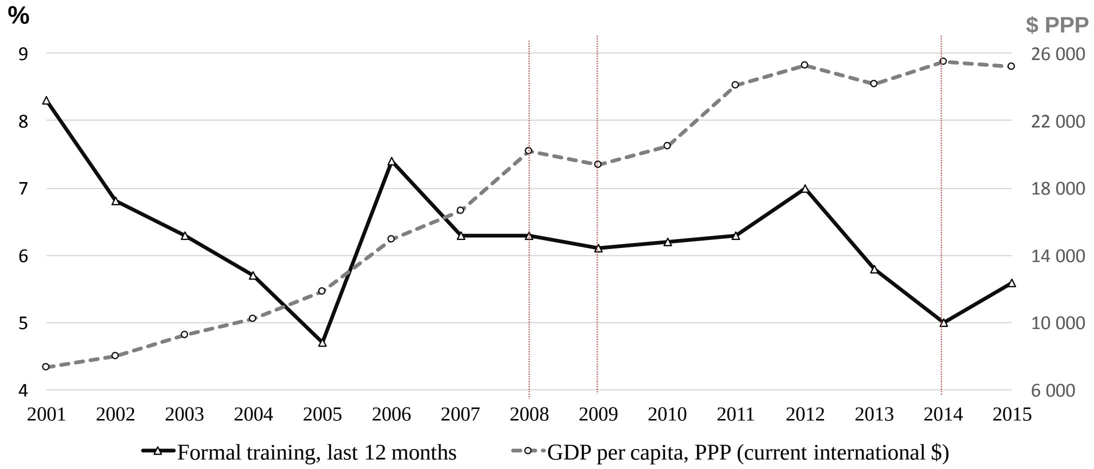
The period 2001–2014 was a time of expansion for higher education in Russia (Gimpelson & Kapeliushnikov, 2016; Kyui, 2016). According to the Organisation for Economic Co-operation and Development (OECD) (2016), Russia has the second highest share of adults attaining tertiary education of all OECD members: the relative share of employees with tertiary education increased notably from approximately 20% in 2000 to more than 30% in 2013 (Gimpelson & Kapeliushnikov, 2016). The growing number of university (tertiary-type A) graduates, more than half of which were enrolled part-time, mainly supported this expansion.

This growth of higher education coincides with the specificity of Russia as a knowledge society; however, Russia has yet to become a knowledge economy. The incidence of training in Russia remained low and stable, and even slightly declined, during 2001–2014. The recent literature on training explains such 'training poverty' in developed nations in terms of 'a falling demand for skills formation', which is 'inherent in a "low skills" trajectory for large swathes of the … economy' (Green et al., 2016, p. 441). In light of this, we assume that the main reason for the low level of training in Russia is a reluctance in parts of the labour market to upgrade skills and innovate. As a result, people employed in certain industries and occupations are unwilling to invest time and effort in acquiring new knowledge and skills. At the same time, employees from more efficient and advanced niches of the labour market tend to develop their skills on a regular basis through their personal experience with training. Thus, the present research focuses on the individual patterns of skills acquisition (Tharenou, 1997, 2001) in Russia during the prosperous years of economic growth between 2001 and 2014.

2. Economic growth and skills acquisition in contemporary Russia

The year 2001 was a tipping point for the Russian economy. By then, the economy had passed through the murky waters of perestroika, which were marked by social instability and economic fluctuations (Gerber & Hout, 1998). In the early 2000s, the economic situation became broadly favourable to the absorption of new capital and expectations; in other words, the new era in the socioeconomic and political life of Russia had begun (Voigt, 2006). By this time, most of the remaining Soviet industrial infrastructure had been reused within the institutions of the market economy. Such institutions were formed as a result of the completed transition to a market economy and the 'great reallocation of human capital' (Sabirianova, 2002). However, this reallocation has not led to widespread skills acquisition in Russia. As Figure 1 shows, the average annual share of workers who received training from 2001 to 2015 was approximately 6% of the working population, a much lower share than in the rest of Europe (Arulampalam, Booth, & Bryan, 2004; Bosch & Charest, 2012; O'Connell, 1999).

Figure 1. Formal Training (solid line) and Gross Domestic Product (GDP) per capita (dashed line) in Russia



Source: Training data retrieved from the Russian Longitudinal Monitoring Survey—Higher School of Economics (RLMS—HSE) data, with representative samples weighted by post-stratification weight. GDP per capita data retrieved from the World Bank.
Notes: % is the percentage of the working population. GDP is gross domestic product. PPP is purchasing power parity.

One of the main reasons for this lack of skills acquisition is that the growth of the Russian economy has been led by natural resources. Gaddy and Ickes (2013) document the abundance of tradable natural resources during Russia's recent economic growth, suggesting that the country's extensive input-driven growth trajectory is a legacy of the late Soviet era. The recent slowdown in growth strongly supports the view that the rapid economic expansion before the crisis was due to external drivers related to the positive situation in foreign markets, namely rising oil and gas prices, rather than the productivity of human capital (Connolly, 2012). Hence, Hypothesis 1 proposes that the *prosperous years of economic growth had no effect on the probability of training.*

Timmer and Voskoboynikov (2014) calculate that the annual average productivity growth of Russia has been only approximately 2.25% since the mid-1990s. In many industries, employees' salaries have grown faster than their productivity, a factor which, among others, may have contributed to the lack of incentive to develop human capital and upgrade the quality of skills. This situation has been widespread; moreover, it has been significantly enhanced by the institutional arrangements of the Russian labour market which has allowed low-productivity firms to survive as well as the hiring of workers with relatively low levels of human capital (Gimpelson & Kapeliushnikov, 2013), thereby producing significant skill mismatches (Demmou & Wörgötter, 2015).

Recent findings suggest that Russian employees were systematically overpaid during the period of economic growth. Because salaries were detached from workers' competencies, we suggest that employees and employers had less of a need to direct their efforts and time to lifelong education and skills. Ultimately, when 'the gap between productivity and wage is independent of the skill level of the worker, the firm has no interest in increasing the worker's skills' (Acemoglu & Pischke, 1999, p. F120). According to Gimpelson and Kapeliushnikov (2013), there is little incentive to upgrade human capital in low-productivity firms and, consequently, among poorly skilled labour. Thus, Hypothesis 2 proposes that *employees from generic (i.e. routine, exchangeable, and disposable) labour are less willing to receive training*.

The findings of Timmer and Voskoboynikov (2014) suggest that the only industries in which human capital can be a growth factor are finance and business services; however, even these industries do not necessarily encourage training because much of their productive performance 'is of some basic catching-up character' (Timmer & Voskoboynikov, 2014, p. S418). With regard to such ambiguity, the industry-specific level is less informative than the occupational structure.

Occupations and occupation-based groupings are of great importance (see Castells, 2010; Goldthorpe, Llewellyn, & Payne, 1987; Wright, 1997), particularly in an advanced industrial society (Grusky, 2001) or an information society (Castells, 2010). In such societies, skilled occupations form the major strata. Thus, skilled employees may seek training to retain their positions in the occupational and status hierarchy (Goldthorpe & McKnight, 2006). Moreover, skilled managers are more important than institutions when a country is absorbing a set of new technologies (Acemoglu, Aghion, & Zilibotti, 2006). In OECD countries, the incidence of training is statistically higher among managers and professionals than other occupations (O'Connell, 1999). Berger, Earle, and Sabirianova (2001) show that during the economic restructuring in Russia (1994–1998), managers and skilled professionals were more likely to undertake training than other occupations. Following this strand of the literature, we control for administrative power at work (i.e. whether employees have any subordinates). Further, in line with the functionalist perspective, hardworking employees should also reveal a positive inclination for training as a criterion for promotion (Ding, Fields, & Akhtar, 1997) or retaining employment in the public sector (Arulampalam et al., 2004; Méndez & Sepúlveda, 2016).

There are also contradictory findings. These relate to demographic differences, with a particular focus on gender. For example, empirical evidence suggests that in advanced industrial societies such as the United Kingdom, the disparities in the incidence of training between men and women are converging (Green & Zanchi, 1997), whereas others argue that these differences are still important and even non-linear (Cho, Kalomba, Mobarak, & Orozco, 2013; Fitzenberger & Muehler, 2015; Polavieja, 2012). Although recent findings on Russia support the former strand of the literature (Anikin, 2017; Berger et al., 2001), it remains important to control for gender because of a gender division of labour among occupations and, in particular, the predominance of women in semi- and low-skilled occupations despite their high levels of education (Anikin, 2012; Klimova, 2012).

Another significant demographic parameter usually considered in models of training is age. With panel studies, the age variable has become an important individual-specific indicator because it enables such studies to capture the ageing effect on workers' investments in human capital. Further, belonging to a specific age group may affect workers' participation in training. For instance, younger employees are more likely to receive training than senior employees. The explicit modelling of this so-called 'cohort effect' enables researchers to capture important differences between individuals of various age groups compared with other socioeconomic indicators.

Following classical studies of human capital (Lemieux, 2006; Mincer, 1962, 1974), researchers suggest tenure (specific work experience) as a relevant determinant of training. The importance of this indicator is even higher in a longitudinal study, enabling researchers to capture whether the experience of growth leads employees to engage in training. In line with prior estimates, we expect that the impact of tenure on training is either negligible or negative (Bartel & Sicherman, 1998; Berger et al., 2001; Loewenstein & Spletzer, 1999). Further, a panel study makes it possible to account for the person-specific cohort effects of tenure. The human resource management routines applied in many organizations are likely to categorize workers into 'tenure groups' and use these categorizations to decide on whether they need formal training.

From this perspective, newcomers, particularly those who have recently changed their occupations, are more likely to receive training. Such a perspective was worthwhile when studying the earlier transition period of the Russian economy (Berger et al., 2001), during which inter-occupational flows were much more intensive (Sabirianova, 2002). However, from 2001 to 2014, occupation–job flows may also have been a significant factor of retraining. Thus, because we do not split training into different components (additional training and retraining), it seems crucial to consider occupation–job mobility and the experience of unemployment.

We disregard the distinction between additional training and retraining since Russia had overcome the turbulent times of restructuring by 2000 (Berger et al., 2001; Gimpelson & Kapeliushnikov, 2016; Gimpelson & Lippoldt, 2001). After the 2000s, the Russian economy had already moved beyond the major phase of transition and had stabilized its institutions and structures. In the earlier stages of transition, by contrast, workers could 'leapfrog' from low- to high-level occupations by acquiring new knowledge and skills on training courses (e.g. some accountants became chief executive officers); however, even then this was not a widespread phenomenon (Gerber & Hout, 2004). In Russia nowadays, these 'jumps' are hardly possible. Since the 2000s, researchers have noted downward intergenerational mobility, while 'social origins' have become a 'more salient factor in sorting workers into privileged and impoverished positions' (Gerber & Hout, 2004).

Based on the literature and empirical results, we can arrange the observed factors of training as shown in Table 1. We should clarify that the list of factors we use to model the probability of training is not the only one possible. For instance, our multivariate analysis excludes several observed factors which have recently been considered to be important determinants of training incidence such as occupation-specific wage differentials, types of employment relationships, employment in the quaternary sector, education and other explicit human capital characteristics which measure cognitive skills such as computer and language skills, organization-specific indicators such as ownership and organizational size, and self-rated health and other self-assessed parameters such as work satisfaction and the opportunities for professional growth associated with a job (Anikin, 2017).

Table 1. Observed and Unobserved Factors of Training, Indicators, and Expectations

Factors	Indicators	Anticipated impact
Single-level part	*(Observed)*	
Socio-demographics	Males	Positive
	Age (years)	Negative
	Residence in cities	Positive
Occupation- and job-specific level	Tenure (years)	Negative
	Qualified non-manual labour	Positive
	Working time, average hours per day	Positive
	Working time, more than eight hours per day	Positive
	Have subordinates	Positive
	Occupation–job mobility	Positive
	Unemployment in a previous year	Negative
	Have second job	Positive
Multilevel part	*(Unobserved)*	
Individuals	Individual-specific variation across persons	Sign.
Occupations	Variation between occupations	Sign.
Time	Variation between years	Insign.

Notes: The abbreviations 'Sign.'/'Insign.' mean that we anticipate a particular component will have a statistically significant (or insignificant) contribution to the variation in training. See Appendix A for an extended description of the variables.

Some of these indicators are included in the measurement of other indicators. For example, the horizontal variability between minor occupations is attributable to sectoral differences, whereas the hierarchical variation across occupations reflects educational and skill-specific differentials. Other variables are omitted because they either produce unacceptable amounts of missing data in the longitudinal panels or contain significant measurement errors (see Appendix A). For instance, self-rated health is sensitive to panel attrition and ageing, an issue covered in more detail in the 'Data' section. Further examples are 'objective' variables such as organizational size and ownership measured on the basis of survey respondents' answers. In this regard, respondents are not always sure about the relevant numbers and actual information related to their employers and real owners (a situation which may occur to employees working in the military-industrial field); thus, such respondents may provide false data unintentionally.

3. Data

This study uses the panel samples of the Russian Longitudinal Monitoring Survey—Higher School of Economics (RLMS—HSE). The survey started in 1995 by the Carolina Population Centre at the University of North Carolina in Chapel Hill, USA and the Institute of Sociology of the Russian Academy of Sciences. The data are primarily maintained and distributed by the National Research University Higher School of Economics in Moscow. The RLMS—HSE data are rigorously described by Kozyreva, Kosolapov, and Popkin (2016).

Since the major focus of the present study is on skills training over the recent period of economic growth, we limit our panel to the 14 rounds between 2001 (round 10) and 2014 (round 23). This selection of data has no missing time points because surveys were conducted once a year in the autumn months. However, we selected respondents who are currently working. Thus, we used 99,101 observations of the 23,870 respondents in the longitudinal data sequence (see Appendix B). This compound data set consists of an unbalanced panel with embedded round gaps. Although it is common practice to eliminate observations within a panel which exhibit gaps in their data sequences (Baum, 2006), this approach may involve a loss of efficiency in coefficient estimation (Biørn, 2016). Thus, we do not eliminate gaps to minimize these losses.

The RLMS—HSE data enable us to model skills training using the following question: 'During the last 12 months, did you take part in professional courses, advanced training, or other any courses, including foreign language classes?' This question is a binary variable with a value of 1 for 'yes' and 0 for 'no' (see Appendix C for this and other variables). Appendix D (Table D2) shows that this variable contains a significant disproportion of answers. Most answers are 'no'; only a few respondents say that they received training during the year. As aforementioned, we apply this question only to the working population.

The RLMS—HSE data do not enable a distinction to be drawn between on-the-job training, periods of training-related unemployment, and affirmative action training. The presented data contain limited direct information about prior periods of unemployment (e.g. 'Did not work in November last year':

see Appendix C); however, we can check the current primary activity of respondents. Based on the chi-square test, we study the relationship between unemployment and training and do not find any statistically significant correlation. Among unemployed Russians (see Appendix D, Table D1), the percentage of those who received training in the prior 12 months does not statistically differ from the percentage of those who did not. Moreover, this finding applies to all the years of monitoring considered in the present study. Appendix D (Table D2) summarizes the data on training in terms of the working and non-working populations.

The International Standard Classification of Occupations (ISCO-88) is used to encode the occupations of working respondents and, consequently, the occupational structure. According to the International Labour Organization, ISCO-88 has four levels of aggregation. The modified version of ISCO-88 contains 430 minor four-digit occupational levels.[1] In line with Castelles (2010), ISCO-88 is used to produce two occupational classes: 'qualified non-manual workers' and 'generic labour'. Aggregated forms of occupations are supposed to represent the basic structural elements of the situation in the labour market. The highest percentage rates of those who received training are among respondents with either skilled or semi-skilled jobs within the non-manual labour and managerial groups. The year-to-year transition matrices show that the average percentage of trained workers is significantly higher among managers, professionals, and semi-professionals, who are combined under the category 'qualified non-manual workers', than 'generic labour' such as office clerks, salespeople, farmers, craft workers, operators and assemblers, and elementary occupations.

As aforementioned, the period 2001–2014 fully covers the recent economic growth of Russia before the crisis which occurred in autumn 2014. Figure 1 shows that the growth trajectory of Russia was not sustainable during these years because the global economic crisis affected the Russian economy in 2009. However, economic growth had fully recovered by 2010. To trace these and other time-specific peculiarities, we utilize a year-specific variable: its 14 values are intended to capture the variations between 2001 and 2014 at the year level which represent the so-called 'period effect'.

Unfortunately, we cannot avoid sample attrition because we are using a panel study of micro-level changes. Here, the data collectors do not usually follow individual household members who have moved away from the original dwelling unit. Although some households and individuals who have moved are followed (as in round VII) to complete the interview, this is rare because true panels are costly to maintain. Thus, attrition is a common issue for most panel studies. With regard to the RLMS—HSE data, 'the influence … of household turnover does not seriously distort the geographic distribution of the sample or its size or household-head characteristics' (Heeringa, 1997). This situation applies even though the influence of sample attrition of the RLMS—HSE rounds for a panel of individual respondents on the percentage of individuals from the Moscow and St. Petersburg regions and more general urban domains is the greatest (Heeringa, 1997). Further, Heeringa (1997) warns researchers that attrition may cause a general ageing effect regarding the panel of individuals and lead to losses of panel respondents from the higher-income groups.

More recent and rigorous studies of sample attrition using RLMS—HSE data (Denisova, 2010; Gerry & Papadopoulos, 2015) confirm that attrition is systematically related to demographic, health, and other socioeconomic characteristics; however, Gerry and Papadopoulos (2015) admit that a carefully specified model can minimize attrition bias. Thus, the authors' preliminary findings support the 'state dependence' hypothesis (Maddala, 1987) and confirm the importance of unobserved individual heterogeneity in longitudinal studies based on RLMS—HSE data. In the present study, we do not test the 'state dependence' hypothesis (Heckman, 1981) because a lagged variable for training leads to a significant amount of missingness; however, we accept the advice to apply models which assess unobserved individual heterogeneity.

4. Methodology

A panel analysis of micro-data objectifies one of the basic issues of econometric analysis, that of the fundamental heterogeneity problem. In longitudinal regression analysis, the parameters of interest may vary, either across individuals or over time, or both. In the canonical econometric literature, variation across individuals is expressed in terms of individual heterogeneity (Hausman & Taylor, 1981) because it may be related to a set of individual-specific features which stay unchanged over time but vary across individuals. The most salient example of such features may be gender or temperament. In the literature on training, unobserved individual heterogeneity, u_i, may relate to unobserved individual abilities (or talents), which are expected be rigid over time. So-called random effects (RE) models are required to capture this heterogeneity.

4.1 Individual heterogeneity

When individual heterogeneity is present, RE models have greater efficiency than counterfactual methods, which are models with fixed, or constant, effects (FE); however, there is a penalty for such efficiency. In general, a model is inconsistent if the revealed heterogeneity is correlated with the covariates. With regard to RE models, which may be considered to be simple examples of multilevel models (MLMs), this issue occurs because of the omission of FE in classical RE models. According to Antonakis, Bendahan, Jacquart, and Lalive (2010), MLM followers are likely to fit the models using random coefficients without checking whether the level 1 variables correlate with FE due to the higher-level entity. Antonakis et al. (2010) warn that one cannot use RE if level 1 variables, x_{it}, are correlated with FE: the FE are an omitted cause. The problem is that the widely used Hausman specification test is scarcely helpful here, as it assumes that the RE estimator is fully efficient; however, this is usually not the case. The latter circumstance may explain why RE models are rejected more often than FE models. For example, the literature on training in Russia often suggests rejecting particular specifications of RE models and accepting less efficient FE models under the assumption that inconsistency exists in RE models per se. In other words, human capital acquisition in Russia is likely to be modelled as a homogeneous phenomenon (Berger et al., 2001; Lazareva, 2006; Sabirianova, 2002). However, if this homogeneity is assumed incorrectly, the FE estimates will be biased.

In this regard, Cameron, Gelbach, and Miller (2008) suggest conducting a panel bootstrap of the Hausman specification test or using the Wooldridge robust version of it. Nonetheless, even these techniques do not solve the issue of endogeneity in RE models: such an issue is not a routine matter that simply needs technical correction. Instead, the recent literature recommends reassessing the RE model critically in terms of omitted variables and the misspecification of disregarded heterogeneity.

'Pure' FE models are somewhat problematic because they only estimate so-called within effects, thereby producing heterogeneity bias. To avoid this problem, all the higher-level entities are included in the model as dummy variables (Allison, 2009). Then, the mean of the higher-level entities is deducted from both sides of the regression equation. This method should free a researcher from estimating a parameter for each higher-level unit.

The growing critique of 'pure' FE models is usually built around Mundlak's (1978) approach, which offers an RE solution for heterogeneity bias by 'attempting to model two processes in one term' (Bell & Jones, 2015, p. 141). With regard to a panel data example, where individuals, *i*, are considered at level 2 as they observed on multiple occasions (level 1), *t*, the Mundlak formulation explicitly models FE by adding one extra term on the right-hand side of the regression equation, thereby averaging each time-variant covariate across time points (see equation (1)):

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_3 x.mean_i + \beta_4 z_i + (u_i + e_{it}). \tag{1}$$

Equation (1) represents a model with a continuous dependent variable, y_{it}, specified in two-level terms; notations are utilized for a panel data structure. With regard to models with binary outcomes, the form of the equation will be the same but with no level 1 error term, e_{it}. Term x_{it} denotes a vector of the time-variant level 1 covariates, whereas $x.mean_i$ denotes a vector of the individual-specific means of x_{it}. β_3 represents the contextual effect because it explicitly models the difference between the within and between effects (Bell, Jones, & Fairbrother, 2017). β_4 accounts for the effects of the time-invariant variables (z_i) which exist at a higher level (*i*). The term u_i represents homogeneous RE at level 2, namely the higher-level (individual-specific) residuals assumed to be normally distributed (see assumption (4)).

Equation (1) is efficient and is recommended for use in repeated cross-sectional studies (Bell & Jones, 2015). With panel data, the contextual effect is not that informative because level 1 units represent occasions, namely individual–time observations; thus, the within and between effects are of more interest. For this reason, Bell and Jones (2015) rearrange equation (1) to obtain an RE within–between (REWB) model:

$$y_{it} = \beta_0 + \beta_1(x_{it} - x.mean_i) + \beta_2 x.mean_i + \beta_4 z_i + (u_i + e_{it}). \tag{2}$$

Equation (2) represents a simplified version of REWB because it still assumes homogeneous RE, u_i, meaning that randomization is applied only to the intercept and that the slopes are kept fixed. In contrast to equation (1), in equation (2), β_1 represents the estimate of the within effect of a time-variant variable, x_{it}, and β_2 stands for the between effect of the same variable. Equation (2) is a general form of a two-level REWB model which considers individuals to be 'clusters'.

Following conventional recommendations (Maddala, 1987), we use a probit link function to specify an REWB model with a binary response because probit models do not have a conditional likelihood. A general form of a probit specification for an REWB model with homogeneous RE is as follows (equation (3) and assumption (4)):

$$y_{it} \sim \text{Binomial}(\text{const}_{it}, \pi_{it})$$

Micro-level:

$$\text{Probit}(\pi_{it}) = \boldsymbol{\beta_{0i}} + \beta_1(x_{it} - x.mean_i) + \beta_2 x.mean_i + \beta_4 z_i$$

Macro-level:

$$\boldsymbol{\beta_{0i}} = \beta_0 + u_i$$

Full-length model:

$$\text{Probit}(\pi_{it}) = \beta_0 + \beta_1(x_{it} - x.mean_i) + \beta_2 x.mean_i + \beta_4 z_i + (u_i) \tag{3}$$

$$\text{var}(y_{it} \mid \pi_{it}) = \pi_{it}(1 - \pi_{it}) / \text{const}_{it}$$

$$[u_i] \sim \text{N}(0, \Omega_u) : \Omega_u = [\sigma^2_{u_i}] \tag{4}$$

The value of the individual-specific variation of a person's unobserved characteristics, $\sigma^2_{u_i}$, is our particular interest.

4.2 Structural heterogeneity

Applied economists assume that individual heterogeneity, which is expected to account for people's unmeasured abilities, absorbs the major proportion of the entire heterogeneity of an outcome. However, sociology suggests that structures may also play a significant role in predicting the agency. Social, cultural, and other contextual differences (e.g. geographical factors) absorb a notable share of the variation of many socioeconomic phenomena (Kim, Mohanty, & Subramanian, 2016); thus, the role of individual abilities is usually overestimated. Further, if one ignores structural heterogeneity, the issue of endogeneity arises. Bell and Jones (2015) call this type of endogeneity 'heterogeneity bias', which affects the parameter estimates if one does not specify the structure-specific variance. Given that an REWB model contains greater flexibility and generalizability, it offers an opportunity to account for not only individual but also structural heterogeneity (Bell & Jones, 2015); thus, it is an REWB model with structural effects.

In the context of this advice, our second model considers individual and structural heterogeneity. Our specification of the REWB model with structural effects considers occupation-specific units and the variation between them. In technical terms, we keep the intercept (the grand mean) random, given both the individual-specific (level 2, '*i*' lower index) and the structural-specific (level 3, '*k*' lower index) variables. Equation (5) presents the model as follows:

$$y_{itk} \sim \text{Binomial}(\text{const}_{itk}, \pi_{itk})$$

Micro-level:

$$\text{Probit}(\pi_{itk}) = \boldsymbol{\beta_{0ik}} + \beta_1(x_{itk} - x.mean_{ik}) + \beta_2 x.mean_{ik} + \beta_4 \text{z}_{ik}$$

Macro-level:

$$\boldsymbol{\beta_{0ik}} = \beta_0 + u_{ik} + v_k$$

Full-length model:

$$\text{Probit}(\pi_{itk}) = \beta_0 + \beta_1(x_{itk} - x.mean_{ik}) + \beta_2 x.mean_{ik} + \beta_4 \text{z}_{ik} + (u_{ik} + v_k) \quad (5)$$

$$\text{var}(y_{itk} \mid \pi_{itk}) = \pi_{itk}(1 - \pi_{itk}) / \text{const}_{itk}$$

$$[u_{ik}] \sim \text{N}(0, \Omega_u) : \Omega_u = [\sigma^2_{u_{ik}}]$$

$$[\,v_k\,] \sim \mathrm{N}(0, \Omega_v) : \ \Omega_v = \left[\sigma^2_{v_k}\right]$$

In equation (5), the structural variance component, ,$\sigma^2_{v_k}$„ is considered to capture the structural variation. Because the multilevel process we address in the REWB model with structural effects does not refer to a clear hierarchy, we allow for cross-classification; in other words, individuals (level 2) can move between occupations (level 3) over time.

4.3 Time-specific heterogeneity

Equation (2) has some useful properties. With regard to a panel with no missingness, if x_{it} denotes age, then $x.mean_i$ would represent individuals' cohort; thus, β_2 contains the estimate of the 'cohort effect'. This point leads to an age–period–cohort (APC) discourse. The final model of interest accounts for time-specific heterogeneity, which is needed to test Hypothesis 1. However, the specification of periods (years) in an APC model is challenging because we cannot use age, period, or cohort as explanatory components in the fixed part of the regression equation. The reason is that age, period, and cohort are linearly related (a circumstance known as the APC identification problem). Although this identification problem is recognized as unsolvable, researchers are still seeking efficient options. For instance, drawing on repeated cross-sectional data, Bell and Jones (2017) propose a hierarchical APC which circumvents the APC identification problem by specifying both the period and the cohort as cross-classified RE. With our panel study, we treat only periods as RE, not cohorts. By doing this, we minimize the amount of unexplained variation in the model, a method which coincides with the general strategy applied by Bell and Jones (2017). Equation (6) presents the model as follows:

$$y_{itkl} \sim \mathrm{Binomial}(\mathrm{const}_{itkl}, \pi_{itkl})$$

Micro-level:

$$\mathrm{Probit}(\pi_{itkl}) = \boldsymbol{\beta}_{\boldsymbol{0ikl}} + \beta_1(x_{itkl} - x.mean_{ikl}) + \beta_2 x.mean_{ikl} + \beta_4 z_{ikl}$$

Macro-level:

$$\boldsymbol{\beta}_{\boldsymbol{0ikl}} = \beta_0 + u_{ikl} + v_{kl} + f_l$$

Full-length model:

$$\mathrm{Probit}(\pi_{itkl}) = \beta_0 + \beta_1(x_{itkl} - x.mean_{ikl}) + \beta_2 x.mean_{ikl} + \beta_4 z_{ikl} + (u_{ikl} + v_{kl} + f_l) \quad (6)$$

$$\mathrm{var}(y_{itkl} \mid \pi_{itkl}) = \pi_{itkl}(1-\pi_{itkl})/\,\mathrm{const}_{itkl}$$

$$[u_{ikl}] \sim \mathrm{N}(0, \Omega_u) : \Omega_u = [\sigma^2_{u_{ikl}}]$$
$$[\,v_{kl}\,] \sim \mathrm{N}(0, \Omega_v) : \Omega_v = [\sigma^2_{v_{kl}}]$$
$$[\,f_l\,] \sim \mathrm{N}(0, \Omega_f) : \Omega_f = [\sigma^2_{f_l}]$$

In the final stage, we estimate a four-level model which comprises three pieces of unexplained variation in training: the individual-specific variance component, $\sigma^2_{u_{ikl}}$; the occupation-specific variance component, $\sigma^2_{v_{kl}}$; and the period (year)-specific variance component, $\sigma^2_{f_l}$. All the listed unexplained components are treated as cross-classified RE.

Cross-classified MLMs are complicated; hence, they are hard to fit using traditional deterministic methods applicable to binary response outcome models, which are (quasi) maximum likelihood methods. For this reason, Bayesian estimation methods are required. The Markov chain Monte Carlo method provides precision-weighted estimates, particularly when we have few units at higher levels (Goldstein, 1995), which is exactly the issue in our final model (we have only 14 years). Ultimately, in cross-class interaction models, even with a larger number of units, the known issues (biased estimates and too short confidence intervals) of deterministic methods become more apparent (Stegmueller, 2013). Hence, we fit the cross-classified REWB probit models using the Markov chain Monte Carlo method, monitoring a chain length of 100,000 and a burn-in length of 500.

5. Results and discussion

We confirm the significance of individual-specific heterogeneity in skills training. As shown in Tables 2 and 3, individual-specific (level 2) variance is significantly different from zero and remains significant in all three models. In Model 1, the variance partition coefficient (VPC), which may be used to identify the amount of variation in the probability of training attributable to differences between the higher-level entities of interest (Browne, Subramanian, Jones, & Goldstein, 2005), is 0.344. This finding indicates that differences between the time-invariant characteristics of individuals explain up to 34.4% of the variation in the probability of formal skills training in Russia, after accounting for

Table 2. Determinants of Skills Training, 2001–2014

	Model 1	[S.E.]	**Model 2**	[S.E.]	**Model 3**	[S.E.]
			Fixed Part			
Constant	-1.574***	[0.040]	-1.609***	[0.051]	-1.611***	[0.055]
Male	-0.165***	[0.021]	-0.135***	[0.026]	-0.135***	[0.025]
Residency						
Town (ref. cat.)						
Cities	0.077***	[0.023]	0.092***	[0.022]	0.093***	[0.022]
Villages	-0.109***	[0.029]	-0.115***	[0.029]	-0.115***	[0.029]

Age$_{within}$	-0.038***	[0.003]	-0.037***	[0.003]	-0.037***	[0.004]
Age$_{between}$-gm	-0.017***	[0.001]	-0.014***	[0.001]	-0.014***	[0.001]
Job-specific						
Tenure$_{between}$ -gm	0.012***	[0.002]	0.003**	[0.002]	0.003**	[0.002]
Subordinates	0.206***	[0.021]	0.233***	[0.023]	0.233***	[0.023]
Working hours$_{within}$	0.012***	[0.004]	0.007**	[0.004]	0.008**	[0.004]
Working hours$_{between}$-gm	0.004	[0.004]	-0.001	[0.004]	-0.001	[0.004]
Overwork	0.054**	[0.024]	0.093***	[0.025]	0.091***	[0.025]
Generic labor	-0.558***	[0.022]	-0.503***	[0.049]	-0.508***	[0.048]
Occupation-job flows						
Changed either job or profession (ref. cat.)						
Profession and job remain the same	-0.119***	[0.032]	-0.140***	[0.031]	-0.138***	[0.032]
Changed profession, but not a job	0.290***	[0.057]	0.282***	[0.056]	0.282***	[0.057]
Changed both job and profession	0.148***	[0.039]	0.168***	[0.038]	0.168***	[0.039]
Did not work in November last year	-0.001	[0.046]	0.023	[0.045]	0.025	[0.045]
Second job	0.318***	[0.035]	0.280***	[0.034]	0.279***	[0.034]
Random parameters						
Level 4: years						
$\sigma^2_{f_l}$					0.003	[0.002]
Level 3: occupations						
$\sigma^2_{v_{kl}}$			0.114	[0.013]	0.114	[0.014]
Level 2: individuals						
$\sigma^2_{u_{ikl}}$	0.525	[0.024]	0.394	[0.020]	0.396	[0.019]
Model diagnostics and No of units						
DIC:	38,872.2		38,232.7		38,210	
pD:	4,795.37		4,202.73		4,216.78	
Units: years (level 4)					14	
Units: occupations (level 3)			430		430	
Units: individuals (level 2)	23,382		23,382		23,382	
Units: occasions (level 1)	95,040		95,040		95,040	

Notes: All three models utilize ‘random intercept’ specification. Individual-specific group means are centred around the grand mean (‘gm’) and do not predict beyond the range of the data (Bell & Jones, 2015). The ‘S.E.’ column presents the standard errors of the estimated parameters. The mean values of the estimated parameters are also presented. The asterisks denote Bayesian p-values (Bp), specified in accordance with the following rule: *** Bp <0.01, ** Bp<0.05, * Bp<0.1. For more details, see Appendices E, F, and G.

important individual characteristics and the within and between effects of time-variant socioeconomic parameters such as age, tenure, and working hours.

This considerable amount of unexplained variation may be omitted if a researcher models the longitudinal likelihood of training within an FE framework. From a theoretical perspective, the unobserved characteristics of individuals are crucial to understand why some workers receive formal training and others do not. This finding notably helps reassess and enrich the dominant view among applied economists that skills training in advanced industrial societies (and in Russia) is a function of employers' inclinations to invest in the human capital of their workers. During the recent economic growth in Russia, workers had opportunities to improve their qualifications—or at least, these opportunities were not heavily restricted by their employers. However, most employees seemed to be reluctant to use these opportunities and develop their skills.

Further, the effect of unobserved individual heterogeneity is robust, although its value decreases when controlling for the other sources of unexplained variation in the model. As shown in Model 2, the estimated value of the variance of unobserved individual heterogeneity, $\sigma^2_{u_{ik}}$, reduces from 0.525 to 0.394 after controlling for the occupational variance component, $\sigma^2_{v_k}$, the value of which is 0.114. Then, the VPC for the individual-specific variance within occupations is 0.261. In other words, individual heterogeneity embraces up to 26.1% of total residual variance after accounting for structural heterogeneity. The VPC for occupation-specific variance is 0.076; thus, 7.6% of the variation in the likelihood of training is related to the inequality between occupations. Variation at both levels explains up to 33.7% of the total variation in the probability of training during 2001–2014. Hence, unobserved individual heterogeneity is the most remarkable source of the unexplained variation of training. Therefore, accounting for this heterogeneity helps uncover the individual-based nature of skills development, thereby explaining the enormous diversity of the training probability among qualified non-manual workers in contemporary Russia (Anikin, 2017).

Although the contribution of structural RE to the likelihood of training is less salient than that of individual RE, Table 3 confirms that the estimated value of the occupation-specific variance component is significantly different from zero. Further, such a remarkable difference in the estimated values of the occupation-specific and individual-specific variance components originates from the different numbers of units at these levels. Indeed, a larger number of units at the individual level boosts the value of the estimated parameter and increases its contribution to the unexplained part of the model.

Could we omit the structural heterogeneity parameter? Our answer to this question is fairly positive because we find a significant reduction in the Bayesian deviance information criterion (DIC) value after accounting for occupation-specific variation in Model 2; indeed, this reduction is from 38,872.2 to 38,232.7, given the same number of observations at level 1 (95,040). Thus, we cannot ignore structural heterogeneity when assessing individual heterogeneity in panel studies of training, otherwise the 'pure' individual-specific RE would be slightly exaggerated. Unfortunately, in 'traditional' RE models, most authors terminate modelling at the individual heterogeneity stage, thereby losing (and leaving unexplained) up to 7.6% of the structural heterogeneity of training in Russia.

Table 3. Significance of Homogeneous RE Using Chi-square Tests

Variance component	**Model 1**	**Model 2**	**Model 3**

Individual-specific $\sigma^2_{u_{ikl}}$	495.436 (p = 0.000)	400.616 (p=0.000)	418.604 (p=0.000)
Occupation-specific $\sigma^2_{v_{kl}}$	-	72.234 (p=0.000)	70.184 (p=0.000)
Year-specific $\sigma^2_{f_l}$	-	-	2.636 (p=0.052)

Notes: The results of the chi-square tests are computed independently for each term in each model; directional H_0 hypotheses are used. The p-values are in parentheses. See Appendix H for further examination about the posterior distributions of the variance components.

The year-specific variation, by contrast, barely increases the fit of the model and leads to a relatively slight reduction in the DIC (from 38,232.7 to 38,210.0). Despite this decrease in the DIC (which can be considered to be a slight improvement of the model), the results of the chi-square test (see Table 3) lead to the acceptance of H_0, assuming $\sigma^2_{f_l} = 0$ at the conventional significance level of 0.1 (p = 0.052). However, we seem to reject the null hypothesis at the significance levels of 0.05 and 0.01. Consequently, the unobserved variance between years barely explains the unobserved variation in training. As shown in Table 2, the estimated value of the parameter $\sigma^2_{f_l}$ is 0.003 and the VPC for the year-specific variation is 0.002; thus, only 0.2% of the variation in the likelihood of training is attributable to the inequality between years. Hence, these results confirm Hypothesis 1, which proposes that the years of economic prosperity had no effect on the probability of training.

With regard to the contribution of the observed characteristics to the variation in training, the main results are as follows. First, we confirm that separation between the within- and between-person effects produce tangible results. As shown in Table 2, three time-variant scale-measured terms must be estimated: age, tenure, and working hours. All three terms represent observed individual-specific time-variant characteristics; however, only two of them are significant determinants of training during the period of interest, namely age and working hours per average day.

The negative effect of the within-person variation of age indicates that while a worker is becoming older (relative to the worker's average age observed during the considered period), the likelihood of training significantly decreases (assuming all other parameters remain unchanged). The time-invariant cohort effect (i.e. the between-person effect centred on the grand mean, $Age_{between}$-*gm*; see Table 2) is also negative and highly statistically significant.

We revert to the models without non-linear (quadratic) effects for these terms and the within-person effect for tenure^2 because all the terms appear to be no different from zero; moreover, they affect the fit with increased complexity and the DIC. This finding indicates that the considered within-person effects of age and working hours seem to have a linear impact on training, whereas a worker's tenure contributes a cohort effect (i.e. the time-invariant between-person effect of tenure).

As shown in Table 2, workers who stay in the same jobs at companies for a greater number of years (considered relatively to the average working experience within the labour market, *Tenure*$_{between}$ *-gm*) demonstrate a lower incidence of training than their less experienced peers. This finding is particularly important because it reveals the nature of skills development in Russian enterprises, which are reluctant to invest in the human capital of senior and experienced workers. In other words, this finding illustrates a stumbling block for lifelong learning in Russia. Unfortunately, existing studies of training in Russia have tended to disregard this effect and focus only on the insignificance of the within-person effect for tenure (Berger et al., 2001). Thus, the *tenure-specific cohort effect* seems to be a significant FE variable omitted from these studies, a situation which could lead to the publication of biased estimates.

In contrast to year-measured working experience, the FE for working hours per average day (*Working hours*$_{between}$*-gm*) is not significant, while the within-person effect (*Working hours*$_{within}$) is positive and statistically significant. Thus, hardworking individuals are more likely to receive training than other workers. Taking into account the positive affect of additional employment and hard work (i.e. working more than eight hours per day), this finding indicates that skills development during 2001–2014 was partly supported by the high demand from Russian enterprises for employees who ‘can work’ (i.e. skilled labour). Moreover, we confirm the proposal that newcomers at work are more likely to receive training than others, particularly if they hold managerial positions or have just changed their occupations. Although we do not know whether these employees received training before or after occupation–job mobility, this finding indicates that occupation–job flows are matched to skills development practices. Ultimately, such a finding challenges the popular viewpoint that occupations and jobs are significantly mismatched with skills in contemporary Russia (Demmou & Wörgötter, 2015).

Finally, our findings support Hypothesis 2, which proposes the crucial role of macro-occupational classes in obstructing individuals from training. As shown in this study, working in the powerless generic labour market significantly decreases the incidence of training. Moreover, employees residing in villages are less likely to receive training than their counterparts living in towns and cities. Recalling the inverse relationship between older cohorts and training, we conclude that older semi- and low-skilled workers residing in rural Russia are a social group whose skills were barely formally developed during the recent years of economic prosperity.

6. Conclusion

Skills training in Russia reflects individual heterogeneity; hence, it may be linked to meritocracy to a greater extent than expected. Individual time-variant and time-invariant determinants of training substantially contribute to skills acquisition, even in societies which have yet to establish knowledge-based economies. The unobserved differences between individuals gain a larger share in the prediction of training, explaining up to 26.1–34.4%. This finding helps explain the immense diversity in the probability of training among qualified non-manual workers in contemporary Russia (Anikin, 2017). From a methodological perspective, this finding justifies the importance of RE models of training which allow for distinct within and between effects. However, skills training remains a strongly confined phenomenon; as a result, it does not have a considerable influence on the Russian economy compared with external and structural factors. For instance, we find that only 0.2% of the variation in training is attributable to the years of economic prosperity after accounting for other important parameters.

Why is the incidence of training in Russia relatively low, even during a growth period? Some scholars consider the problem to be significantly hampered by specific institutional arrangements which disparage productivity based on human capital and thus result in significant skill mismatches (Demmou & Wörgötter, 2015). Another possible explanation comes from the finding that organizations have low incentives to invest in workers' human capital because of high personnel turnover (Travkin & Sharunina, 2016). Our study considerably reassesses these views. The structural determinants of training are clearly revealed in the significant role of the occupational structure in contributing to the concept of the industrial society and our understanding of the social origins of skills in such a society (Green, 2013). Further, the differences between occupations significantly predict the probability of training, capturing up to 7.6% of the variation in the likelihood of training.

The negative prediction of training is explained to a large extent by the numerous 'bad' jobs accepted by workers in the 'generic labour' category, the categorization of employees by enterprises, resulting in skills development discrimination against older cohorts, and the rural location of labour. By contrast, the positive prediction of training is more likely to be embedded in a post-industrial context. This study shows that the probability of training can be determined by qualification-level matching of workers with 'good' jobs and the usefulness of employees as represented by those who work more than eight hours per day, apply themselves, and seek additional employment. Such a finding demonstrates the indirect link between training and labour market demand during the recent economic growth.

Our findings highlight the crucial role of the individual in skills development during the years of economic prosperity, although these years per se barely affected training. With regard to policy recommendations, our results help reassess the existing discourse on skills development in contemporary Russia. Nowadays, experts develop recommendations at the organizational level and seek efficient mechanisms which encourage businesses to develop their employees' skills. Despite the significance of such a policy stream, policymakers and experts should perhaps switch their focus from organizations to individuals and develop instruments which enhance the personal inclinations of workers, particularly older cohorts, towards skills acquisition.

† This paper was supported by the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2019 'Mobilization and improvement in human capital of the middle

class and valnurable groups'. I am also grateful to the Joint Usage and Research Center of the Institute of Economic Research, Kyoto University, for the financial support for this article.

Notes

[1] The modified version of ISCO-88 is applied to match the reality of the Russian labour market and eliminate classification mismatches. Regarding mismatches in the occupational coding of RLMS—HSE, see Sabirianova (2002). In the modified version of ISCO-88, we classify professionals as specialists who have a university degree or a related equivalent. By this, we mean that the number of years of education in Russia is less important than formal credentials. 'Managers' are those with more than five employees under their direction. Those managers with fewer than five subordinates are treated as supervisors and coded as a separate category among professionals. Those professionals who have no university degree or a related equivalent are encoded as a separate category among 'semi-professionals'. The latter are filtered accordingly in relation to the level of education required for each group (tertiary education: unfinished undergraduate or vocational training). Some minor occupations are directly recoded as lower occupations because of the specifics of their work and value in the labour market. For more details, see Anikin (2012).

[2] We omit the within effect from the 'fixed part' of the regression equation in accordance with the corresponding property of REWB models (Bell et al., 2017).

References

Acemoglu, D., Aghion, P., & Zilibotti, F. (2006) "Distance to frontier, selection, and economic growth", *Journal of the European Economic Association*, Vol. 4, No. 1, 37-74.

Acemoglu, D., & Pischke, J. S. (1999) "Beyond Becker: Training in imperfect labor markets", *The Economic Journal*, Vol. 109, No. 453, F112-F142.

Allison, P. D. (2009) *Fixed effects regression models*, London: Sage.

Alvesson, M., & Benner, M. (2016) "Higher education in the knowledge society: Miracle or mirage?", in J. Frost, F. Hattke, & M. Reihlen eds., *Multi-level Governance in Universities* (pp. 75-91), New York: Springer.

Anikin, V. A. (2012) "The modernization potential of the professional structure of the employed population of Russia", *Sociological Research*, Vol. 51, No. 6, 44-80. doi:10.2753/SOR1061-0154510605

Anikin, V. A. (2017) "Occupational propensity for training in a late industrial society: Evidence from Russia," *International Journal of Training and Development*, Vol. 21, No. 4, 249-284.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010) "On making causal claims: A review and recommendations," *The Leadership Quarterly*, Vol. 21, No. 6, 1086-1120.

Arulampalam, W., Booth, A. L. & Bryan, M. L. (2004) "Training in Europe," *Journal of the European Economic Association*, Vol. 2, No. 2- 3, 346-360.

Bartel, A. P. & Sicherman, N. (1998) "Technological change and skills acquisition of young workers," *Journal of Labor Economics*, Vol. 16, 718-755.

Baum, C. F. (2006) *An Introduction to Modern Econometrics Using STATA*, Texas: Stata Press.

Bell, A. & Jones, K. (2015) "Explaining fixed effects: Random effects modelling of time-series cross-sectional and panel data," *Political Science Research and Methods*, Vol. 3, No. 1, 133-153.

Bell, A. & Jones, K. (2017) "The hierarchical age–period–cohort model: Why does it find the results that it finds?," *Quality & Quantity*, 1-17.

Bell, A., Jones, K. & Fairbrother, M. (2017) "Understanding and misunderstanding group mean centring: A commentary on Kelley et al.'s dangerous practice," *Quality & Quantity*, 1-6. doi:10.1007/s11135-017-0593-5

Berger, M. C., Earle, J. S. & Sabirianova, K. Z. (2001) "Worker training in a restructuring economy: Evidence from the Russian transition," in Polachek, S. ed., *Worker Wellbeing in a Changing Labor Market* (research in labor economics, volume 20) (pp. 159-190): Emerald Group Publishing Limited.

Biørn, E. (2016) *Econometrics of Panel Data: Methods and Applications*, Oxford University Press.

Bosch, G. & Charest, J. (2012) *Vocational Training: International Perspectives*, New York: Taylor & Francis.

Browne, W. J., Subramanian, S. V., Jones, K. & Goldstein, H. (2005) "Variance partitioning in multilevel logistic models that exhibit overdispersion," *Journal of the Royal Statistical Society: Series A* (Statistics in Society), Vol. 168, No. 3, 599-613.

Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2008) "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, Vol. 90, No. 3, 414-427.

Castells, M. (2010) *The Rise of the Network Society* (the information age: Economy, society and culture, volume 1) (2nd ed.). Singapore: Wiley-Blackwell.

Cho, Y., Kalomba, D., Mobarak, A. M. & Orozco, V. (2013) "Gender differences in the effects of vocational training: Constraints on women and drop-out behavior," *World Bank Policy Research Working Paper*, Washington, DC: World Bank.

Connolly, R. (2012) "Climbing the ladder? High-technology export performance in emerging Europe," *Eurasian Geography and Economics*, Vol. 53, No. 3, 356-379.

Demmou, L. & Wörgötter, A. (2015) "Boosting productivity in Russia: Skills, education and innovation," OECD Economics Department Working Papers, Vol. 1189. doi:http://dx.doi.org/10.1787/5js4w26114r2-en

Denisova, I. A. (2010) "Adult mortality in Russia: A microanalysis," *Economics of Transition*, Vol. 18, No. 2, 333–363.

Ding, D., Fields, D. & Akhtar, S. (1997) "An empirical study of human resource management policies and practices in foreign-invested enterprises in China: The case of Shenzhen special economic zone," *International Journal of Human Resource Management*, Vol. 8, No. 5, 595-613.

Fitzenberger, B. & Muehler, G. (2015) "Dips and floors in workplace training: Gender differences and supervisors," *Scottish Journal of Political Economy*, Vol. 62, No. 4, 400-429.

Gaddy, C. G. & Ickes, B. W. (2013) "Russia's dependence on resources," in Alexeev, M. & Weber, S. eds., *The Oxford Handbook of the Russian Economy*, New York: Oxford University Press.

Gerber, T. P. & Hout, M. (1998) "More shock than therapy: Market transition, employment, and income in Russia, 1991–1995," *American Journal of Sociology*, Vol. 104, No. 1, 1-50.

Gerber, T. P. & Hout, M. (2004) "Tightening up: Declining class mobility during Russia's market transition," *American Sociological Review*, Vol. 69, No. 5, 677-703.

Gerry, C. J. & Papadopoulos, G. (2015) "Sample attrition in the RLMS, 2001–10," *Economics of Transition*, Vol. 23, 425–468.

Gimpelson, V. & Kapeliushnikov, R. (2013) "Labor market adjustment: Is Russia different?," in Alexeev, M. & Weber, S. eds., *The Oxford Handbook of the Russian Economy*, Oxford University Press.

Gimpelson, V. & Kapeliushnikov, R. (2016) "Polarization or upgrading? Evolution of employment in transitional Russia," *Russian Journal of Economics*, Vol. 2, No. 2, 192-218.

Gimpelson, V. & Lippoldt, D. (2001) *The Russian Labor Market: Between Transition and Turmoil*, Boston: Rowman & Littlefield.

Goldstein, H. (1995) "Hierarchical data modelling in the social sciences," *Journal of Educational and Behavioral Statistics*, 201-204.

Goldthorpe, J. H., Llewellyn, C. & Payne, C. (1987) *Social Mobility and Class Structure in Modern Britain* (2nd ed.), Oxford: Clarendon Press

Goldthorpe, J. H. & McKnight, A. (2006) "The economic basis of social class," in Morgan, S. L., Grusky, D. B. & Fields, G. S. eds., *Mobility and Inequality: Frontiers of Research in Sociology and Economics* (pp. 109-137), California: Stanford University Press.

Green, F. (2013) *Skills and Skilled Work: An Economic and Social Analysis*, Oxford: Oxford University Press.

Green, F., Felstead, A., Gallie, D., Inanc, H. & Jewson, N. (2016) "The declining volume of workers' training in Britain," *British Journal of Industrial Relations*, Vol. 54, 422-448.

Green, F. & Zanchi, L. (1997) "Trends in the training of male and female workers in the United Kingdom," *British Journal of Industrial Relations*, Vol. 35, No. 4, 635-644.

Grusky, D. B. (2001) "The past, present, and future of social inequality," in Grusky, D. B. ed., *Social Stratification: Class, Race, and Gender in Sociological Perspective* (Second ed., pp. 3-51), Boulder: Westview Press.

Hausman, J. A. & Taylor, W. E. (1981) "Panel data and unobservable individual effects," *Econometrica*, Vol. 49, No. 6, 1377-1398.

Heckman, J. J. (1981) "Heterogeneity and state dependence," in *Studies in Labor Markets* (pp. 91-140). University of Chicago Press.

Heeringa, S. G. (1997) Sample attrition, replenishment, and weighting in rounds V-VII, Retrieved from The Russia Longitudinal Monitoring Survey - Higher School of Economics website: http://www.cpc.unc.edu/projects/rlms-hse/project/samprep

Kim, R., Mohanty, S. K. & Subramanian, S. V. (2016) "Multilevel geographies of poverty in India," *World Development*, Vol. 87, 349-359.

Klimova, A. (2012) "Gender differences in determinants of occupational choice in Russia," *International Journal of Social Economics*, Vol. 39, No. 9, 648-670.

Kozyreva, P., Kosolapov, M. & Popkin, B. M. (2016) "Data resource profile: The Russia Longitudinal Monitoring Survey—Higher School of Economics (RLMS-HSE) phase II: Monitoring the economic and health situation in Russia, 1994–2013," *International Journal of Epidemiology*, Vol. 45, No. 2, 395-401.

Kyui, N. (2016) "Expansion of higher education, employment and wages: Evidence from the Russian transition," *Labor Economics*, Vol. 39, 68-87.

Lazareva, O. V. (2006) "On-the-job training in Russia: Determining factors and return," (in Russian) Vol. 06/05, Moscow: EERC.

Lemieux, T. (2006) "The "mincer equation" thirty years after schooling, experience, and earnings," in Grossbard, S. ed., *Jacob Mincer a Pioneer of Modern Labor Economics* (pp. 127-145), New York: Springer.

Loewenstein, M. A. & Spletzer, J. R. (1999) "Formal and informal training: Evidence from the NLSY," *Research in Labor Economics*, Vol. 18, 403-438.

Maddala, G. S. (1987) "Limited dependent variable models using panel data," *Journal of Human Resources*, Vol. 22, No. 3, 307-338.

Méndez, F. & Sepúlveda, F. (2016) "A comparative study of training in the private and public sectors: Evidence from the United Kingdom and the United States," Contemporary Economic Policy, Vol. 34, No. 1, 107-118.

Mincer, J. (1962) "On-the-job training: Costs, returns, and some implications," *The Journal of Political Economy*, Vol. 70, No. 5, 50-79.

Mincer, J. (1974) *Schooling, Experience, and Earnings*, New York: National Bureau of Economic Research, Columbia University Press.

Mundlak, Y. (1978) "On the pooling of time series and cross section data," *Econometrica*, Vol. 46, No. 1, 69-85.

O'Connell, P. J. (1999) *Adults in Training: An International Comparison of Continuing Education and Training*, Paris: OECD.

Obizhaeva, A. (2016) "The Russian rubble crisis of December 2014," *Voprosy Ekonomiki*, Vol. 5.

OECD (2016) *Russian Federation Education at a Glance: OECD indicators*, Paris: OECD Publishing.

Polavieja, J. G. (2012) "Socially embedded investments: Explaining gender differences in job-specific skills," *American Journal of Sociology*, Vol. 118, No. 3, 592-634.

Sabirianova, K. Z. (2002) "The great human capital reallocation: A study of occupational mobility in transitional Russia," *Journal of Comparative Economics*, Vol. 30, No. 1, 191-217.

Schmidbauer, H., Rösch, A., Uluceviz, E. & Erkol, N. (2016) "The Russian stock market during the Ukrainian crisis: A network perspective," *Czech Journal of Economics and Finance*, Vol. 66, No. 6, 478-509.

Stegmueller, D. (2013) "How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches," *American Journal of Political Science*, Vol. 57, No. 3, 748-761.

Tharenou, P. (1997) "Organisational, job, and personal predictors of employee participation in training and development," *Applied Psychology*, Vol. 46, No. 2, 111-134.

Tharenou, P. (2001) "The relationship of training motivation to participation in training and development," *Journal of Occupational and Organizational Psychology*, Vol. 74, No. 5, 599-621.

Timmer, M. P. & Voskoboynikov, I. B. (2014) Is mining fuelling long- run growth in Russia? Industry productivity growth trends since 1995, *Review of income and wealth*, Vol. 60, No. 2, 398-422.

Travkin, P. V. & Sharunina, A. (2016) "The returns to training in Russia: A difference-in-differences analysis," *International Journal of Training and Development*, Vol. 20, No. 3, 262-279.

Voigt, P. (2006) "Russia's way from planning toward market: A success story? A review of economic trajectories, transition progress and Putin's merits," *Post-Communist Economies*, Vol. 18, No. 2, 123-138.

Wright, E. O. (1997) *Class Counts: Comparative Studies in Class Analysis*, Cambridge: Cambridge University Press.

Appendix A

Description of Main Variables

Variable name	Type	Description
Response variable		
Training	Binary [0; 1]	Courses for the improvement of professional skills or any other courses over the last 12 months 0- No; 1- Yes

Variable name	Type	Description
Single-level independent factors		
Socio-demographics		
Gender	Binary [0; 1]	0- Female; 1- Male
Residency	Nominal [1; 3]	1- City; 2- Town; 3- Village
Age	Scale	
Age squared	Scale	
Occupation- and job-specific level		
Tenure	Scale	Year started the primary job subtracted from the year of survey
Tenure squared	Scale	
Occupational class	Binary [0; 1]	0- Generic labour: office clerks, sales workers, famers, craft workers, operators and assemblers, and elementary occupations 1- Qualified non-manual labour: managers, professionals, and semi-professionals;
Have any subordinates	Binary [0; 1]	0- No; 1- Yes
Working time, hours in average workday	Scale	
Working time, more than eight hours in average workday	Binary [0; 1]	0- No; 1- Yes
Occupation-job flows since November last year	Nominal [1; 6]	1- Profession and place of work remain the same; 2- Changed profession, but not place of work; 3- Changed place of work, but not profession; 4- Changed both place of work and profession; 5- Changed either place of work or profession; 6- Did not work in November last year
Second job	Binary [0; 1]	0- No; 1- Yes
Higher level entities		
Level 2: Individuals	Nominal [1; 23,870]	Unique longitudinal person ID
Level 3: Occupations	Nominal [1; 430]	Occupations coded via 4-digit code as per ISCO-88
Level 4: Years	Nominal [1; 14]	Years (waves) of survey covering the period 2001–2014

Appendix B

Variables and Missingness

Variables, short labels	Missing, N	Missing time points

Selected variables		
Gender	0	
Residency	0	
Age	3	
Training	94	
Second job	110	
Changed job since November last year	169	
Subordinates	180	
Occupational structure (ISCO-88)	245	
Occupational class	789	
Tenure	997	
Working hours	2,958	
Omitted variables		
Job satisfaction	5,563	2001
Ownership, government	12,463	
Ownership, foreign enterprise	12,516	
Ownership, Russian enterprise	13,731	
Official employment	13,867	2001
Economic sector / industry	17,008	2001-2003
Organizational size	33,628	
Use of personal computer at work	39,604	
Lagged training	52,525	

Source: The RLMS—HSE data, panel samples.

Appendix C

Description of Waves, 2001-2014

Year	All observations	Observations for working respondents

	Frequency	**Percent**	**Frequency**	**Percent**
2001	12,121	5.4	4,871	4.9
2002	12,523	5.6	5,102	5.2
2003	12,656	5.6	5,282	5.3
2004	12,641	5.6	5,339	5.4
2005	12,237	5.4	5,245	5.3
2006	14,689	6.5	6,547	6.6
2007	14,505	6.4	6,589	6.7
2008	14,026	6.2	6,476	6.5
2009	13,991	6.1	6,392	6.4
2010	21,343	9.5	9,703	9.8
2011	21,993	9.8	9,842	9.9
2012	22,534	10.0	9,995	10.1
2013	21,753	9.7	9,655	9.7
2014	18,372	8.2	8,063	8.2
Total	225,384	100.0	99,101	100.0

Source: The RLMS—HSE data, panel samples.

Appendix D

Table D1. Proportions of working and non-working population, 2001-2014, %

	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**

Employed	40.3	40.3	40.6	42.4	42.5	44.3	45.4	46.2	46.0	47.0	47.8	47.9	48.2	47.6
Out-of-Labour Force	52.6	52.7	52.6	50.0	50.2	49.3	49.3	48.9	50.1	48.3	47.3	47.7	47.8	47.4
Unemployed, the RLMS-HSE data	7.1	7.0	6.8	7.6	7.3	6.4	5.3	4.9	3.9	4.7	4.9	4.4	4.0	5.0
Unemployment, the official rate*	10.6	9.0	7.9	8.2	7.8	7.1	7.1	6.0	6.2	8.3	7.3	6.5	5.5	5.5

Source: The RLMS—HSE data, representative samples; The official rate of unemployment is compiled from the Federal State Statistics Service data, see elsewhere:

http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/

Notes: The RLMS—HSE data are weighted by the post-stratification sampling weight (provided with the data). Out-of-Labour Force includes pensioners and students.

Table D2. Proportions of working and non-working population within Russians who received formal training in the same year, 2001-2014, %

	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Employed	71.4	73.7	77.2	72.7	66.3	82.2	79.2	76.8	79.0	81.8	85.3	85.8	88.1	84.1
Out-of-Labour Force	22.9	21.5	17.6	23.9	26.0	15.0	17.6	18.7	16.2	12.6	10.8	10.0	9.1	12.5
Unemployed	5.7	4.8	5.2	3.4	7.7	2.8	3.2	4.5	4.8	5.6	3.9	4.2	2.8	3.4

Source: The RLMS—HSE data, representative samples

Notes: The RLMS—HSE data are weighted by the post-stratification sampling weight (provided with the data).

Appendix E

Model 1, Estimated Parameters

Parameter	**Posterior mean**	**S.E.**	**Credible Interval 2.5%**	**Credible Interval 97.5%**	**ESS**	**Bayesian p-value**
			Fixed Part			
Constant	-1.574	0.040	-1.652	-1.496	504	0.000
Male	-0.165	0.021	-0.207	-0.123	5,173	0.000
Residency						
Town (ref. cat.)						
Cities	0.077	0.023	0.033	0.121	3,270	0.000
Villages	-0.109	0.029	-0.166	-0.051	4,916	0.000
$\text{Age}_{\text{within}}$	-0.038	0.003	-0.043	-0.032	21,295	0.000
$\text{Age}_{\text{between}}$-gm	-0.017	0.001	-0.019	-0.015	5,775	0.000
Job-specific						
$\text{Tenure}_{\text{between}}$ -gm	0.012	0.002	0.009	0.015	5,127	0.000
Subordinates	0.206	0.021	0.164	0.247	6,759	0.000
Working $\text{hours}_{\text{within}}$	0.012	0.004	0.004	0.019	11,103	0.001
Working $\text{hours}_{\text{between}}$-gm	0.004	0.004	-0.003	0.012	5,105	0.120
Overwork	0.054	0.024	0.006	0.102	4,683	0.013
Generic labour	-0.558	0.022	-0.601	-0.517	4,297	0.000
Occupation-job flows						
Changed either job or profession (ref. cat.)						
Profession and job remain the same	-0.119	0.032	-0.183	-0.055	712	0.000
Changed profession, but not a job	0.290	0.057	0.179	0.401	3,570	0.000
Changed both job and profession	0.148	0.039	0.071	0.224	1,322	0.000
Did not work in November last year	-0.001	0.046	-0.090	0.089	1,905	0.486
Second job	0.318	0.035	0.250	0.385	14,002	0.000
			Random Part			
Level 2: individuals						
$\sigma^2_{u_i}$	0.525	0.024	0.478	0.571	738	
			Model diagnostics and number of units			
Units: individuals (level 2)	23,382					
Units: occasions (level 1)	95,040					
Estimation:	MCMC					
DIC:	38,872.2					
pD:	4,795.4					
Burn-in:	500					
Chain Length:	100,000					

Appendix F

Model 2, Estimated Parameters

	Posterior mean	**S.E.**	**Credible Interval 2.5%**	**Credible Interval 97.5%**	**ESS**	**Bayesian p-value**
	Fixed Part					
Constant	-1.609	0.051	-1.708	-1.510	414	0.000
Male	-0.135	0.026	-0.185	-0.085	3,540	0.000
Residency						
Town (ref. cat.)						
Cities	0.092	0.022	0.049	0.135	3,624	0.000
Villages	-0.115	0.029	-0.171	-0.058	5,375	0.000
$\text{Age}_{\text{within}}$	-0.037	0.003	-0.043	-0.032	20,907	0.000
$\text{Age}_{\text{between}}$-gm	-0.014	0.001	-0.016	-0.012	5,775	0.000
Job-specific						
$\text{Tenure}_{\text{between}}$ -gm	0.003	0.002	-0.000	0.006	5,417	0.026
Subordinates	0.233	0.023	0.188	0.278	6,228	0.000
Working $\text{hours}_{\text{within}}$	0.007	0.004	-0.000	0.015	10,750	0.032
Working $\text{hours}_{\text{between}}$-gm	-0.001	0.004	-0.009	0.006	4,845	0.394
Overwork	0.093	0.025	0.044	0.141	4,442	0.000
Generic labour	-0.503	0.049	-0.600	-0.407	674	0.000
Occupation-job flows						
Changed either job or profession (ref. cat.)						
Profession and job remain the same	-0.140	0.031	-0.200	-0.079	1,070	0.000
Changed profession, but not a job	0.282	0.056	0.172	0.391	4,453	0.000
Changed both job and profession	0.168	0.038	0.094	0.241	2,086	0.000
Did not work in November last year	0.023	0.045	-0.065	0.111	2,761	0.301
Second job	0.280	0.034	0.214	0.347	13,900	0.000
	Random Part					
Level 3: occupations						
$\sigma^2_{v_k}$	0.114	0.013	0.087	0.140	8,226	
Level 2: individuals						
$\sigma^2_{u_{ik}}$	0.394	0.020	0.355	0.433	588	
	Model diagnostics and number of units					
Units: occupations (level 3)	430					
Units: individuals (level 2)	23,382					
Units: occasions (level 1)	95,040					
Estimation:	MCMC					
DIC:	38,232.744					
pD:	4,202.7					
Burn-in:	500					
Chain Length:	100,000					

Appendix G

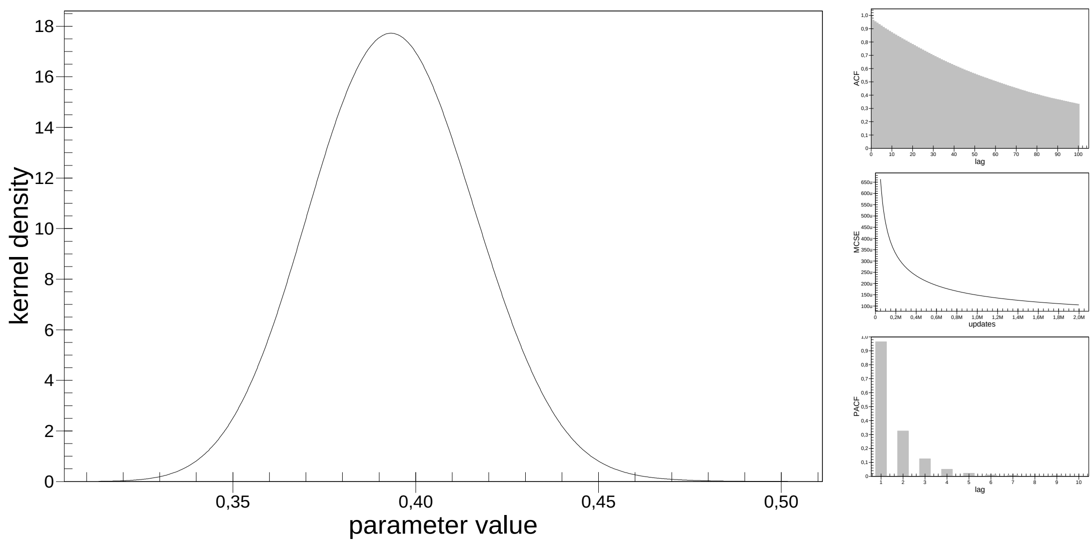
Model 3, Estimated Parameters

	Posterior mean	**S.E.**	**Credible Interval 2.5%**	**Credible Interval 97.5%**	**ESS**	**Bayesian p-value**
			Fixed Part			
Constant	-1.611	0.055	-1.719	-1.505	283	0.000
Male	-0.135	0.025	-0.186	-0.087	3,451	0.000
Residency						
Town (ref. cat.)						
Cities	0.093	0.022	0.050	0.137	3,562	0.000
Villages	-0.115	0.029	-0.172	-0.057	5,090	0.000
$\text{Age}_{\text{within}}$	-0.037	0.004	-0.044	-0.030	8,828	0.000
$\text{Age}_{\text{between}}$-gm	-0.014	0.001	-0.017	-0.012	5,630	0.000
Job-specific						
$\text{Tenure}_{\text{between}}$ -gm	0.003	0.002	-0.000	0.006	5,165	0.029
Subordinates	0.233	0.023	0.187	0.279	6,069	0.000
Working $\text{hours}_{\text{within}}$	0.008	0.004	-0.000	0.015	10,765	0.030
Working $\text{hours}_{\text{between}}$-gm	-0.001	0.004	-0.008	0.007	4,926	0.413
Overwork	0.091	0.025	0.042	0.139	4,324	0.000
Generic labour	-0.508	0.048	-0.600	-0.414	753	0.000
Occupation-job flows						
Changed either job or profession (ref. cat.)						
Profession and job remain the same	-0.138	0.032	-0.202	-0.075	655	0.000
Changed profession, but not a job	0.282	0.057	0.171	0.392	2,781	0.000
Changed both job and profession	0.168	0.039	0.092	0.245	1,126	0.000
Did not work in November last year	0.025	0.045	-0.064	0.114	1,827	0.291
Second job	0.279	0.034	0.212	0.346	14,092	0.000
			Random Part			
Level 4: years						
σ^2_{fl}	0.003	0.002	-0.001	0.006	17,832	
Level 3: occupations						
$\sigma^2_{v_{kl}}$	0.114	0.014	0.087	0.141	7,518	
Level 2: individuals						
$\sigma^2_{u_{ikl}}$	0.396	0.019	0.358	0.433	659	
			Model diagnostics and number of units			
Units: years (level 4)	14					
Units: occupations (level 3)	430					
Units: individuals (level 2)	23,382					
Units: occasions (level 1)	95,040					
Estimation:	MCMC					
DIC:	38,209.9					
pD:	4,216.8					
Burn-in:	500					
Chain Length:	100,000					

Appendix H

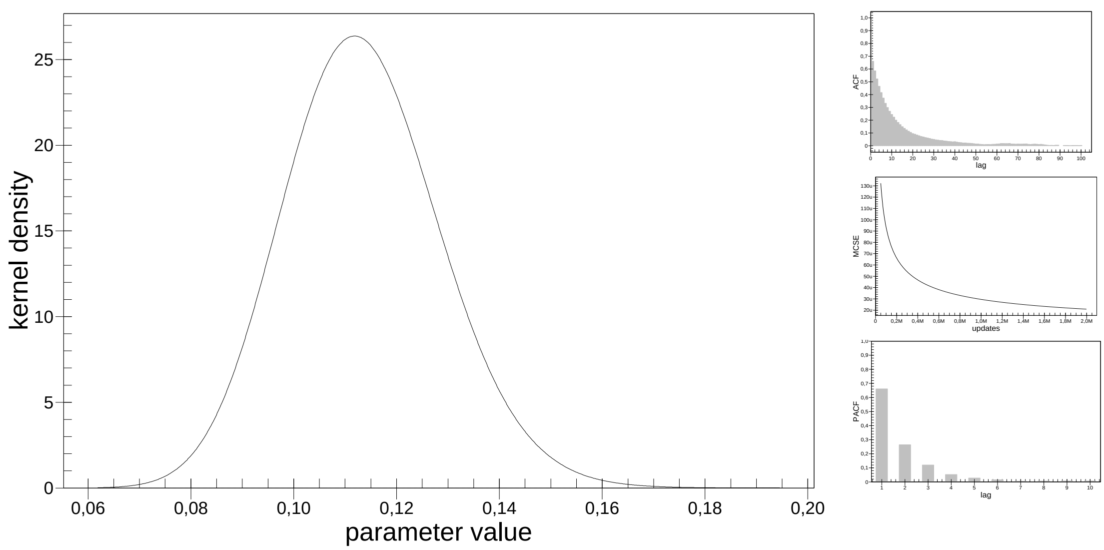
Variance Components, Key MCMC Diagnostics (Based on Model 2)

Unobserved individual-specific variance $\sigma^2_{u_{ik}}$



Posterior mean = 0.394 (0.000), SD = 0.020,
mode = 0.393. ESS = 588

Unobserved occupation-specific variance $\sigma^2_{v_k}$



Posterior mean = 0.114 (0.000), SD = 0.013,
mode = 0.112. ESS = 8,226

Notes: SD = Standard Deviation. ESS = Effective sample size. ESS is a parameter of Bayes diagnostics that used as a criterion for a sufficient number of MCMC simulations. It shows the 'restored' number of units of distribution of a parameter of interest. It is conventional practice to terminate simulation chain when ESS exceeds 250.