Post-enlargement return migrants' earnings premium: Evidence from Latvia:

Mihails Hazans[♣]
mihazan@lanet.lv

University of Latvia and BICEPS
September 2008

The paper exploits a recent survey of over ten thousand economically active residents of Latvia; about 5% of respondents have worked abroad over the last three years, while 12% have family members with such experience. Post-enlargement labor migration from Latvia has been predominantly unskilled, yet return migrants when compared to stayers are, on average, more educated and less likely to work as unskilled manuals at their current or last job. Instrumental variable evidence suggests that return migrants are neither positively nor negatively selected in terms of earnings. Both earnings functions (corrected for selection into employment and for non-response on the wage question) and propensity score matching results suggest that return migrants command a substantial (15% on average) earnings' premium after controlling for worker demographic characteristics and education, as well as foreign and unemployment experience of family members. Accounting for variety of job characteristics leaves the unexplained gap unchanged among all workers (about 15%) and among women (7 to 8%), while narrows it down from 20 to 15% for men. Furthermore, we show that these earnings gains of the average return migrant can be viewed as being fully caused by foreign experience. Detailed decomposition of the gap, as well as its analysis by skill group, ethnicity, and across earnings distribution is performed. In the upper quartile of earnings distribution the unexplained gap in favor of return migrants exceeds 40%, while in the bottom quartile it is negative. In conclusion, we discuss supply and demand side effects behind the earnings gap between returnees and stayers.

Keywords: return migration, selection, earnings gap, propensity score matching, causal effect, decomposition

JEL classification: F22, J61, J31, C14

[♣] The paper benefited from comments by the participants of the meeting of the Marie Curie research training network „Transnationality of Migrants" in Riga, 26-27 October 2007. I am particularly grateful to Simone Bertoli who read the previous version and made a extremely useful comments on several issues, including the role of endogenous control variables in interpretation of matching results.

1 Introduction

In the new EU member states, post-enlargement outflow of the labor force to EU-15 resulted in shortage of workers. The Baltic countries, especially Lithuania and Latvia, experienced the highest rates of labor outflow among EU-8 (see Figure 1 below and Table 1 in Hazans 2007a)[1]. Return migration – and hence understanding return migration – is of critical importance for these countries. Interestingly, only the Baltics outflow rates started to decline in 2006 (Figure 1).

Scholars working on migration issues have paid considerable attention to the phenomenon of return migration. Theoretical models of return migration in the life-cycle planning and human capital investment framework have been proposed by Djajic and Milbourne (1988), Galor and Stark (1991), Stark (1995), Stark et al. (1997), Dustmann (1997a, 1997b, 1999), Zahniser and Greenwood (1998), Iara (2006), Dustmann and Weiss (2007). In these models, returning is driven by higher marginal utility of consumption in the home country than in the host country, by price differnces, or by expected gains at home from human capital accumulated abroad.

Dustmann (1993, 1997b) found that temporary and permanent migrants differ in human capital investment in the host country. Dustmann (1996) surveys European history of return migration and studies the determinants of returning. Dustmann and Kirchkamp (2001), Dustmann (2003), Gundel and Peters (2008) study the duration of stay in the host country.

There are theoretical reasons to expect that return migrants earn at home more than otherwise similar stayers. The literature (Dustmann 1997b; Co, Gang, and Yun 2000; Barrett and O'Connell 2001; Iara 2006) has pointed out to reasons related to human capital, positive selection, and signaling. First, general and/or specific human capital accumulated abroad might pay off. Second, return migrants might be positively selected on some characteristics desirable for employers (but usually unobserved for the researcher), e.g. initiative, motivation, adaptability. Third, employers might perceive working abroad as a signal or either higher productivity or presence of (initially unobservable) desirable characteristics. In this paper we suggest and explore

[1] This is consistent with pre-enlargement earnings in these countries being lower than in other accession countries. On the other hand, Hazans (2003) based on cross-country comparison of inter-regional migration rates, suggested that inherent mobility of labor force in the Baltic countries is higher than in Poland, Hungary, Czech Republic, and Slovakia.

also other reasons. Due to savings from higher earnings abroad, return migrants can afford to search for a job longer; moreover, by revealed preference, they are likely to place more value on wages (vs. other job attributes) than stayers. Both factors lead to higher reservation wages, hence to higher earnings. Finally, compared to otherwise similar stayers, return migrants are more confident and more likely to try high-end vacancies, which also leads to better outcomes. On the other hand, one can suggest (along the lines of Barrett and O'Connell 2001) also stories supporting a negative premium for having worked abroad. On the human capital part, recent domestic experience might be more valuable than foreign experience. On the signaling part, employers might see migration experience as a signal of some characteristics undesirable for the given firm (e.g. excessive risk taking or lack of reliability), or as a signal of being unsuccessful in the local labor market.

Thus, empirical evidence on earnings of return migrants is not only of practical but also of theoretical interest. The two items of interest are:

(i) The wage gap between the return migrants and otherwise similar stayers;

(ii) The wage differential which is 'caused' by migration experience (the 'average treatment effect on treated', ATET).

While some estimates of (i) and (ii) may coincide under certain assumptions, these two concepts differ in interpretation and econometric methodology involved (see next section for details)[2]. While the ATET allows for causal inference, it is also more demanding in terms of data and assumptions. The interpretation of the wage gap is less straightforward in our context. As mentioned above, the gap might include: (a) the selection effect due to unobservables, (b) the signaling part of the migration effect, (c) the returns to accumulated abroad unobservable[3] components of human capital, (d) the reservation wage effect, and (e) the labor market segmentation effect. Here (b), (c), (d), and (e) are also parts of the ATET, but (a) is not. On the other hand, if only post-migration characteristics are observed, the gap excludes the effect of change in observable characteristics (e.g. language skills or family status) caused by migration, while ATET includes this effect. Despite ambiguous interpretation, the wage gap between return migrants and stayers is of interest from the policy perspective, as well of for comparisons across studies.

[2] Needless to say, both the values and interpretation of the ATET and of the wage gap depend heavily on the set of control variables.

[3] Or not included among the controls for whatever reason.

However, compared to the huge empirical literature on migration in general, relatively few papers provide evidence on labor market outcomes of return migrants in their home countries. Zahniser and Greenwood (1998) find that for return migrants in Mexico, an additional year of U.S. experience yields a return that is eight times higher than that of an additional year of Mexican experience. Results of Co *et al* (2000) suggest that in Hungary, there is a premium to work experience abroad for women, while for men, the return to working abroad is not generally significant. In Ireland, by contrast, male college graduates who emigrated for labor-related reasons, after returning earn 15 percent more than otherwise similar stayers, while no wage premium is found for female returning migrants, as well as for persons of both genders who emigrated for non-labor reasons (Barrett and O'Connell 2001[4]); De Coulon and Piracha (2005), as well as Kilic *et al.* (2007) find that in Albania return migrants, compared to stayers, are more likely to be self-employed or business owners. De Coulon and Piracha (2005) also find that return migrants enjoy higher returns to being managers or self-employed than stayers do; returnees earn, on average, about 25% more than they would have if stayed at home (when disaggregated by skill or age group, the gain is 50-70%). Yet return migrants are negatively selected in terms of earnings: Would they stay at home, their earnings would be much lower than that of similar stayers. Iara (2006) employs data from survey of youth conducted in 2003 in 13 countries of Central and Eastern Europe which have all by now joined the EU; she finds that return migrants are positively selected and earn, after correction for selectivity, by about 30% more than similar stayers; by the same margin, their earnings exceed their counterfactual earnings would they stay.

Thus, existing evidence suggest that positive wage premium exists for return migrants in most cases[5]. But this evidence is far from systematic. One obvious reason for scarce evidence on returning migrants is lack of good data. Until very recently, most researchers have used data collected in the host countries. When data from source countries are used, the focus is usually on remittances or migration intentions, while the number of return migrants in the sample is typically not sufficient for obtaining

[4] Barrett and O'Connell (2001) present some [inconclusive] evidence that reported results are not biased by selection effect.
[5] Cohen and Haberfeld (2001), by contrast, document that prime-age Israeli men returning from US in 1978-1983 earned 8% less than otherwise similar stayers (without selectivity correction).

reliable estimates. For instance, Co *et al.* (2000) have just 76 male wage earners and 36 female wage earners in their sample of return migrants; Iara (2006) has 93 employed return migrants dispersed over 13 countries; Barrett and O'Connell (2001) have 158 return migrants, but their results differ significantly by gender and reason for emigration, and the underlying sub-samples are also quite small.

This paper is one of the first studies on post-enlargement return migrants in the new EU member states. It takes advantage of a big (over 10 thousand respondents) survey of economically active residents of Latvia, with a rich set of controls and about five hundred return migrants in the sample.

Our paper offers also a methodological contribution by combining instrumental variable (switching regression) techniques, Oaxaca-Blinder decomposition, and propensity score matching to compare earnings or return migrants with that of stayers[6]. We also include correction for non-response on wage question, usually ignored in empirical studies; this correction appears to be significant and uncorrelated with correction for selection into employment. The instrumental variable evidence suggests that return migrants are neither positively nor negatively selected in terms of earnings. Both earnings functions and matching results suggest that return migrants command a substantial (15% on average) and statistically significant earnings' premium. This premium is larger for men, but is significant also for women.

The rest of the paper is organized as follows. Section 2 briefly outlines the propensity score matching methodology and discusses its appropriateness in the given context. Details on data used in the paper are explored in Section 3. Section 4 describes the estimation results. Section 5 concludes by discussion possible reasons, both on the supply and on the demand side, behind the earnings gap between returnees and stayers. We focus on the reasons which are not related to transferability of specific skills but consistent with the assumptions justifying the application of propensity score matching.

[6] The previous migration literature has used for this purpose OLS, treatment effects, switching regression, and maximum likelihood estimates.

2 Methodological issues

To facilitate the discussion, let us present the relevant sequential-decision model with partial observability (Maddala (1983; Section 9.6) which incorporates the main outcomes underlying data at hand. Let earnings W in the home country be determined by $w|M=m = \beta_m Z + u_m$ (m=0, 1), where $w = \log(W)$, Z includes relevant worker's and (maybe) job characteristics, M is the indicator of being a return migrant ($M = 1$ for return migrants, $M = 0$ for stayers), and u_m is idiosyncratic error (unobserved earning ability); to keep notation simple we omit the subscript denoting individual hereafter. Let I_k (k = 1, 2, 3, 4) be indicator variables for participation in paid employment (I_1), reporting earnings in the survey (I_2), having foreign work experience (I_3), returning to the home country (I_4). These indicators are related to corresponding propensities by, say, probit models: $I_k = 1$ iff $Y_k = \theta_k X_k + \varepsilon_k > 0$. The X's are vectors of characteristics, the thetas are unknown parameters, and the epsilons are error terms. Earnings w are observed only if $I_1 = I_2 = 1$[7]; individuals with ($I_3 = 1$, $I_4 = 0$) are not observed, i. e. the sample is truncated. Plausibly, all epsilons are correlated with u_m and with each other, perhaps with some exceptions in specific applications.

While the theoretical models of return migration mentioned in Section 1 include the determinants of decision to return and some model also the migration decision, empirical applications usually do not account for the fact that these two are separate decisions with non-trivial correlation. Instead, analysis focuses on selection into being a return migrant, thus assuming (sometimes implicitly) that return migrants are randomly selected from the pool of migrants – an assumption which contradicts to most of the theoretical models. The only partial exception we know of is Zahniser and Greenwood (1998); they use data collected in both Mexico and US, and model selection into work in US, but mention, without details, that results are similar when distinction between stayers and return migrants is modeled as well. Selection into employment is accounted for in Zahniser and Greenwood (1998), Co et al. (2000),

[7] Typically, I_2 is observed only if $I_1 = 1$; in our application, however, only those unemployed who have worked no more than a year ago are included in the survey, and they are asked about earnings in the last job, so I_2 is fully observed, although we do not use past earnings as such. Partial observability of I_2 would not affect the following discussion.

and Iara (2006), but not in Barrett and O'Connell (2001) or De Coulon and Piracha (2005); none of these papers accounts for selection into reporting earnings.

As far as methodology is concerned, Zahniser and Greenwood (1998) estimate a single wage equation with two standard (Heckman 1979) selectivity corrections (not tested for independence though) and endogenous (fitted) foreign experience; Co *et al.* (2000) apply ML; De Coulon and Piracha (2005) - endogenous switching model (Maddala 1983) and nonparametric methods; Iara (2006) - endogenous switching model with double selection (Co et al. 2005).

This paper suggests a hybrid approach: the selections into employment and reporting earnings are addressed by a traditional two-step procedure (Berman and Wolfe 1984; Tunali 1986), while propensity score matching (PSM) is used to compare earnings of stayers and returners.

We start with the simpler of the two measures of return migrants' earnings premium, the unexplained gap $G = G_{|X}$ between earnings of return migrants and stayers with the same characteristics; here X is a vector (set) of control variables on which the gap is conditioned. We define the gap and other measures in terms of w; in empirical application, these measures will be converted to percentage changes in W.

According to Frölich (2007, Theorem 1), the question "How much would be stayers' earn on average if their characteristics would be distributed as those of return migrants?" can be answered using consistent PSM estimator:

$$\mathrm{E}[w \mid M = 0, F(X) = F(X \mid M = 1)] = \mathrm{E}[w \mid M = 0, F(p(X)) = F(p(X) \mid M = 1)].$$

Here $p(x) = \Pr(M = 1 \mid X = x)$ is the propensity score (which can be estimated e.g. by a probit model), and $F(\ )$ stands for cumulative distribution function. Hence estimating the covariate-adjusted unexplained gap

$$G = G_{|X} = \mathrm{E}[w \mid M = 1] - \mathrm{E}[w \mid M = 0, F(X) = F(X \mid M = 1)] \qquad (1)$$

amounts to application of PSM to estimating difference in mean *w* between return migrants and stayers. The calculations do not involve 'counterfactuals', and no assumptions are needed about the outcome variable, *w*, except for existence of first moments involved[8]. See Frölich (2007) for an example of using *G* as an alternative to the unexplained (discrimination) part of the Oaxaca – Blinder decomposition[9], further references and discussion. In the context or return migration, another non-parametric approach of DiNardo *et al.* (1996) has been used by De Coulon and Piracha (2005). See Appendix 1 for details on PSM.

Given that earnings' data are only available for employed individuals who have chosen to answer the wage question in the survey, a correction for sequential double selection is called for[10]. Hence, instead of raw data on *w* in (1) we use the prediction from wage equation augmented with selectivity correction terms λ_1 (for selection into paid employment) and λ_2 (for answering the wage question in the survey)[11]:

$$w|(I_1 = 1, I_2 = 1, M=m) = \beta_m Z + \delta_{1m}\lambda_{1m}(X_1) + \delta_{2m}\lambda_{2m}(X_2) + u_m \quad (m=0, 1) \quad (2)$$

The set of controls *Z* might be the same as or different from *X* used for matching; typically, one would want to get accurate predictions, so *Z* should include all observable determinants of earnings. The choice of *X*, on the other hand, depends on desired interpretation of the gap *G* and can vary from just human capital endowment to all worker and job characteristics. To avoid weak identification, X_1 and X_2, the determinants of wage employment and reporting earnings in the survey, each should include at least one strong instrument excluded from *Z* (together at least two such

[8] As a technical precondition for PSM one has to assume that $0 < p(x) < 1$ for any relevant vector x of values of the control variables (common support assumption; if it is not fulfilled, PSM has to be performed on the common support only).

[9] In its simplest form, this decomposition is given by $E[w|M=1] - E[w|M=0] = \beta_0(E[Z|M=1] - E[Z|M=0]) + (\beta_1 - \beta_0)E[Z|M=1]$, where the last term is the unexplained gap. See Oaxaca and Ransom (1994), Neumann and Oaxaca (2004) for further details and extensions.

[10] Behrman, Wolfe and Tunali (1980) were first to consider selection into reporting earnings; they extend Heckman's (1979) two-step method for estimating a model with sequential double selection, assuming independence of the two decisions; Maddala (1983; Section 9.6) presents two-step and maximum likelihood (ML) estimates for this and related models. When the two selection decisions are correlated, Tunali (1986) derives a two-step estimator, while Co *et al.* (2000) derive ML estimator.

[11] See section 4 below for details on the selection equations. Apart from selectivity correction, another reason to use predictions rather than raw wages in practice is the fact that earnings are often reported in surveys as interval data (in such cases the wage equation is estimated by interval regression method).

instruments). In general, the correction terms should be estimated from a bivariate probit model, but they reduce to inverse Mills ratios from usual probits when the two selection rules are independent (Berman and Wolfe 1984; Tunali 1986), as appears to be the case in our application. [12]

Let us turn now to the [causal] effect of work abroad on return migrants' earnings,

$$ATET = \mathrm{E}[w^1 \mid M = 1,\ X^1] - \mathrm{E}[w^0 \mid M = 1,\ X^0]\,. \qquad (3)$$

Here random variables w^0 and w^1 (respectively, X^0 and X^1) refer to situation in absence and presence of [actual or potential] foreign work experience and are related to the observed variables by the so-called observational rule

$$w = (1 - M)w^0 + Mw^1\,,\ X = (1 - M)\,X^0 + M\,X^1\,.$$

Workers' personal characteristics which can be affected by (or, in other words, are endogenous with respect to) foreign work experience in a systematic (i.e., distributional) way include language skills and marital status (other examples are discussed in section 4); obviously, some job attributes might be affected as well. If all variables in X are supposed to be exogenous with respect to M, then $X^1 = X^0$.

Let us define the *migration selection effect*, S, as expected, in absence of foreign work experience, difference between earnings of the return migrants and stayers with the same [potential] post-migration characteristics:

$$S = E[w^0 \mid M = 1,\ X^0] - E[w^0 \mid M = 0, F(X^1) = F(X^1 \mid M = 1)]. \qquad (4)$$

Replacing X^1 with X^0 in (4) gives an alternative version of S (denoted S^0).

Assuming common set of controls, X, for $ATET$, G, and S, simple manipulations give:

$$ATET = G - S + \gamma\,\mathrm{E}[X^1 - X^0 \mid M = 1],\ \gamma = d\mathrm{E}[w^0 \mid M = 0, X]\big/ d\,X\ , \qquad (5)$$

[12] Combining (2) with a matching estimator of (1) provides a formal way to avoid the ambiguity in decomposing the selectivity correction terms in (2) into explained and unexplained part (see Neumann and Oaxaca 2004): the identification is achieved by choosing the set of control variables for matching.

where γ is the coefficient vector in an [auxiliary] regression of *stayers*' predicted earnings from (2) on the set of controls *X* used for matching (if *X* is the same as *Z* in (2), and the earnings function is not corrected for selectivity, then $\gamma = \beta_0$). Equation (5) includes two statements. First, the part of unexplained gap which is due to unobserved differences between stayers and return migrants is not *caused* by migration experience. Second, the change in expected (in absence of foreign experience) earnings associated with migration-driven *systematic* shift in observed characteristics is a part of the causal effect of migration on earnings not captured by the unexplained gap. Note that when the set of controls widens, *G* and *S* move in the same direction, so *ATET* does not necessarily change.

Equation (5) modifies results by Lechner (2008), who has clarified the role of endogenous controls in application the PSM estimators and provided sufficient conditions for *ATET* = *G* (recall that *G* is estimable without restrictive assumptions). On the other hand, one can combine (5), (2), and information about the signs and/or sizes of the selection effect, as well as of the plausible changes in the observed variables which are endogenous to *M*, to get upper and/or lower bounds on *ATET*.

To fix the ideas, we start with the case when the selection into employment and reporting earnings is ignored, and earnings are reported precisely rather than in interval form, so that Lechner's (2008) results apply to (log) potential earnings of a randomly picked individual. Then, *ATET* = *G* if the following two assumptions (where 'characteristics' refer to control variables used for matching) hold:

> **Assumption A.1:** Return migrants, would they stay, would have the same distribution of earnings as stayers whose [potential] post-migration characteristics would be similar to observed characteristics of return migrants;
> **Assumption A.2:** Expected *average* earnings of stayers with characteristics similar to either observed characteristics of return migrants or their net characteristics are the same.

Note that the selection effect *S* = 0 due to **A.1**, while **A.2** removes the last term on the RHS of (5). **A.1** and **A.2** correspond to „CIA-P for Y^0 conditioning on X^1" and „E.2

with s=0" in Lechner (2008). If the version of (5) with S^0 instead of S is used, one can replace **A.1** with the following

> **Assumption A.0:** Return migrants, would they stay, would have the same distribution of earnings as stayers with [observable] characteristics similar to net (of any migration-driven change) characteristics of return migrants.

In this case, **A.2** should refer to 'distribution" (rather than 'expected average') earnings of stayers. Assumptions **A.0** and modified **A.2** (a combination not used in this paper) correspond to „CIA-P for Y^0 conditioning on X^0 " and „E.1 with s=0" in Lechner (2008).

In case when selection into employment and reporting earnings is corrected (see (2)), „earnings" in assumptions **A.1**, **A.2** should be understood as „earnings of employees who have answered respective question in the survey". Plausibly, this should not affect the validity of assumptions. Below, we discuss practical applicability of **A.1** (or **A.0**) and **A.2**.

Applicability of A.1 and A.2: Case 1. Suppose that X includes only worker's characteristics (and no job attributes), but conditional on these characteristics, dometsic earnings and propensity to „migrate and return" are independent: return migrants are neither positively nor negatively selected in terms of dometsic earnings. This is a rather strong but not completely unrealistic assumption (which appears to hold in our application). It can be tested (in practice, in form **A.1** easier than in form **A.0**) by instrumental variable methods, see section 18.4 in Wooldridge (2002) for details. Alternatively, one would need a specially designed survey which provides (maybe through registry data) also pre-migration earnings of return migrants (and, ideally, some psychological characteristics of respondents, maybe for a smaller sample). In addition, if X includes some characteristics potentially endogenous to migration, pre-migration values of those are needed to test **A.2**, unless there is some external evidence that these characteristics are in fact exogenous. On the other hand, even rough upper bounds on migration-driven changes in characteristics are sufficient if it turns out that the total effect of these change on earnings is negligible (this is

certainly the case if each of the suspicious variables does not in fact affect earnings; but it might happen also that some changes are good for earnings, while other are bad, and they compensate each other).

Applicability of A.1 and A.2: Case 2. Suppose the set of controls *X* is rich and includes characteristics of the worker, her family, and her job. Then, **A.1 (**and, perhaps less plausibly, **A.0**) might hold even when return migrants are negatively or positively selected (conditional on just observed personal characteristics), but any (dis)advantage in ability to earn the returnees might have over stayers in terms of unobservables is either exhausted via choosing (or being chosen for) a job with certain observed characteristics or captured by family-related instruments. In this case some of the variables in *X* are likely endogenous to migration, so one again would like to have their pre-migration values (at least imputed) to test **A.2** . But, as argued above, rough upper bounds on migration-driven changes might be sufficient.

3 Data

This paper exploits a representative survey of over 10 000 economically active residents[13] of Latvia. The data were collected between October 2006 and March 2007 in the framework of one of the projects of the National Program of Labor Market Studies [14]. 4.8% of respondents have worked abroad during the last three years, while 12% have a family member who has been working abroad during the three years prior to survey or was working abroad during the survey. 8.0% are potential emigrants (among the answers to the question "What are you ready to do in order to get a suitable job?" picked "Take a job abroad"). Thus we have sufficiently large samples of return migrants, as well as stayers with key characteristics similar to that of migrants.

Another advantage of having a large dataset is possibility to use more detailed control variables than found in most studies. We estimate two broad types of models: (i) with only basic demographic and human capital controls (in particular, we distinguish eight education categories, four levels of urbanization, and five regions); (ii) with enterprise and job characteristics (we distinguish 27 [two-digit] groups of occupations, 25

[13] By design, long-term unemployed, as well as self-employed, have been excluded from the survey.
[14] The field work was commisioned by the Latvian University of Agriculture and conducted by *GFK Custom Research Baltic*

sectors, four types of employers (central government institutions, local government institutions, NGOs, and others), four firm size categories, and five contract types, including work without written contract).

Latvia has a large (more than 40%) share of minority population (mostly Russian speaking), significant proportion of which does not hold Latvian citizenship (see section 3.7 in Zimmerman *et al.* (2008) for details), and hence was not covered directly by the removing the legal restrictions on working in UK, Ireland, and Sweden after EU enlargement in 2004. Hence we include dummies for being ethnic non-Latvian and for not holding Latvian citizenship.

Most of the recent labor migration from Latvia was to UK and Ireland. Hence, to disentangle the wage effect of migration experience as such from possible effect of English language skills, one needs information on respondents' proficiency in English language. On the other hand, both Latvian language skills and Russian language skills are important determinants of wages (see Hazans 2007b for evidence on impact of language skills on wages in Latvia). The data used in this paper included information on adequacy of respondents' skills in Latvian, Russian, and English languages for the jobs they occupy.

In addition, the data includes the following variables which are of interest in the migration context or can be used as instruments:

- Whether some of the respondent's family members were working abroad at the time of survey;
- Whether some of the respondent's family members have been working abroad during the last three years;
- Whether some of the respondent's family members have been unemployed during the last three years;
- The way of finding respondent's current job in Latvia.
- Whether the respondent is a trade union member.

Main limitation of our data is the absence of detailed information about experience abroad (duration of work abroad, the host country, and occupation). Arguably, these

limitations are not likely to compromise our results. First, according to other sources, the pool of return migrants is not too heterogeneous in these respects. Indeed, vast majority of recent temporary migrants have worked in UK or Ireland, the largest two countries which have opened their labor markets immediately after enlargement. According to a survey conducted in 2006, 56% or male return migrants and 61% of their female counterparts have spent abroad no more than a year, and 18-19% - one to two years (University of Latvia 2007, Figure 4.3). Most of the Latvian migrants in UK and Ireland, as well as most of the return migrants, have been employed abroad in various manual or low skilled non-manual occupations (Home Office *et al.* 2008, Table 7 and Chart 9 [15]; University of Latvia 2007, Figure 4.5).

4 Results

This section starts with comparing descriptive statistics of return migrants and stayers. We proceed with discussion of the earnings functions, presenting instrumental variable evidence which suggests that return migrants are neither positively nor negatively selected in terms of earnings. Next, we focus on selectivity corrections related to employment and reporting earnings; results depend heavily on gender as well as on whether or not job attributes are controlled for. After this preparatory work we discuss the PSM results on the unexplained earnings gap between return migrants and stayers, as well as conditions under which this gap can be interpreted as (or allows estimating) the causal effect of migration experience on earnings (ATET). In the latter stage, when analyzing the variables potentially affected by migration experience, we use auxiliary PSM of return migrants with potential migrants who have not worked abroad migration.

Descriptive statistics. Table 1 reports mean values of demographic, human capital, and key job characteristics of return migrants and stayers. By design, the results are representative for the part of economically active population including employees and job-seekers unemployed for less than one year. Females constitute 43% of return migrants, compared to 54% among stayers. As expected, return migrants are much

[15] When interpreting Chart 9, one has to take into account that the majority of workers in the administration, business & management sector work for recruitment agencies and could be employed in a variety of occupations; as seen from Annex A in Home Office *et al.* (2008), they are mainly low skilled.

younger: mean age difference is 7 years (34 vs. 43) for men and 11 years (33 vs. 44) for women. A closer look reveals that cohorts aged 15-24, 25-34, 35-44, and 45-54 have shares 17 to 18% each among stayers, while for return migrants these shares are 37%, 24%, 18%, and 10% respectively; 29% of stayers and less than 12% of returnees are older than 54 years. About half of returnees live with partners; among stayers this proportion exceeds 60%. While above figures are to some extent in line with the conventional story that most migrants are young, male, and single, the differences between return migrants and stayers are smaller than one could expect.

Among return migrants 36% are Russian-speakers, while 12% do not hold Latvian citizenship; among stayers both proportions are higher (41% and 20%), and the gaps are more pronounced for women than for men. This is in line with the absence of legal work provisions for non-citizens of Latvia in UK and Ireland (plausibly, women are less inclined to work abroad illegally).

Return migrants are much less than stayers concentrated in Riga and surrounding region, where out-migration rate was lower than elsewhere because of higher wages. On the other hand, 28% of return migrants and only 11% of stayers are found in Zemgale region; this discrepancy is too big to be explained only by high out-migration rate in Zemgale, but is in line with the fact that recent wage growth here was especially strong (see Hazans 2007b), and earnings-wise this region was not far behind the Riga region at the time of the survey. Assuming that most of the returnees settle in the region they left, this suggests that *return migrants are not randomly selected from the pool of migrants*[16]. Return migrants are more mobile than stayers also internally: 19% of them commute outside the district of residence, as opposed to 7% among stayers.

As mentioned at the end of Section 3, post-enlargement labor migration from Latvia has been predominantly low skilled, yet return migrants when compared to stayers are significantly less likely to have just primary education and more likely to have tertiary education. Return migrants are also less likely to work as unskilled manuals or to

[16] See Hazans and Philips 2009 for detailed comparison of out-migration and return migration rates across regions and types of settlement.

work without written contract[17]; 32% of return migrants and just 13% of stayers have supervisory responsibilities. This suggests that disproportionally many return migrants have above-average human capital and labor market outcomes.

On the other hand, returnees are by a wide margin more likely than stayers to have a fix-term contract, to be unemployed or to have a family member with unemployment experience during the last three years. It is hard to say to what extent this can be qualified as evidence for below-average labor market standing which triggered out-migration and to what extent it is caused by migration experience (for instance, we do not know whether unemployment spells of respondent's family members preceded the period of work abroad). The former interpretation seems justified for the respondents which during the last three years have been without work for more than a year or have been registered unemployed: more than 70% of them were employed during the survey, but, according to estimated earnings function, were paid 15% less than otherwise similar workers without 'stigma'. On the other hand, current unemployment or fixed-term job might well be voluntary (looking for a good job while spending money earned abroad; working on a fixed-term contract while planning another period of work abroad).

Return migrants and stayers very similar in terms of distribution among economic sectors, except for public services, where stayers have a larger share. Finally, in line with the well-known network effect (see e.g. Carrington *et al.* 1996), 45% of return migrants have family members working abroad recently or during the survey, while among stayers this proportion is just 10%.

Selectivity. To test whether the return migrant differ from stayers in terms of unobserved productivity, as well as to correct earnings functions for selection into employment and for reporting earnings (9.4% in our sample are unemployed, while 8.8% have not answered the question about earnings in the current or last job[18]), we apply instrumental variable techniques.

[17] In Table 1, job-related variables refer to the current job for employed respondents and to the last job for the unemployed; results change very little if only employed are used.
[18] The question on earnings asks about „average monthly earnings during the last year". Answers to this question given by those unemployed during the survey have not been used for wage regressions.

The first-stage probits for selection into being a return migrant are later also used to generate propensity scores for matching. These models (of which we report only the one without job controls for the pooled sample, see Table 2) should be seen as descriptive only rather than structural; some of the control variables (including family members' foreign experience) are clearly endogenous, but this is not a problem as long as we are not interested in the coefficient estimates. Results (consistent with the descriptives discussed above) suggest that, other personal (not job) characteristics being equal, likelihood to be a return migrant is larger for men than for women, as well as for singles as compared to those living with a partner; strongly declines with age; increases with the level of education; is smaller among non-citizens; increases with degree of urbanization; is larger for those whose family members have worked abroad or have been unemployed during the last three years.

As Appendix 2 and Tables 2-4 show in detail, return migrants are neither positively nor negatively selected in terms of earnings. This is true for men, women, as well as in the pooled sample[19], both in the reduced form models without job attributes and when job attributes are controlled for. In the latter case, however, the corrections for selection into employment and for responding to the question on earnings cannot be neglected. As expected, the correction for non-response to the earnings question is positive (non-responders earn more than similar responders). The correction for being unemployed is negative for men but positive for women.

Unexplained earnings gaps. In absence of the selection effect, estimated coefficients on return migrant dummy in the earnings functions presented in Table 4 are unbiased, indicating that return migrants earn, on average, 16% more than stayers with similar demographic characteristics and education (we distinguish eight categories); for men this gap amounts to 21%, while for women it is just 10%. Wooldridge (2002: Section 18.4) provides various technical assumptions which justify interpreting these gaps as caused by migration experience. A more straightforward and transparent approach is based on PSM estimates of the unexplained gap combined with formula (5).

[19] Occupational segregation by gender is modest in Latvia (Hazans 2005), hence we estimate both pooled models and the ones split by gender.

The PSM estimates for various subpopulations are found in Table 5. For each subpopulation, we estimate two earnings gaps between return migrants and stayers: (i) the gap unexplained by worker and family characteristics; (ii) the gap unexplained by worker, family, and job characteristics (the latter estimates are discussed in detail further on). Each gap is estimated twice (with similar results): using two nearest neighbors matching, as well as local linear regression matching. For all workers and for men the estimates are similar to the ones implied by Table 4 (as well as to Oaxaca – Blinder type estimates reported in Table 7). For women, the PSM estimates of the gap are lower (about 6% when only worker characteristics are controlled), although not statistically significant from the earnings function-based estimates.

The earnings functions suggest that among men, as well as among all workers, the unexplained earnings gap between return migrants and stayers does not narrow down after controlling for a variety of job attributes: tenure, two-digit occupation, sector of economic activity, adequacy of Latvian, Russian, and English language skills for the current job, tenure, ownership sector, presence and type of contract, size of the enterprise, way of finding the current job. For women, the evidence from the earnings functions is inconclusive (the gap tends to decline when job characteristics are controlled, but the errors are too big to make a conclusion). However, matching results presented in Table 5 suggest that controlling for job characteristics leaves the unexplained gap virtually unchanged among all workers (about 15%) and among women (7 to 8%), while narrows it down (statistically insignificantly) from 20 to 15% among men. Plausibly, better labor market outcomes which return migrants have (on average) in some respects are out-weighted by less favorable outcomes for the same or other return migrants in other respects. However, male return migrants seem to find somewhat better jobs than stayers with similar basic characteristics.

Causality. The above exogeneity result suggests that the migration selection effect $S = 0$ in formula (5). Hence, to estimate the causal effect of return migration on earnings, we only need to assess the last term in (5), the change in expected, in absence of foreign experience, average earnings associated with migration-driven systematic shift in observed characteristics. In fact, as shown in Table 6 and Appendix 3 (based on matching return migrants with stayers who plan to move), this change is negligible in both settings: when only worker and family characteristics or also job

characteristics are controlled (the size of estimated change is less than 1% in all cases except male workers with job attributes controls, when it is +3%). This is despite substantial effects of migration on family related variables and union membership (which do not transmit to earnings), as well as on language skills, way of finding job, commuting distance (all leading to small gains in earnings), job tenure, enterprise size (both leading to small losses), contract type (small gains within genders but a loss among all workers), and ownership sector (positive among women but negative among men). To sum up, conditional on worker's demographic characteristics and education, as well as migration and unemployment experience of family members, the *average* causal effect of foreign experience on return migrants' earnings equals the PSM estimate of the unexplained gap: 14 to16% when men and women are pooled together, 20 to 25% among men, and 6% among women. This includes gains from better jobs, as well as from better pay for similar jobs. When job characteristics are controlled, Table 5 estimates adjusted for the small changes found in Table 6 (panel B) suggest that the causal effect is 13 to15% for all workers, 18 to 20% among men, and about 7% among women.

Estimates by ethnicity and skills. Table 5 presents also PSM estimates of return migrants' earnings premium by ethnicity and skills. Premiums among ethnic Latvians and non-Latvians are virtually identical.

When matching on worker characteristics only, the estimated premium is somewhat higher (20.4%) among workers with education beyond secondary level than among the rest of the workers (15.8%), but this difference is not significant either. When job attributes are added, the gap widens to 24.3% among the more skilled workers but narrows down to 11.7% among the less skilled, so that the difference becomes significant. In other words, more educated return migrants have jobs with somewhat worse characteristics than stayers with similar basic characteristics, but enjoy high premium (24%) compared to stayers with similar jobs. Less educated returnees, by contrast, have somewhat better jobs than stayers with similar basic characteristics, but for the same job they are paid just 12% more than similar stayers.

Finally, we find that among manual workers the return migrants enjoy much higher earnings premium than among non-manual workers (the latter category includes

managers, professionals, associated professionals, technicians, clerks, as well as service, shop and market workers[20]). This does not contradict to the estimates by education, because many of non-manual workers have secondary education.

Detailed decomposition of the earnings gaps. Table 7 compares stayers' and return migrants' earnings functions and presents detailed Oaxaca-Blinder decomposition[21] of the earnings gaps when only basic characteristics are controlled. The unexplained gap calculated by the Oaxaca-Blinder method are similar to the 'naive' estimates from Table 4 (and to the PSM estimates): about 20 log points for men and about 10 points for women, while overall observed gaps are 31 and 26 points, respectively. There are very few statistically significant differences in returns to [basic] observed characteristics. For both genders, stayers living in the rural areas are disadvantaged in terms of earnings, but for return migrants it is not the case. A more detailed look reveals one likely reason: 30% of return migrants living in the countryside work outside the districts they live in, while among stayers this proportion is just 8%. However, type of settlement and region of residence contribute virtually nothing to the total unexplained gap for men; for women, region helps to close the gap by 3 points.

The largest part of the unexplained gap (19 points for men and 14 points for women) comes from factors not included in the model, which are captured by the base earnings (earnings of a 41 years old reference person. i.e. someone with professional secondary education, ethnic Latvian, Latvian citizen, living in rural area of Riga region). For both genders, the age effect narrows the unexplained gap by 9 to10 points: return migrants' age-earnings profiles are steeper and peak later than that of stayers, and this hurts the bulk of returnees which are young. Effects of ethnicity and citizenship contribute to the unexplained gap 3.5 points for men and 11.4 points for women. While among stayers there is no ethnic wage effects once citizenship is controlled for, return migrant females of non-Latvian ethnicity earn 18% more than their otherwise similar ethnic Latvian counterparts; on the other hand, females without

[20] In the previous version of the paper service, shop and market workers were classified as manual, in which case the premiums among the two groups are similar.
[21] We apply *normalized regression* approach by Yun (2005), which identifies deviations from [non-weighted] mean level, thus avoiding the ambiguity related to fixed reference group.

Latvian citizenship face much more pronounced earnings disadvantage among return migrants than among stayers. Finally, the differences in returns to education (mostly not significant though) make the unexplained gap wider by 5.1 points for men but narrower by 3.4 points for women.

As a robustness check we have assessed contribution of each of the main demographic factors to PSM estimates of the unexplained gaps by excluding one factor at a time (in which case the results do not add up to the total gap of course), as well as by including them sequentially (in which case a decomposition is possible but depends on the sequence of inclusion), and conclusions are qualitatively similar to the ones reported above.

Our preferred specification (Table 4, model B) does not control for supervisory responsibilities. One can argue that presence of such responsibilities is as much an outcome as a control variable, and it is clearly endogenous to earnings. When such control is added (Table 4, model C), the estimated earnings gap between return migrants and other workers stays unchanged for men, but disappears for women and declines by 3 points in the pooled sample.

Heterogeneity. The PSM estimates of the unexplained gaps (but not of the causal effect) can be easily performed at various points of the distribution of earnings (see Frölich 2007). Results presented in Table 8 reveal a very pronounced heterogeneity in the unexplained premium earned by return migrants. In the upper quartile of earnings distribution this premium (conditioned on demographics and education) exceeds 40% among all workers, as well as among all above mentioned subpopulations split by gender, ethnicity, education, or occupation; for some of these subpopulations the unexplained gap in favor of return migrants in the top quartile is even larger: about 50% among ethnic minorities and non-manual workers, and about 60% among manual workers and among workers with education beyond secondary. On the other hand, in the bottom quartile of earnings distribution the gap is negative for all subpopulations. At the 10^{th} percentile, return migrants earn 20% less than stayers with similar characteristics; this gap is relatively stable across subpopulations, varying from –17% among ethnic Latvians and workers with secondary or lower education to –25% among women and manual workers to –31% among non-manual workers. In

other words, inequality in earnings is much more pronounced among return migrants than among stayers: results presented in Table 8 imply that 90/10 ratio among return migrants is more than twice as large as among otherwise similar stayers.

Comparing the gaps with and without job controls (Table 8) leads to a conclusion that in every subpopulation the advantage of return migrants over their stayers counterparts at the top of earnings distribution stems mainly from being better paid for similar jobs. On top of this, the top earners among return migrants who are males, ethnic Latvians, or non-manual workers, have better job types than their stayer counterparts from the same groups (to a much smaller extent, this applies also to females, manual workers, and highly educated workers). Likewise, in all subpopulations considered, at the low end of earnings distribution return migrants are paid less than stayers with similar jobs; among women, ethnic Latvians, and low-educated workers the low-paid return migrants have also slightly worse job types than their stayer counterparts.

5 Conclusion

This paper investigates post-enlargement return migrants' earnings premium in Latvia by combining instrumental variables, propensity score matching, and decomposition techniques. Results suggest that return migrants are neither positively nor negatively selected in terms of earnings. Nevertheless, after controlling for worker demographic characteristics and education, as well as foreign experience and unemployment experience of family members, return migrants command a substantial earnings' premium: about 15% on average, more than 20% among men, and 6% among women. Accounting for variety of job characteristics leaves the unexplained gap unchanged among all workers, extends it slightly among women, while narrows it down to 15% among men.

In both settings (with and without job controls) these earnings gains of the average return migrant can be viewed as being fully caused by foreign experience. In other words, the average return migrant earn 15% more than he/she would earn in absence of foreign experience. There is, however, a great deal of individual heterogeneity: In the upper quartile of earnings distribution the unexplained gap in favor of return

migrants exceeds 40%, while in the bottom quartile it is negative, reaching –20% at the 10th percentile. This pattern persists, with slight variation, in subpopulations split by gender, ethnicity, occupation, or education.

Most of the previous authors ascribe wage premiums earned by return migrants to specific skills transferred from the host country. Here we discuss some alternative explanations. On the supply side, the following reasons could explain why return migrants might find better paid jobs (even if they have no productivity advantage over stayers):

(i) They can afford to search for a job longer. According to a recent study, 70% of return migrants spend earnings from abroad for everyday needs (University of Latvia 2007, Table 4.15).

(ii) They are more confident and "aim higher". In our sample, 17.3% or return migrants and just 6.6% of stayers have used such 'high end' job search methods as competition for the vacancy, Internet, and private recruitment firms. On the other hand, more than 20% of 328 return migrants (mostly from UK and Ireland) interviewed in a recent study, indicated increased self-respect and initiative as important gains from foreign experience;

(iii) They value wages (vs. other job attributes) higher than stayers.

Regarding more traditional human capital and signaling explanations, we would like to emphasize some which are also consistent with the version of conditional independence assumption used in this paper, as well as with the absence of selection. On the demand side, employers might believe that migrants are positively selected even if it is not the case on average. One can also suggest a combined supply-demand effect: some return migrants bring back better work habits and/or organizational (rather than specific!) skills, which are observed and valued by employers (this hypothesis is supported by the fact that return migrants are much more likely than otherwise similar stayers to have supervisory responsibilities).

References

Abadie A., Drukker D., Herr J. L. , Imbens G. W (2004). Implementing matching estimators for average treatment effects in Stata. *Stata Journal* 4(3): 290-311.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated Normal. *Econometrica* 41: 997-1016

Barrett A., O'Connell P. J. (2001). Is There a Wage Premium for Returning Irish Migrants? *Economic and Social Review* 32(1): 1-21.

Behrman, J. R., Wolfe, B. L. (1984). Labor force participation and earnings determinants for women in the special conditions of developing countries. *Journal of Development Economics*, 15(1-3): 259-288

Behrman J. R., Wolfe B. R., Tunali I. (1980). Determinants of women's earnings in a developing country: A double selectivity extended human capital approach. *Institute for Research on Poverty Discussion Paper* no. 596-80. Madison: University of Wisconsin.

Becker S. O., Ichino A (2002). Estimation of average treatment effects based on propensity scores. The Stata Journal 2(4): 358-77.

Bertoli S. (2008). The impact of migration on poverty in Ecuador: Erring on the side of excessive optimism? Paper presented at the *Migration and Development Conference*. Lille, June 26-28, 2008.

Blundell R., Dearden L., Sianesi B. (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society*, Series A 168: 473-512.

Caliendo M., Kopeinig S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, Vol. 22, No. 1: 31-72.

Carrington W. J., Detragiache E., Vishwanath T. (1996). Migration with Endogenous Moving Costs. *American Economic Review*, Vol. 86, No. 4, pp. 909-930.

Chiswick B. R., Miller P. W. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, 13(2): 246-288

Co C.Y., Gang I. N., Yun M.-S. (2000). Returns to returning. *J of Population Economics* 13 (1): 57-79.

Co C.Y., Gang I. N., Yun M.-S. (2005). Self-Employment and Wage Earning in Hungary. *Review of Development Economics* 9 (2): 150-165

Cohen J., Haberfeld Y. (2001). Self-selection and return migration: Israeli-born Jews returning home from the United States during the 1980s. *Population Studies,* 55: 79–91.

De Coulon A., Piracha M. (2005). Self-selection and the performance of return migrants: the source country perspective. *J of Population Economics* 18 (4):779–807

DiNardo J., Fortin N., Lemieux T. (1996) Labor market institutions and the distribution of wages,1973–1992: a semi-parametric approach. *Econometrica* 64(5):1001–1044

Djajic S., Milbourne R. (1988). A General Equilibrium Model of Guest-Worker Migration: A Source-Country Perspective. *Journal of International Economics*, 25, 335-351.

Dustmann C. (1993) Earnings adjustment of temporary migrants. *J. of Population Economics* 6 (2): 153-168

Dustmann C. (1996). Return Migration: The European Experience. *Economic Policy*, (22): 214-245.

Dustmann C. (1997a) Return Migration, Uncertainty, and Precautionary savings. *Journal of Development Economics,* 52, 295-316

Dustmann C. (1997b). Differences in the Labor Market Behavior Between Temporary and Permanent Migrant Women. *Labour Economics*, 4(1): 29-46.

Dustmann C. (1999). Temporary Migration, Human Capital, and Language Fluency of Migrants. *Scandinavian Journal of Economics*, 101(2): 297-314.

Dustmann C. (2003). Return Migration, Wage Differentials, and the Optimal Migration Duration. *European Economic Review,* 47 (2): 353-67.

Dustmann C., Kirchkamp, O. (2001). The Optimal Migration Duration and Activity Choice After Re-migration. *IZA Discussion Paper* No. 266

Dustmann C., Weiss Y. (2007). Return Migration: Theory and Empirical Evidence from the UK. *British Journal of Industrial Relations*, 45 (2): 236-256.

Frölich M. (2005). Matching estimators and optimal bandwidth choice. *Statistics and Computing* 15: 197–215

Frölich M. (2007). Propensity score matching without conditional independence assumption--with an application to the gender wage gap in the United Kingdom. *Econometrics Journal* 10(2):359-407.

Galor O., Stark O. (1991). The Probability of Return Migration, Migrants' Work Effort, and Migrants' Performance. *Journal of Development Economics*, 35:339-405.

Gundel S., Peters H. (2008). What Determines the Duration of Stay of Immigrants in Germany? Evidence from a Longitudinal Duration Analysis. *SOEP* paper No. 79.

Hazans M. (2003). Determinants of Inter-Regional Migration in the Baltic Countries, ZEI Working Paper No. B17-2003:1-35. Available at SSRN: http://ssrn.com/abstract=742184

Hazans M. (2007a). Looking for the Workforce: The Elderly, Discouraged Workers, Minorities and Students in the Baltic Labour Markets. *Empirica*, 34(4): 319-349.

Hazans, M. (2007b). Coping With Growth and Emigration: Latvian Labor Market Before and After EU Accession. Available at SSRN: http://ssrn.com/abstract=971198

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equations System. *Econometrica* 46: 931–960.

Heckman J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47:153-161.

Heckman J.J., Ichimura H., Todd P.E. (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65, 261-294.

Heckman J.J., LaLonde R.J., Smith J.A. (1999). The economics and econometrics of active labor market programs. In: *Handbook of labor economics*, Vol.3A, Ch. 31. Amsterdam: North-Holland.

Home Office – Border and Immigration Agency – Dept. for Work and Pensions – HM Revenue and Customs (2008). Union Accession Monitoring Report: A8 countries May 2004 – June 2007.
http://www.bia.homeoffice.gov.uk/sitecontent/documents/aboutus/Reports/accession_monitoring_report/report12/may04jun07.pdf?view=Binary

Iara A. (2006). Skill Diffusion in Temporary Migration? Returns to Western European Working Experience in the EU Accession Countries. *Centro Studi Luca d'Agliano Development Studies Working Paper* No. 210 Available at SSRN: http://ssrn.com/abstract=918037

Kilic T., Carletto C., Davis B., Zezza, A. (2007). Investing Back Home: Return

Migration and Business Ownership in Albania. *World Bank Policy Research Working Paper No. 4366* Available at SSRN: http://ssrn.com/abstract=1017172

Lechner M. (2008). A note on endogenous control variables in causal studies. *Statistics & Probability Letters*, Vol. 78, No. 2: 190-195.

Leuven E., Sianesi B (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html

Maddala G.S. (1983). Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press.

Muehler G., Beckmann M., Schauenberg B. (2007). The Returns to Continuous Training in Germany: New Evidence from Propensity Score Matching Estimators. *ZEW Discussion Paper* No. 07-048.

Nakosteen R.A., Zimmer M. (1980). Migration and Income: The Question of Self-Selection. *Southern Economic Journal*, Vol. 46, No. 3 (Jan., 1980), pp. 840-851

Neumann S., Oaxaca R.L. (2004). Wage decompositions with selectivity-corrected wage equations: A methodological note. *Journal of Economic Inequality* **2:** 3–10.

Oaxaca R.L., Ransom M.R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61 (1): 5-21

Rosenbaum P., Rubin D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70: 41-50.

Stark O., Helmenstein C., Yegorov Y (1997). Migrants' Savings, Purchasing

Power Parity, and the Optimal Duration of Migration. International Tax and Public Finance, 4 (3), 307-24.

Stark O. (1995). Return and Dynamics: The Path of Labor Migration when Workers Differ in their Skills and Information is Asymmetric," *Scandinavian Journal of Economics*, 97(1): 55-71.

Tunali I. (1986) A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Remigration. In: Ronald G. Ehrenberg (ed.). *Research in Labor Economics*, 8(B): Conn.: JAI Press: 235-283.

University of Latvia (2007). Geographical mobility of population. Research report. University of Latvia: Riga, 240 pp. (in Latvian).

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yun, M.-S. (2005), „Normalized Equation and Decomposition Analysis: Computation and Inference," *IZA Discussion Paper* No 1822.

Zahniser S. S., Greenwood M.J. (1998). Transferability of Skills and the Economic Rewards to U.S. Employment for Return Migrants in Mexico. In: U.S. Commission on immigration reform. *Migration Between Mexico and the United States: Binational Study*. Volume 3, 1133-1152. http://www.utexas.edu/lbj/uscir/binpapers/v3b-5zahniser.pdf

Zanutto E. L. (2006). A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data. *Journal of Data Science*, 4: 67-91.

Zimmerman K. F., Kahanec M., Constant A., DeVoretz D., Gataullina L., Zaiceva A. (2008). Study on the social and labour market integration of ethnic minorities. *IZA Research report* No. 16. Bonn: IZA

Appendix 1. The Propensity Score Matching Methodology and its Application

The idea of the PSM approach is to find, for every 'treated' individual (here: return migrant), one or more 'matches' from the control group ('non-treated') with similar observed characteristics. The unexplained gap (or, under stricter assumptions, the ATET) can be estimated as the difference between mean (predicted) earnings of return migrants and their matched stayer counterparts. The theorem of Frölich (2007) mentioned in section 2 extends the results of Rosenbaum and Rubin (1983), who were first to show that matching on the whole set of control variables can be replaced by matching on the propensity score. The probit model serves only for 'balancing' distributions of control variables in the two sub-populations (treated and non-treated), and there is no need to have an instrument exogenous with respect to the outcome variable.

Initially, applications of PSM in economics were limited to evaluating labor market policies, in which case the interest is in ATET rather than just the unexplained gap, and more assumptions are needed (see Heckman, Ichimura, and Todd (1998), Heckman, LaLonde, and Smith (1999) for details). More recently, however a wider range of applications emerged, including among others:

- Returns to education or training (Blundell et al 2005; Muehler *et al* 2007);
- Gender wage gap (Zanutto 2006; Frölich 2007);
- Effect of remittances on poverty (Bertoli 2008).

Apart from relaxing the assumptions necessary to justify the PSM approach, Frölich (2007) has developed a weighted PSM estimator which can be used under non-random sampling (e.g. stratified household surveys or oversampling of treatment participants). Moreover, Frölich's results allow estimating the unexplained gap at various places of the distribution of the outcome variable.

There are several technical methods of performing matching. Caliendo and Kopeinig (2008) provide a very good practical guide. *Stata*® software for PSM has been developed by Becker and Ichino (2002), Leuven and Sianesi (2003), and Abadie et al. (2004). The main results reported in this paper (Table 5) have been obtained by matching on common support. Following suggestions in the literature (Frölich 2005;

Caliendo and Kopeinig 2008) and taking into account very large size of our control group, we have chosen the following two methods: (a) two-nearest-neighbors method[22]; (b) local linear regression matching[23]. Scores (predicted probabilities to be a return migrant) are generated by probit models[24]. The model for being a return migrant with worker characteristics and family instruments is found in Table 2; similar models for subpopulations, as well as models which control also for job characteristics, are available from the author on request. These models serve for balancing the covariates among return migrants and the control groups; they are purely descriptive and do not pretend on revealing causal relationships.

Quality of match is fairly good: after matching there is no significant difference in means for any of the matching variables; standartized bias (see e.g. Caliendo and Kopeinig 2008) after matching is below 5% for most variables and below 10% for almost all variables (there are few cases with standartized bias between 10 and 15% for female sub-sample, but in these cases absolute difference in means is very small).

[22] Using just the nearest neighbor works fine in the pooled sample, but does not sufficiently reduce the bias for some of the subpopulations. On the other hand, further increase of number of neighbors used does not notably change the results.

[23] The 'optimal' bandwidth (varying by sub-sample from 0.0017 to 0.0071) for estimating (Epanechnikov) kernel density of *p*-score in the control group has been used. Given that these widths are close to zero, and sample size is about 10,000, there is no not need to modify local linear regressions with the ridge terms (see Frölich 2005 for details).

[24] Taking into account that data come from a stratified survey, and following Frölich (2007), we have used survey weights when estimating propensity scores and gaps; results change very little when only gaps are weighted. Hence the probits can be used for descriptive purposes as well. Unweighted results are similar, while corresponding estimates of the unexplained earnings gaps are hilgher by one or two ponits.

Appendix 2. Selectivity corrections and exogeneity tests

In this appendix we deal with correcting earnings functions for selection into employment and for reporting earnings, as well as describe tests for exogeneity of the return migrants dummy. Table 2 presents the three probit selection models for the pooled sample.

Error correlations (presented in Table 3) are estimated from seemingly unrelated probits with exclusion restrictions (although bivariate probits without restrictions give virtually the same results). Other things equal, lack of Latvian citizenship has a strong negative effect on propensity to be a return migrant, while it has no effect on propensity to be employed or on reporting earnings. Belonging to ethnic minority has a negative effect on employability, but no effect in the other two models. The effect of living with a partner is negative on return migration, positive on employment, and absent on reporting earnings. The effects of living with children or alone are positive on employment but absent in the other two models.

The important finding is that, both in the models without job attributes and when these attributes are controlled for, unobserved factors affecting employment and reporting earnings are uncorrelated for both genders, as well as for the pooled sample. Hence the inverse Mills ratios can be used as selectivity correction terms in the earnings functions (Berman and Wolfe 1984; Tunali 1986), as long as each of the two selection models has its own significant instrument which does not affect earnings (after accounting for observed characteristics). Moreover, errors in the return migration model are uncorrelated with the ones in the reporting earnings model, and very weakly negatively correlated with errors in the employment model: without job-related controls, ρ is -0.110 for men and -0.165 for women; with job controls correlations disappears for men but survives for women.

We deal with the issues of selection into employment and non-reporting earnings first; testing exogeneity of the return migrant dummy follows shortly, but *a priori* the respective coefficients in Table 4 do not pretend to be unbiased. The dummy for having a family member with recent foreign work experience (respectively,

unemployment experience) has a strong negative effect in the employment model (respectively, strong positive effect on reporting earnings), see Table 2 for the pooled sample.[25] These dummies are valid instruments: Wald test does not reject the null that in the earnings functions both coefficients are jointly zeros, with *p*-value 0.46 in the pooled sample, 0.26 for men, and 0.86 for women, see Table 4. Hence, as explained above, amending the earnings functions with two inverse Mills ratios is sufficient to remove the bias associated with selection into employment and with non-response to the earnings question in the survey.

In the pooled and men's samples, these ratios are jointly insignificant in the earnings functions, with *p*-values 0.80 and 0.90 respectively; dropping them almost does not affect other coefficients and the error variance. We conclude that for the model without job attributes selection is not an issue in these cases. In line with the literature, selection seems somewhat more important among females: the coefficient on the Mills ratio associated with the response to the earnings question is positive (as expected) and significant at 10% (even at 5% when the second, insignificant, Mills ratio is dropped); joint insignificance of both Mills ratios is close to being rejected (*p*-value 0.11). However, other coefficients and the error variance again change very little when selection is ignored.

When job attributes are included, corrections for selection into employment and for responding to the question on earnings cannot be neglected. As expected, the correction for non-response to the earnings question is positive[26] (non-responders earn more than similar responders) and quite sizable. For instance, mean value of the non-selection hazard is 0.178 for employed men, who have answered the wage question and 0.368 for those who have not. Hence the predicted log(wage) of responder exceeds that of otherwise similar non-responder by 0.625*(0.363 - 0.177) = 0.124. The lower bound of the confidence interval for the error correlation between earnings function an the selection equation is 0.79 for men, 0.05 for women, and 0.47 in the pooled sample, while the upper bound exceeds unity.

[25] Results by gender are available on request.
[26] The lower bound of the confidence interval for the error correlation between the earnings and the selection equations is 0.79 for men, 0.05 for women, and 0.47 in the pooled sample, while the upper bound exceeds unity in all three cases.

The correction for being unemployed is negative for men but positive for women (significant in both cases). Calculations similar to the one above shows that expected wage of a (short-term) unemployed man (respectively, woman) is, on average, 10% below (respectively, 6% above) that of otherwise similar employed person. Positive sign for women should not come as a surprise in a very tight and ever tightening (at the survey time) labor market with sever shortage of unskilled workers: women who have chosen to search longer (recall, however, that only those unemployed for less than one year have been surveyed), plausibly could afford waiting for a good offer due to support from a partner or another family member.

The above results suggest that sample selection can (respectively, cannot) be ignored when testing exogeneity of the return migrant dummy, *M*, in the earnings function without (respectively, with) job attributes. Foreign experience and unemployment experience of family members are very significant determinants of being a return migrant (see Table 2), and we again use both as instruments (results are similar, with slightly larger errors, when just foreign experience is used). For the models with selectivity corrections (i.e. when job attributes are controlled or in the women's sample) we have enough additional instruments, because in these cases the dummies for living with children and living alone (in the women's sample – also living with a partner) are excluded from the earnings function.

We have performed three types of tests:

[1] A simple Hausman-like test of H_0: $\rho = 0$ based on IV estimate of the model {$w = \beta X + \alpha M + u$, $M = \gamma X + \theta Z + v$, $\text{corr}(u, v) = \rho$} with the linear probability model for *M* in the first stage;

[2] A test of H_0: $\rho = 0$ based on Procedure 18.1 in Wooldridge (2002), where the fitted probabilities from the first-stage probit model are used as instruments for *M* in the second stage (importantly, this test does not require the probit model to be correct, i.e. there are no distributional assumptions involved);

[3] A test of H_0: $\rho_1 = \rho_2 = 0$ based on Procedure 18.4 in Wooldridge (2002) built on Heckman (1978) two-step estimator for the switching regression model $w = \beta X + \alpha M + \delta M(X - \overline{X}) + \rho_1 M\Lambda_1 + \rho_2(1-M)\Lambda_2$, where Λ_1 and Λ_2 are migrants'

and stayers' non-selection hazards from the first-stage probit. Interaction terms reduce error variance and account for possible non-constant selection bias.

Each of the three tests has been performed twice, using in the first stage either just one instrument (foreign experience of a family member) or adding also unemployment experience, giving six specifications. In none of the specifications exogeneity of the return migrant dummy comes even close to being rejected (*p*-values from specification [3] with two instruments in the first stage are shown in Table 4).

We conclude that, conditional on demographic characteristics and education, return migrants are neither positively nor negatively selected in terms of earnings. This is true also when detailed job attributes controls are included in the model, although in this case the earnings functions should be corrected for selection into employment and for response to the earnings question.

Appendix 3. Causal effect of return migration on family and implied effect on earnings expected in absence of foreign experience

The aim of this Appendix is to show that stayers' expected average earnings would change by less than 1% if they would experience a change in observed characteristics equal to migration-driven systematic shift in characteristics of return migrants. This allows dropping the last term from formula (5) when estimating the causal effect of migration on earnings.

The characteristics in question are those used for matching stayers and return migrants when estimating the earnings gap. Let us start with the case when matching is by worker's and family characteristics, while job characteristics are not controlled (estimates "Matched [b]" in Table 5). The control variables which could be affected by migration experience include living arrangements (with partner; with children), as well as unemployment and migration experience of family members. Indeed, some partnerships might not survive time apart; on the other hand, return migrants might be financially better than stayers prepared to marry, so the sign of the migration effect on living arrangements is ambigous *a priori* (as shown below, the first story line appears to be closer to reality). The other two effects should be positive: Network effect

suggests a positive causality from migration experience of a person to incidence of such experience among his relatives, while remittances might induce voluntary separations of migrants' family members from their jobs and unemployment spells later on. Migration-driven changes (i.e., the *ATET*s) in the four above mentioned variables have been estimated by PSM, using stayers who plan to move as a control group; such choice justifies assuming away the selection effects, at least in the first approximation. The set of variables used for this [auxiliary] matching includes gender, age, education (eight categories), ethnicity (four groups), region (5), and degree of urbanization (4), variables which are themselves either strictly exogenous or change very little on average in the short run, especially if only changes caused by foreign experience are considered. As for regions, we have evidence from another survey based on a sample of 187 return migrants who have worked in UK or Ireland; just 6% have changed region of residence during the three years before survey.

The PSM estimates reported as Δx in Table 6 confirm very strong (almost 20 percentage points on average) and significant positive effects of return migration on foreign and unemployment experience of family members. The effects on living with partner or with children are smaller in size and both negative, the former significant only among women, while the latter – among men. If, despite our choice of the control group, there are some selection effects left, they, plausibly, have the same signs as estimated gaps in all four cases, so the causal effects would be smaller.

Next, we use stayers' expected earnings $E[\ln W | M = 0]$ (obtained from regressions with worker and job controls) corrected for selection into paid employment and reporting earnings in the survey, see Appendix 2 for details. These are regressed on the set of control variables used for matching stayers to return migrants when estimating the earnings gap; estimated coefficients are reported in Table 6 (panel A) as γ; only coefficients on variables suspected to be affected by foreign experience in a systematic way are shown. The overall average effect of change in these variables on expected earnings is then calculated as (sum of) $\gamma\Delta x$. The estimated effects are negligible: –0.1% on average, –0.6% for men, and 0.2% for women. Conservative upper and lower bounds obtained by keeping only positive or only negative $\gamma\Delta x$ (thus accounting e.g. for possible selection effects) are also within ±1% range. This

concludes the proof of our claim that the effect of endogenous control variables on earnings can be ignored when the earnings gaps are conditioned only on worker characteristics.

The case when job characteristics are controlled as well is handled in a similar way. Our working assumption, based on the nature of post-enlargement migration, is that the *distribution* of return migrants among economic sectors and occupations is not different from what it would have been in absence of foreign experience (at the individual level some professional mobility could have been triggered though). Yet there is a range of job characteristics which could have been affected in a systematic way: (i) method of finding the job; (ii) language skills (Latvian, Russian, and English) required for the job and adequacy of worker's language skills; (iii) commuting distance; (iv) job tenure; (v) enterprise size and ownership sector; (vi) contract type; (vii) union membership. To save space, we report only estimated (causal) effects on earnings $\gamma\Delta x$ for each of these factors (Table 6, panel B).

It appears that migration driven changes in each of these factors, except contract type and union membership, have statistically significant effect (positive for factors (i) - (iii), negative for factors (iv) and (v)) on expected earnings. However, these effects are fairly small and sum up to 0.8% on average, 3% for men, and –0.5% for women. Moreover, conservative upper and lower bounds obtained by keeping only positive or only negative $\gamma\Delta x$ are also small compared to estimated unexplained earnings gaps : 3.0% and –2.1% on average, 5.3% and -1.7% for men, 2.8% and -3.4% for women.

Table 1: Mean characteristics of economically active population: Return migrants vs. stayers, 2006/Q4-2007/Q1

	All			Men			Women		
Have worked abroad?	No	Yes	Dif.	No	Yes	Dif.	No	Yes	Dif.
Sample size	9708	469		4062	253		5646	216	
Female	0.542	0.432	***	0.000	0.000		1.000	1.000	
Age	42.57	33.37	***	41.19	34.05	***	43.74	32.47	***
Living with a partner	0.619	0.513	***	0.662	0.563	***	0.583	0.447	***
Living with children	0.475	0.421	**	0.425	0.416		0.517	0.428	**
Ethnic minority	0.447	0.406		0.443	0.417		0.451	0.392	
Interviewed in Russian	0.415	0.362	**	0.420	0.385		0.412	0.330	**
Non-citizen	0.199	0.122	***	0.209	0.161	*	0.190	0.071	***
Completed education									
Less than secondary	0.188	0.132	***	0.230	0.149	***	0.153	0.109	*
Secondary	0.607	0.583		0.624	0.621		0.592	0.532	
Tertiary	0.203	0.282	***	0.143	0.230	***	0.254	0.351	***
Lives in: Riga	0.336	0.266	***	0.326	0.250	**	0.345	0.286	
Another big city	0.172	0.233	***	0.167	0.245	***	0.177	0.217	
Small city	0.183	0.205		0.186	0.243	*	0.181	0.156	
Rural area	0.308	0.297		0.321	0.263	*	0.297	0.341	
Riga region	0.489	0.350	***	0.483	0.335	***	0.494	0.369	***
Vidzeme region	0.105	0.086		0.102	0.112		0.106	0.052	***
Latgale region	0.159	0.196	*	0.162	0.195		0.157	0.196	
Zemgale region	0.113	0.275	***	0.116	0.255	***	0.110	0.302	***
Kurzeme region	0.135	0.094	***	0.137	0.103	*	0.133	0.082	***
Union member	0.096	0.061	***	0.064	0.066		0.124	0.055	***
Employed	0.911	0.838	***	0.917	0.862	**	0.905	0.807	***
Job characteristics									
Highly skilled non-manual	0.251	0.278		0.187	0.228		0.305	0.344	
Low-skilled non-manual	0.255	0.271		0.103	0.122		0.385	0.467	**
Skilled manual	0.459	0.457		0.508	0.477		0.417	0.432	
Unskilled manual	0.185	0.133	***	0.237	0.191	*	0.141	0.058	***
Has direct subordinates	0.134	0.317	***	0.148	0.320	***	0.123	0.313	***
Permanent contract	0.725	0.629	***	0.687	0.575	***	0.757	0.699	
No written contract	0.113	0.080	**	0.142	0.106		0.088	0.046	***
Works in another district	0.068	0.185	***	0.092	0.219	***	0.047	0.140	***
Needs just native language	0.418	0.301	***	0.475	0.349	***	0.370	0.238	***
Works in Agriculture	0.086	0.066		0.114	0.089		0.063	0.035	*
Works in Industry	0.170	0.189		0.212	0.232		0.135	0.133	
Works in Construction	0.113	0.134		0.226	0.226		0.018	0.013	
Works in Market services	0.442	0.471		0.350	0.379		0.520	0.592	*
Works in Public Services	0.178	0.121	***	0.089	0.055	**	0.254	0.207	
During the last three years:									
The person has been without work for more than a year or has been registered unemployed	0.143	0.270	***	0.135	0.277	***	0.150	0.262	***
A family member has been unemployed	0.155	0.483	***	0.171	0.487	***	0.142	0.478	***
A family member has been (or is) working abroad	0.103	0.447	***	0.093	0.420	***	0.112	0.481	***

Notes: By design, the sample includes only employees or job-seekers unemployed for less than one year. For the unemployed, job-related variables refer to the last job; results change very little if only employed are used.
***, **, * - the null hypothesis of equality of means is rejected at the 1%, 5%, 10% level, respectively.

Table 2: Selection equations (seemingly unrelated probits without job attributes, pooled sample)									
Dependent variable Y	Has worked abroad during the last 3 years			Employed wage earner			Reports earnings (in the current or last job)		
Pr(Y=1)	0.048			0.906			0.912		
	Coef.	S.E.	Sig.	Coef.	S.E.	Sig.	Coef.	S.E.	Sig.
Female	-0.270	0.053	***	-0.166	0.043	***	0.216	0.040	***
Education (vs. secondary professional)									
Basic or less	-0.388	0.125	***	-0.473	0.068	***	0.074	0.074	
Unfinished secondary	-0.068	0.117		-0.453	0.073	***	0.131	0.087	
Secondary general	0.148	0.073	**	-0.048	0.057		0.177	0.058	***
Unfinished tertiary	0.395	0.111	***	0.028	0.115		-0.212	0.101	**
Tertiary (first level)	0.266	0.113	**	0.114	0.108		-0.318	0.091	***
Tertiary (second level)	0.302	0.081	***	0.276	0.075	***	-0.182	0.059	***
Master or doctor degree	0.280	0.194		1.398	0.348	***	-0.357	0.133	***
Age	-0.015	0.002	***	0.065	0.009	***	-0.014	0.009	
Age squared/100				-0.079	0.010	***	0.023	0.011	**
Lives with a partner	-0.167	0.055	***	0.115	0.053	**			
Lives with children				0.163	0.048	***			
Lives alone				0.186	0.073	**			
Ethnic minority				-0.142	0.046	***			
Non-citizen	-0.216	0.082	***						
During the last three years:									
Has worked abroad				-0.250	0.088	***			
A family member has been (or is) working abroad	0.741	0.063	***	-0.319	0.061	***			
A family member has been unemployed	0.522	0.061	***				0.188	0.058	***
Constant	-1.533	0.130	***	0.395	0.181	**	1.358	0.194	***
Other controls	Region (5); Degree of urbanization (4)								
Pseudo R2	0.1929			0.0985			0.0407		
Sample size	10177			10177			10177		
Notes: Robust standard errors are reported in all tables hereafter. ***, **, * - estimates different from zero at the 1%, 5%, 10% level, respectively. Omitted regressors (exclusion restrictions) are very insignificant if included.									

Table 3: Error correlations in seemingly unrelated probits, by gender									
	Men and women (N=10177)			Men (N=4315)			Women (N=5862)		
Controls	Worked abroad - Employed	Worked abroad - Reports earnings	Employed - Reports earnings	Worked abroad - Employed	Worked abroad - Reports earnings	Employed - Reports earnings	Worked abroad - Employed	Worked abroad - Reports earnings	Employed - Reports earnings
Basic (as in Table 2)	-0.134*** (0.045)	-0.055 (0.044)	-0.030 (0.038)	-0.110 (0.069)	-0.061 (0.058)	0.001 (0.055)	-0.165*** (0.057)	-0.030 (0.069)	-0.040 (0.053)
Basic + Job attributes	-0.077 (0.047)	-0.037 (0.045)	-0.016 (0.041)	-0.030 (0.072)	-0.054 (0.061)	0.019 (0.063)	-0.156 (0.244)	-0.016 (0.068)	-0.027 (0.053)
Notes: Results are virtually unchanged if bivariate probits without exclusion restrictions are applied.									

Table 4: Earnings functions corrected for selection into paid employment and non-reporting earnings									
	All			Men			Women		
Log (monthly earnings)	Coef.	S.E.	Sig.	Coef.	S.E.	Sig.	Coef.	S.E.	Sig.
	Model A: Without job attributes controls								
Female	-0.390	0.021	***						
Education (vs. secondary professional)									
Basic or less	-0.397	0.032	***	-0.386	0.045	***	-0.382	0.041	***
Unfinished secondary	-0.288	0.032	***	-0.275	0.053	***	-0.275	0.038	***
Secondary general	-0.115	0.020	***	-0.077	0.027	***	-0.113	0.025	***
Unfinished tertiary	0.142	0.036	***	0.132	0.079	*	0.177	0.037	***
Tertiary (first level)	0.226	0.039	***	0.079	0.054		0.277	0.042	***
Tertiary (second level)	0.339	0.024	***	0.241	0.043	***	0.362	0.027	***
Master or doctor degree	0.626	0.055	***	0.652	0.110	***	0.588	0.062	***
Age	0.027	0.004	***	0.027	0.005	***	0.028	0.006	***
Age squared/100	-0.042	0.005	***	-0.045	0.007	***	-0.040	0.007	***
Lives with a partner	0.032	0.014	**	0.098	0.031	***	--	--	--
Lives with children	0.038	0.015	***	0.069	0.029	**	--	--	--
Ethnic minority	-0.004	0.016		-0.003	0.027		-0.009	0.020	
Non-citizen	-0.036	0.019	*	0.011	0.029		-0.073	0.023	***
Region (vs. Riga region)									
Vidzeme	-0.106	0.025	***	-0.157	0.049	***	-0.057	0.026	**
Latgale	-0.290	0.043	***	-0.389	0.069	***	-0.243	0.032	***
Zemgale	-0.038	0.024		-0.029	0.041		-0.045	0.027	*
Kurzeme	-0.129	0.024	***	-0.130	0.037	***	-0.105	0.029	***
Has worked abroad (*M*)	0.140	0.037	***	0.184	0.053	***	0.080	0.048	*
Non-selection hazard: paid employment (*Mills1A*)	0.103	0.154		0.092	0.241		0.137	0.176	
Non-selection hazard: reports earnings (*Mills2A)*	0.038	0.318		0.104	0.340		0.527	0.313	*
Other controls	Degree of urbanization (4 categories)								
H0: *Mills1A* = *Mills2A*= 0	Prob > F = 0.800			Prob > F = 0.899			Prob > F = 0.119		
H0: *M* exogenous	--			--			Prob > F = 0.384		
	Model A without Mills1A, Mills2A								
Has worked abroad (*M*)	0.150	0.033	***	0.190	0.047	***	0.098	0.042	**
H0: Omitted family related variables = 0 if included	Prob > F = 0.460			Prob > F = 0.263			Prob > F = 0.861		
H0: *M* exogenous	Prob > F = 0.763			Prob > F = 0.601			Prob > F = 0.754		
	Model B: With job attributes controls (excl. supervisory responsibilities)								
Has worked abroad (*M*)	0.155	0.033	***	0.187	0.046	***	0.055	0.041	
Mills1B (paid employment)	-0.185	0.110	*	-0.430	0.142	***	0.242	0.121	**
Mills2B (reports earnings)	0.455	0.132	***	0.652	0.154	***	0.275	0.130	**
H0: *M* exogenous	Prob > F = 0.596			Prob > F = 0.778			Prob > F = 0.332		
Additional controls: Tenure (5 categories); Type of contract (7); Sector of activity (25); Ownership sector (4); Enterprise size (4); Occupation (27); Job location vs. residence (3); Need to use languages on the job; Latvian, Russian, and English language skills (3 categories each); Way of finding the job (6)									
	Model C: With job attributes and supervisory responsibilities controls								
Has worked abroad (*M)*	0.125	0.032	***	0.196	0.046	***	0.028	0.041	
Mills1C (paid employment)	-0.195	0.106	*	-0.659	0.136	***	0.232	0.125	**
Mills2C (reports earnings)	0.488	0.129	***	0.640	0.153	***	0.307	0.129	**
H0: *M* exogenous	Prob > F = 0.559			Prob > F = 0.956			Prob > F = 0.290		
Sigma (*A/ B/ C*)	0.462 / 0.417 / 0.412			0.494 / 0.442 / 0.440			0.419 / 0.371 / 0.367		
Sample size	8447			3525			4922		
Notes: ***, **, * - estimates different from zero at the 1%, 5%, 10% level, respectively. Non-selection hazards are derived from models reported in Table 2 and similar ones by gender (with extra controls in Models B, C; in women's sample, both family related instruments are significant and kept in the earnings reporting model).									

Table 5: Matching results. Outcome variable: Predicted Log (earnings) [a]

Population	Sample	Return migrants	Control group [d]	Difference [d]	Difference [e]	S.E. [d]	Earnings gap, % [d, e]	N [f]
All	Unmatched	5.420	5.135	0.285		0.034	33.3	9,689
	Matched [b]	5.423	5.290	0.132	0.152	0.036	14 to 16	464
	Matched [c]	5.420	5.276	0.144	0.126	0.037	13 to 15	458
Men	Unmatched	5.617	5.314	0.303		0.045	35.4	4,055
	Matched [b]	5.617	5.436	0.182	0.224	0.047	20 to 25	249
	Matched [c]	5.612	5.474	0.138	0.156	0.048	15 to 17	247
Women	Unmatched	5.176	4.984	0.192		0.049	21.2	5,634
	Matched [b]	5.176	5.117	0.059	0.058	0.047	6	206
	Matched [c]	5.183	5.116	0.067	0.079	0.052	7 to 8	195
Secondary or lower education	Unmatched	5.278	5.034	0.244		0.045	27.6	7,278
	Matched [b]	5.278	5.132	0.146	0.162	0.044	16 to 18	288
	Matched [c]	5.280	5.169	0.111	0.133	0.048	12 to 14	287
Tertiary or unfinished tertiary education	Unmatched	5.701	5.436	0.265		0.056	30.3	2,398
	Matched [b]	5.701	5.516	0.186	0.214	0.061	20 to 24	161
	Matched [c]	5.698	5.480	0.218	0.259	0.053	24 to 30	173
Manual workers	Unmatched	5.460	5.103	0.358		0.062	43.0	4,210
	Matched [b]	5.460	5.203	0.257	0.245	0.063	28 to 29	178
	Matched [c]	5.484	5.261	0.223	0.218	0.067	24 to 25	178
Non-manual workers	Unmatched	5.371	5.168	0.203		0.045	22.5	5,094
	Matched [b]	5.371	5.266	0.105	0.101	0.048	11	247
	Matched [c]	5.368	5.293	0.074	0.090	0.050	8 to 9.5	243
Ethnic Latvians	Unmatched	5.427	5.123	0.305		0.046	35.6	6,079
	Matched [b]	5.427	5.258	0.169	0.169	0.048	18.4	287
	Matched [c]	5.425	5.295	0.130	0.148	0.049	14 to 16	295
Ethnic minorities	Unmatched	5.417	5.150	0.267		0.053	30.6	3,605
	Matched [b]	5.417	5.256	0.161	0.177	0.054	17 to 19	164
	Matched [c]	5.411	5.275	0.136	0.163	0.061	14 to 18	163

Notes: [a] Earnings are predicted separately for return migrants and stayers by gender (or, when relevant, by education). The earnings functions (same for both matching procedures [b], [c]) control all worker and job characteristics mentioned below, as well as presence of supervisory responsibilities, and are corrected for selection into paid employment and for non-reporting earnings. [b] Matching on common support by worker's demographic characteristics and education, as well as foreign experience and unemployment experience of family members. [c] Matching on common support by worker's demographic characteristics and education; tenure (5 categories); type of contract (7); sector of activity (25); ownership sector (4); enterprise size (4); occupation (27); job location vs. residence (3); way of finding the job (6); need to use languages on the job (yes/no); Latvian, Russian, and English language skills (3 categories each), as well as foreign experience and unemployment experience of family members. [d] Two nearest-neighbors matching; with [e] Local linear regression matching; the 'optimal' bandwidth (varying by sub-sample from 0.0017 to 0.0071) for estimating density of *p*-score in the control group has been used. Both in [d] and [e], *p*-weights are applied when estimating propensity scores and gaps; results change very little when only gaps are weighted. [f] Sample size on common support for the control group (before matching) and for the return migrants.

Table 6: Migration impact on family and job characteristics: Matching estimates and implied effect on earnings expected in absence of foreign experience

	A. Family characteristics only								
	Men and women			Men			Women		
x	Δx	γ	$\gamma\Delta x$	Δx	γ	$\gamma\Delta x$	Δx	γ	$\gamma\Delta x$
x_1: Lives with a partner	-0.038 (0.033)	0.029 (0.007)	-0.001	-0.004 (0.042)	0.116 (0.012)	-0.001	-0.077 (0.050)	0.001 (0.007)	-0.000
x_2: Lives with children	-0.051 (0.032)	0.037 (0.006)	-0.002	-0.052 (0.017)	0.046 (0.010)	-0.002	-0.025 (0.048)	0.010 (0.007)	-0.000
x_3: A family member worked(-s) abroad	0.183 (0.028)	0.033 (0.009)	0.006	0.216 (0.039)	0.011 (0.016)	0.002	0.165 (0.046)	0.065 (0.010)	0.011
x_4: A family member has been unemployed	0.193 (0.028)	-0.023 (0.008)	-0.004	0.189 (0.036)	-0.030 (0.012)	-0.006	0.298 (0.046)	-0.030 (0.009)	-0.009
Pos. effects on E[ln W]			0.006			0.002			0.011
Neg. effects on E[ln W]			-0.007			-0.009			-0.009
Total effect on E[ln W]			-0.001			-0.006			0.002
(s.e.), as if Δx = const			(0.002)			(0.004)			(0.003)

	B. Job and family characteristics (by factor)					
	Men and women		Men		Women	
x (# categories)	$\gamma\Delta x$	s.e.[a]	$\gamma\Delta x$	s.e.[a]	$\gamma\Delta x$	s.e.[a]
Adequacy of language skills: English (3)	0.0112	(0.0004)	0.0188	(0.0006)	0.0168	(0.0004)
Latvian & Russian (3 each)	0.0048	(0.0007)	0.0026	(0.0003)	0.0048	(0.0005)
Need languages other than native (2)	0.0038	(0.0003)	0.0013	(0.0001)	-0.0006	(0.0001)
Way of finding job (6)	0.0057	(0.0007)	0.0142	(0.0010)	-0.0141	(0.0003)
Job location vs. residence (3)	0.0038	(0.0003)	0.0125	(0.0005)	-0.0033	(0.0003)
Ownership sector (4)	0.0011	(0.0002)	-0.0006	(0.0002)	0.0015	(0.0002)
Tenure (5)	-0.0126	(0.0007)	0.0031	(0.0004)	-0.0124	(0.0005)
Enterprise size (4)	-0.0063	(0.0004)	-0.0071	(0.0004)	-0.0036	(0.0004)
Contract type (7)	-0.0014	(0.0009)	0.0004	(0.0006)	0.0013	(0.0003)
Union member (2)	0.0000	(0.0003)	-0.0003	(0.0001)	0.0004	(0.0001)
Family (see x_1 - x_4 above)	-0.0008	(0.0010)	-0.0091	(0.0013)	0.0032	(0.0008)
Total positive effects on E[ln W]	0.0304	(0.0010)	0.0529	(0.0013)	0.0280	(0.0011)
Total negative effects on E[ln W]	-0.0211	(0.0016)	-0.0171	(0.0013)	-0.0339	(0.0012)
Total effect on E[ln W]	0.0076	(0.0019)	0.0290	(0.0018)	-0.0051	(0.0072)

	Matching used to derive Δx		
Control variables	Gender, age, education (8 categories), ethnicity (3), region (5), urbanization (4)		
N (Return migrants)	456	237	213
N (Stayers, plan to move)	663	286	377
	Stayers' auxiliary earnings functions used to derive $\gamma = d\mathrm{E}[\ln W \mid M = 0, X] / d X$		
Control variables	**A:** Those used for matching, and family; **B:** Also those listed above in column x, occupation (27), and sector of activity (25)		
N	9689	4055	5634
R-squared (**A / B**)	0.7154 / 0.9269	0.6686 / 0.9678	0.7297 / 0.9837

Notes: The table allows estimating the last term in formula (5). Migration effects Δx are estimated by PSM on common support (two nearest-neighbors, with p-weights). Expected earnings E[ln W] used in derivation of γ are obtained from regressions with worker and job controls corrected for selection into paid employment and reporting earnings in the survey (see Appendix 2 for details). Significance of family members' experience variables x_3, x_4 in the auxiliary earnings functions is entirely due to the selection effects. [a] Robust standard errors assuming Δx = const.

Table 7: Decomposition of earnings gap between return migrants and stayers without job attributes controls

	Men			Women		
Have worked abroad?	No	Yes	Dif.	No	Yes	Dif.
Education						
Basic or less	-0.436	-0.147		-0.440	-0.654	
Unfinished secondary	-0.294	-0.306		-0.353	-0.230	
Secondary general	-0.112	0.031		-0.213	-0.378	
Secondary professional	-0.033	-0.037		-0.074	-0.157	
Unfinished tertiary	0.109	0.224		0.072	0.134	
Tertiary (first level)	0.063	-0.034		0.239	0.246	
Tertiary (second level)	0.206	0.267		0.295	0.258	
Master or doctor degree	0.651	0.002	***	0.517	0.780	
Age – 41	-0.010	-0.004	*	-0.005	-0.003	
[(Age – 41) squared]/100	-0.045	-0.068		-0.035	-0.061	
Lives with a partner [a]	0.090	0.156		--	--	
Lives with children [a]	0.065	0.016		--	--	
Ethnic minority [a]	0.003	-0.060		-0.011	0.166	**
Non-citizen [a]	0.013	-0.024		-0.068	-0.367	*
Region						
Riga region	0.134	0.198		0.089	0.062	
Vidzeme region	-0.021	0.013		0.021	-0.009	
Latgale region	-0.223	-0.355		-0.112	-0.119	
Zemgale region	0.106	0.114		0.050	-0.065	
Kurzeme region	0.004	0.031		-0.048	0.131	
Degree of urbanization						
Capital city	0.067	0.048		0.118	-0.016	
Big city	0.052	-0.134		0.013	-0.099	
Small city	-0.009	0.026		-0.038	0.031	
Rural area	-0.110	0.060	*	-0.093	0.084	**
Constant: ln(Base earnings)	5.468	5.663	***	5.087	5.225	**
Sample size	3525			4922		
	Gaps in ln(earnings)			***Gaps in ln(earnings)***		
Factors	[1]	[2]	Total	[1]	[2]	Total
Education	0.053	0.051	0.104	0.107	-0.034	0.073
Age	0.078	-0.097	-0.020	0.051	-0.085	-0.035
Living arrangements	-0.007	0.009	0.002	--	--	--
Ethnicity and citizenship	-0.001	0.035	0.034	0.009	0.114	0.123
Region	-0.016	0.006	-0.010	-0.007	-0.031	-0.038
Degree of urbanization	-0.001	0.007	0.006	-0.002	-0.005	-0.007
Base earnings [b]	0.000	0.194	0.194	0.000	0.139	0.139
Total	0.105	0.204	0.310	0.158	0.098	0.256

Notes: Normalization is performed as in Yun(2005), to allow detailed decomposition. However, coefficients on binary variables (marked [a]) are shown in non-transformed form. [b] Base earnings refer to a worker aged 41, with professional secondary education, ethnic Latvian, Latvian citizen, living in rural area of Riga region.

***, **, * - Equality of normalized coefficients is rejected at the 1%, 5%, 10% level, respectively.

[1] - characteristics' effect (the explained gap), $\beta_0(E[Z \mid M=1]-E[Z \mid M=0])$.

[2] – the coefficients' effect (the unexplained gap), $(\beta_1-\beta_0)E[Z \mid M=1]$.

Table 8: Unexplained earnings gap between return migrants and stayers: PSM estimates at various points of earnings distribution

Percent

Workers	Job controls	Percentiles				
		10	25	50	75	90
All	No	-19.8	-3.4	13.9	39.7	75.7
	Yes	-23.6	-5.7	18.4	45.3	65.9
Men	No	-20.0	-8.4	20.0	46.9	88.2
	Yes	-20.8	-9.5	14.6	33.2	68.1
Women	No	-26.1	-14.1	6.8	39.4	57.7
	Yes	-19.0	-10.9	5.0	36.3	47.6
Education ≤ secondary	No	-17.1	-4.1	12.2	43.7	60.6
	Yes	-23.7	-9.2	9.1	40.0	60.1
Education > secondary	No	-25.7	-10.8	23.3	58.5	102.6
	Yes	-19.7	2.7	22.2	63.1	94.2
Manual workers	No	-24.5	-3.5	34.9	57.9	101.0
	Yes	-23.1	-4.2	28.5	56.0	93.6
Non-manual workers	No	-30.7	-10.2	11.6	52.2	73.0
	Yes	-28.6	-4.0	4.7	37.5	56.0
Ethnic Latvians	No	-16.5	-7.6	23.5	40.8	69.5
	Yes	-25.6	-8.4	20.6	34.6	55.0
Ethnic minorities	No	-21.9	3.1	8.6	47.9	85.2
	Yes	-24.8	1.2	10.6	41.2	84.7

Notes: Results obtained by comparing, on common support, [*p*-weighted] percentiles of distributions of return migrants' earnings and covariate-adjusted PSM estimate of stayers' earnings. Two nearest-neighbors matching has been applied; see details in Table 5 (including Notes).

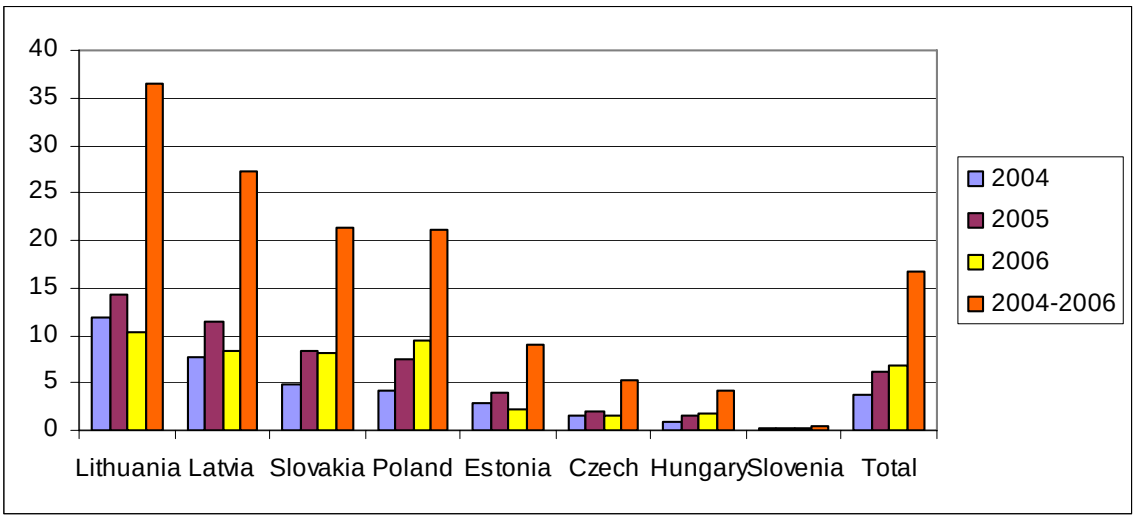

Figure 1. Workers from the new EU member states registered in UK, 2004-2006 (per 1000 economically active population of the sending country in 2004).
Source: Home Office et al. (2008) and own calculation.