WPS8652

Policy Research Working Paper 8652

Does Job Polarization Explain the Rise in Earnings Inequality?

Evidence from Europe

Maurizio Bussolo
Ivan Torre
Hernan Winkler

WORLD BANK GROUP
Europe and Central Asia Region
Office of the Chief Economist
November 2018

Abstract

Earnings inequality and job polarization have increased significantly in several countries since the early 1990s. Using data from European countries covering a 20-year period, this paper provides new evidence that the decline of middle-skilled occupations and the simultaneous increase of high- and low-skilled occupations are important factors accounting for the rise of inequality, especially at the bottom of the distribution. Job polarization accounts for a large share of the increasing inequality between the 10th and the 50th percentiles, but it explains little or none of the increasing inequality between the 50th and 90th percentiles. Other important developments during this period, such as changing wage returns, higher educational attainment, and increased female labor force participation, account for a small portion of the changes in inequality.

This paper is a product of the Office of the Chief Economist, Europe and Central Asia Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/research. The authors may be contacted at mbussolo@worldbank.org, itorre@worldbank.org, and hwinkler@worldbank.org.

Does Job Polarization Explain the Rise in Earnings Inequality?

Evidence from Europe*

Maurizio Bussolo[γ]

Ivan Torre[¥]

Hernan Winkler[§]

JEL Codes: D31, J21, J24, J31

Keywords: Inequality, Polarization, Decomposition

---

* The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

γ The World Bank, e-mail: mbussolo@worldbank.org

¥ The World Bank, e-mail: itorre@worldbank.org

§ The World Bank, e-mail: hwinkler@worldbank.org

1. INTRODUCTION

Income inequality rose steadily in Europe over the past decades. While the share of total income accruing to the richest 10 percent was less than 30 percent in 1970, it reached 35 percent in 2010 (Piketty and Saez, 2014). Accordingly, the Gini coefficient experienced a significant increase in most European economies since the 1980s (OECD, 2017; Cingano, 2014). This was also particularly true for labor income inequality, whose Gini index rose on average by 5 points across Europe between 1990 and 2015 (Figure 1). There is a large body of empirical literature documenting the negative impacts of rising income inequality on a host of socio-economic outcomes such as health, safety, social mobility, cohesion and economic growth.[1]

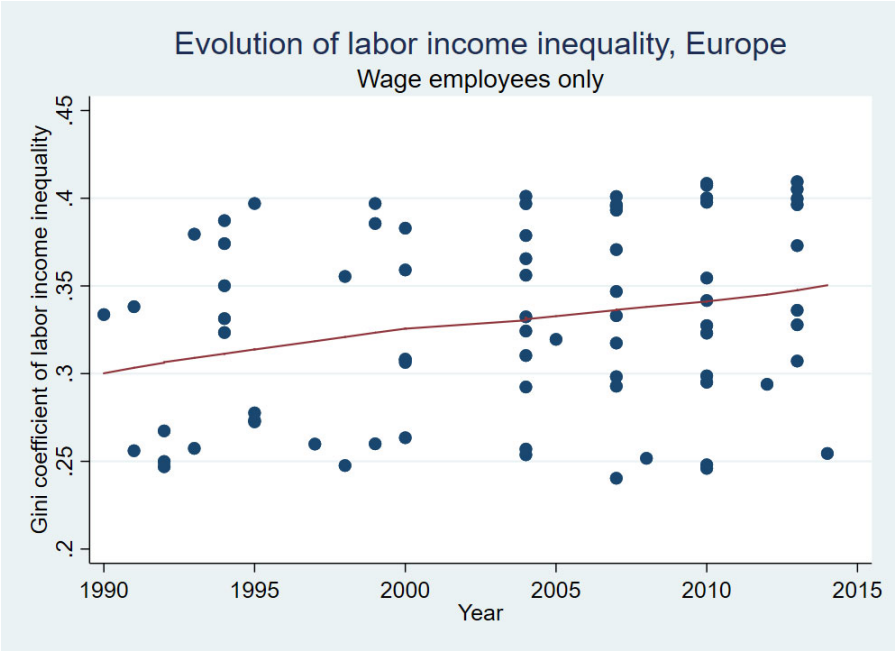
Several arguments have been put forward as potential drivers of the rise in inequality, including skill-biased technology adoption (Acemoglu, 1998), institutional change (Piketty and Saez, 2006) and globalization (Jaumotte and Papageorgiou, 2013; Meschi and Vivarelli, 2009). The rise in labor income inequality coincided with another important development: the increase in job polarization (Figure 2). This process, which is characterized by a decline in middle-skill jobs and a simultaneous rise in low- and high-skill jobs, has been pervasive in developed and developing countries.[2] Goos, Manning and Salomons (2014) find that job polarization in Europe was driven by routine-biased technological change and offshoring, as both forces disproportionately lowered the demand for middle-skilled workers, who tend to be in occupations intensive in routine tasks. At the same time, this process increased the relative demand of high-skilled and low-skilled workers, who tend to be in occupations intensive in non-routine cognitive and manual tasks, respectively. The displacement of workers from middle skill jobs to high- and low-skill jobs may have an impact on the wage distribution, since both the volume of individuals with low earnings and the volume of individuals with high earnings will increase, leading to an overall higher wage inequality. In fact, Goos and Manning (2007) find that job polarization can explain most of the increase in wage inequality experienced by the United Kingdom between the 1970s and 1990s. However, the empirical link between income inequality and job polarization has not been systematically investigated

[1] See, for example, Pickett and Wilkinson (2015); Enamorado et al. (2016); Aaberge et al. (2002); Winkler (2016); Cingano (2014).

[2] See Autor, Katz and Kearney (2006) for the US, Goos, Manning and Salomons (2009) for Europe and World Bank (2016) for developing countries.

yet[3]. Part of the difficulty of such an investigation is due to the lack of panel data that would allow to 'follow' displaced routine workers to their new occupation or out of employment.

Figure 1

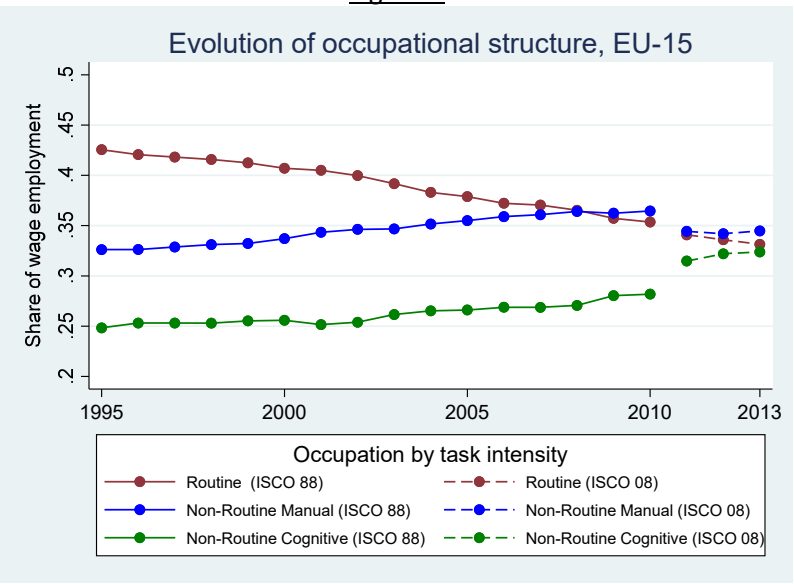


Note: this figure plots the Gini coefficient of wage employees' labor income inequality in several European countries in different years between 1990 and 2015. Each point represents a country-year observation. The red line represents the smoothed locally weighted regression (lowess) with a bandwidth of 0.99. Countries included are Austria, Denmark, France, Germany, Ireland, Italy, the Netherlands, Norway, Poland, the Slovak Republic, Slovenia, Spain, and the United Kingdom. Source of microdata are LIS-harmonized household surveys.

This paper attempts to fill this gap and provides new evidence that the rise in job polarization accounts for a significant share of the increase in earnings inequality observed in a group of European countries over a 20-year time span. To estimate the extent to which labor market polarization accounts for the increase in inequality, and to overcome the lack of availability of panel data, we use a decomposition approach that works with repeated cross sections. Decomposition methods are typically used to disentangle the importance of different factors in accounting for changes in an outcome of interest. These methods have been used extensively in labor economics to understand the drivers of the gender gap, wage inequality, poverty rates and productivity dispersion.[4]

[3] Acemoglu and Autor (2011) propose a novel *theory* to account for the polarizations of occupations observed in the US job market. Their theory, by adding multiple tasks, expands the usual two dimensions labor market approach of the skilled-unskilled relative demand and supply (the Tinbergen race of education and technology).
[4] See, for example, Korkeamäki and Kyyrä (2006); Blau and Kahn (2005); Azevedo et a. (2013), and; Faggio, Salvanes and Van Reenen (2010).

Figure 2



Note: this figure shows the evolution of the occupational structure of wage employment of the aggregated EU-15 countries (excluding Finland, Germany and Sweden) from 1995 to 2013. Occupations are classified according to their task intensity – See Appendix 2 for a detailed description of the classification. The red line indicates the share of wage employment in occupations intensive in routine tasks; the blue line indicates the share of wage employment in occupations intensive in non-routine, manual tasks and the green line indicates the share of wage employment in occupations intensive in non-routine, cognitive tasks. The source is the annual EU Labor Force Survey. Because of a change in the classification of occupations used by the survey (From ISCO 88 to ISCO 08) there is a break in the data after 2011.

Our decomposition method follows Bourguignon and Ferreira (2005). We first estimate an occupational choice model linking individual characteristics to occupations, and a set of Mincer equations for each type of occupation and worker, for the initial and final year of a period running from the early 1990s to 2013. We then carry out a set of counterfactual simulations, whereby we alternatively switch parameters and variables' values in the final year with those of the initial year to assess the importance of each factor. Each counterfactual creates a simulated distribution of wages, which is "in between" the initial and the final distribution. In this way, the full change between the initial and final distribution, in terms of inequality or other indicators, can be decomposed in various parts.

More specifically, the decomposition technique assumes that the earnings distribution changes because of the shifts in the structure of occupations and because of the changes in the returns paid to specific occupations and workers. In addition, the structure of occupation changes because of variations in the characteristics of individual workers (education, age, gender amongst other) and the "utility" returns paid to them. Thus, to isolate the inequality impact of the polarization of the occupations, we proceed in stages.

In a first stage, we assume that the occupation polarization is due only to changes of the parameters ("utility" returns) linking individual characteristics to specific occupations. To do that individual characteristics are kept at the levels they had in the initial year, but the parameters are changed from those of the initial year to those of the final year. This simulation assumes, for example, that college graduates at the end of the period have a different probability than college graduates at the beginning of the period to be in high-skill occupations, but that the share of college graduates in the pool of workers is constant in the two periods. This simulated occupation structure is then used to calculate a simulated ("intermediate-parameters") earnings distribution. This is done by recalculating the earnings for each worker who has switched occupation with the returns to the occupations kept equal to those of the initial year.

In a second stage, we assume that the occupational structure changes only because the characteristics of the labor force are shifting. This exercise produces another simulated ("intermediate-characteristics") earnings distribution and accounts for the share of occupational changes and earnings inequality driven by factors such as changing age and gender composition of the labor force or educational upgrading, under the assumption that the probability of having a certain occupation conditional on individuals' characteristics is the same as in the initial year. In a final stage, the model allows to simulate changes in earnings driven directly by changes in returns to specific occupations and workers ("intermediate-returns").

Comparing the position of these three simulated intermediate earnings distributions – the "parameters", the "characteristics", and the "returns" one – with respect to the initial and the final distribution allows to decompose the total distributional change into parts that are associated to each of these specific factors.

Using repeated cross-sections of household surveys for Germany, Poland and Spain from the early 1990s to 2013, this paper shows that this period was characterized by a significant increase in earnings inequality and polarization of occupations. We find that holding everything else constant, the occupational *parameters* account for at least 44 percent of the increase in the P90/10 earnings ratio over the considered period and group of countries. In terms of the P50/10 earnings ratio, the rise of low-skill occupations and the decline of mid-skill ones account for more than its full observed rise; so the model over-estimates the role of occupational change for the increase of inequality at the bottom part of the distribution. In contrast, the decomposition model shows that simulated occupational shifts explain none or very little of the inequality increase at the top of the income distribution. These findings confirm those

of Goos and Manning (2007) for the United Kingdom. Other important labor market developments that took place during this period such as changes in wage returns to education, changing occupation wage premia, higher female labor force participation, a rapidly aging labor force and human capital upgrading account for a lesser fraction of the rise in inequality.

An interesting feature of the decomposition model is that it allows to generate synthetic-panels and makes possible to follow displaced routine workers. Therefore, we also investigate the profiles of workers who are more likely to lose their routine jobs. We find that many of them move to unemployment or to low-skilled jobs.[5] Only a minority of them move to a non-routine cognitive occupation. This is consistent with the predictions of Autor (2010) for the US. Educational attainment is an important factor at explaining the direction of the transition, as more educated workers are more likely to move to non-routine cognitive occupations, while the least ones tend to move to non-routine manual occupations or unemployment. Workers in routine occupations who change status are, in general, younger than those who do not.

This paper focuses on Europe, and more specifically on Germany, Poland and Spain, for three reasons. First, given the long-standing emphasis on wage and employment protection policies in European economies, this paper shows that inequality may still increase because of changes in the employment structure, even as wages and employment rates are not shifting. Second, the countries selected represent three broad EU regions, namely Northern, Eastern and Southern Europe. These regions have different levels of economic development, trajectories, labor market institutions and policies, thereby any common patterns among them would add robustness and external validity to our results. Third, they have reliable and comparable household surveys over a time span long enough to capture changes in inequality and job polarization. Appendix 4 extends the analysis to four non-EU countries which have witnessed a different evolution of their labor market: Georgia, the Kyrgyz Republic, the Russian Federation and Turkey.

The rest of this paper is structured as follows. Section 2 describes the methodology. Section 3 describes the data sources and provides descriptive statistics. Section 4 presents the results, and Section 5 concludes.

[5] Since the model does not distinguish between unemployment and inactivity, we use the term "unemployed" to group both.

2. METHODOLOGY

2.1 MODEL STRUCTURE

As in the Oaxaca-Blinder approach, the method used here decomposes the observed change in the distribution of earnings between two periods (*t* and *t'*) in separate components. Given that earnings are obtained as the multiplication of the quantity of specific assets, or characteristics, times their related returns, the method accounts for an 'assets' and a 'returns' component in the decomposition of the total change. It does so by generating intermediate earnings distributions. The 'assets' one uses the initial year returns with the final year quantities, and vice versa for the 'returns' one. Following Bourguignon and Ferreira (2005) this section offers a formal presentation of this method.

We define $f^{\tau}(y)$ (where $\tau=t$ or t') as the marginal (density) distribution of the joint distribution $\varphi^{\tau}(y, X)$ where X is a vector of observed individual or household characteristics (such as occupation, education, age, gender), and $g^{\tau}(y|X)$ as the distribution of earnings conditional on X:

$$f^{\tau}(y) = \iint_{C(X)} g^{\tau}(y|X)\ \chi^{\tau}(X)\ dX \qquad (1)$$

Where the summation is over the domain $C(X)$ on which X is defined and $\chi^{\tau}(X)$ is the distribution of all elements of X at time τ. We can then express the change from $f^{t}(y)$ to $f^{t'}(y)$ as a function of the change of the two distributions appearing in equation (1): the distribution of earnings conditional on characteristics X, $g^{\tau}(y|X)$, and the distribution of these characteristics, $\chi^{\tau}(X)$. To do so, we define the following counterfactual experiment:

$$f_g^{\,t\to t'}(y) = \iint_{C(X)} g^{t'}(y|X)\ \chi^{t}(X)\ dX \qquad (2)$$

This distribution represents what would have been observed at time *t* if the distribution of returns conditional on characteristics, $g^{\tau}(y|X)$, had been that observed in time *t'*. Similarly, we can define the following counterfactual experiment:

$$f_\chi^{\,t\to t'}(y) = \iint_{C(X)} g^{t}(y|X)\ \chi^{t'}(X)\ dX \qquad (3)$$

This distribution represents what would have been observed at time *t* if the distribution of characteristics, $\chi^{\tau}(X)$, had been that observed in time *t'*. Note that this distribution could also have been obtained starting from period *t'* and replacing the conditional earnings distribution of that period by the one observed in period *t*. The following identities can be defined:

$$f_g^{\,t\to t'}(y) \equiv f_\chi^{\,t'\to t}(y) \ \text{ and } \ f_\chi^{\,t\to t'}(y) \equiv f_g^{\,t'\to t}(y) \qquad (4)$$

We can thus decompose the observed distributional change $f^{t'}(y) - f^{t}(y)$ into

$$f^{t'}(y) - f^{t}(y) = [\, f_g^{\;t\to t'}(y) \; - f^{t}(y)] \; + [f^{t'}(y) - f_g^{\;t\to t'}(y)] \qquad (5)$$

The first term on the right-hand side of equation (5) describes the way in which the distribution of earnings has changed over time because of the change in the distribution conditional on characteristics *X*. It shows how the same distribution of characteristics -that of period *t*- would have resulted in a different earnings distribution had the conditional distribution $g^{\tau}(y|X)$ been that of period *t'*. To see that the second term is indeed the effect of the change in the distribution of characteristics X that took place between times *t* and *t'*, we can use equation (4) and rewrite the decomposition as:

$$f^{t'}(y) - f^{t}(y) = [\, f_g^{\;t\to t'}(y) \; - f^{t}(y)] \; + [f^{t'}(y) - f_\chi^{\;t'\to t}(y)] \qquad (6)$$

The change in the distribution of earnings over time can be decomposed in a *rewards* and a *characteristics* component, the first and the second term of equation (6), respectively. This is similar to a standard Oaxaca-Blinder decomposition but, instead of referring just to the means of the distributions, this decomposition refers to the full distributions. The only restrictive property of this decomposition method is its path dependence: changing the conditional income distribution from the one observed in *t* to the one observed in *t'* does not have the same effect on the distribution when this is done with the distribution of characteristics *X* observed in *t*, as when *X* is observed in *t'*.

For the purpose of our study, we rely on a parametric representation of the distributions used for defining counterfactuals. Moreover, to better understand the role played by changes in occupational structure, we will model occupational choice rather than treating occupations as a simple individual characteristic. Suppose we can partition the vector X of characteristics in *(O, W)*, where *O* is the occupation of the individual and *W* are other exogenous individual and household characteristics. A general parametric representation of the conditional functions $g^{\tau}(y|O,W)$ and $h^{\tau}(O|W)$ relate the earnings *y* to occupation *O* and characteristics W, on the one hand, and relates occupations *O* to characteristics *W*, on the other hand, according to some predetermined functional form. These relationships may be denoted as follows:

$$y = G[O, W, \varepsilon;\; \Omega_\tau]$$

$$O = H[W, \eta;\; \Phi_\tau]$$

Where Ω_τ and ϕ_τ are sets of parameters: we will call Ω_τ the set of *returns to characteristics* and ϕ_τ the set of *occupation structural parameters* that link individual characteristics *W* to occupation *O* – they represent the conditional correlations of characteristics to occupational choice and can reflect, for instance, the

technological equilibrium matching skills to occupations. Also ε and η are random variables and play a role similar to the residual term in standard regressions as they represent the dispersion of earnings *y* and occupations *O* for given values of *(O,W)* and *W* respectively. They are also assumed to be distributed independently of these characteristics according to density functions $\pi^{\tau}(\ )$ and $\mu^{\tau}(\ )$. With this parametrization, the marginal distribution of earnings in period τ may be written as follows:

$$f^{\tau}(y) = \int_{G(O,W,\varepsilon;\,\Omega_{\tau})=y} \pi^{\tau}(\varepsilon)\, d\varepsilon \ \times \ \left[\int_{H(W,\eta,\Phi_{\tau})=O} \mu^{\tau}(\eta)\, d\eta\right] \Psi^{\tau}(W)\, dO\, dW \quad (7)$$

Counterfactuals may be generated by modifying some or all of the parameters in sets Ω_{τ} and ϕ_{τ} the distributions $\pi^{\tau}(\ )$ and $\mu^{\tau}(\ )$, or the joint distribution of exogenous characteristics $\Psi^{\tau}(W)$. These counterfactuals may be defined as follows:

$$D[\Psi,\pi,\eta;\Omega,\Phi] = \int_{G(O,W,\varepsilon;\,\Omega)=y} \pi(\varepsilon)\, d\varepsilon \ \times \ \left[\int_{H(W,\eta,\Phi)=O} \mu(\eta)\, d\eta\right] \Psi(W)\, dO\, dW \quad (8)$$

Where any of the three distributions π( ), μ( ), Ψ( ) and the two sets of parameters Ω and ϕ can be observed at time *t* or *t'*. For instance, $D[\Psi_t, \pi_t, \mu_t; \Omega_{t'}, \phi_t]$ refers to the distribution of earnings obtained by applying to the population observed at time *t* the set of returns to characteristics of time *t'* while keeping constant the distribution of the random residual term ε and all that is concerned with the variables *O* and *W*. Thus, the contribution of the change in parameters from Ω_t to $\Omega_{t'}$ may be measured by the difference between $D[\Psi_t, \pi_t, \mu_t; \Omega_{t'}, \phi_t]$ and $D[\Psi_t, \pi_t, \mu_t; \Omega_t, \phi_t]$, which is $f^{t}(y)$. But, of course, other decomposition paths may be used. For instance, the comparison may be performed using the population at time *t'* as reference, in which case the contribution of the change in parameters Ω would be given by $D[\Psi_{t'}, \pi_{t'}, \mu_{t'}; \Omega_{t'}, \phi_{t'}] - D[\Psi_{t'}, \pi_{t'}, \mu_{t'}; \Omega_t, \phi_{t'}]$. In order to address this problem, our strategy will be to consider both paths and estimate the "average" contribution of – referred to as a Shapley-value approach.

In our study, the decomposition path of the changes between $f^{t}(y)$ and $f^{t'}(y)$ we will use will be the following:

$$\begin{aligned} f^{t'}(y) - f^{t}(y) = {} & \{f^{t'}(y) - D[\Psi_{t'},\pi_{t'},\eta_{t'};\Omega_{t'},\Phi_t]\} + \\ & \{D[\Psi_t,\pi_{t'},\eta_{t'};\Omega_{t'},\Phi_t] - D[\Psi_{t'},\pi_{t'},\eta_{t'};\Omega_{t'},\Phi_t]\} + \\ & \{D[\Psi_t,\pi_{t'},\eta_{t'};\Omega_t,\Phi_t] - D[\Psi_t,\pi_{t'},\eta_{t'};\Omega_{t'},\Phi_t]\} + \\ & \{D[\Psi_t,\pi_{t'},\eta_{t'};\Omega_t,\Phi_t] - f^{t}(y)\} \quad (9) \end{aligned}$$

Where the first term in curly braces represents the contribution of changes in occupational choice as represented by occupation structural parameters Φ, the second term in curly braces represents the

contribution of exogenous characteristics *W*, the third term in curly braces represents the contribution of returns to characteristics Ω and the remaining term is the residual. In section 2.2 we detail the exact functional forms we use to carry out this decomposition analysis.

Note that when carrying out the counterfactual analysis for ϕ and *W* we are also simulating counterfactual distribution of occupations. In fact, we can define

$$f^{\tau}(O) = \iint_{C(W)} h^{\tau}(O|W)\ \Psi^{\tau}(W)\, dW \tag{10}$$

In a similar way as we defined the earnings distribution in equations (1). Thus, we can simulate the following two occupations distributions:

$$f_h{}^{t\to t'}(O) = \iint_{C(W)} h^{t'}(O|W)\ \Psi^{t}(W)\, dW \tag{11}$$

$$f_W{}^{\mathrm{t}\to t'}(O) = \iint_{C(W)} h^{\mathrm{t}}(O|W)\ \Psi^{t'}(W)\, dW \tag{12}$$

Where (11) represents the distribution of occupations that would have been observed at time *t* if the distribution of occupations conditional on characteristics *W* had been that observed in time *t'*. On the other hand (12) represents the distribution of occupations that would have been observed at time *t* if the distribution of exogenous characteristics *W* had been that observed in time *t'*. Based on the identities of (4), this distribution in (12) is identical to the one that would have been observed at time t' if the distribution of occupations conditional on characteristics *W* had been that observed in time t. Note also that this is the distribution of occupations underlying the counterfactual earnings distribution in the first term in curly braces in equation (9).

2.2 IMPLEMENTING THE DECOMPOSITION

The main objective of our work is to decompose the change of the earnings distribution and, in particular, to assess the impact of the shift of the structure of occupations on the total change of earnings. In the empirical implementation of the method, the discrete earnings distribution can be summarized as a list of earnings such that:

$$F(y_i, \dots, y_n) = F\left(\sum_{k=1}^{4} I_i^k\, y_i^k, \dots, \sum_{k=1}^{4} I_n^k\, y_n^k\right)$$

Where y_i are the earnings of individual *i*, which are made up of the earnings the individual gets in each occupation *k* (y_i^k). The indicator function I_i^k takes a value of 1 if the individual is employed in occupation

k and zero otherwise. We restrict individuals to be employed in only one occupation at the time and there are four possible occupations (see section immediately below).

a. Occupational choices

Individuals are allocated to occupations according to the following model:

$$I_i^k = 1 \; if \; W_i \Phi^k + \varepsilon_i^k > Max(0, W_i \Phi^m + \varepsilon_i^m), k = 1, \dots, K, \forall m \neq k \qquad (13)$$

$$I_i^k = 0 \text{ for all } k = 1, \dots, K \text{ if } Z_i \Phi^k + \varepsilon_i^k \leq 0 \text{ for all } k = 1, \dots, K$$

Where Z_i is a vector of individuals characteristics and Φ^k is a vector of coefficients for each occupation, k; and ε_i^k is a vector of random variables identically and independently distributed across individuals and activities according to the law of extreme values. The intuition behind this model is that individual i chooses occupation k if the utility associated from being in such occupation, $Z_i \Phi^k + \varepsilon_i^k$, is greater than that associated from every other occupation. Without imposing more structure to the model and ignoring the dynamic aspect of occupational choices, this model fails to capture the actual process by which individuals choose an occupation as in the Roy model. Thereby, we argue that instead of estimating occupational choices, we model the conditional distributions of occupations based on individual characteristics such as education, age, gender, region and area. In other words, our model tries to *account* for occupational choices rather than modeling their causal determinants.

We estimate model (13) using a multinomial logit considering four mutually exclusive occupations:

1: Not working
2: Non-routine, manual task intensive occupation
3: Routine task intensive occupation
4: Non-routine, cognitive task intensive occupation

b. Earnings equations

In the next steps, we estimate earnings equations for each occupation k using a log-linear Mincerian model:

$$\ln\left(y_i^k\right) = X_i \Omega^k + \epsilon_i^k \qquad (14)$$

Where X_i is a vector of individual characteristics such as individual characteristics such as education, age, gender, region, area and sector of economic activity; Ω^k is a vector of coefficients and ϵ_i^k is a random

variable assumed to be distributed identically and independently across individuals according to the standard normal distribution. We estimate equation (14) by ordinary least squares.

c. Decomposition approach

We estimate models (13) and (14) for the first and final year, and simulate the impact of occupational changes by substituting the estimated parameters for one year with the parameters of the other year. We then use this hypothetical income to calculate a series of distributional statistics and compare them against those estimated using the actual income data.

- **C.1** Accounting for the impact of occupational changes

To carry out this simulation, we assign the estimated coefficients of equation (13) in year t' to the household survey in year t. To allow individuals to change occupations in the simulation, we need the residual terms ε_i^k of the multinomial logit in equation (13), which are unobserved. Following Inchauste et al. (2014) and Train and Wilson (2008), we draw the residuals from an extreme value distribution in a way that is consistent with observed choices. The simulated earnings for individual i are given by:

$$\widetilde{y_i}^{t\to t',\Phi} = \sum_{k=1}^{4} y_{i,t}^k \widetilde{I_{i,t}^k}(Z_{i,t}, \widetilde{\Phi^{k,t'}}, \varepsilon_i^k)$$

The difference between this simulated distribution and the earnings distribution observed in year t is accounted by the change in the occupational structure, which in turn is due to changes in structural parameters (Φ) of the occupation model. This is the main focus of the decomposition in this paper. Note that the occupation structure can also shift because the exogenous characteristics change. However, while the impact of shifting characteristics can be used to decompose the distribution of occupations, one cannot separate the *direct* impact of shifting characteristics on earnings from the *indirect* impact of shifting characteristics on occupations and from these on earnings. This simultaneous direct and indirect impacts of shifting characteristics are thus jointly accounted in step C.2 below as follows.

- **C.2** Accounting for the impact of changes in relevant exogenous characteristics

In order to study the role played by changes in individual and household characteristics which are *exogenous* to the occupational choice model (such as education, gender and age structure of the population) we perform a reweighting exercise as the one proposed by Bourguignon et al. (2008). First of all, we split exogenous characteristics (Z_i for the occupational choice model and X_i for the earnings equation) into a group of relevant, common characteristics (W_i: education, gender and age), which are

the focus of our exercise, and remaining specific characteristics (R_i^Z for the occupational choice model and R_i^X for the earnings equation). Any sample distributional statistic *G* is a function of the individuals' income ($y_{i,t}$) and their corresponding sample weight ($\omega_{i,t}$). Our exercise consists in modifying the weights of year *t* so the joint distribution of the relevant exogenous characteristics (W_i) match that of year *t'*. In other words, if in year *t'* the average years of schooling are lower than in year *t*, we then modify the weights of year *t* so the sample of that year has the same average years of schooling than that of year *t'*. To do this simultaneously for all the set of relevant exogenous characteristics we use the cross-entropy approach (Wittenberg, 2010). The simulated earnings of individual *i* are given by:

$$\widetilde{y_i}^{t\to t',W} = \sum_{k=1}^{4} \widetilde{I_{i,t}^k}(\, R_{i,t}^Z, \widetilde{W_{i,t'}}, \Gamma^{k,t}, \varepsilon_{i,t}^k)\, \widetilde{y_{i,t}^k}(R_{i,t}^X, \widetilde{W_{i,t'}}, \Omega^{k,t}\, \epsilon_{i,t}^k)$$

As for the parametric simulations mentioned above, this reweighting method is path dependent. We take this into account by also performing the exercise in reverse order – choosing year *t'* as baseline and reweighting the sample of that year by weights that simulate the joint distribution of year *t*. We then report the average difference in the distributional statistics between year *t* and year *t'* using both reweighting orders. Note that this exercise entails changing sample weights but it keeps individual characteristics and parameters unchanged.

- **C.3** Accounting for the impact of changes in occupational wage premia

Up to this point we accounted for the changes of the earnings due to counterfactual simulations where the parameters of the occupation model and the exogenous characteristics of the individuals have been allowed to change, but in which the wages paid to specific occupations have not been modified. This last step in the decomposition is performed by using the wage equation. More specifically, the simulated earnings for individual *i* are given by:

$$\widetilde{y_i}^{t\to t',\Omega} = \sum_{k=1}^{4} I_{i,t}^k\, \widetilde{y_{i,t}^k}(X_{i,t}, \widetilde{\Omega^{k,t'}}, \epsilon_{i,t}^k)$$

Since the results of these simulations (C.1 and C.2) will depend on the year chosen as the baseline, we also run them in reverse order, that is assigning the coefficients of year t to the characteristics of year t'.

- Decomposition order

One of the caveats of our methodology is its path dependency on the decomposition order. That is, the results change whether one performs first the simulation on occupational changes, on wage premia or on the exogenous characteristics. Since the main focus of our analysis is occupational changes, the first

counterfactual simulation we carry out is the one corresponding to occupations (**C.1**). That is, we will attribute to changes in occupations the difference in earnings between the counterfactual simulation and the actual earnings distribution:

$$\Delta^{occ} = y_{i,t} - \widetilde{y}_{\iota}^{\,t\to t',\Gamma}$$

$$\Delta^{occ} = y_{i,t} - \sum_{k=1}^{4} y_{i,t}^{k}\, \widetilde{I_{\iota,t}^{k}}(\, Z_{i,t}, \widetilde{\Gamma^{k,t'}}, \varepsilon_{i}^{k})$$

Preserving the changes resulting from this first simulation, we then move on to the reweighting of exogenous characteristics (**C.3**). We will attribute to the changes in these characteristics the difference between the previous counterfactual simulation and the one corresponding to the reweighting exercise. That is:

$$\Delta^{exc} = \widetilde{y}_{\iota}^{\,t\to t',\Gamma} - \widetilde{y}_{\iota}^{\,t\to t',\Gamma,\mathrm{W}}$$

$$\Delta^{exc} = \sum_{k=1}^{4} y_{i,t}^{k}\, \widetilde{I_{\iota,t}^{k}}\left(\, Z_{i,t}, \widetilde{\Gamma^{k,t'}}, \varepsilon_{i}^{k}\right) - \sum_{k=1}^{4} \widetilde{I_{\iota,t}^{k}}(\, R_{i,t}^{Z}, \widetilde{\mathrm{W}_{\iota,t'}},, \widetilde{\Gamma^{k,t'}}, \varepsilon_{i,t}^{k})\, \widetilde{y_{\iota,t}^{k}}(R_{i,t}^{X}, \widetilde{\mathrm{W}_{\iota,t'}}, \Omega^{k,t}\, \epsilon_{i,t}^{k})$$

We then move to the wage premia simulation (**C.2**). We will attribute to the changes in these premia the difference between the simulation in the previous step and the one corresponding to the wage premia simulation. That is:

$$\Delta^{wpr} = \widetilde{y}_{\iota}^{\,t\to t',\Gamma,\mathrm{W}} - \widetilde{y}_{\iota}^{\,t\to t',\Gamma,\mathrm{W},\Omega}$$

$$\Delta^{wpr} = \sum_{k=1}^{4} \widetilde{I_{\iota,t}^{k}}\left(\, R_{i,t}^{Z}, \widetilde{\mathrm{W}_{\iota,t'}},, \widetilde{\Gamma^{k,t'}}, \varepsilon_{i,t}^{k}\right) \widetilde{y_{\iota,t}^{k}}\left(R_{i,t}^{X}, \widetilde{\mathrm{W}_{\iota,t'}}, \Omega^{k,t}\, \epsilon_{i,t}^{k}\right)$$
$$- \sum_{k=1}^{4} \widetilde{I_{\iota,t}^{k}}\left(\, R_{i,t}^{Z}, \widetilde{\mathrm{W}_{\iota,t'}},, \widetilde{\Gamma^{k,t'}}, \varepsilon_{i,t}^{k}\right) \widetilde{y_{\iota,t}^{k}}\left(R_{i,t}^{X}, \widetilde{\mathrm{W}_{\iota,t'}},, \widetilde{\Omega^{k,s}}\, \epsilon_{i,t}^{k}\right)$$

Lastly, the unexplained part, which can be attributed to changes in the non-common exogenous characteristics (R_{i}^{Z}, R_{i}^{X}) and unobserved variables (ε_{i}^{k}, ϵ_{i}^{k}) corresponds to the difference between the last simulation and the actual earnings in year t' :

$$\Delta^{unx} = \widetilde{y}_{\iota}^{\,t\to t',\Gamma,\mathrm{W},\Omega} - y_{i,t'}$$

Note that, in a repeated cross-section setting, $y_{i,t'}$ is unobserved because individuals are not followed across years. Thus, the unexplained part of changes in earnings will only be possible to estimate for aggregate, anonymous quantiles of the distribution.

3. DATA AND DESCRIPTIVE STATISTICS

3.1 DATA SOURCES

Most of the empirical work done on labor markets in Europe uses the harmonized EU-LFS quarterly labor force survey. This survey represents an invaluable data source for labor economists. However, public access microdata do not include information on earnings of workers. We thus use household surveys harmonized by the Luxembourg Income Study (LIS) center. These surveys include information on both employment characteristics and earnings of individuals, allowing to carry out the decomposition analysis detailed before. LIS harmonizes different household surveys to a common standard to assure comparability. In this work we use the German Socio-Economic Panel editions of 1994 and 2013, the Household Budget Survey of 1992 and the EU-SILC (Statistics and Income Living Conditions) edition of 2013 for Poland, and the Household Budget Survey of 1990 and the EU-SILC edition of 2013 for Spain[6].

The main variables of interest of our analysis are occupations and labor related earnings. With respect to occupations, we classify them into three categories based on their most intensive task, using O*NET task content information: 1) routine task intensive jobs; 2) non-routine, manual task intensive jobs; 3) non-routine, cognitive task intensive jobs. In Appendix 2 we provide a detailed description of how we construct this classification. With respect to labor related earnings, we use annual earnings coming from wage employment. Due to the limitations that household surveys usually have in correctly capturing self-employed income, we exclude self-employed from our analysis. Self-employed represent between 9% and 15% of the total employment in the countries included in our work. Moreover, we restrict our sample to individuals aged between 18 and 64.

Appendix 1 provides more details on the surveys used and the construction of the variables relevant to our analysis.

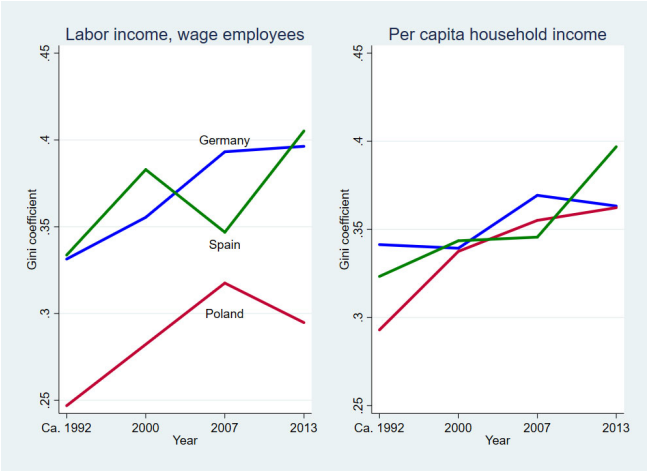
[6] The analysis is also done for Georgia, the Kyrgyz Republic, the Russian Federation and Turkey. See Appendix 4 for data sources and analysis.

3.2 DESCRIPTIVE STATISTICS

The evolution of earnings inequality

During the last 25 years or so, from the beginning of the 1990s to 2013, inequality increased significantly, as shown in figure 3. The Gini coefficient for labor incomes rose by about 8 points for Germany and Spain and 5 points for Poland. The increase is slightly smaller for per capita household total income – 7 points in Poland and Spain and 3 points in Germany – possibly reflecting the equalizing effects of taxes and transfers.

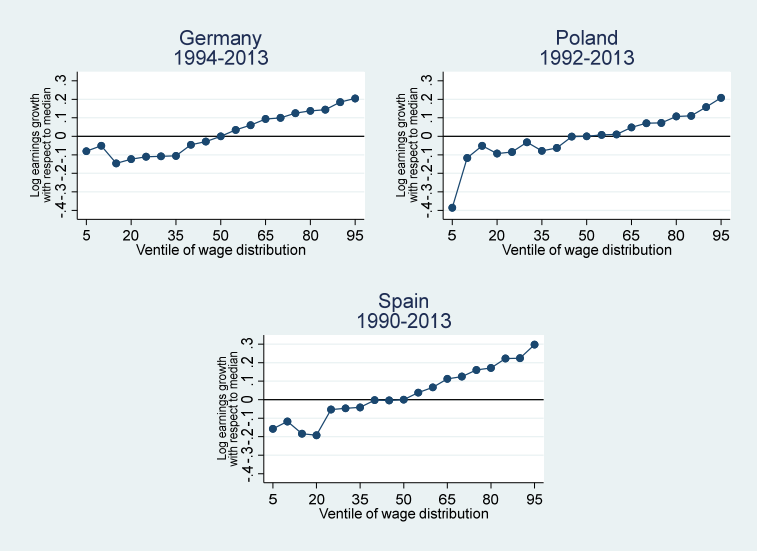
Figure 3 – Evolution of inequality



Source: own elaboration based on Harmonized LIS household surveys. Note: this figure shows the evolution of the Gini coefficient of labor income (only wage employees, excluding self-employed) and of per capita household income (monetary) for three countries. Initial year is 1994 for Germany, 1992 for Poland and 1990 for Spain.

The growth incidence curve depicted in Figure 4 confirms that the increase in inequality was not driven by a particular group. The pattern of income growth during the periods was consistently regressive in Germany, Poland and Spain, with richer percentiles experiencing higher income growth than the poorer ones.

Figure 4 – Change in wages by quantiles of wage distribution



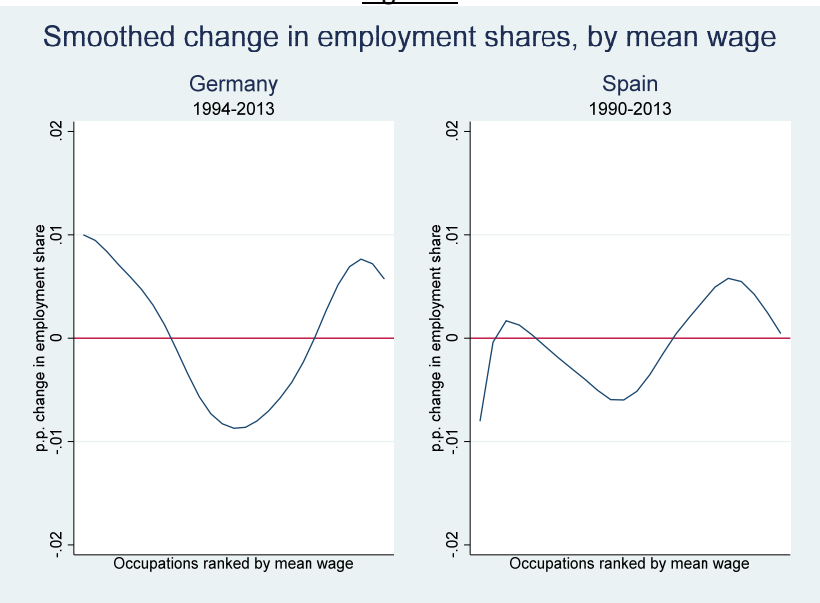
These set of figures plot the log change in wages for the different ventiles of the wage distribution from the initial period to the final period. Changes are expressed with respect to the median, whose change is normalized to zero. Only the wage of wage workers is considered in this analysis.

Occupational changes

Job polarization has been defined as the clustering of jobs at the extremes of the distribution of occupations and, correspondingly, a hollowing of the middle portion of this distribution. This definition entails establishing an ordering of the occupations. This ordering can be obtained according to: (a) the wage or skill level of a specific occupation, or (b) the intensity of certain types of tasks.

Since the correlation between wages and years of schooling (a proxy of skill level) is typically stable over time, some authors (Goos, Manning and Salomons, 2014) have used mean wages as the ranking variable of occupations. Following this approach, we order occupations at the 2-digit level of the ISCO 88 classification according to their mean wage in the initial year, and then plot the change in their shares of total employment in the following 20 or so years. Poland is excluded as the microdata do not allow such level of disaggregation for the initial and final years. The results are shown in Figure 5, which highlights evidence of job polarization for Germany and, to a lesser extent, for Spain.

Figure 5



This figure plots the percentage point change in employment shares from the initial year (1990 for Spain and 1994 for Germany) to the final year (2013) by occupations ranked according to their mean wage in the initial year. The changes are plotted by a locally weighted smoothing regression. Occupations are aggregated to the 2-digit level of the ISCO 88 classification.

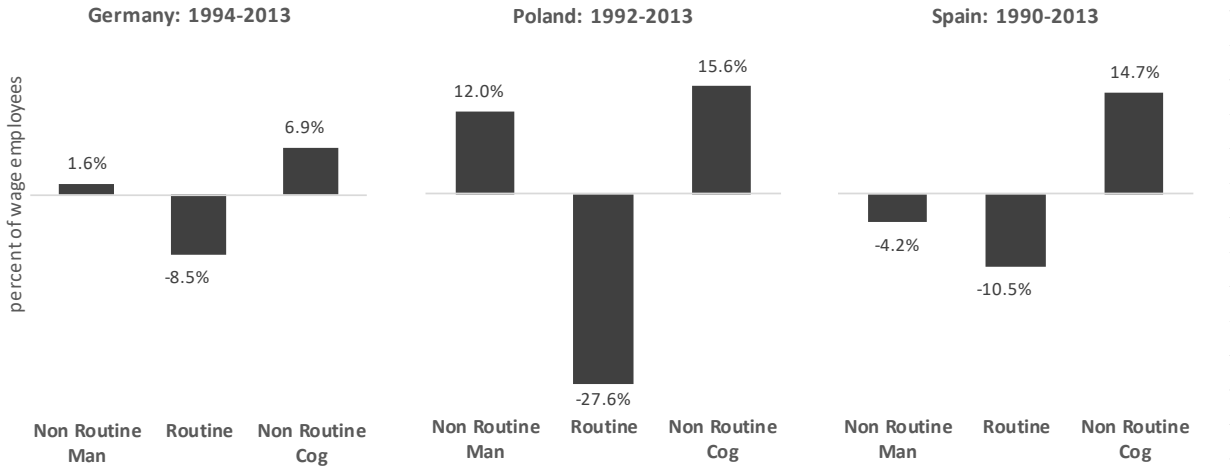
This ordering of occupations is country-specific: an occupation may be paid highly in a country but lowly in another, and vice versa. To determine a ranking of the occupations that is common across countries, we use the approach based on the task content.

According to Acemoglu and Autor's (2011) conceptual framework, occupations can be classified into three categories: occupations relatively intensive in routine tasks, occupations relatively intensive in non-routine cognitive tasks and occupations relatively intensive in non-routine manual tasks. Note that any occupation implies carrying out both routine and non-routine tasks and both cognitive and manual tasks since these are not mutually exclusive; nonetheless, an ordinal index based on the relative intensity of these tasks can be constructed.[7] Occupations intensive in routine tasks tend to be mid-skill occupations, while those intensive in non-routine tasks tend to be on either ends of the skill distribution, cognitive non-routine task intensive ones at the upper end and manual ones at the lower end. Within this framework, job polarization is thus represented by a decrease of the employment share of the routine-intensive occupations and an increase of the non-routine ones. Changes in the employment share of each of the three occupation categories for each country are shown in Figure 6, which confirms that job polarization

[7] For more details of how the index is calculated for the various occupations see Appendix section 2 .

is present in Germany, Poland and up to a certain degree also in Spain. In these three countries de-routinization seems to be a common trend.

Figure 6: Change in employment shares, by country and occupation

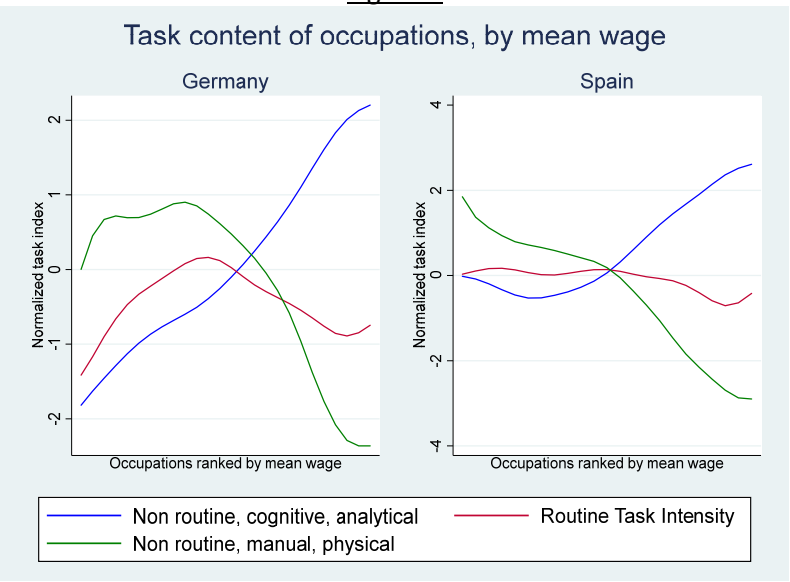


Source: Authors calculations using LIS harmonized household surveys. Note: This figure shows the change, in percentage points, of the share of total employment (wage employees, excluding self-employed) over a period of about twenty years of three occupations categories: non-routine manual task intensive occupations, routine ones, and non-routine cognitive tasks intensive ones. The initial and final years depends on data availability. For more details on the construction of the occupation categories please see the appendix.

The two approaches presented above do not produce the exact same ranking of occupations, however the two rankings are not far apart. This is because, as illustrated in Figure 5, there is a correlation between the intensity of task contents of occupations and wages paid in these occupations. Using the normalized task index of occupations, Figure 7 shows that, for Germany and Spain, the intensity of non-routine cognitive tasks is higher for jobs at the top of the wage distribution while routine task intensity is higher in the middle of the distribution. Low pay jobs are more intensive in non-routine manual occupations.

In sum, no matter what the criteria for ordering of occupations is used – either the average wage paid in each occupation, or the intensity of their task content – there is evidence of a hollowing of the middle of the distribution of occupations, or of its polarization. For the three countries under study, jobs entailing a high degree of routine tasks and paid around the middle of the wage scale are declining.

Figure 7



This figure plots three task content indices of occupations ranked by their mean wage in the initial period as in Figure 3 XXX. The three task content indices (intensity in non-routine, cognitive, analytical tasks; intensity in non-routine, manual, physical tasks; routine task intensity) are normalized to their economy-wide means. The indices are plotted by a locally weighted smoothing regression. Occupations are aggregated to the 2-digit level of the ISCO 88 classification

4. RESULTS

4.1 ACCOUNTING FOR THE INCREASE IN EARNINGS INEQUALITY

In all three countries, changes in the occupational structure seem to be a major factor – vis-à-vis changes of assets and of returns – behind the increasing inequality observed during the period, particularly at the bottom of the wage distribution. This main finding of the decomposition exercise is in line with what Goos and Manning (2007) found for the United Kingdom, while it differs slightly from the result of Acemoglu and Autor (2011) who showed that, for the US, polarization in occupations was related to increased wages both at the bottom and at the top of the distribution.

A scenario where the characteristics/assets of individuals and their relative returns remain the same but where the occupational structure is shifting would register an increase of the P90/10 ratio (the ratio between the average earnings of the 90th and 10th) equivalent to 44 percent of the total increase in Germany, 91 percent in Poland, and 312 percent in Spain, as shown in column (1) of Table 1. So, indeed, occupational changes have a significant distributional impact on earnings. More in detail, the hollowing of the middle jobs seems to generate more inequality pressure at the bottom half of the earning range. This scenario accounts for 211, 184, and 772 percent of the total change of the P50/10 ratio in Germany,

Poland and Spain, respectively. Conversely, occupational shifts at the top half of the distribution account for much smaller shares (see the P90/50 ratios in column (1) of table 1).

There are two reasons for these differential polarization impact on the top and bottom part of the earnings distribution. The first is technical and has to do with the fact that the occupational change simulated with a change of the parameters is not equivalent to the full change, and the second is because most of the displaced middle skilled workers tend to move down in the distribution. A more detailed explanation is provided below.

Table 1 – Decomposition results

Country	Inequality measure, labor income	Change 1990s-2013	Percentage explained by: **Occup.** (C1, section 2.2) (1)	**Charact.** (C2, section 2.2) (2)	**Returns** (C3, section 2.2) (3)
Germany	Gini coefficient	0.058	-1	20	30
	P90/10	3.148	44	14	185
	P90/50	0.404	-6	13	21
	P50/10	0.311	211	23	818
Poland	Gini coefficient	0.078	42	-4	27
	P90/10	0.938	91	-10	59
	P90/50	0.303	26	-9	30
	P50/10	0.208	184	-4	79
Spain	Gini coefficient	0.068	92	-21	51
	P90/10	2.672	312	-10	158
	P90/50	0.466	39	-19	60
	P50/10	0.442	772	86	114

Source: Authors' calculations based on LIS harmonized surveys. Note: the percentages shown in columns (1), (2) and (3) are derived using the distributions of earnings simulated according to the methods described in the respective sub-sections of section 2.2. The change in the inequality indicators corresponds to the ones estimated in the working sample and may differ from those of the whole sample of wage workers. Observations with missing data in some of the relevant variables are excluded. The change in the Gini coefficient of labor income for the whole set of wage workers (including those with missing information) was 0.065 in Germany, 0.060 in Poland, and 0.071 in Spain

The education, age and gender structure of the labor force underwent significant transformations between the early 1990s and 2013. For example, in our sample, the share of individuals with tertiary education expanded substantially in Poland, where it went from 7% to 20%, and in Spain where it went from 12% to 32%; but not so much in Germany where it increased from around 22% to 27%. Another scenario, where only assets/characteristics of individuals are changing, is therefore simulated to account for the distributional impact on earnings of these transformations. Note that, when characteristics are varying, there are two effects on earnings. One effect is direct; for example, a larger number of well-

educated workers translates into a larger group of people earning higher wages[8]. The second effect is indirect and operates through occupational shift. A large share of the increased pool of well-educated people, as in the previous example, will find jobs in non-routine occupations and benefit from both an education and an occupation premium. These two effects cannot be disentangled and their joint impact is shown in column (2) of table 1. In short, changes in individuals' characteristics, in isolation, account for a minor share of the overall changes in inequality.

A final scenario is simulated to assess the importance of changes in the returns of occupations (and individual characteristics). Changes of inequality driven by the returns account for a non-trivial fraction of the overall inequality increase, but this fraction is smaller than that accounted for the polarization of occupations. According to column (3) of Table 1, changes in returns to characteristics explain between 30 to 50 percent of the increase in the Gini coefficient. Comparing the results of the P90/50 with those of P50/10 illustrates that changes in inequality were driven by both a strong increase in relative earnings at the top, and a stronger, disproportionate decline in relative earnings at the bottom of the distribution of earnings. To a large extent, this was driven by the increase in the returns to education among non-routine, cognitive task intensive occupations, more prevalent at the top of the wage distribution.[9]

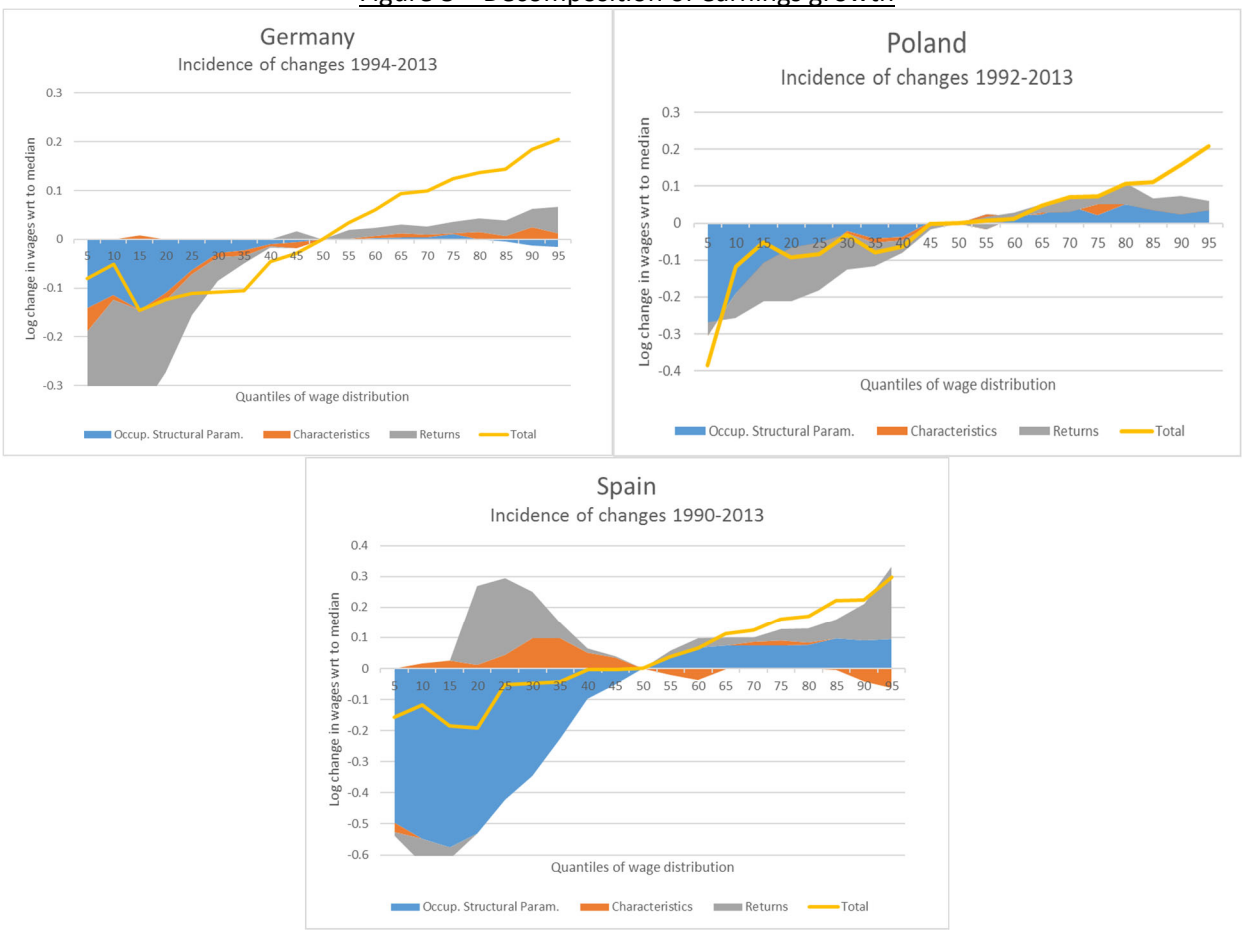
The summary results of table 1 can be extended to the full distribution by plotting a series of growth incidence curves (GIC), as done in Figure 8. For each percentile – ordered from the poorest 1 percent to the richest 1 percent – these curves illustrate the change of earnings from the initial year to the final year, i.e. the full observed change, as well as the change that is due to a single component of the decomposition. In all three countries, the 'full' GIC (the yellow line in the figure) is upward sloped indicating that the earnings distribution underwent a regressive adjustment. In addition, changes in occupations, individual characteristics and wage returns over-account for the relative decline of earnings at the bottom of the distribution. This means that factors not included in the model (like, for instance, within-firm or within-sector changes in returns to individual characteristics) partially offset the role of these changes in driving inequality at the bottom of the distribution. In particular, the results highlight the magnitude of occupational changes as a driver of earnings inequality at the bottom of the earnings distribution. In contrast, the simulations account for almost all the increase in relative earnings at the top of the distribution in Poland and Spain. In Germany, the simulations account for only a small share of the increase

[8] In this scenario, returns are unchanged.

[9] According to the estimates of the Mincer equations in Tables A.2 and A.3 in the appendix, the returns to college education in those occupations increased by between 8 and 23 log points in Spain and Germany over the period under analysis, while they remained stable or slightly decreased in Poland.

of inequality at the top, which implies that other unobserved factors were more important in driving this change.

Figure 8 – Decomposition of earnings growth



Source: Authors' calculations based on LIS harmonized surveys.

4.2 UNDERSTANDING THE ROLE OF OCCUPATIONAL CHANGE

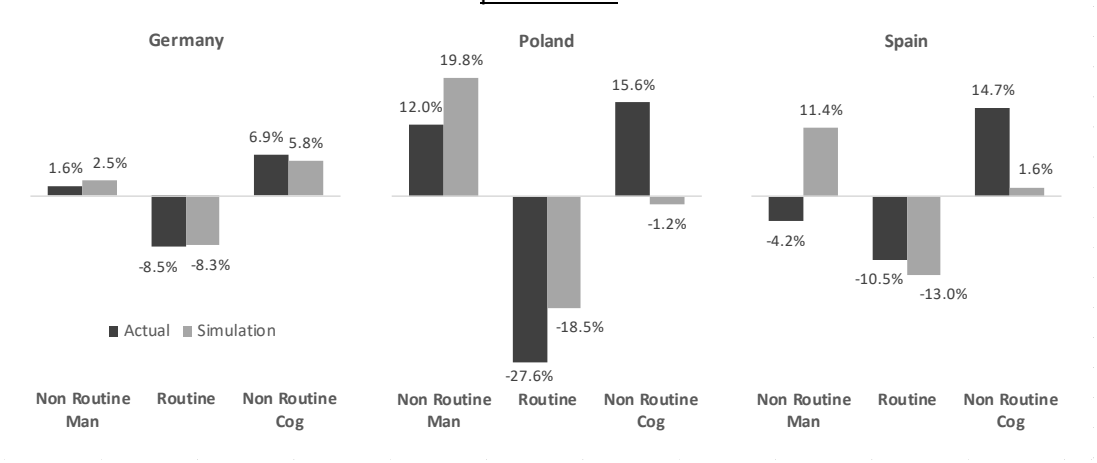
In the decomposition model used here, simulating the transformation of the structure of occupations is done in two steps. In the first, the characteristics of individuals are not changed while the parameters linking these to occupations are modified; in the second, the characteristics are altered while the parameters are held fixed[10]. Together these two steps account for the full change in the distribution of occupations. However, while the effect on the distribution of earnings of the partial occupational change due to the first step can be identified and measured, the effect of the second step cannot be separated from the direct impact (on earnings) of changing characteristics. In any case, by showing the two partial

[10] Formally, the first step generates the distribution of occupations defined in equation 11 while the second step generates the distribution defined in equation 12.

occupational changes, this section aims at clarifying the drivers of occupational change and thus link these drivers to the inequality of earnings.

The first partial occupational shift is depicted in Figure 9 alongside the observed full shift. In Germany, the simulation and the observed data look very similar. This means that changes in structural parameters account for the bulk of the occupational change. In the case of Poland and Spain, the decline in routine intensive occupations is closely approximated, but the growth in non-routine, manual task intensive jobs is overestimated. Interestingly, the shift in parameters plays a negligible role in accounting for the observed growth of jobs intensive in non-routine cognitive tasks. In these two countries, the impact of the parameters-related polarization of jobs on earnings must be taken with a grain of salt, given that the polarization simulated with just the shift in parameters (that reported in column (1) of table 1) does not correspond so closely with the observed polarization.

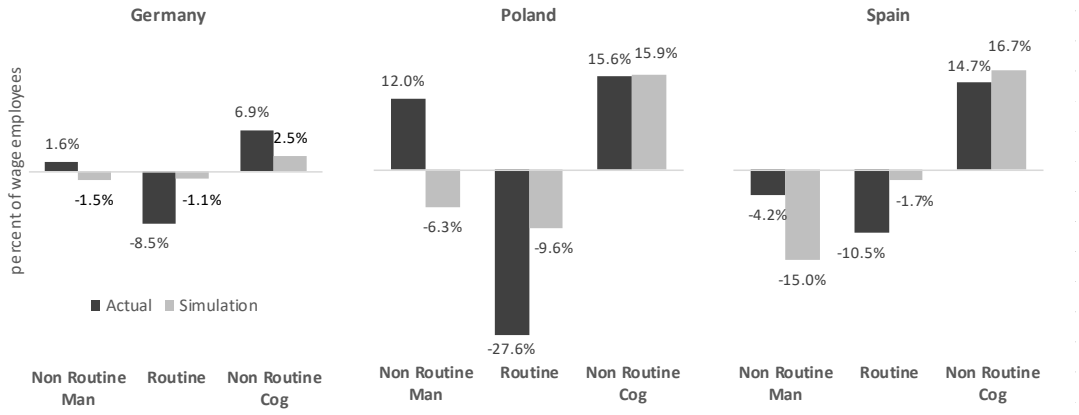
Figure 9 – Simulation of occupational change, holding constant all variables except occupation structural parameters



Note: this figure compares the observed change in occupational structure (dark colors) with the one in the counterfactual scenario (light colors) where the occupation structural parameters in 2013 are the same as in the 1990s.

Figure 10 illustrates the second partial occupational change, the one accounted for by changes in individual characteristics. As expected, in the case of Germany, the simulated occupational change is very small. In the case of Poland and Spain, this second simulation, by underestimating the growth in non-routine, manual task intensive skills, offsets the overestimation of the first (parameters-related) simulation. At the same time, this second simulation matches quite well the increase in the share of non-routine, cognitive task intensive jobs.

Figure 10 – Simulation of occupational change, holding constant all variables except individual characteristics (education, age, gender structure)



Note: this figure compares the observed change in occupational structure (dark colors) with the one in the counterfactual scenario (light colors) where the average individual characteristics (education, age, gender structure) in 2013 are the same as in the 1990s.

In sum, these simulations highlight that the shift of the parameters linking given characteristics to given occupations is a prominent driver for the reduction of routine workers for all the countries. In other words, the simulations suggest that no matter how much the characteristics of the working population had changed, if these parameters had remained unchanged across time, there would have been less hollowing out of the middle occupations.

Conversely, the expansion of the bottom and the top portions of the distribution of occupation seems driven by different factors in the different countries. In Germany, shifts of parameters are still playing the main role, but in Poland and Spain education upgrading and changes in other characteristics mattered for growth of non-routine cognitive jobs at the top; while a combination of both explains the increase, at the bottom, for non-routine manual jobs.

Given these results, and specifically that the occupational parameters are the main driver of the drop of the middle routine intensive occupations, we can finally use the model to generate a pseudo-panel and thus, ultimately, to tell us where the displaced routine workers go and what happens to their earnings.

A brief explanation is helpful here. As shown above, for all the three countries, a shift of the parameters of the occupation model generates a counterfactual where there is no de-routinization or, equivalently, a counterfactual where more individuals are employed in middle routine jobs. But in the real world, where the parameters did change, these counterfactual routine workers end up in different occupations. This procedure allows 'following' these individuals, from their counterfactual routine occupations to their observed non-routine occupations.

Two main findings are highlighted by this exercise and illustrated in tables 2 and 3 below. Firstly, most displaced routine workers end up either not employed or in non-routine manual jobs and only a few can move up to non-routine cognitive jobs. In addition, disproportionally more women, and younger and less educated people fall down in the distribution of occupations when de-routinization hits. Both these shifts translate into large losses in terms of earnings, as shown in figure 11, and ultimately explain the rising inequality of the distribution of earnings.

More in details, Table 2 describes for the three countries, where each individual of a group of 100 counterfactual routine workers actually ends up in the 2013 observed data. Consider, for example, the case of Germany: 15 workers are now not employed, 8 of them are in non-routine manual jobs, and only 6 are in non-routine cognitive jobs; and up to 71 remain in routine jobs. For all the countries the movement into non-routine manual intensive jobs is larger than the movement into non-routine cognitive task intensive jobs, which is consistent with the predictions of Autor (2010) for the US.

Table 2 – Flows out of routine-task intensive occupations

	Observed occupations				Counter-factual
	Not employed	Non-routine, manual	Non-routine, cognitive	Routine	All in routine
	Movers			Stayers	
Germany	15	8	6	71	100
Poland	10	15	9	66	100
Spain	24	11	3	62	100

Source: Authors' calculations. Note: This table presents the observed occupations of 100 individuals who would be, in the counterfactual scenario, employed in routine task intensive occupations. For example, the first column shows individuals who are 'not employed' but that in the counterfactual simulation (obtained by shifting the parameters of the occupational model from their current 2013 values to the values they had at the beginning of the 1990s) would have been employed in routine task intensive occupations. The second column shows individuals who are employed in 'non-routine, manual task intensive occupations' but would have been employed in routine task intensive occupations in the counterfactual scenario, and so on.

The characteristics of displaced routine workers who end up jobless or in lower occupations are quite different from the characteristics of those who remain in routine jobs or of those who move up to better paid cognitive jobs. In Germany, as shown in table 3, the share of female workers is higher among the displaced routine workers who move into non-routine manual and cognitive occupations, while the share of males is higher among those who become unemployed. In contrast, in Poland and Spain displaced workers are more likely to be female regardless of their new status when compared to those who kept their routine occupation in the simulation. In Germany, Poland and Spain, routine workers who move into

non-routine cognitive occupations have more years of education, while those who move into non-routine manual occupations have roughly the same level of education than those who stay in routine jobs. Those who move from routine occupations into unemployment have a substantially lower educational attainment. For instance, a routine worker who does not change occupations in Germany has 12 years of education, while one that moves into unemployment has only 8.9 years of education. In Germany and Spain, routine workers who move into any other status including unemployment are younger than those who stay in routine jobs. The same holds in Poland, although the share of routine workers who move into non-routine cognitive positions tends to be older than the rest.

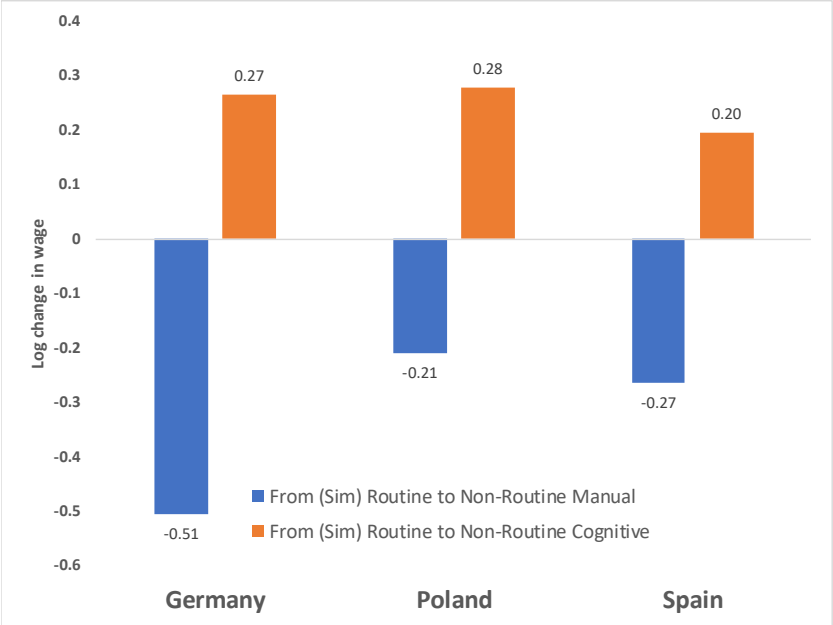
Table 3 – Flows out of routine-task intensive occupations, individual characteristics

Country	Characteristic	Observed occupations			
		Not employed	Non-routine, manual	Non-routine, cognitive	Routine
		Movers			Stayers
Germany	Percentage of women	37.8	83.6	56.6	48.6
	Years of schooling	8.9	11.1	13.3	12.0
	Age	39.3	42.1	40.5	43.9
Poland	Percentage of women	64.7	49.1	41.0	25.5
	Years of schooling	9.8	12.2	14.2	11.8
	Age	35.4	39.0	39.0	40.2
Spain	Percentage of women	38.0	45.1	33.3	37.8
	Years of schooling	11.8	12.0	14.4	12.1
	Age	38.9	39.3	41.9	39.9

Note: this table shows a set of descriptive characteristics of workers grouped according to their observed occupation in contrast with their simulated routine occupation. So, for example, the first column shows individuals who in the simulation would have been employed in routine occupations but that, because of the change in the parameters of the occupation model, are now not employed. The other columns are interpreted in the same way.

Occupational changes have significant implications for earnings. Figure 11 shows the simulated change in the wage of workers who in the simulation would have been in routine jobs but then move into non-routine manual or cognitive occupations. As expected, the change in earnings is significant. For example, a displaced German routine worker moving to a non-routine manual occupation is expected to experience a close to 50 percent decline in wages; in Poland and Spain the corresponding losses would be of about 20 and 30 percent, respectively. Clearly, gains in wages are shown for the fewer displaced workers who move up in the occupation distribution.

Figure 11



Note: this figure shows the difference between the actual wages in the observed occupations of individuals who, in the simulation, would have earned wages in routine occupations. The blue bars show the loss in terms of lower wages (log change in wage can be interpreted as a percentage difference) for the "movers" in non-routine, manual task intensive. The orange bars show the gains for the "movers" in non-routine, cognitive task intensive jobs. Note that, as shown in table 2, fewer workers move to these higher paid occupations.

5. CONCLUSIONS

This paper provides new evidence that the rise in job polarization accounts for a significant share of the increase in earnings inequality observed in a group of European countries over a 20-year time span. Using decomposition techniques, it shows that changes in occupation structural parameters -which explain a great deal of the polarization in occupations- account for a part of the observed increase in earnings inequality that is larger than that related to changes in the returns or individual characteristics, such as educational attainment. While these countries experienced a rise in inequality both at the top and the bottom of the earnings distribution, changes in occupation structural parameters mostly explain the latter.

The fact that the match between individual characteristics and occupations, a correlation which is captured by parameters of the occupational model, has changed considerably in the last twenty years may be due to different reasons – technological change being a leading explanation. De-routinization and new technologies may have changed the types of tasks performed by middle and high paying occupations,

and thus may have made the skills of workers holding a high school diploma or less no longer sufficient to be employed in these occupations.

These results show that labor market inequalities can emerge even when wage and employment protection legislation limit the extent of wage flexibility or job losses. Given that the process of job polarization appears to be ongoing, these results raise concerns about the future path of inequality in Europe.

Finally, it is important to mention some caveats of the methodology used in this paper.[11] First, decomposition techniques do not allow to show general equilibrium effects as they just generate statistical counterfactuals. For example, if the occupational structure changed, this would likely change the wage returns as well, something that our model cannot capture. Second, while our results help uncover a robust connection between job polarization and earnings inequality, they do not address what are the causes behind it. Thereby, understanding the role of technological change as a causal factor of job polarization and earnings inequality in a general equilibrium setting remains an important area of future research.

[11] See Fortin, Lemieux and Firpo (2010) for a review of decomposition methods in economics.

REFERENCES

Aaberge, R., Björklund, A., Jäntti, M., Palme, M., Pedersen, P. J., Smith, N., & Wennemo, T. (2002). Income inequality and income mobility in the Scandinavian countries compared to the United States. Review of Income and Wealth, 48(4), 443-469.

Apella, Ignacio and Gonzalo Zunino (2017) "Cambio tecnológico y mercado de trabajo en Argentina y Uruguay. Un análisis desde el enfoque de tareas", Serie de informes técnicos del Banco Mundial en Argentina, Paraguay y Uruguay n. 11, World Bank, Montevideo, Uruguay.

Autor, David, H., Lawrence F. Katz, and Melissa S. Kearney. "The polarization of the US labor market." *American economic review* 96.2 (2006): 189-194.

Acemoglu, Daron. "Why do new technologies complement skills? Directed technical change and wage inequality." *The Quarterly Journal of Economics* 113.4 (1998): 1055-1089.

Goos, Maarten, Alan Manning, and Anna Salomons. "Job polarization in Europe." *American economic review* 99.2 (2009): 58-63.

Jaumotte, F., Lall, S., & Papageorgiou, C. (2013). Rising income inequality: technology, or trade and financial globalization?. *IMF Economic Review*, *61*(2), 271-309.

Korkeamäki, Ossi, and Tomi Kyyrä. "A gender wage gap decomposition for matched employer–employee data." *Labour Economics* 13.5 (2006): 611-638.

Blau, Francine D., and Lawrence M. Kahn. "Do cognitive test scores explain higher US wage inequality?." *Review of Economics and statistics* 87.1 (2005): 184-193.

Azevedo, J. P., Azevedo, J. P., Inchauste, G., Olivieri, S., Saavedra, J., & Winkler, H. (2013). Is labor income responsible for poverty reduction? A decomposition approach.

Giulia Faggio, Kjell G. Salvanes, John Van Reenen; The evolution of inequality in productivity and wages: panel data evidence, *Industrial and Corporate Change*, Volume 19, Issue 6, 1 December 2010, Pages 1919–1951, https://doi.org/10.1093/icc/dtq058

Meschi, Elena, and Marco Vivarelli. "Trade and income inequality in developing countries." *World development* 37.2 (2009): 287-302.

Piketty, T., & Saez, E. (2006). The evolution of top incomes: a historical and international perspective. *American economic review*, *96*(2), 200-205.

Autor, David. "The polarization of job opportunities in the US labor market: Implications for employment and earnings." Center for American Progress and The Hamilton Project (2010).

Autor, David and Daron Acemoglu (2011) "Skills, Tasks and Technologies: Implications for Employment and Earnings" in Handbook of Labor Economics, Vol. 4, Elsevier B.V.

Autor, David and David Dorn (2013) "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Job Market" in *American Economic Review*, vol. 103 (6): 1553-1597

Autor, David; Frank Levy and Richard J. Murmane (2003) "The Skill Content of Recent Technological Change: An Empirical Exploration" in *Quarterly Journal of Economics*, vol. 116 (4): 1279-1333

Bourguignon, François and Francisco H.G. Ferreira (2005). "Decomposing changes in the distribution of household incomes: methodological aspects" in Bourguignon, F; F. Ferreira and N.Lustig (eds.) The microeconomics of income distribution dynamics in East Asia and Latin America, World Bank, Washington, DC.

Bourguignon, François; Francisco H.G. Ferreira, and Phillippe G. Leite (2008) "Beyond Oaxaca–Blinder: Accounting for differences in household income distributions." In *The Journal of Economic Inequality* vol. 6(2): 117-148.

Cingano, F. (2014), "Trends in Income Inequality and its Impact on Economic Growth", OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing. http://dx.doi.org/10.1787/5jxrjncwxv6j-en
Cingano, Federico. "Trends in income inequality and its impact on economic growth." (2014).

Di Carlo, Emanuele; Salvatore Lo Bello; Sebastian Monroy-Taboada; Ana Maria Oviedo; Maria Laura Sanchez Puerta and Indhira Santos (2016) "The Skill Content of Occupations across Low and Middle Income Countries: Evidence from Harmonized Data", IZA Discussion Paper Series No. 10224, Bonn.

Enamorado, T., López-Calva, L. F., Rodríguez-Castelán, C., & Winkler, H. (2016). Income inequality and violent crime: Evidence from Mexico's drug war. *Journal of Development Economics*, *120*, 128-143.

Goos, Marten and Alan Manning (2007) "Lousy and Lovely Jobs: the Rising Polarization of Work in Britain" in *Review of Economics and Statistics*, vol. 89(1): 118-133

Goos, Marten; Alan Manning and Anna Salomons (2009) "Job Polarization in Europe" in *American Economic Review: Papers and Proceedings* vol. 99(2): 58-63

Goos, Marten; Alan Manning and Anna Salomons (2014) "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring" in *American Economic Review*, vol. 104(8): 2509-2526

Hardy, Wojciciech; Roma Keister and Piotr Lewandowski (2016) "Technology or Upskilling? Trends in Task Composition of Jobs in Central and Eastern Europe", IBS Working Paper Series 1/2016, Warsaw

IBS-Institute for Structural Research (2015) Occupation classification crosswalks: data an codes.

Inchauste, Gabriela; João Pedro Azevedo; Boniface Essama-Nssah; Sergio Olivieri; Trang Van Nguyen; Jaime Saavedra-Chanduvi, and Hernan Winkler (2014) Understanding Changes in Poverty. World Bank, Washington, DC

Maloney, William F. and Carlos Molina (2016) "Are Automation and Trade Polarizing Developing Country Labor Markets, Too?" World Bank Policy Research Working Paper 7922, World Bank, Washington, DC.

OECD (2017) "UNDERSTANDING THE SOCIO-ECONOMIC DIVIDE IN EUROPE"

Pickett, Kate E., and Richard G. Wilkinson. "Income inequality and health: a causal review." Social science & medicine 128 (2015): 316-326.

Piketty, T., & Saez, E. (2014). Inequality in the long run. Science, 344(6186), 838-843.

Train, Kenneth and Wesley W. Wilson (2008) "Estimation on stated-preference experiments constructed from revealed-preference choices" in *Transportation Research Part B: Methodological*, vol. 42(3): 191-203.

Winkler, Hernan. "The Effect of Income Inequality on Political Polarization." Mimeo.

Wittenberg, Martin (2010) "An introduction to maximum entropy and minimum cross-entropy using Stata" in *The Stata Journal*, vol. 10(3): 315-330.

World Development Report (2016) Digital Dividends, World Bank, Washington, DC

Appendix 1: Data and variable definitions

A.1 Sources

	Baseline			Final		
	Year	Survey and observations	Harmonization	Year	Survey and observations	Harmonization
Germany	1994	German Socio-Economic Panel *17812 obs.*	LIS	2013	German Socio-Economic Panel *41657 obs.*	LIS
Poland	1992	Household Budget Survey *18807 obs.*	LIS	2013	EU-SILC *102780 obs.*	LIS
Spain	1990	Household Budget Survey *72119 obs.*	LIS	2013	EU-SILC *31622 obs.*	LIS

A.2 Variables

Employment status: three categories: regular wage employment, self-employment, out of employment (out of labor force and unemployed)

Occupation: ISCO88 or ISCO08 occupation code for primary job grouped into three categories. Check Appendix section 2 for full description of occupation categories

Wage: annual labor market incomes expressed in local currency units, constant prices of the final year.

Education: maximum level of education attained, ISCED three categories (low for primary or no education; medium for secondary education; high for tertiary education or more)

Appendix 2: Construction of occupation categories

Grouping occupations according to their task content implies making a decision on which task dimension to prioritize over others. As the potential number of tasks by which an occupation can be characterized is very large, we rely on pre-built task content indices by IBS (2015) which originate from O*NET[12] and follow Acemoglu and Autor (2011). There are six task content indices: i) non-routine, cognitive, analytical; ii) non-routine, cognitive, personal; iii) routine, cognitive; iv) routine, manual; v) non-routine, manual, physical; vi) non-routine, manual, personal. Additionally, indices iii) and iv) can be combined into a routine task intensity (RTI) index based on Autor, Levy and Murnane (2003). Each occupation at the ISCO 88 4-digit level (unit group titles) has a value in every task content index. For the purpose of this work we aggregate occupations at the ISCO 88 2-digit level (sub-major group titles) by taking a simple average of the indices of the unit groups included in the corresponding sub-major group. This is done in order to have a common aggregation level across countries since not all the surveys record occupations at the 4-digit level.

For the ISCO 88 classification we have in total 27 sub-major occupation groups and we will split them into three groups according to the following algorithm. First of all, we rank the 27 groups according to the RTI index and we create our first category -occupations intensive in routine tasks- by choosing the top third (9 groups) which have the highest value for the index. We are left with 18 sub-major occupation groups which we will split in two according to their value of the non-routine, cognitive, analytical index[13]. The top half which has the highest values of the non-routine, cognitive, analytical index are classified into our second category -occupations intensive in non-routine, cognitive tasks- and the remaining bottom half is classified into our third category -occupations intensive in non-routine, manual tasks. Table A.1 presents a statistical summary of the categories. Note that our categorization of occupations is based on the relative intensity of some tasks: non-routine, manual, physical task content is high in both the first and third groups, but the first group has also high routine task intensity whereas the third group has a low value for routine tasks. In this sense, the first group in relatively more routine-intensive than the third group, which is relatively more intensive in non-routine, manual, physical tasks.

[12] A caveat of using O*NET data is the implicit assumption that the task content of each occupation is the same across all the countries – and, in particular, that is the one of each occupation in the United States, for which O*NET was specifically constructed. There is evidence that the types of tasks performed by the same occupation (e.g. an office clerk) differ across countries (Di Carlo et al., 2016).

[13] Results practically do not change if we use the non-routine, cognitive, personal index.

Table A.1 – Summary statistics of occupation categories

	Occupations intensive in routine tasks	**Occupations intensive in non-routine, cognitive tasks**	**Occupations intensive in non-routine, manual tasks**
RTI index	1.930	0.188	0.079
O*NET task content indices (average)			
Routine, manual	9.308	6.336	8.191
Routine, cognitive	9.929	8.973	8.495
Non-routine, cognitive, personal	8.538	10.635	8.734
Non-routine, cognitive, analytical	8.651	11.105	8.120
Non-routine, manual, physical	10.867	7.952	11.309
Non-routine, manual, personal	2.905	3.513	3.037
Examples (ISCO 88 sub-major groups)	Office clerks (41), Metal, machinery and related trades workers (72), Stationary-plan and related operators (81)	Corporate managers (12), Physical, mathematical and engineering science professionals (21), Life science and health associate professionals (32)	Personal and protective services workers (51), Sales and services elementary occupations (91), Drivers and mobile-plant operators (83)

Source: own elaboration based on IBS (2015)

This classification is possible when occupation data are available at the ISCO 2-digit level. For Poland (1992 survey) these data are only available at the ISCO 1-digit level (major groups). In this case the first occupation category (occupations intensive in routine tasks) comprises ISCO major groups 4, 7 and 8; the second occupation category (occupations intensive in non-routine, cognitive tasks) comprises ISCO major groups 1, 2 and 3; the third occupation category comprises ISCO major groups 5, 6 and 9.

Appendix 3: Results of the Mincer Equation

Table A.2 – Point estimates of tertiary-secondary education wage premium, household heads

	Initial year			Final year			Difference year		
	NR, M	R	NR, C	NR, M	R	NR, C	NR, M	R	NR, C
Germany	0.239	0.196	0.080	0.187	0.151	0.315	-0.052	-0.045	0.236
Poland	-0.050	0.336	0.382	0.202	0.181	0.277	0.252	-0.155	-0.105
Spain	0.248	0.090	0.152	0.275	0.057	0.261	0.027	-0.033	0.109

This table shows the point estimate of the tertiary-secondary education wage premium (i.e. the log difference of the return to tertiary education and the return to secondary education) for household heads in each occupation category: Non-routine, Manual task intensive occupations (NR, M), Routine task intensive occupations (R) and Non-routine, Cognitive task intensive occupations (NR, C). Estimates come from a standard Mincer equation.

Table A.3 – Point estimates of tertiary-secondary education wage premium, spouses

	Initial year			Final year			Difference year		
	NR, M	R	NR, C	NR, M	R	NR, C	NR, M	R	NR, C
Germany	0.212	0.021	0.318	0.103	0.235	0.479	-0.110	0.214	0.161
Poland	-0.166	0.320	0.250	0.144	0.142	0.253	0.310	-0.178	0.003
Spain	0.004	0.017	0.439	0.159	0.275	0.523	0.155	0.258	0.084

This table shows the point estimate of the tertiary-secondary education wage premium (i.e. the log difference of the return to tertiary education and the return to secondary education) for spouses in each occupation category: Non-routine, Manual task intensive occupations (NR, M), Routine task intensive occupations (R) and Non-routine, Cognitive task intensive occupations (NR, C). Estimates come from a standard Mincer equation.

Appendix 4: Post-Soviet countries and Turkey

In this section we replicate the analysis for a set of four non-EU countries belonging to Europe and Central Asia. These are Georgia, the Kyrgyz Republic, the Russian Federation and Turkey. The surveys used are detailed in Table A.4.

Table A.4 – Sources for non-EU countries

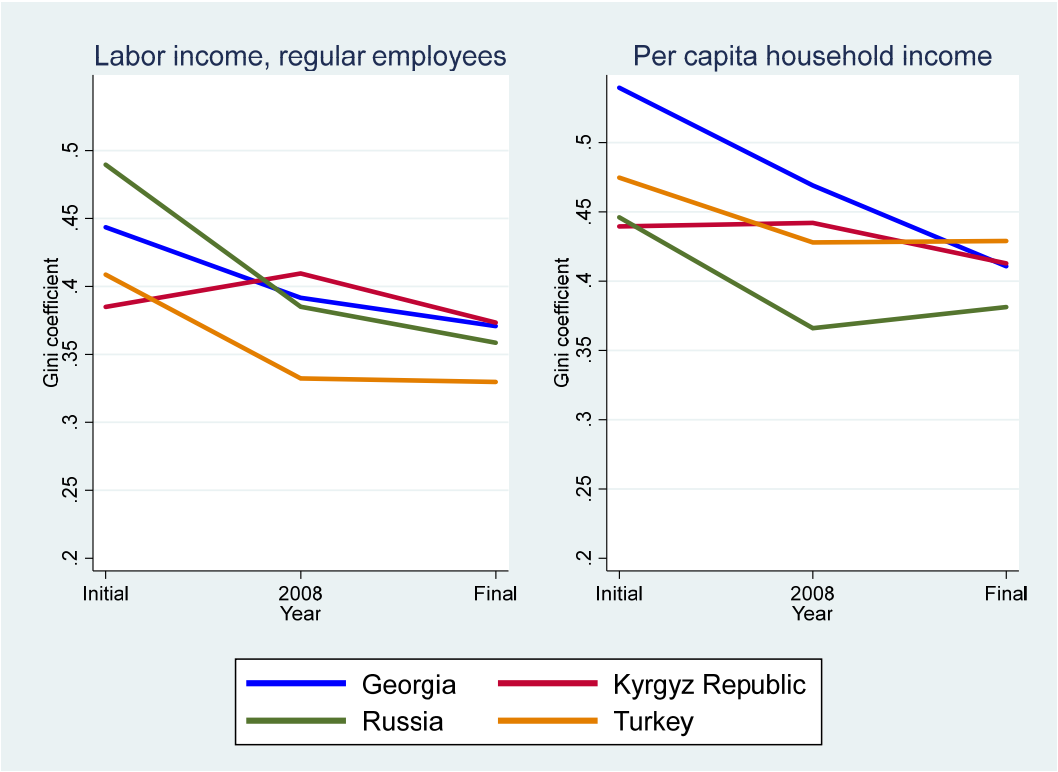
	Baseline			Final		
	Year	Survey and observations	Harmonization	Year	Survey and observations	Harmonization
Georgia	2002	Household Integrated Survey *40050 obs.*	ECAPOV	2015	Household Integrated Survey *38130 obs.*	ECAPOV
Kyrgyz Republic	2004	Kyrgyz Household Integrated Survey *21176 obs.*	ECAPOV	2014	Kyrgyz Household Integrated Survey *20094 obs.*	ECAPOV
Russia	1994	Russia Longitudinal Monitoring Survey *11280 obs.*	None	2014	Russia Longitudinal Monitoring Survey *18365 obs.*	None
Turkey	2002	Household Labor Force Survey *300689 obs.*	None	2013	Household Labor Force Survey *502426 obs.*	None

A4.1 DESCRIPTIVE STATISTICS

The evolution of earnings inequality

Figure 2 presents the evolution of the Gini coefficient for labor income of regular employees and for per capita household income in the four non-EU countries of our study. This measure for labor income experienced a significant decrease during this period, of about 13 points in Russia, 8 points in Turkey, 7 points in Georgia and 1 point in the Kyrgyz Republic and Poland. The decrease is of similar magnitude for per capita household income – 13 points in Georgia, 6 points in Russia, 5 points in Turkey and 2 points in the Kyrgyz Republic.

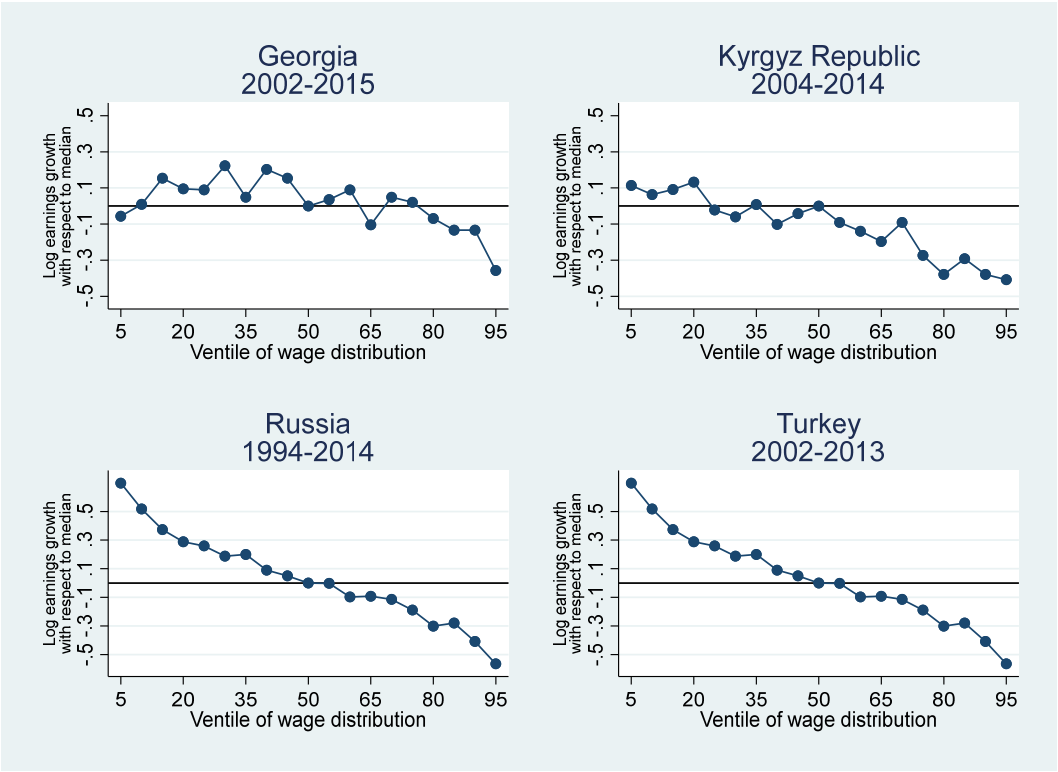
Figure A.1 – Evolution of inequality



Source: own elaboration based on household surveys. This figure shows the evolution of the Gini coefficient of labor income (only regular employees, excluding self-employed) and of per capita household income (monetary) for four countries. Initial year is 2002 for Georgia, 2004 for the Kyrgyz Republic, 1998 for Russia and 2002 for Turkey. Final year is 2015 for Georgia, 2014 for the Kyrgyz Republic and Russia, and 2013 for Turkey.

Figure A.2 shows the rate of income growth - relative to that of the median percentile- along the wage distribution and confirms that the decrease in inequality was not driven by a particular group. In fact, the patterns of income growth during the periods was consistently progressive in Georgia, the Kyrgyz Republic, Russia and Turkey, with richer percentiles experiencing lower income growth than the poorer ones.

Figure A.2 – Change in wages by quantiles of wage distribution

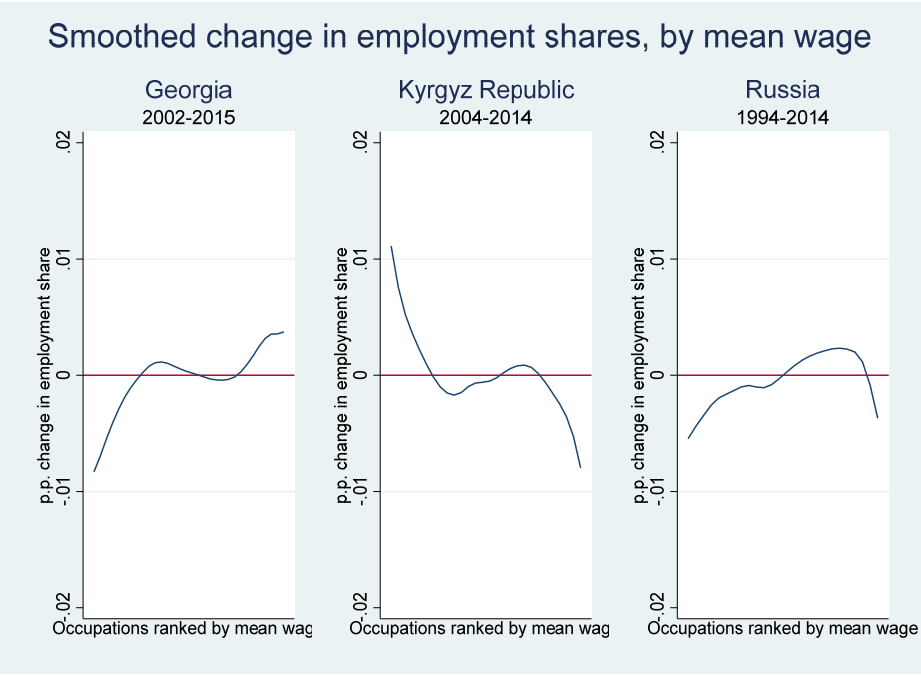


These set of figures plot the log change in wages for the different ventiles of the wage distribution from the initial period to the final period. Changes are expressed with respect to the median, whose change is normalized to zero. Only the wage of regular employees is considered in this analysis.

Occupational changes

In order to investigate the presence -or not- of job polarization we order occupations at the 2-digit level of the ISCO 88 classification according to their mean wage in the initial year and plot their change in the employment share in the time period that followed. Turkey is excluded as the microdata does not allow such level of disaggregation for the initial and final year. As Figure A.3 shows, there is no evidence of job polarization for any of the three countries depicted: employment growth is not concentrated at the extremes of the wage distribution, like in the United States, but in different parts of the distribution – the middle and top in Georgia, the bottom in the Kyrgyz Republic, the middle in Russia.

Figure A.3

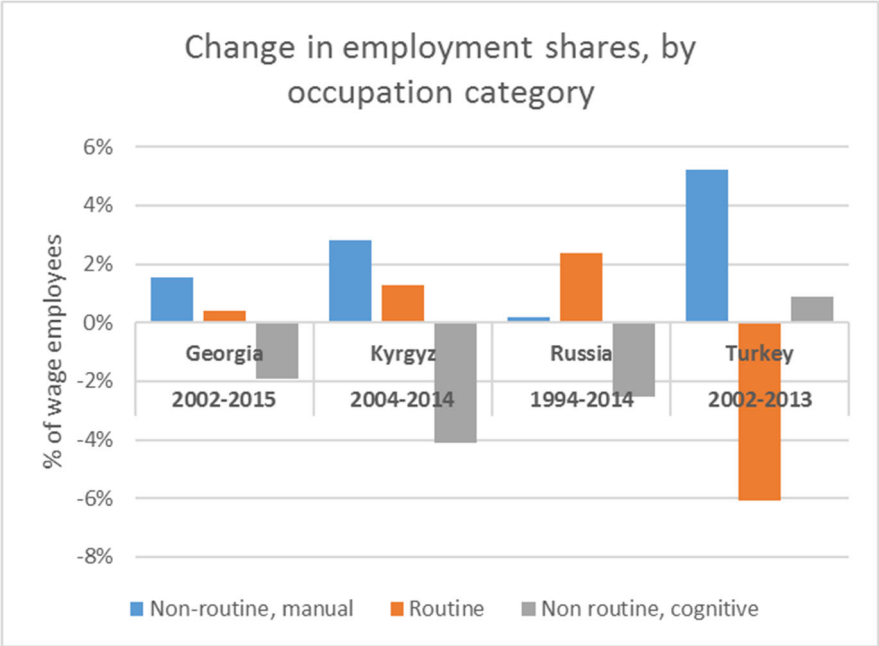


This figure plots the percentage point change in employment shares from the initial year (2002 for Georgia, 2004 for the Kyrgyz Republic and 1994 for Russia) to the final year (2015 for Georgia, 2014 for the Kyrgyz Republic and Russia) by occupations ranked according to their mean wage in the initial year. The changes are plotted by a locally weighted smoothing regression. Occupations are aggregated to the 2-digit level of the ISCO 88 classification.

As wages vary across countries, this ordering of occupations is country-specific: an occupation may be highly paid a country but low paid in another and vice versa. Thereby, Figure A.3 is not informative of what type of occupations are increasing or decreasing in relative importance from a cross-country point of view. A different way of looking at changes in occupational structure that allows for a cross-country comparison is to group occupations according to their task content. Following Acemoglu and Autor (2011)'s conceptual framework, we classify occupations into three categories: occupations relatively intensive in routine tasks, occupations relatively intensive in non-routine cognitive tasks and occupations relatively intensive in non-routine manual tasks. Note that any occupation implies carrying out both routine and non-routine tasks and both cognitive and manual tasks since they are not mutually exclusive: as described more in detail in the Appendix section 2, we group occupations according to the relative intensity of these tasks. Since occupations intensive in routine tasks are considered mid-skill occupations, job polarization has also been defined as the decline in the employment share of these occupations vis-à-vis an increase in the employment share of occupations intense in non-routine, cognitive tasks -high skill jobs- and in non-routine, manual tasks -low skill jobs-. In Figure A.4 we present the changes in the

employment share of each of the three occupation categories for each country. It confirms that job polarization is not present in either Georgia, the Kyrgyz Republic or Russia since routine task intensive occupations are actually growing as share of employment. Only in Turkey do we see a pattern similar to that of EU countries, albeit with a smaller growth of non-routine, cognitive task intensive occupations and a higher growth of occupations intensive in non-routine, manual tasks.

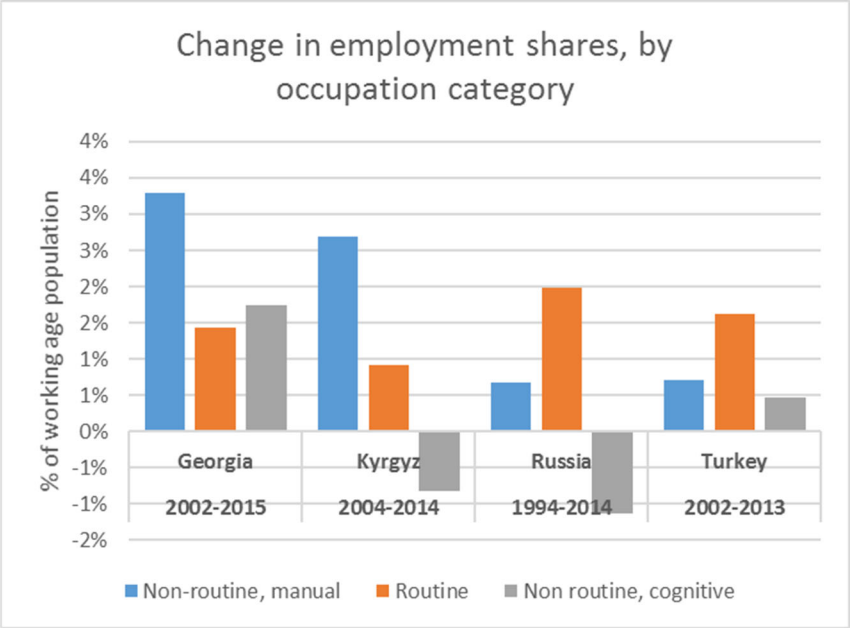
Figure A.4



This figure shows the change, in percentage points, of the share of employment (regular employees, excluding self-employed) over a period of ten or twenty years of the three occupations categories: in blue, occupations relatively intensive in non-routine, manual tasks; in orange, occupations relatively intensive in routine tasks; in grey, occupations relatively intensive in non-routine, cognitive tasks. The time period depends on data availability. For more details on the construction of the occupation categories please see the appendix.

These figures do not account for the important increase in wage employment in these countries in the period over analysis – increases of 1 percent of the working age population in Russia to 7 percent in the Kyrgyz Republic. If we plot the change in employment shares expressed as percentage of working age population, only in Kyrgyz and Russia do we see a decrease in the absolute number of workers in non-routine, cognitive task intensive occupations.

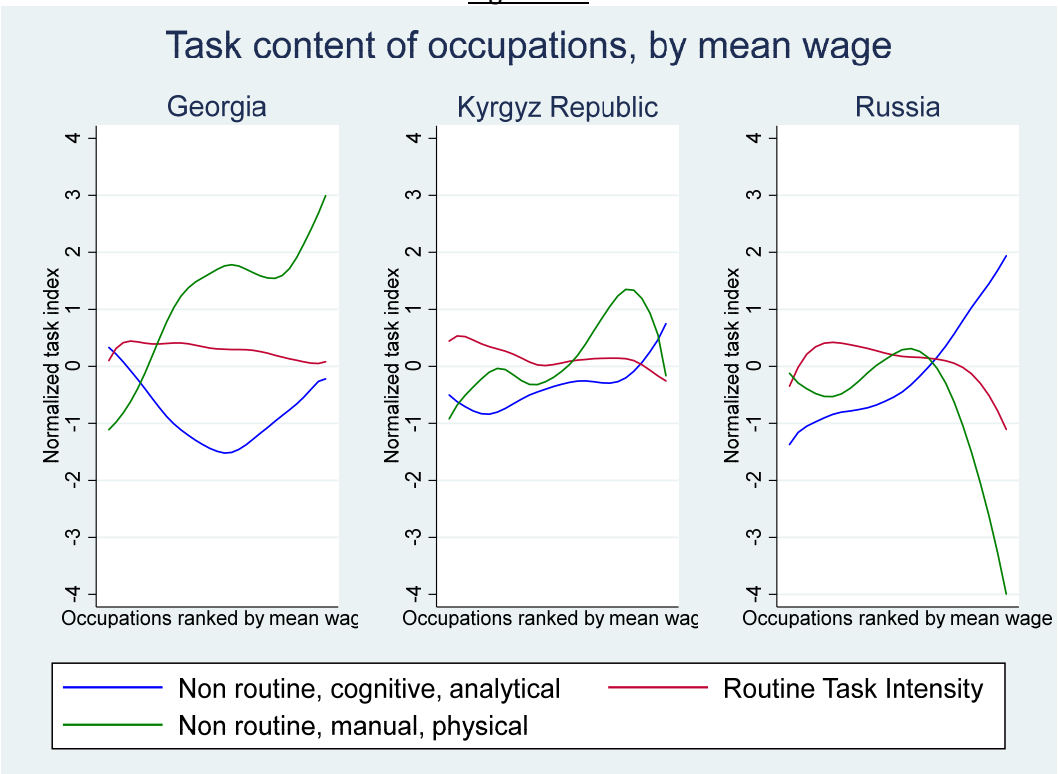
Figure A.5



This figure shows the change, in percentage points, of the share of working age population over a period of ten or twenty years employed in three occupations categories: in blue, occupations relatively intensive in non-routine, manual tasks; in orange, occupations relatively intensive in routine tasks; in grey, occupations relatively intensive in non-routine, cognitive tasks. The time period depends on data availability. For more details on the construction of the occupation categories please see the appendix

Figure A.6 presents the correlation between the task content of occupations and wages in the initial year. Following Acemoglu and Autor (2011), we use the task content of occupations according to the O*NET database and apply it to the data from Georgia, the Kyrgyz Republic and Russia. While this last country shows a pattern roughly similar to that of EU countries -non-routine, cognitive task intensity being higher at the top-, the figures for Georgia and the Kyrgyz Republic suggest a very unusual correlation of task to wages in the initial year, with non-routine, manual task intensity being higher at high paid occupations.

Figure A.6

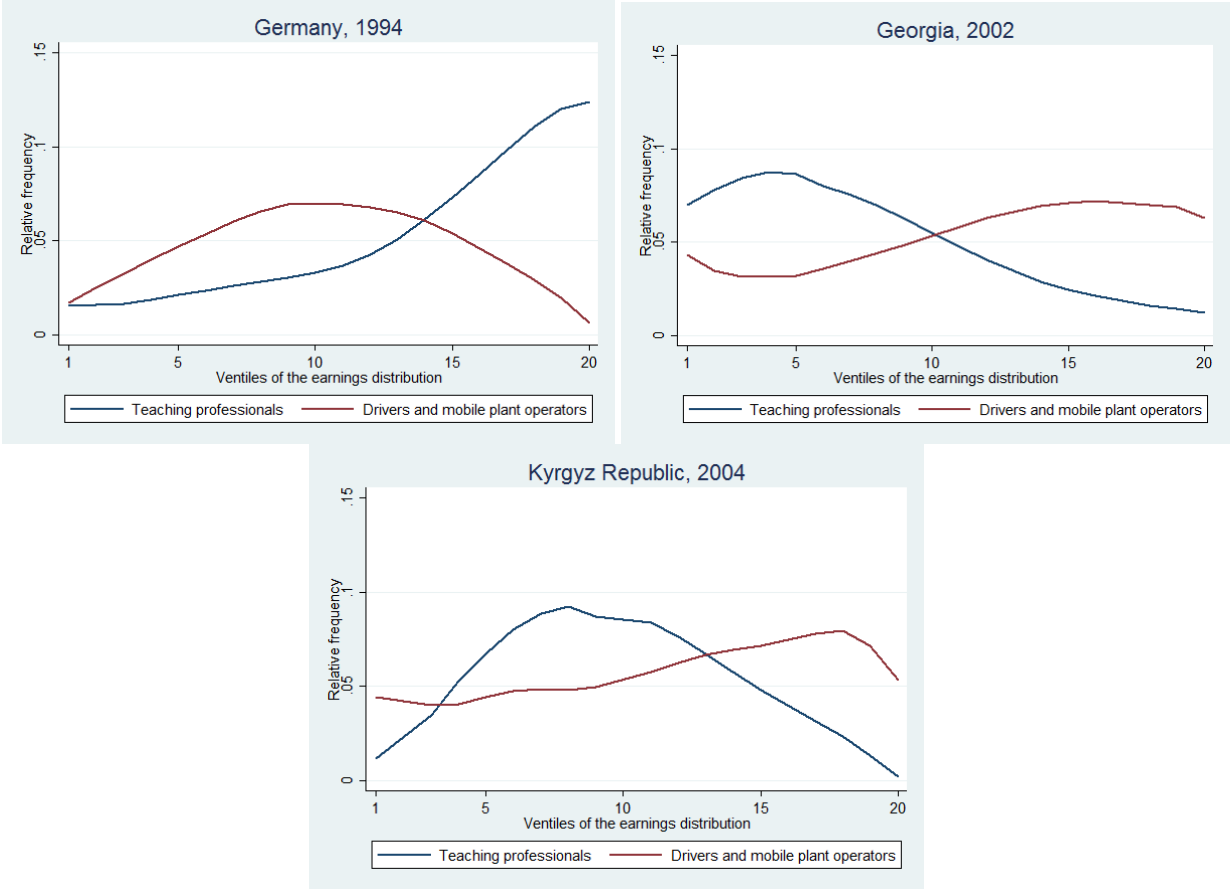


This figure plots three task content indices of occupations ranked by their mean wage in the initial period as in Figure 3. The three task content indices (intensity in non-routine, cognitive, analytical tasks; intensity in non-routine, manual, physical tasks; routine task intensity) are normalized to their economy-wide means. The indices are plotted by a locally weighted smoothing regression. Occupations are aggregated to the 2-digit level of the ISCO 88 classification

Heterogeneity in wage compensation of occupations across Europe and Central Asia

In order to further investigate this pattern, in Figure A.7 below we plot the distribution of teaching professionals (ISCO code 23, relatively intense in Non-routine, Cognitive tasks) and drivers and mobile plant operators (ISCO code 83, relatively intense in Routine tasks) by ventile of the overall wage distribution of the economy for the initial years of our analysis in Georgia, the Kyrgyz Republic and Germany.

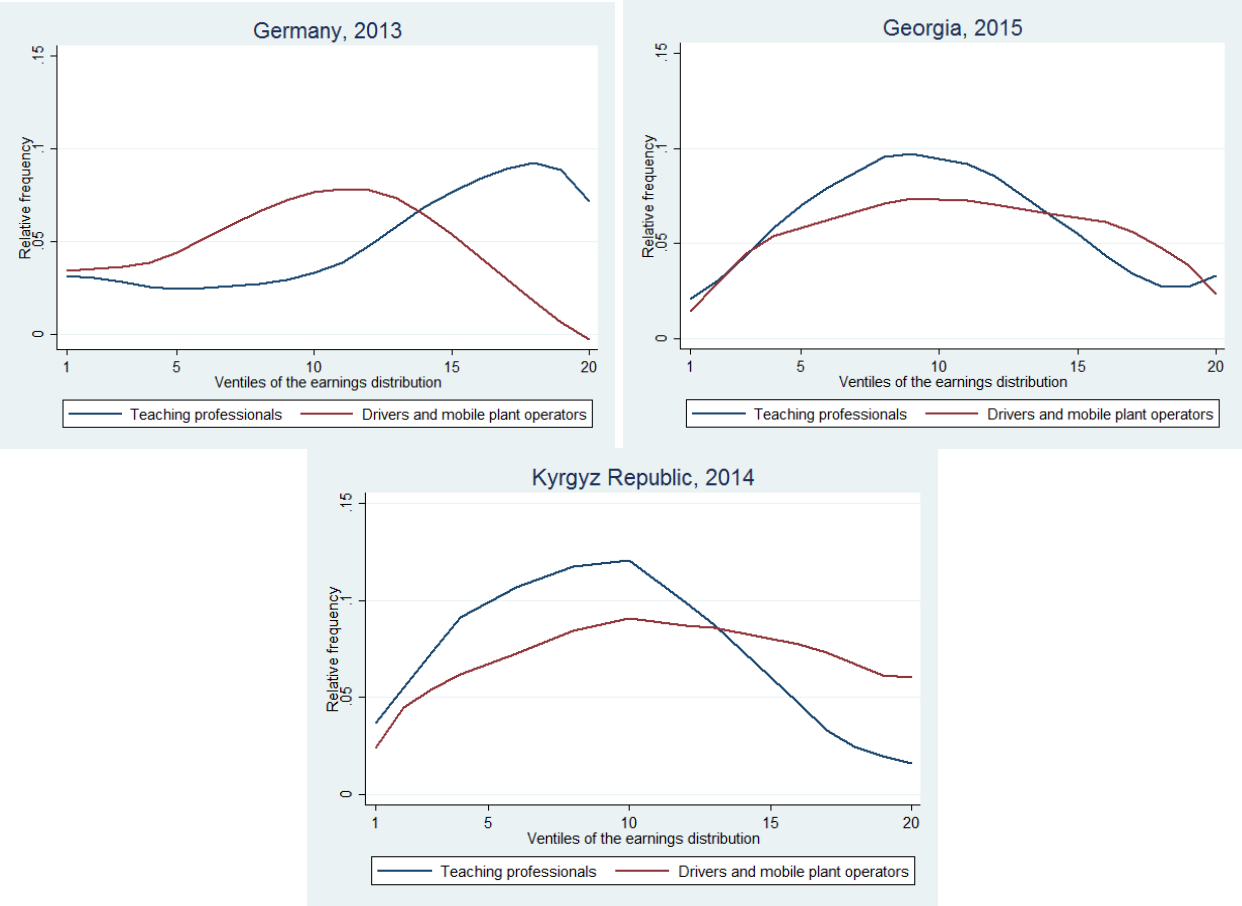
Figure A.7 – Distribution of Teaching Professionals and Drivers and Mobile Plant Operators in the earnings distribution, initial year



This figure plots the relative distribution of teaching professionals (ISCO code 23) and drivers and mobile plant operators (ISCO code 83) on the overall earnings distribution in the initial year of the analysis. All curves are smoothed by a locally weighted regression. All values include self-employed. Similar patterns are observed when excluding self-employed.

The difference in the compensations is strong: while teaching professionals are concentrated in the upper part of the earnings distribution in Germany, in Georgia and the Kyrgyz Republic they are rather found in the bottom half. Conversely, drivers and mobile plant operators are found in the middle of the distribution in Germany, while in Georgia and the Kyrgyz Republic they are located in the upper part of the earnings distribution. This difference most probably owes to the fact that teaching professionals are prevalently employed by the public sector, differently to drivers which work relatively more in the private sector. Public sector employees are notably less well paid than private sector employees in these countries, particularly during the early transition years.

Figure A.8 – Distribution of Teaching Professionals and Drivers and Mobile Plant Operators in the earnings distribution, final year



This figure plots the relative distribution of teaching professionals (ISCO code 23) and drivers and mobile plant operators (ISCO code 83) on the overall earnings distribution in the final year of the analysis. All curves are smoothed by a locally weighted regression. All values include self-employed. Similar patterns are observed when excluding self-employed.

In Figure A.8 we plot the same distributions in the final year of our sample (2013 in Germany, 2014 in the Kyrgyz Republic and 2015 in Georgia). While in Germany the pattern is roughly similar to one twenty years earlier, in Georgia and the Kyrgyz Republic there has been a slight convergence in the distribution of both teaching professionals and drivers and mobile plant operators. However, the distribution of teaching professionals is still skewed to the left with respect to drivers and mobile plant operators, which are comparatively more prevalent in the top half of the earnings distribution.

A4.2 RESULTS

A4.2.1 ACCOUNTING FOR THE DECREASE IN EARNINGS INEQUALITY

Table A.5 displays the change in earnings inequality accounted for different factors. It shows the change in the Gini coefficient and the ratio between the average earnings of the 90th and 10th (P9010), 90th and

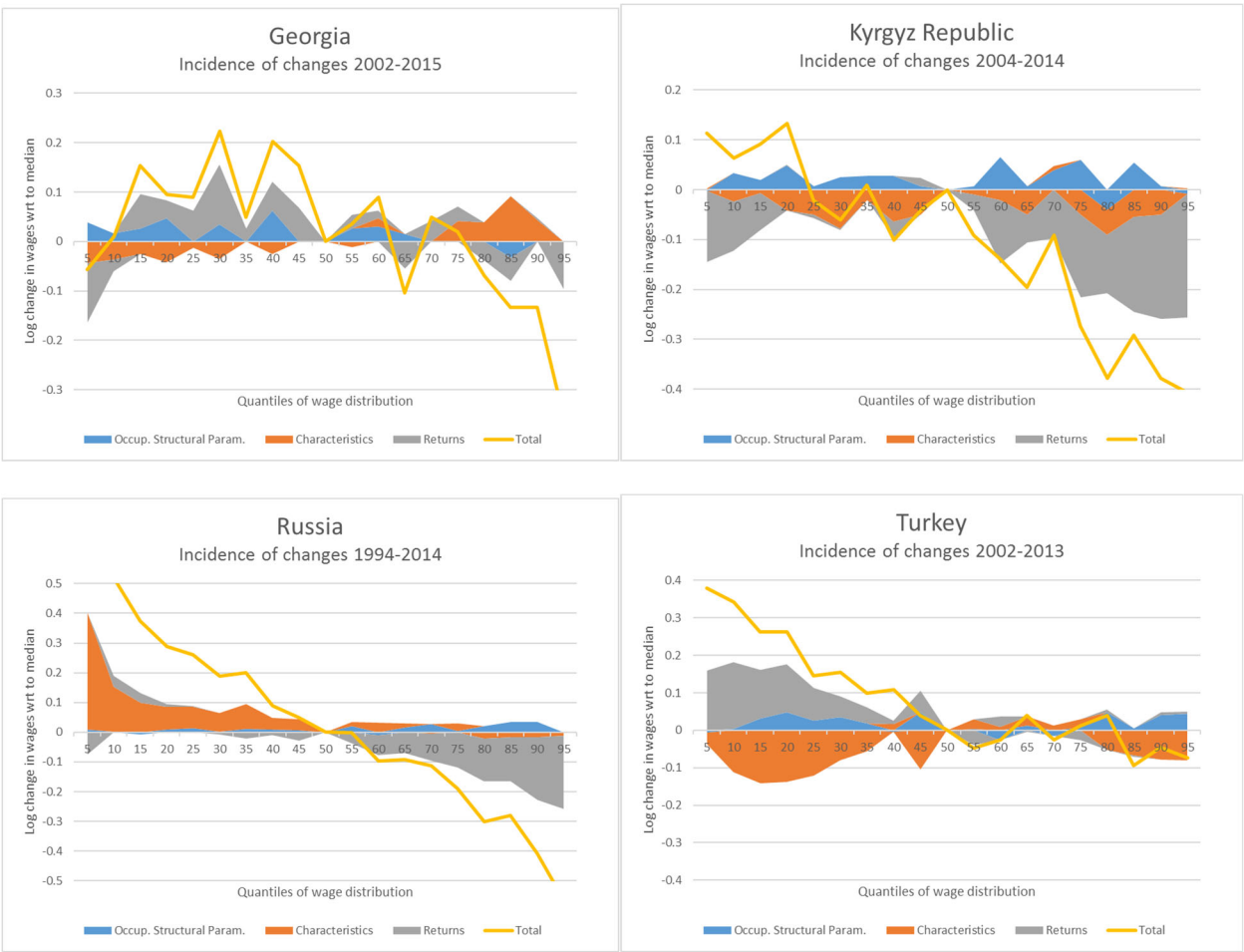
50th (P9050) and 50th and 10th (P5010) percentiles of the earnings distribution. In all four countries, changes in the occupational structure seem not to be an important factor behind the decreasing inequality observed during the period. It accounts for almost a negligible part of the change in the Gini coefficient. Wage returns to characteristics seem to be a more relevant factor, although in no case except the Kyrgyz Republic (where overall change in inequality was slim) does it represent more than 30% of the observed decrease. The result of the Mincer equations (not shown here) suggest that declining returns to education may be driving this result. Changes in individuals' characteristics account for a small share of the overall changes in inequality, and in Georgia and Turkey they would have actually resulted in an increase in inequality instead of the decrease that was observed. Overall, however, the model leaves an important part of the change in earnings inequality unexplained by the variables included in the simulation.

Table A.5 – Decomposition results

Country	Inequality measure, labor income	Change 1990s-2013	Percentage explained by: **Occup.** (C1, section 2.2) (1)	**Charact.** (C2, section 2.2) (2)	**Returns** (C3, section 2.2) (3)
Georgia	Gini coefficient	-0.075	4	-14	20
	P90/10	-1.025	12	-62	-52
	P90/50	-0.357	0	-36	-7
	P50/10	-0.026	170	-380	-355
Kyrgyz Republic	Gini coefficient	-0.013	-2	141	119
	P90/10	-2.380	7	7	18
	P90/50	0.303	-1	16	54
	P50/10	0.208	53	-39	-209
Russia	Gini coefficient	0.068	-1	12	26
	P90/10	2.672	-5	15	28
	P90/50	0.466	-10	2	54
	P50/10	0.442	0	28	7
Turkey	Gini coefficient	0.068	2	-11	26
	P90/10	2.672	-7	-5	38
	P90/50	0.466	-82	163	-14
	P50/10	0.442	0	-35	51

Figure A.9 shows the graphical version of the simulation results. In all four countries, changes in occupations explain very little of the change in earnings. Individual characteristics and wage returns under-account for the relative decline of earnings at the top of the distribution in all the countries except Turkey. This means that unobserved factors not included in the model partially offset the role of these changes in driving inequality at the top of the distribution.

Figure A.9 – Decomposition of earnings growth



A4.2.2 UNDERSTANDING OCCUPATIONAL CHANGE

Occupational change in post-Soviet countries and Turkey during the last decade has two distinctive features: on the one hand, an increase in participation rates, which have brought "new" workers into all occupational categories, and on the other hand a relative growth of routine task intensive and particularly non-routine, manual task intensive jobs at the expense of a decrease in the share of employment of non-routine, cognitive task intensive occupations. In table 5 we present the distribution of those employed in non-routine, manual task intensive jobs and routine task intensive jobs according to their occupational category in the counterfactual scenario where occupational structure parameters are the same as in the initial period. In this way, we are able to simulate the "occupation of origin" of those who are presently employed in those two occupation categories.

Table A.6 – Flows into non-routine, manual and routine intensive occupations

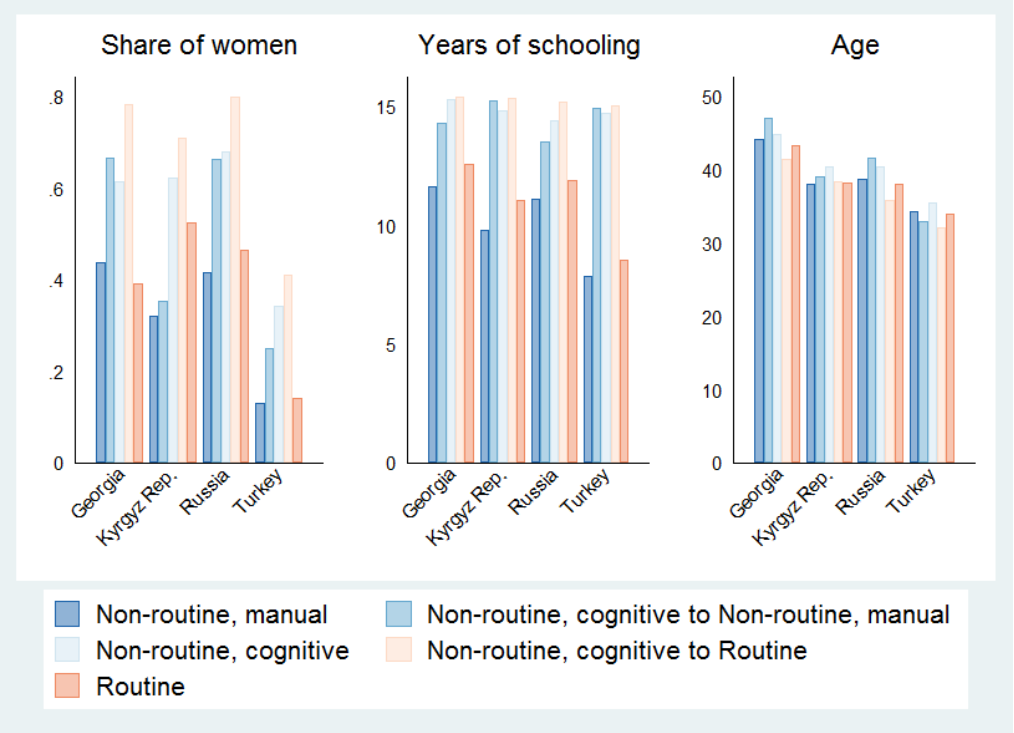
	Occupation in counterfactual scenario (assuming occupational structure parameters as in the initial year)				Actual change in share of employment, p.p. of working age population
Actual occupation observed in final year	Not employed	Non-routine, manual	Routine	Non-routine, cognitive	
Non-routine, manual	"New Entrants"	"Stayers"	"Movers"		*Non-routine, manual*
Georgia	34.0%	57.4%	2.8%	5.8%	3.3%
Kyrgyz Rep.	21.0%	69.4%	2.2%	7.4%	2.7%
Russia	17.0%	69.5%	5.5%	8.0%	0.7%
Turkey	31.7%	57.1%	5.0%	6.2%	3.3%
Routine	"New Entrants"	"Movers"	"Stayers"	"Movers"	*Routine*
Georgia	28.8%	1.5%	63.6%	6.1%	1.4%
Kyrgyz Rep.	17.1%	4.1%	70.8%	7.9%	0.9%
Russia	14.0%	2.3%	68.3%	15.4%	2.0%
Turkey	18.1%	0.0%	77.6%	4.2%	2.5%

This table presents different statistics on the flows into non-routine, manual task intensive and routine task intensive occupations. The first four columns describe the distribution of the individuals presently employed in non-routine, manual task intensive occupations (first four rows) and in routine task intensive occupation (second four rows) according to their occupation in the counterfactual scenario that assumes structural parameters to be the same as those in the initial sample year. The first column shows those individuals who, according to our analytical model, would have been out of employment in the counterfactual scenario. The second column shows those individuals who would have been employed in non-routine, manual task intensive occupations had the occupational structural parameters been the same as in the initial year. The third column shows those individuals who would have been employed in routine task intensive occupations and the fourth column shows those who would have been employed in non-routine, cognitive task intensive occupations in the counterfactual scenario. Lastly, the fifth column shows the change in the share of employment (over the working age population) between the initial year (2002 for Georgia, 2004 for the Kyrgyz Republic, 1994 for Russia and 2003 for Turkey) and the final year (2015 for Georgia, 2014 for the Kyrgyz Republic and Russia, 2013 for Turkey) of the corresponding occupation categories.

The first pattern that emerges from the figures presented in Table A.6 is the relevance that have those who have moved from out of employment into both occupational categories – a reflection of the important increase in participation rates. Around a third of those that are presently working in non-routine, manual task intensive occupations in Georgia and Turkey -the countries where that occupational category presented the highest growth, more than three percentage points of the working population during the period under analysis- would have been out of the labor force had the occupational structure been the same as in the early 2000s. For the Kyrgyz Republic and Russia that same figure is around 20%. In the case of those presently employed in routine task intensive occupations, the percentage of individuals who would have been out of the labor force is lower in all the cases – below 20% in all countries except Georgia, where it is slightly lower than 30%. A second common pattern is that, considering those who moved from employed occupation categories, the share of "movers" from non-routine, cognitive

task intensive jobs is the largest, even in Georgia and Turkey where the actual share of non-routine, cognitive jobs increased as percentage of the working age population. It appears that the growth of non-routine, manual task intensive jobs and routine intensive jobs has been fueled by individuals coming from out of the labor force or from non-routine, cognitive task intensive occupations, with limited movement between the two growing categories. In Figure A.10 we present some descriptive statistics of the "movers" from non-routine, cognitive task intensive occupations in comparison to those of the "stayers" in that same category and in the two growing categories – non-routine, manual task intensive and routine task intensive occupations.

Figure A.10 – Characteristics of "stayers" and "movers": Georgia, Kyrgyz Rep., Russia and Turkey



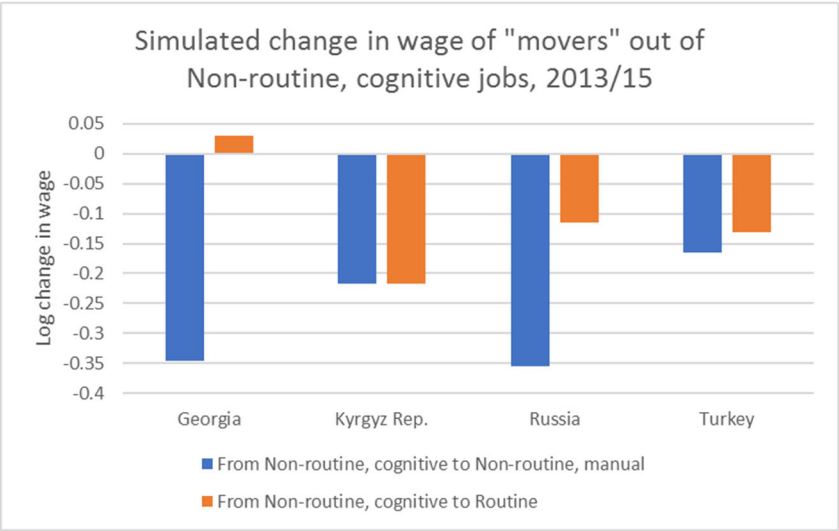
This figure shows a set of descriptive characteristics of the individuals in the final year sample. The dark blue bar shows the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in non-routine, manual task intensive occupations. The light blue bars show the average characteristics of individuals who in the actual data are employed in non-routine, manual task intensive occupations but in the counterfactual scenario are employed in non-routine, cognitive task intensive jobs ("movers" from non-routine, cognitive task intensive to non-routine, manual task intensive occupations). The lightest blue bars show the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in non-routine, cognitive task intensive occupations ("stayers" in non-routine, cognitive task intensive occupations). The light orange bars show the average characteristics of individuals who in the actual data are employed in routine task intensive occupations but in the counterfactual scenario are employed in non-routine, cognitive task intensive occupations ("movers" from non-routine, cognitive task intensive to routine task intensive occupations). The dark orange bar shows the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in routine task intensive occupations.

Non-routine, cognitive task intensive jobs have in all four countries a share of women higher than the two other occupation categories and, thus, the "movers" out of those jobs are made up of a higher share of women than those in non-routine, manual task intensive and routine task intensive jobs. The share of women in "movers" out of non-routine, cognitive task intensive jobs is above 50% in all the countries except Turkey, where it is between 20% and 30%. In this sense, within-employed occupational change appears to be predominantly female driven in these countries.

In terms of schooling, a similar pattern arises: as expected, the years of education of both "stayers" in non-routine, cognitive task intensive jobs and "movers" out of that category are higher than in the other two, growing occupational categories. In particular, this suggests that "movers" into non-routine, manual task intensive and routine task intensive jobs may be overskilled with respect to the incumbents in those occupations. Lastly, with respect to age all occupational categories appear to have a similar age profile, without significant differences between them.

The consequence in terms of labor income of the occupational change we have just described can be seen in the Figure A.11, where we have simulated the difference in wages that "movers" have between the actual salary they receive in either non-routine, manual task intensive occupations or routine task intensive occupations and the salary they would have received in non-routine, cognitive task intensive jobs.

Figure A.11



This figure shows the difference between the actual wage and the counterfactual wage for the group of "movers" out of non-routine, cognitive jobs in the final year sample. The blue bars show the difference between the wage "movers" in non-routine, manual task intensive jobs receive and the wage they would have received if they had been employed in a non-routine, cognitive task intensive occupation. The orange bars show the difference between the wage "movers" in routine task intensive jobs receive

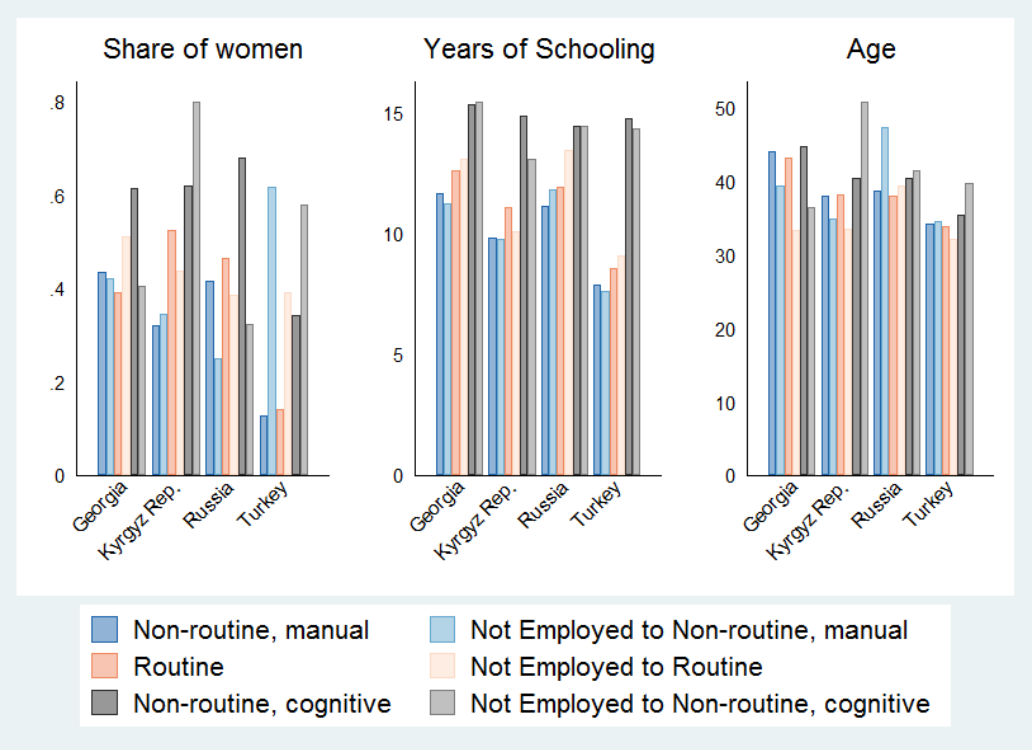
and the wage they would have received if they had been employed in a non-routine, cognitive task intensive occupation. See the methodological appendix for more details about the estimation of counterfactual wages.

Differently to the case of Germany, Poland and Spain, were a part of the "movers" from de-routinization gained income and another part lost income, in post-Soviet countries and Turkey most of the "movers" actually lose income. Only in Georgia do "movers" from non-routine, cognitive task intensive jobs gain a slight amount of income by moving into routine, task intensive jobs. In the rest of the cases, the loses vary between about 10% of the wage to more than 30%. This is not surprising given the fact that, as we saw in the previous paragraph, "movers" from non-routine, cognitive task intensive jobs are 1) more educated than the "incumbents" in the categories they move to; 2) move into categories with lower returns to education[14]. This evidence suggests that the most educated in these four countries are the relative losers of occupational change.

As we saw in Table A.6, the role of "new entrants" into the labor force is relevant in routine task intensive jobs and especially in non-routine, manual task intensive occupations. In Figure A.12 we compare the average characteristics of "new entrants" into the three occupational categories with the same characteristics of the incumbents in those categories, i.e. the "stayers".

[14] The log difference in wages between tertiary education and secondary education for household heads is highest in non-routine, cognitive occupations than in the other two categories in the four countries.

Figure A.12 – Characteristics of "new entrants" and incumbents: Georgia, Kyrgyz Rep., Russia and Turkey



This figure shows a set of descriptive characteristics of the individuals in the final year sample. The dark blue bar shows the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in non-routine, manual task intensive occupations. The light blue bars show the average characteristics of individuals who in the actual data are employed in non-routine, manual task intensive occupations but in the counterfactual scenario are not employed ("new entrants" into non-routine, manual task intensive occupations). The dark orange bar shows the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in routine task intensive occupations. The light orange bars show the average characteristics of individuals who in the actual data are employed in routine task intensive occupations but in the counterfactual scenario are not employed ("new entrants" into to routine task intensive occupations). The dark gray bars show the average characteristics of individuals who in the actual data and in the counterfactual scenario are employed in non-routine, cognitive task intensive occupations. The light grey bars show the average characteristics of individuals who in the actual data are employed in non-routine, cognitive task intensive occupations but in the counterfactual scenario are not employed ("new entrants" into non-routine, cognitive task intensive occupations).

First of all, with respect to gender, heterogeneous patterns emerge: the share of women among new entrants is considerably higher than that of incumbents in Turkey, suggesting a strong increase in female labor participation rate. In Russia, the opposite is true: the share of women is lower among "new entrants" in relation to incumbents in all three occupational categories: the increase in participation rates appears to be driven by men. In Georgia and the Kyrgyz Republic, the gender profile is mixed, with no clear pattern. With respect to education, new entrants appear in all cases to match the level of education of incumbents, suggesting that new entrants select into occupations where people of the same schooling profile as them are employed. Lastly, with respect to age, new entrants appear to be younger than incumbents in Georgia

and partly in the Kyrgyz Republic[15], while they appear to be older than incumbents in Russia, with no clear age pattern in Turkey. Summing up, the evidence suggests that the increase in participation rates appears to be driven by young men and women in Georgia and the Kyrgyz Republic, by older men in Russia and by women in general in Turkey.

[15] To the point that agricultural familiar or unpaid employment is classified as out of employment in our model, this may actually reflect a move of young people from that type of employment to actual wage employment.