Is it worth coming back?

Wage returns of higher education graduates' mobility patterns

Very preliminary draft – not for citation

Ana Beatriz Rocha[1]

Ricardo Biscaia[2]

Abstract

Is it worth coming back for graduates who study away from home? Using the Portuguese EUROGRADUATE survey, we compare back movers, who return home after studying elsewhere, with early movers, who remain in their study region, and repeat movers, who relocate to a third region for two distinct cohorts, one and five years after graduation. To provide a differential analysis, an entropy-balanced technique is used, and the resulting wage differentials are decomposed between geographical sorting, graduate characteristics and the residual. The wage decomposition shows that the answer depends on the comparison. Against early movers, geographic sorting explains most of the gap, as early movers disproportionately stay in Lisbon and Porto, the labour markets with the deepest concentration of skilled jobs, so the apparent penalty mostly reflects location, not the choice to return. Against repeat movers, geographic sorting plays no role, as graduate characteristics such as family background and prior education account for the gap instead. Only the five-year after graduation cohort comparison with repeat movers shows a significant adjusted disadvantage, about 8%, concentrated in the lower and middle of the wage distribution and turning positive at the 90th percentile. One year after graduation, no adjusted cost is detectable. Back movers are also disproportionately drawn from peripheral and insular regions, while Lisbon graduates rarely return, largely because they rarely leave.

Keywords: graduate mobility; return migration; Portugal; regional mobility; wage returns; early-career graduates

JEL codes: R23, I23, J24, J61

[1] CIPES – Centre for Research in Higher Education Policies, Matosinhos; NIPE – Centre for Research in Economics and Management, University of Minho, Braga.

[2] CIPES – Centre for Research in Higher Education Policies, Matosinhos; CEF.UP – Centre for Economics and Finance at University of Porto; Porto, Portugal

Introduction

In many European countries, higher education is geographically distributed, while economic activity is not. This creates a structural tension for graduates: the region where they study, the region where opportunities concentrate, and the region they come from need not coincide. How graduates navigate that tension matters not only for individual labour-market outcomes but also for the spatial distribution of human capital. Yet wage differences across these mobility trajectories are difficult to interpret. They may reflect differences associated with the mobility sequence itself, or they may arise because different trajectories sort graduates into labour markets with different wage structures.

Portugal provides a particularly useful setting in which to examine this distinction. Higher education is spread across the territory, but economic activity is strongly concentrated in the two main metropolitan areas, sharpening the trade-off between returning home and remaining in higher-opportunity destinations. Against this background, the paper pays particular attention to graduates who return home after studying away, asking how much of their observed wage disadvantage reflects geographical sorting, observed graduate characteristics, and how much persists as an adjusted differential within comparable labor markets.

The main research question is "how do returns between graduates in higher education differ between graduates with different mobility patterns?" More specifically, we focus on the outcomes of those who returned to their home region after studying abroad. To do so, we use the Portuguese version of the Eurograduate dataset, which contains comprehensive information on Portuguese higher education graduates. The sample covers 13% of the graduate population in the academic years of 2016/2017 and 2020/2021, enquired at October 2022, therefore corresponding to two graduate cohorts 5 and 1 year after completing their studies. Beyond knowing the origin region, HEI region, and job region at NUTS III level, the Eurograduate has an important set of sociodemographic and educational information, allowing us to better understand both the determinants of wages and isolate the effect of mobility.

We address this question empirically in two steps. We first estimate OLS wage regressions with a full five-category mobility typology to benchmark the association between mobility trajectories and log hourly wages. We then restrict the sample to graduates who studied outside their home region and use entropy balancing to construct two pairwise comparisons: (1) back movers versus early movers, and (2) back movers versus repeat movers, decomposing the raw wage gap. Both cohorts are estimated separately throughout.

The results show that geographic sorting into or out of metropolitan labour markets is a statistically significant part of the wage gap associated with returning home, specifically when the comparison is with graduates who remained in the study region - in both cohorts - but not when the comparison is with graduates who moved on to a third region; family background and educational trajectory are significant in all four cohort-comparison cells. A residual wage disadvantage that survives adjustment for observed characteristics appears only in the comparison of the 2016-2017 cohort with graduates who continued to a third region; looking at the distribution, the penalty is concentrated in the lower and middle of the wage distribution and reverses at the top. The regional pattern of returning is itself far from uniform, as the share of graduates who return home ranges from under 5% to over 44% across NUTS III home regions, lowest in the Lisbon metropolitan area.

Mobility trajectories, spatial sorting, and expected wage mechanisms

As raised correctly by Rodríguez-Pose and Tselios (2010), “In principle, the individual pecuniary returns to education should be independent of whether the individual is a local or a migrant. Individuals with similar levels of qualifications working in similar jobs should expect to earn a similar wage regardless of the region or country of origin.” (Rodríguez-Pose and Tselios, 2010, p. 413). However, some specific factors associated with migration could be relevant.

Migration could primarily be sought to improve living conditions, of which wages are a part. This, in turn, leads to a self-selection of mobility, which leads those who are more responsive to economic incentives (Nakosteen et al., 2008) and more immune to the psychological costs of migration to be more prone to move (Rodriguez-Pose & Tselios, 2010). More specifically, for labor market purposes, mobility might be associated with a better matching between qualifications and job demands, as the willingness to move allows for screening a higher number of jobs (Jauhiainen, 2011). As moving entails costs, it is expected that movers will only accept migrating job opportunities with considerable wage premiums. Arguably, the most important disadvantages of migration pertain to information, cultural, and legal barriers that migrants might face and specifically hinder their opportunities in the labor market. However, these are mostly related to international mobility, as internal migrants face few or none of those disadvantages.

Higher education graduates are perhaps the main population to assess the existing wage premium after mobility. Graduates, on the one hand, are expected to benefit more from an increased pool of job offers to better match their specific skills with their jobs, and on the other hand, are more willing to move, perhaps having lower thresholds to trigger a change of region, or even country. A compelling argument is in Mitze and Javakhishvili-Larsen (2020), since they find wage premiums for higher education graduates when changing between Danish municipalities, and find that this effect is stronger when compared to graduates from lower levels of qualification. Other studies address the effects of moving on wages for this population. Lehmer and Ludsteck (2011) show that the premium from changing region does not come only from changing jobs. They analyse, for Germany, the premium of changing regions compared to changing jobs within the same region, and conclude that there are additional wage-premium effects of region change beyond job change, and that this premium becomes fully effective after four years in the new region and is more pronounced for young graduates. Using a multi-European dataset, Rodríguez-Pose and Tselios (2010) found no significant differences in interregional mobility for a set of European countries. Across all levels of education, the authors find no significant differences between the wages of those who have not migrated and those who have migrated locally, especially after controlling for regional variables such as the educational endowment and wages per capita in the regions and neighbouring regions.

Mobility between the place of higher education and the job is, however, not the first relevant mobility a higher education graduate might do before reaching the labour market. Mobility between its original residence/secondary school and an HEI is another very relevant movement with potential consequences for labour market outcomes, particularly wages (Faggian et al., 2006). Knowledge of the local labour markets and cultures might be amplified by those intend to work in the same region that they had already studied and had always lived on, providing a more effective cost of job search (Sjaastad, 1962), while this can only be partially acquired if the graduate only studied in that region, but is not originally from there. Being non-local in HE also means a natural predisposition for moving, since the graduate had already moved once and does not have origin ties to the place they studied. Rehák and Eriksson (2020) show non-local graduates are more likely to migrate after graduation, in the Swedish context.

This sorting of graduates across study locations matters for wages because destination labour markets are not economically equivalent. A large literature on agglomeration economies shows that human capital and productivity are unevenly distributed across cities, and that graduates entering thicker, metropolitan labour markets can access better job matches and higher returns to skill than those entering thinner, peripheral ones (Berry & Glaeser, 2005; Faggian & McCann, 2006). Any wage difference associated with a given mobility trajectory can therefore reflect the type of destination labour market it leads to, rather than the act of moving itself.

Perhaps most relevant in the mobility to study in higher education is the sorting of graduates. Many Higher Education systems are stratified, meaning students select those HEIs they consider most prestigious, generating a concentration of talent that provides benefits for these students. As is often the case, these institutions are usually located where most economic indicators are favourable, leading students to move already to those places where they can make contact with richer and wider labour markets. Therefore, studying only the movement between HEI and job omits that a first-stage of movement, already leaning on towards the labour market, had already occurred in access to HE. An important reference for this argument is Ahlin et al. (2018), who, also in the Swedish context, study both mobilities and conclude that the sorting of the most high-ability graduates happens at the choice of study location, not at the job location. As the job location effect is mainly driven by job characteristics, the authors find that excluding Stockholm – the capital region – leads the job location effect to be non-significant.

This, in theory, makes it essential when studying wage effects of mobility to either consider both types of mobility simultaneously, or to have methods that allow controlling for selection mechanisms of HEI locations. Regarding the former, the most common approach in the literature is the usage of categorical classifications that include both types of mobilities, usually with all combinations from home-to-study and study-to-work (e.g. Kazakis and

Faggian (2017); Di Cintio and Grassi (2013); Jewell and Faggian (2014), Zhao and Hu (2019). That is also the approach followed by this paper.

A complementary strand instead measures mobility directly as the physical distance between residence, study, and work locations rather than through discrete trajectory categories; using this approach for Dutch graduates, Venhorst and Cörvers (2018) find that the positive wage returns to internal migration identified by OLS become statistically insignificant once self-selection is corrected for with an instrumental-variables strategy, illustrating how sensitive migration wage premiums can be to the treatment of self-selection.

In order to consider both patterns for mobility before and after higher education, Di Cintio and Grassi (2013) refined the typology from Faggian et al. (2006), and distinguished graduates among 5 categories: stayers (did not move either to study or work); early movers (moved to study, worked in the same region as HE); late movers (did not move to study, moved to work); back movers (moved to study and returned home to work); and movers (moved to study, and moved to work, but not for the home region). The geographic units used to define migration were tested with NUTS I and NUTS II regions. Using propensity score matching, the authors conclude that "early movers" earn 2,4% more than "stayers," and "late-movers" earn 13,9% more than "stayers." Returning home after studying in another region does not provide any premium compared to those who stayed from the start. Returning home after studying provides a wage penalty of 2,7%.

Kazakis and Faggian (2017) analyzed wage premiums for US labor markets using a similar division of mobility patterns, although with different names. In contrast with Di Cintio and Grassi (2013) and Jewell and Faggian (2014) for the UK, they conclude that the "late-movers" incur a wage penalty of 17,3% compared to "stayers", while the results for "movers" and "back-movers" are confirmed with wage premiums of 7,1% and 7,6%, respectively. The contrasting result emerges after controlling for the results using a first-stage selection equation, where the authors prove that higher wages in "late movers" are due to the mobility of the best workers, who seem to be relatively more concentrated on the "late movers" group. A similar approach from Zhao and Hu (2019) yielded similar patterns for the Chinese case.

Distinguishing back movers from graduates who continue moving onward to a third region is not only an empirical convenience but also reflects a conceptual distinction in the migration literature between return and onward, or repeat, migration. These two forms of continued mobility are associated with different search processes and can select different types of movers, since onward migrants may still be engaged in sequential labour-market search while return migrants are, by definition, no longer searching in the same location (Newbold, 2001; Hunt, 2004). Comparing back movers separately with early movers, who remain in the study region, and with repeat movers, who move on again, therefore tests conceptually distinct alternatives rather than a single generic non-return outcome.

These results show that mobility patterns capture meaningful differences in wage outcomes that are worth exploring in Portugal, whose interior-coastal divide and uneven, but reputationally stratified, distribution of higher education institutions make it a natural setting for this typology.

We pay particular attention to the effects of returning to a home region after studying elsewhere, since some NUTS III regions have no higher education institution of their own: for graduates from these regions, "returning home" and "having studied away" are not independent choices but two sides of the same constraint. What, then, are the labour-market outcomes for graduates working back in the region they came from, once they received their education elsewhere?

Conceptual framework

In this study, graduate mobility is modelled as two linked location choices, conditional on participation in and conclusion of higher education studies, as in Di Cintio and Grassi (2013). First, the individual studies in its home region or elsewhere. Second, conditional on the realised study location and educational trajectory, the graduate is observed working in the study region, the home region, or a third region.[3]

Let H_i, S_i, and W_i denote graduate i's home, higher education, and current workplace regions. Their combination defines five mutually exclusive trajectories: Stayers ($H_i = S_i = W_i$); Early movers ($H_i \neq S_i = W_i$); Late movers ($H_i = S_i \neq W_i$); Back movers ($H_i = W_i \neq S_i$); and Repeat movers ($H_i \neq W_i \neq S_i$). Although the OLS analysis describes all five trajectories, the counterfactual analysis focuses on graduates who studied outside their home region. Conditional on $A_i = \mathbf{1}\{S_i \neq H_i\} = 1$, these graduates follow one of three observed workplace paths:

$$J_i \in \{h, s, o\},$$

where h denotes returning to the home region (Back movers), s remaining in the study region (Early movers), and o moving onward to a third region (Repeat movers). The third region is individual-specific and therefore represents an onward-mobility path rather than a common geographical destination.

Workplace location is not randomly assigned, since graduates choose where to work partly in response to wage offers and partly because of unobserved features, costs, and preferences. Therefore, the characteristics of the workers observed in a given location depend on the returns available in the alternatives. This motivates adjusting for the observable determinants of location choice.

For the empirical analysis below, it is useful to partition the adjustment vector into two blocks, the geographic and the "other characteristics" blocks. The geographic block collects home and study metropolitan status, local higher education access, and home-to-study distance: the channel through which the spatial concentration of economic

[3] This is an analytical sequence rather than a complete chronology, as the data do not require higher education or the observed job to have been the individual's immediately preceding activity.

activity should operate. The "other characteristics" block collects personal and family background together with the educational trajectory realised before workplace choice. This partition motivates the sequential decomposition, which separates how much of any observed wage disadvantage traces to the geographic block from how much traces to the graduate characteristics block.

Empirical Strategy

The empirical analysis translates the two counterfactual wage comparisons into two pairwise designs, both restricted to graduates who studied outside their home region: back versus early movers, and back versus repeat movers. The main outcome is the logarithm of hourly wages. As in Di Cintio and Grassi (2013), the estimated wage differentials refer to employed graduates in the analytical sample; results are therefore conditional on employment and do not capture effects of mobility paths on the probability of employment. All analyses are conducted separately for the t+1 and t+5 cohorts.

We begin by describing the wage distributions of the five mobility configurations and estimating cohort-specific OLS wage regressions, which provide unadjusted and progressively adjusted benchmarks for the association between mobility trajectories and log hourly wages. With stayers as the omitted category, the back-versus-early and back-versus-repeat wage differences are obtained directly from the corresponding mobility-category coefficients. We compare specifications with home- and study-region fixed effects, with and without a wild cluster bootstrap, which is necessary given the small number of home- and study-region clusters.

Common support and entropy balancing

To analyse the effects of returning to the home region, we build comparison groups on common support, separately for back-versus-early and back-versus-repeat movers within each cohort. For each comparison c in {s,o}, we estimate the probability of being observed as a back mover using characteristics determined before the workplace-location decision; the intersection of the two groups' propensity-score ranges defines common support, and observations outside it are excluded to limit extrapolation. Figure A1 and Figure A2 in the Appendix show the resulting propensity-score overlap and common-support bounds for each cohort and comparison. The propensity score is used only to define support, not as the final weighting estimator.

Within the common support, we preprocess the data using entropy balancing (Hainmueller, 2012), which assigns non-negative weights to the comparison group so that the first three moments of the continuous adjustment variables and the category shares of the categorical variables match those of back movers, who retain unit weight. Early movers are reweighted in the return-versus-study-region comparison, and repeat movers in the return-versus-onward comparison. We favour entropy balancing over alternative preprocessing techniques[4] for three reasons. First, it attains exact balance on the targeted moments by construction, rather than balance that must be checked ex post and pursued by trial and error across specifications. Second, it preserves information in the preprocessed data: weights vary smoothly across units, so no observation is discarded for want of a close match. Third, because it is a preprocessing step rather than an estimator, the resulting weights carry into the various specifications and treatment effects reported below.

Because both comparisons condition on having studied away, the balancing variables include characteristics determined before higher education, the study region, and the educational trajectory realised before workplace choice; current-job characteristics (occupation, sector, contract type, part-time status) are excluded from the preferred specification because they may be determined jointly with the observed workplace path.

Wage gap decomposition

This decomposition follows the reweighting approach of DiNardo et al. (1996). We construct a counterfactual distribution by reweighting the comparison group to match the back movers' observed characteristics, and compare descriptive statistics of the resulting distributions directly.

Let $\mathcal{B}_c$ and $\mathcal{C}_c$ denote, respectively, the back-mover and comparison-group observations retained on common support for comparison c, with $N_{B,c}$ and $N_{C,c}$ the corresponding sample sizes, y_i the log hourly wage, and w_{ic} the entropy weight assigned to comparison observation i. We define

$$\hat{\mu}_{B,c} = \frac{1}{N_{B,c}} \sum_{i \in \mathcal{B}_c} y_i \,, \qquad \hat{\mu}_{C,c} = \frac{1}{N_{C,c}} \sum_{i \in \mathcal{C}_c} y_i,$$

and the entropy-standardised comparison mean as

$$\hat{\mu}^{EB}_{C \to B,c} = \frac{\sum_{i \in \mathcal{C}_c} w_{ic}\, y_i}{\sum_{i \in \mathcal{C}_c} w_{ic}}.$$

Descriptively, $\hat{\mu}^{EB}_{C \to B,c}$ is the comparison-group mean wage after standardising its observed pre-workplace characteristics to those of supported back movers. The raw wage gap can then be written as the exact accounting identity

$$\hat{G}^{raw}_c = \hat{\mu}_{B,c} - \hat{\mu}_{C,c} = \left(\hat{\mu}^{EB}_{C \to B,c} - \hat{\mu}_{C,c}\right) + \left(\hat{\mu}_{B,c} - \hat{\mu}^{EB}_{C \to B,c}\right) = \hat{G}^{comp}_c + \hat{G}^{res}_c.$$

[4] Mahalanobis distance matching, genetic matching, entropy balancing, matching or weighting on propensity score, and Mahalanobis distance matching on the estimated propensity score and orthogonalized covariates.

The observed-composition component, $\hat{G}_c^{comp}$, measures how much the comparison-group mean changes once reweighted to the back movers' profile; a negative value means back movers' characteristics are associated with lower wages under the comparison group's wage structure. The residual, $\hat{G}_c^{res}$, compares the back-mover mean with the entropy-standardised comparison mean; a negative value means back movers earn less than observationally comparable graduates on the alternative path.

This reweighting decomposition is related to, but distinct from, the coefficient-based Oaxaca-Blinder approach. Under consistency, overlap, and conditional mean independence of the alternative wage given the adjustment variables, $\hat{G}_c^{res}$ has an average-treatment-effect-on-the-treated interpretation; because these assumptions are strong and the analysis conditions on employment, we interpret it as an adjusted wage contrast rather than a causal return penalty.

All components are additive in log points; percentage differences, reported as $100[\exp(\hat{G}) - 1]$, are not. The share of the raw gap accounted for by observed composition is calculated on the log scale as $\hat{G}_c^{comp}/\hat{G}_c^{raw}$ and interpreted only when the components have economically meaningful signs.

Standard errors and percentile confidence intervals come from a nonparametric cluster bootstrap that resamples study-region clusters with replacement, re-estimating the propensity-score model, common-support bounds, entropy weights, and decomposition components in every replication; replications where the design stage does not converge are excluded.

Because of the specific interest in the spatial concentration of economic activity, we further split $\hat{G}_c^{comp}$ into a geographic and an "other characteristics" component. Within the same common-support sample, entropy balancing is applied twice from the same unweighted comparison group - once targeting only the geographic block, once targeting the full adjustment set - so that the raw gap decomposes as $\hat{G}_c^{raw} = \hat{G}_c^{geo} + \hat{G}_c^{other} + \hat{G}_c^{res}$, where $\hat{G}_c^{geo}$ is the geography-only-weighted comparison mean minus the unweighted mean, $\hat{G}_c^{other}$ is the fully-weighted mean minus the geography-only-weighted mean, and $\hat{G}_c^{res}$ keeps its definition above.

We complement the mean decomposition with the same reweighting logic applied to unconditional wage quantiles (P10, P25, P50, P75, P90), to assess whether any adjusted gap is uniform across the wage distribution or concentrated at particular points.

Data

Data source and survey design

This study draws on data from the EUROGRADUATE pilot survey, a cross-national initiative covering 17 countries in the European Higher Education Area. Specifically, we use the Portuguese dataset provided by the Directorate-General for Education and Science Statistics (DGEEC). The reference population consists of graduates from the 2016/2017 and 2020/2021 cohorts in the Portuguese higher education system. The survey provides detailed information on graduates' academic trajectories and labour market outcomes, including degree-related characteristics, employment conditions, and geographic location at different stages of the education-to-work transition. The questionnaire was administered between November 2022 and March 2023 following a census-based approach. In total, 18,280 valid responses were obtained, corresponding to 12.3% of Portuguese higher education graduates (Rodrigues et al. 2024).

Although both cohorts were surveyed within the same data collection framework, they entered the labour market under markedly different macroeconomic conditions. Graduates from the 2016/2017 cohort transitioned into employment during a period of economic recovery, whereas those from the 2020/2021 cohort did so amid the COVID-19 shock. The two cohorts are also observed at different points in time after graduation, with the earlier cohort measured at a longer post-graduation horizon than the later one. In addition, they may differ in observable composition, including field of study and degree structure.[5] For these reasons, we estimate all models separately by cohort.

The dataset provides detailed information on study location, current job location, demographic characteristics, academic background, occupations, and sector.

Sample construction

The empirical analysis focuses on employed graduates for whom the main outcome and mobility variables can be constructed. The analytical sample is obtained in several steps. Starting from the full set of valid survey responses, we first restrict the sample to respondents who report being employed in Portugal at the time of the survey. We then retain observations with valid information on earnings and weekly working hours, which are required to compute hourly wages. Next, we require non-missing information on the three locations used to define mobility status: the place of residence during secondary education, which we use as a proxy for home region; the location of the HEI attended; and the current work location. Finally, we restrict the sample to observations with non-missing values for the set of core control variables included in the baseline specifications. Appendix Table A1 reports these successive sample restrictions in detail.

Key variable definitions

The main outcome variable is the logarithm of hourly wage. It is computed primarily using reported actual weekly hours of work. In the small number of cases in which actual hours are missing but contractual hours are available, we use contractual hours as a proxy to retain otherwise valid observations. Because contractual hours may differ from effective working time, we assess the robustness of the results by excluding these cases in alternative specifications.

[5] For further details on the questionnaire and survey methodology, see Rodrigues et al. (2024).

The survey identifies four NUTS 3 locations for each graduate. Let H_i denote the origin region, proxied by the region of residence during secondary school;[6] S_i the region of the higher education institution attended; R_i the present region of residence; and W_i the present workplace region. Following the sequential graduate-mobility literature, which traces graduates from domicile to higher education and then to employment (Faggian et al. 2006; Di Cintio and Grassi 2013; Kazakis and Faggian 2017), the baseline classification traces $H_i \rightarrow S_i \rightarrow W_i$. Present residence R_i is retained as a conceptually separate dimension and must not be interpreted as the labour-market location. Definitions and construction details for the remaining adjustment variables are reported in Table A2.

Descriptive statistics

Error! Reference source not found. presents descriptive statistics by mobility trajectory and cohort. The group with the higher average log hourly wages, but not the highest wage dispersion, is the *early movers* in both cohorts. *Back movers* present the lowest average wages. In both cohorts, the percentage of female graduates is higher among *back movers*.

Mobility type	Log hourly wage	Age	Female (%)	Portuguese citizenship (%)	Final grade	CTeSP (%)	Master's (%)
Panel A. 2016-2017 cohort							
Stayers	2.088	31.727	59.710	99.539	14.697	3.221	31.942
	(0.546)	(7.880)	(49.059)	(6.777)	(1.679)	(17.661)	(46.636)
Early movers	2.166	30.799	60.541	99.591	14.827	0.946	43.108
	(0.550)	(6.433)	(48.909)	(6.389)	(1.699)	(9.686)	(49.556)
Late movers	2.085	30.542	56.745	100.000	14.734	4.497	34.690
	(0.558)	(6.142)	(49.596)	(0.000)	(1.659)	(20.746)	(47.649)
Back movers	1.961	31.467	67.075	99.660	14.706	3.448	32.481
	(0.584)	(7.980)	(47.020)	(5.822)	(1.657)	(18.257)	(46.856)
Repeat movers	2.086	30.328	59.218	99.624	14.519	2.048	39.106
	(0.498)	(5.789)	(49.189)	(6.126)	(1.624)	(14.178)	(48.844)
Cohort total	2.076	31.268	60.862	99.623	14.702	2.909	34.805
	(0.553)	(7.346)	(48.811)	(6.131)	(1.671)	(16.808)	(47.640)
Panel B. 2020-2021 cohort							
Stayers	1.920	27.985	63.266	99.527	15.101	2.832	37.626
	(0.497)	(8.397)	(48.217)	(6.863)	(1.692)	(16.591)	(48.454)
Early movers	1.959	27.358	61.959	99.655	15.234	1.822	48.861
	(0.494)	(7.432)	(48.576)	(5.869)	(1.709)	(13.383)	(50.016)
Late movers	1.927	27.432	54.300	99.000	14.842	2.457	41.523
	(0.483)	(7.510)	(49.876)	(9.962)	(1.648)	(15.500)	(49.337)
Back movers	1.869	28.463	68.898	99.233	14.906	4.104	36.393
	(0.491)	(8.611)	(46.316)	(8.727)	(1.644)	(19.848)	(48.139)
Repeat movers	1.919	27.975	58.471	99.578	15.029	1.033	45.248
	(0.503)	(7.947)	(49.328)	(6.489)	(1.803)	(10.122)	(49.825)
Cohort total	1.918	27.921	62.905	99.461	15.062	2.693	40.281
	(0.495)	(8.180)	(48.310)	(7.324)	(1.697)	(16.189)	(49.051)

Table 1. Descriptive statistics by mobility trajectory and cohort.

[6] Residence during secondary education is not necessarily the parental home and may therefore measure origin with some error.

Notes: Means are reported with standard deviations in parentheses. Citizenship percentages use valid citizenship responses only; missing citizenship does not restrict the sample. Hourly wages are winsorised at the cohort-specific 1st and 99th percentiles before taking logarithms. Source: Own computations using EUROGRADUATE 2022.

Across both cohorts, the unconditional distribution of back movers is shifted to the left relative to the other mobility groups (see Figure 1). This first unconditional look already points to a back mover wage penalty relative to early movers, repeat movers, late movers, and even stayers.

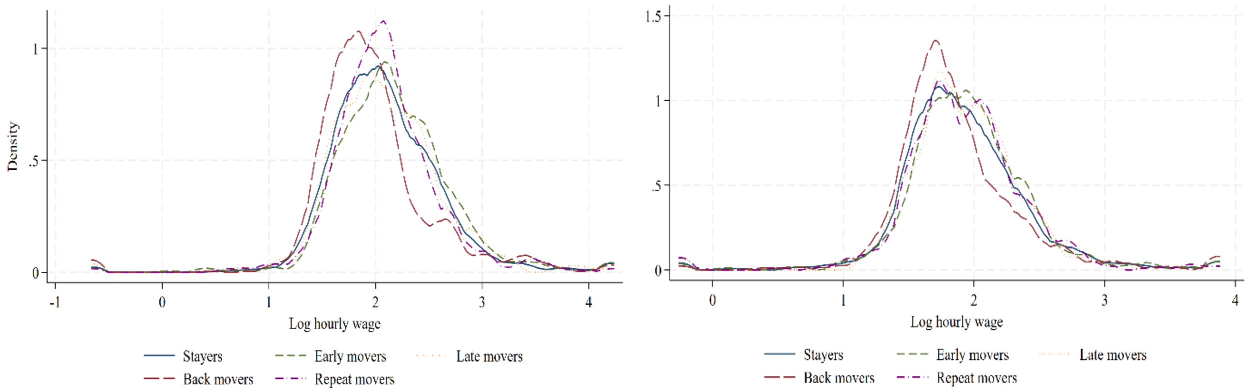

Figure 1. *Wage densities across mobility paths, 2016–2017 cohort on the left, 2020-2021 cohort on the right.*

Figure 2 shows that Lisbon is an extreme case, as 86.7% of graduates from Lisbon are stayers, compared to only 4.3% who are back movers. Açores and Madeira, by contrast, have the highest shares of back movers (42.8% and 38.2%, respectively). The result for the islands is as expected, since the limited supply of higher education there pushes the majority of graduates to leave home to study, while a comparatively large share subsequently return. Alentejo has the lowest share of stayers (16.7%) and the highest share of early movers (41.1%), so most graduates from this region leave early rather than staying at home or returning later. The relationship between the stayer share and the back mover share is negative overall but not linear across all seven regions. Peripherality, rather than the retention rate itself, appears to be the dominant factor behind a region's back mover share. Repeat movers are rare everywhere and almost absent in Lisbon (1.7%).

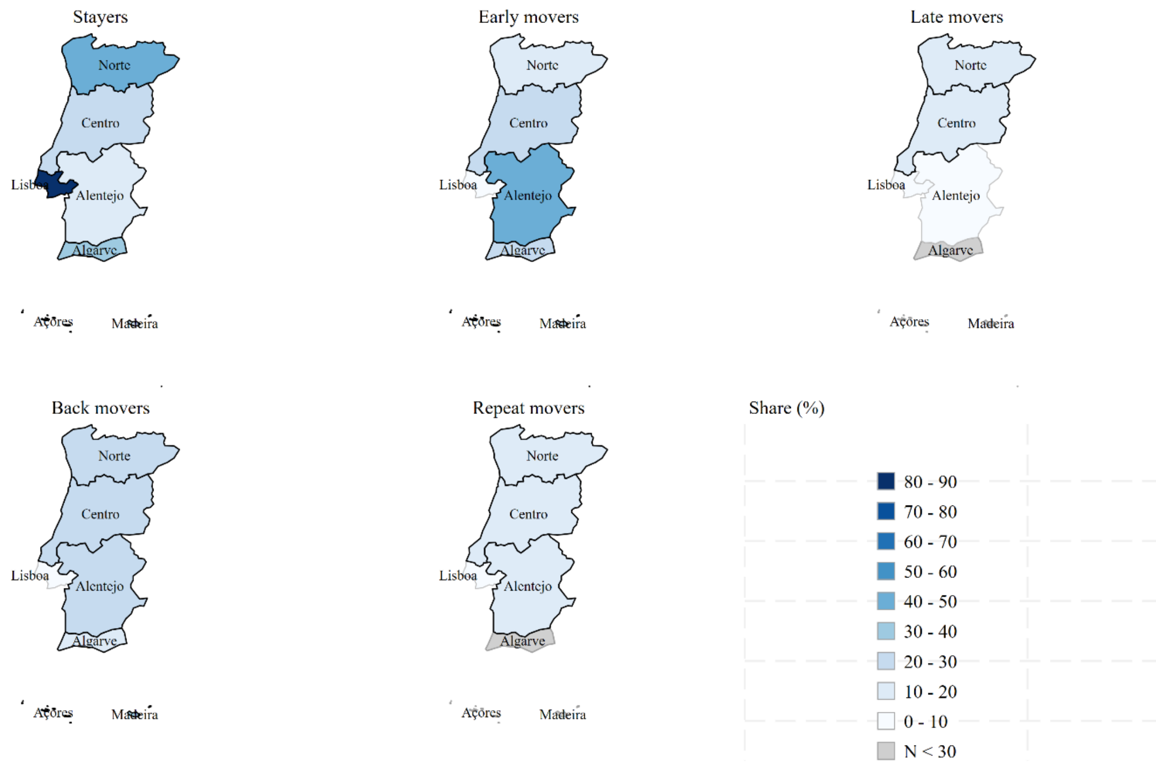

Figure 2*. Where each mobility type concentrates, by home region.*

Notes: Values are shares of each region's own eurograduate population, not of a national total. Panels report, for each of the five mobility trajectories (Stayers, Early movers, Late movers, Back movers, Repeat movers), that trajectory's share of all analytical-sample graduates from the same home NUTS II region. The 2016-17 and 2020-21 cohorts are pooled for visualization purposes. Mobility trajectories are defined from home, higher education institution (HEI), and job locations at NUTS III level, home NUTS II is used only to aggregate these NUTS III defined trajectories for regional display. All five panels share an identical color scale (10-percentage-point class breaks), so shading is directly comparable across mobility types. Grey cells denote home NUTS II x mobility-type cells with fewer than 30 analytical-sample graduates, excluded for disclosure control. Açores and Madeira are rescaled and repositioned as conventional cartographic insets; their true geographic location and size are not used for shading. Source: Own computations using the EUROGRADUATE survey.

Results

OLS benchmarks

Error! Reference source not found. shows results very similar to Kazakis and Faggian (2017), who also estimate an OLS model as a benchmark. *Late movers* have almost non-significant estimated coefficients across the different columns, whereas repeat movers have a significantly positive coefficient. In this baseline framework, the difference between *stayers* and *back movers* appears negligible, particularly once pre-mobility characteristics and home region are accounted for. Therefore, once we partially condition on a set of covariates, being a back mover versus a stayer shows no meaningful wage difference, while being an *early mover* seems to bring a clear advantage for graduates in both cohorts.

Focusing on our main groups of comparison, we verify that the back mover - early mover difference and the back mover - repeat mover difference is negative and precisely estimated across almost all specifications. Only when controlling for fixed effects using wild robust clustered standard errors does the back mover - early mover lose significance. Nevertheless, the back mover - repeat mover remains relevant with a penalty of 0.106 of returning home instead of moving onward to a third region.

	(1)	(2)	(3)	(4)	(5)
	Raw	X_i^0	$X_i^0 + L_i$	+ Home/Study FE	+ FE/Wild cluster
Panel A: 2016–2017 cohort					
Early movers	0.078***	0.146***	0.071***	0.077***	0.077*
	(0.023)	(0.026)	(0.027)	(0.029)	(0.030)
Late movers	-0.003	0.032	0.050*	0.057*	0.057
	(0.028)	(0.028)	(0.028)	(0.029)	(0.053)
Back movers	-0.127***	-0.041*	-0.035	-0.020	-0.020
	(0.023)	(0.023)	(0.023)	(0.027)	(0.036)
Repeat movers	-0.002	0.083***	0.073***	0.086***	0.086*
	(0.024)	(0.027)	(0.026)	(0.029)	(0.036)
back mover - early mover	-0.205***	-0.187***	-0.106***	-0.097***	-0.097
SE: *back - early*	0.028	0.028	0.029	0.031	0.090
back mover − repeat mover	-0.125***	-0.124***	-0.108***	-0.106***	-0.106**
SE: *back- repeat*	0.029	0.029	0.028	0.029	0.043
Predetermined controls X_i^0	No	Yes	Yes	Yes	Yes
Education and HEI trajectory L_i	No	No	Yes	Yes	Yes
Home- and study-region fixed effects	No	No	No	Yes	Yes
Wild-cluster bootstrap	No	No	No	No	Yes
Home-region clusters					25
Study-region clusters					22
Observations	4,847	4,847	4,847	4,847	4,847
Adj. R-squared	0.011	0.073	0.112	0.111	0.111
Panel B: 2020–2021 cohort					
Early movers	0.039**	0.084***	0.040*	0.050**	0.050
	(0.019)	(0.022)	(0.022)	(0.024)	(0.024)
Late movers	0.007	0.023	0.030	0.047*	0.047
	(0.026)	(0.025)	(0.025)	(0.026)	(0.045)
Back movers	-0.051***	0.005	0.011	0.025	0.025
	(0.019)	(0.020)	(0.020)	(0.023)	(0.023)
Repeat movers	-0.001	0.048*	0.041	0.064**	0.064
	(0.025)	(0.026)	(0.026)	(0.027)	(0.042)
back mover - early mover	-0.090***	-0.078***	-0.029	-0.025	-0.025
SE: *back - early*	0.023	0.023	0.024	0.025	0.043
back mover − repeat mover	-0.050*	-0.043	-0.030	-0.039	-0.039
SE: *back - repeat*	0.028	0.027	0.027	0.027	0.043
Predetermined controls X_i^0	No	Yes	Yes	Yes	Yes
Education and HEI trajectory L_i	No	No	Yes	Yes	Yes
Home- and study-region fixed effects	No	No	No	Yes	Yes
Wild-cluster bootstrap	No	No	No	No	Yes
Home-region clusters					25
Study-region clusters					21
Observations	5,273	5,273	5,273	5,273	5,273
Adj. R-squared	0.002	0.077	0.116	0.128	0.128

Table 2. OLS benchmarks for log hourly wages by cohort

Notes: Each panel is estimated separately. All five columns use the common complete-case marker defined by the fixed-effects specifications. Stayers are the omitted mobility category. The dependent variable is the winsorized log hourly wage. Column (2) adds the predetermined controls: secondary-school grade, sex, parental education, age and age squared, local HEI accessibility, and metropolitan home status. Column (3) adds degree level, tertiary final grade, university and public-HEI status, field of study, and metropolitan study status. Columns (4) and (5) add home and study NUTS 3 fixed effects, which absorb the two metropolitan indicators. Columns (1)-(4) use heteroskedasticity-robust standard errors and conventional p-values. Column (5) reports two-way clustered standard errors by home and study NUTS III. Because there are only 21-22 study-region clusters and cluster sizes are unbalanced, the coefficient stars in column (5) use a null-imposed wild cluster bootstrap with 9,999 replications, two-way error clustering, and bootstrap clustering by study NUTS III. The bootstrap is run from the algebraically equivalent regression with explicit home- and study-region indicators. Current-job characteristics and current-job-region fixed effects are excluded because workplace location defines the mobility path. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *Source:* Own computations using the EG22 survey.

OLS estimates by home region group

Table 3 disaggregates the same five-category comparison by home region and highlights where the back-mover disadvantage is concentrated. While early movers seem to have an advantage in AMP and the interior of Portugal compared to stayers, the same is not true for the other regions. At the same time, all types of non-returning movers seem to be beneficial for the graduates that came from the interior part, as the magnitudes are higher among repeat movers and late movers. The Porto Metropolitan Area (AMP) stands out, since back movers from Porto earn 29.2 log points less than early movers from Porto in the 2016-2017 cohort, the single largest back-early gap of any region in either cohort. The Islands group shows a marginally significant back-early gap that changes sign across cohorts (-19.3 log points in 2016-2017, +18.7 in 2020-2021), but given the small underlying samples (212 and 237 observations), we read this as sampling noise rather than a stable regional effect.

Mobility trajectory	AMP	AML	Litoral	Islands	Other mainland
Panel: 2016-2017 cohort					
Early movers	0.278** (0.131)	-0.057 (0.173)	0.074 (0.050)	0.031 (0.145)	0.135** (0.053)
Late movers	-0.027 (0.059)	-0.158* (0.086)	0.132*** (0.042)	--	0.160*** (0.051)
Back movers	-0.015 (0.104)	0.053 (0.180)	0.041 (0.043)	-0.162 (0.120)	0.003 (0.048)
Repeat movers	0.073 (0.148)	-0.218 (0.250)	0.094** (0.044)	-0.004 (0.163)	0.157*** (0.042)
Back movers - Early movers	-0.292** (0.121)	0.110 (0.107)	-0.034 (0.049)	-0.193* (0.109)	-0.133 (0.047)
Back movers - Repeat movers	-0.088 (0.122)	0.271 (0.187)	-0.053 (0.049)	-0.158 (0.169)	-0.155 (0.042)
Observations	752	1,217	1,216	212	1,450
Adj. R-squared	0.134	0.061	0.154	0.094	0.114
Panel: 2020-2021 cohort					
Early movers	0.153* (0.092)	0.101 (0.180)	0.012 (0.052)	-0.218* (0.131)	0.096** (0.045)
Late movers	0.112** (0.048)	-0.100* (0.058)	0.037 (0.041)	-0.288 (0.418)	0.162** (0.066)
Back movers	0.126 (0.078)	0.002 (0.179)	0.024 (0.038)	-0.031 (0.090)	0.022 (0.043)
Repeat movers	0.109 (0.120)	-0.129 (0.194)	0.102** (0.048)	-0.108 (0.108)	0.058 (0.047)
Back movers - Early movers	-0.027 (0.085)	-0.100 (0.111)	0.012 (0.049)	0.187* (0.102)	-0.074 (0.033)
Back movers - Repeat movers	0.017 (0.085)	0.130 (0.105)	-0.078 (0.049)	0.076 (0.097)	-0.036 (0.040)
Observations	879	1,510	1,244	237	1,403
Adj. R-squared	0.139	0.141	0.113	0.058	0.095

Table 3. OLS mobility wage differentials, by home region.

Notes: Cells report OLS coefficients with heteroskedasticity-robust standard errors in parentheses. Stayers are the omitted mobility category. Each regression includes the full observed background, home-geography, education, and HEI controls used in the main OLS benchmark; covariates without variation inside a column are omitted. Litoral combines coastal mainland NUTS 3 regions outside AMP/AML (Alto Minho, Cavado, Algarve, Região de Aveiro, Região de Coimbra, Região de Leiria, Oeste, and Alentejo Litoral); Islands combines Azores and Madeira, and Other mainland contains the remaining interior regions. The coastal/interior split is descriptive and follows the conventional Portuguese litoral/interior geographic distinction rather than an official NUTS-level classification; cell sizes are small in several groups and are reported without suppression because this table is descriptive. Home and study NUTS 3 fixed effects are excluded to retain a common specification. *p<0.10, ** p<0.05, *** p<0.01. Source: Own computations using the EUROGRADUATE survey.

Table 3's home-region breakdown is still unconditional, since it does not tell us whether back movers and their comparison groups are similar on characteristics that predate the workplace decision. The remainder of the paper narrows to the two focal comparisons (1) back versus early movers, (2) back versus repeat movers and asks that question directly, first by checking common support and covariate balance, then by decomposing the adjusted wage gap itself.

Overlap (common-support) and covariate balance

We next focus on the two counterfactual comparisons relevant for graduates who studied outside their home region: returning home rather than remaining in the study region; and returning home rather than moving onward to a third region. For each comparison and cohort, we restrict the data to empirical common support in the estimated probability of being a back mover and entropy-balance the comparison group on pre-workplace characteristics. Only a small number of observations are removed by the common-support rule (Table 4, column 4), so the main conclusions are not driven by extensive trimming. Nevertheless, the groups look quite different before weighting.

Depending on the comparison, between 10 and 21 of the 27 measured characteristics exceed the conventional 0.10 imbalance threshold. The largest differences concern metropolitan study and origin, distance travelled to higher education, local access to institutions, and educational trajectories. A raw wage comparison would therefore mix the return path with substantial geographic and educational sorting.

Comparison	Back N	Control N	Trimmed B/C	Max abs. SMD raw	Max abs. SMD weighted	Control ESS	Max. weight
2016-2017							
Back-Early	896	720	3 / 20	0.7630	0.00003	326.5	9.74
Back-Repeat	893	534	6 / 3	0.3118	0.00008	268.3	15.30
2020-2021							
Back-Early	926	848	0 / 30	0.6614	0.00001	390.4	8.91
Back-Repeat	915	477	11 / 7	0.2362	0.00012	297.5	12.35

Table 4. Common support and covariate balance summary, by cohort and comparison.

Notes: Back N and Control N are observations retained on empirical common support; Trimmed B/C reports the number of back-mover and comparison-group observations excluded by the support rule. Max abs. SMD raw and weighted are the largest absolute standardized mean difference across the 27-variable balancing set before and after entropy weighting, respectively (see Tables A3 and A4 for the full covariate-by-covariate breakdown). Control ESS is the effective sample size of the reweighted comparison group; Max. weight is the largest entropy weight assigned to a single comparison-group observation. Source: Own computations using the EUROGRADUATE 2022 survey.

After entropy weighting, every measured difference is effectively zero (see Figure 3). The adjusted wage comparison is made between groups that are very similar with respect to the observed adjustment variables. The gain in comparability comes with a loss of precision. The weighted comparison groups have effective sample sizes of about 268 to 390, and a relatively small number of observations receive sizeable weights. The figure makes the change in comparability visible, supporting the interpretation that any remaining wage gap is not being generated by the measured composition differences listed in Tables A3-A4. It cannot, however, rule out differences in other variables not present in the dataset and relevant for labour market performance, such as unobserved ability, preferences, networks, or wage offers.

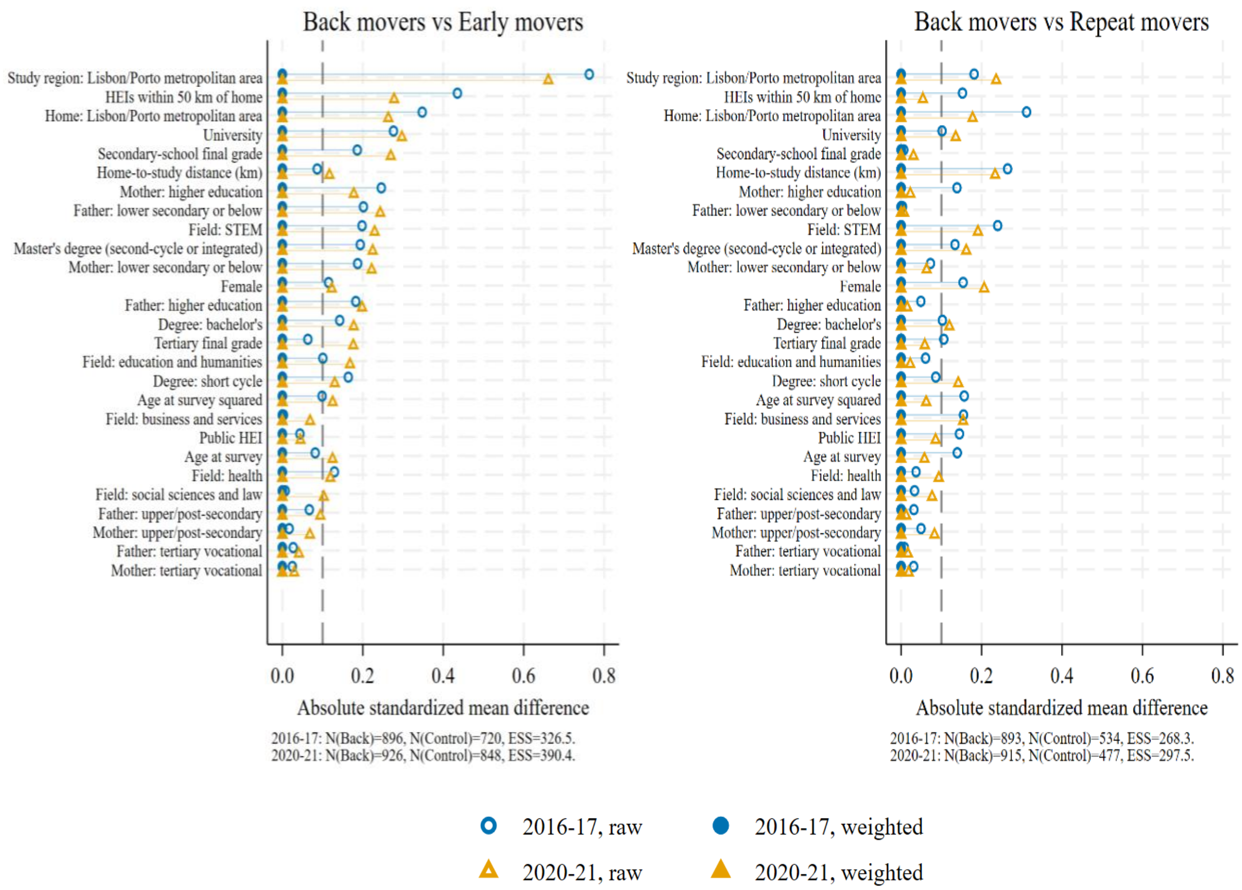

Figure 3. *Absolute standardized mean differences before and after entropy weighting, by cohort and comparison.*

Notes: Circles are the 2016-17 cohort and triangles are the 2020-21 cohort; hollow markers show the raw (unweighted) difference and filled markers show the entropy-weighted difference. Each row is one balancing-set covariate, sorted by the larger of the two cohorts' raw imbalance. Dashed line: |SMD| = 0.1.

Wage decomposition after entropy balance

The sequential decomposition (Table 5) splits the observed-composition term into a geographic-sorting component, graduate observed characteristics, and the unexplained part, assessed here using Wald p-values from the bootstrap standard errors for consistency with the OLS tables above.

Comparison	Component	Estimate	Bootstrap SE
2016-2017			
	Raw gap	-0.198***	0.049
Early movers	Geographic sorting	-0.059*	0.031
	Graduate observed characteristics	-0.086***	0.027
	Residual gap	-0.052	0.043
	Raw gap	-0.130***	0.030
Repeat movers	Geographic sorting	-0.018	0.013
	Graduate observed characteristics	-0.035*	0.020
	Residual gap	-0.077**	0.030
2020-2021			
	Raw gap	-0.087***	0.025
Early movers	Geographic sorting	-0.046**	0.023
	Graduate observed characteristics	-0.045*	0.023
	Residual gap	0.005	0.033
	Raw gap	-0.048	0.039
Repeat movers	Geographic sorting	-0.002	0.014
	Graduate observed characteristics	-0.044**	0.022
	Residual gap	-0.003	0.031

Table 5. Sequential entropy-balanced decomposition of the back-mover wage gap.

Notes: Raw gap is the unweighted difference in mean log hourly wages, Back movers minus the comparison group. Geographic sorting is the geography-only-weighted comparison mean minus the unweighted comparison mean; other observed sorting is the fully-weighted comparison mean minus the geography-only-weighted mean; the residual gap is the Back-mover mean minus the fully-weighted comparison mean. The three components sum exactly to the raw gap. Bootstrap SEs come from a full-pipeline cluster bootstrap (999 replications) resampling study-region NUTS III clusters, re-estimating the propensity score, common support, both entropy-balancing calls, and the decomposition in every replication. * p<0.10, ** p<0.05, *** p<0.01 (Wald test on the bootstrap SE). Source: Own computations using the EUROGRADUATE 2022 survey.

The sign is negative across all results. Back movers have a penalty against both comparison groups. Raw gap is the difference between back movers and the comparison group, and it varies from 5% to approximately 18% of gross wage disadvantage depending on the comparison group and cohort.

If we compare back movers with early movers, we see a large gap, and this gap is mostly explained by geographic sorting and graduate observed characteristics. In 2016-17, geographic sorting (-0.059) plus other observed sorting (-0.086) sum to -0.145 of the -0.198 total (about 73%), leaving a small, statistically insignificant residual (-0.052). By 2020-21, the covariates account for essentially all of the gap, as the residual is +0.005, statistically indistinguishable from zero. Relative to graduates who moved away early and did not return, there is no evidence of an unexplained wage cost to being a back mover; the wage difference seems to be entirely compositional (region of origin, family background, and the other balanced characteristics).

The pattern reverses with repeat movers, most clearly in 2016-17. Geographic sorting is negligible, and other observed sorting is modest (-0.035), but the residual is the largest component and the only one in the entire table that is statistically significant (-0.077). More than half of the raw gap (59%) remains unexplained even after adjusting for geography and the other observed characteristics. This is the one cell where a wage gap persists after full adjustment for the observed covariates. If being a back mover comes with a “natural” penalty, it can be seen in the unexplained part.

Geographic sorting is close to zero in the repeat-mover comparisons (-0.018 and -0.002) but substantial in the early-mover comparisons (-0.059* and -0.046**). This is consistent with the spatial pattern already documented (Figure 2). If repeat movers have a regional-origin composition similar to back movers' (both concentrated in peripheral/island regions), the geographic component has little left to explain in that specific comparison.

Graduate observed characteristics are the only components significant in all four rows; this is the most robust and consistent driver of the gap, regardless of cohort or comparison group. The raw back-vs-repeat-mover gap falls from -0.130 (significant) to -0.048 (not significant, with a standard error nearly as large as the point estimate: 0.039).

Distributional extension

The mean decomposition can hide offsetting patterns at different points of the wage distribution. Table 6 repeats the composition/residual split (without the geographic/other-sorting refinement, to keep the bootstrap tractable) at the 10th, 25th, 50th, 75th, and 90th percentiles of the unconditional wage distribution, for the same four cohort-comparison cells. Figure 4 plots the same estimates with percentile bootstrap confidence intervals.

For the 2016-2017 cohort, the residual is negative from P10 through P75 in both comparisons, reaching -9.5 to -11.8 log points at the median, and turns positive at P90 but is not statistically significant. The composition component is negative and relatively stable across the distribution in both comparisons, consistent with the mean result.

For the 2020-2021 cohort, the pattern is noisier. Read together with Table 5, the distributional results show that graduates who return home and would otherwise be in the middle of the wage distribution bear the largest adjusted disadvantage, whereas those who would otherwise be top earners do not. This pattern is more consistent with a loss of access to the highest-paying job matches available only in larger labour markets than with a uniform "penalty" attached to the return decision itself.

Component	P10	P25	P50	P75	P90
Panel A. 2016-2017: Back movers vs Early movers					
Raw gap	-0.120*** (0.036)	-0.192*** (0.026)	-0.226*** (0.034)	-0.288*** (0.024)	-0.113* (0.059)

Composition	-0.101*** (0.032)	-0.169*** (0.036)	-0.131*** (0.033)	-0.174*** (0.034)	-0.185*** (0.045)
Residual	-0.020 (0.035)	-0.023 (0.041)	-0.095** (0.038)	-0.113*** (0.040)	0.071 (0.062)
Panel B. 2016-2017: Back movers vs Repeat movers					
Raw gap	-0.130*** (0.038)	-0.170*** (0.028)	-0.156*** (0.030)	-0.170*** (0.036)	-0.004 (0.061)
Composition	-0.018 (0.038)	-0.065** (0.028)	-0.038 (0.029)	-0.093** (0.037)	-0.080* (0.045)
Residual	-0.112** (0.051)	-0.105*** (0.033)	-0.118*** (0.040)	-0.077* (0.045)	0.076 (0.060)
Panel C. 2020-2021: Back movers vs Early movers					
Raw gap	-0.103*** (0.033)	-0.085*** (0.026)	-0.165*** (0.025)	-0.125*** (0.034)	0.000 (0.040)
Composition	-0.081 (0.054)	-0.071*** (0.021)	-0.127*** (0.027)	-0.092*** (0.034)	-0.029 (0.039)
Residual	-0.022 (0.063)	-0.015 (0.022)	-0.038 (0.031)	-0.033 (0.039)	0.029 (0.054)
Panel D. 2020-2021: Back movers vs Repeat movers					
Raw gap	-0.041 (0.034)	-0.053* (0.028)	-0.129*** (0.033)	-0.105*** (0.037)	0.000 (0.046)
Composition	-0.041 (0.031)	-0.012 (0.028)	-0.053 (0.034)	-0.022 (0.026)	0.000 (0.042)
Residual	0.000 (0.044)	-0.041 (0.033)	-0.076* (0.042)	-0.083** (0.037)	0.000 (0.057)

Table 6. Decomposition across the wage distribution.

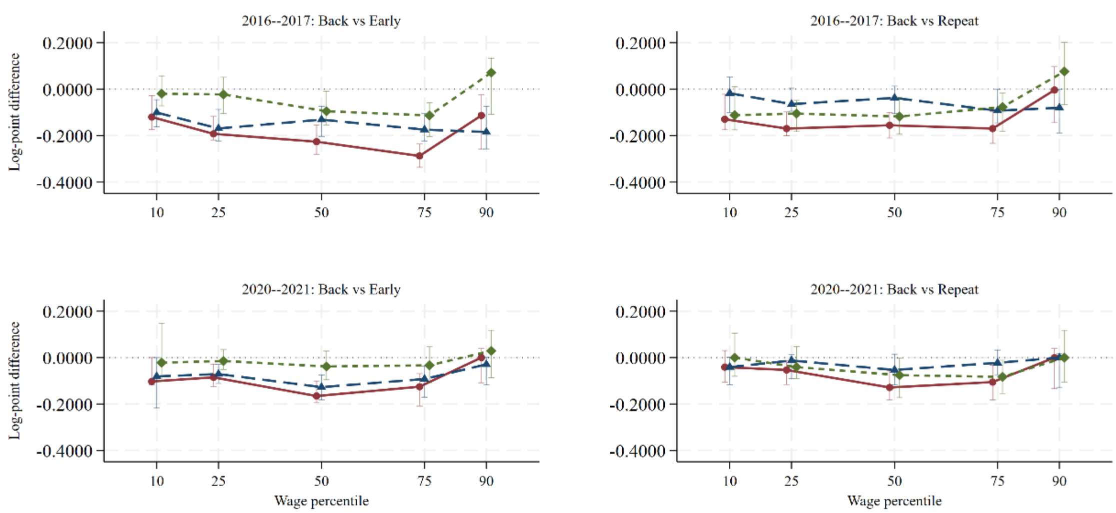

Figure 4.Raw gap, composition, and residual across the wage distribution (P10-P90), by cohort and comparison.

Notes: Capped lines are percentile-bootstrap 95% confidence intervals. Source: Own computations using the EUROGRADUATE 2022 survey.

Conclusion

Is it worth coming back? Using the Portuguese EUROGRADUATE survey and an entropy-balanced sequential decomposition, we find that the answer depends on the counterfactual. Relative to graduates who stayed in the study region, geographic sorting is a statistically detectable part of the back-mover disadvantage in both cohorts. Early movers disproportionately remain in Lisbon and Porto, the two labour markets with by far the deepest concentration of skilled jobs, and part of what looks like a "return penalty" is simply the wage structure of the places early movers stay in, rather than a penalty attached to returning as such. Relative to graduates who moved to a third region, geographic sorting plays essentially no role; repeat movers are not disproportionately drawn to the metropolitan core, and what distinguishes back movers instead is family background and prior educational trajectory, a pattern that holds for both comparisons and both cohorts.

Only one of the four cohort-comparison cells - the 2016-2017 comparison with repeat movers - leaves a statistically significant adjusted residual, on the order of an 8% wage disadvantage for otherwise comparable graduates. The disadvantage is concentrated in the lower and middle of the wage distribution and turns positive at the 90th percentile, so returning home is not uniformly costly. It is costliest for graduates who would otherwise be in the middle of the distribution, and not for the highest earners, who appear to negotiate comparable wages wherever they work. Five years after graduation, returning home carries a modest and unevenly distributed wage cost when the alternative was continued mobility; one year after graduation, in the middle of the pandemic labour market, no such adjusted cost is detectable in either comparison.

Back movers are not geographically neutral, as they are disproportionately drawn from peripheral and insular regions such as the Açores and Madeira, while graduates from Lisbon rarely return, largely because they rarely leave in the first place. This composition matters for the wage results in Table 5. Once graduates' geographic origin and other observed characteristics are accounted for, the wage gap between back movers and early movers disappears, showing no evidence of a net wage cost to returning relative to those who moved away and stayed away. Relative to repeat movers, however, part of the gap remains unexplained, at least in the earlier cohort, indicating that whether returning is associated with a wage cost depends on the comparison group, not on returning as such.

Three limitations qualify these conclusions. First, our estimates are conditional on employment and cannot speak to whether mobility trajectories also affect the probability of finding work; a graduate who returns home and remains unemployed is invisible to a wage comparison that only observes the employed. Second, the adjusted residual is a wage contrast between observationally comparable graduates on a common support of pre-workplace characteristics, and not necessarily a causal return penalty, as characteristics such as unobserved ability, preferences, networks, and outside options can still differ systematically between back movers and their comparison groups. Third, our empirical strategy is static and treats each graduate's trajectory as a single realised path; repeat movers in particular plausibly reflect an ongoing sequential job search that a two-period comparison cannot fully capture. Extending the decomposition to a formal sensitivity bound on selection on unobservables, and tracking graduates over a longer horizon to see whether the 2020-2021 cohort's null result is a pandemic-specific artefact or a genuine cohort difference, are natural next steps.

References

Ahlin, L., Andersson, M., & Thulin, P. (2018). Human capital sorting: The “when” and “who” of the sorting of educated workers to urban regions. Journal of Regional Science, 58(3), 581–610. DOI: https://doi.org/10.1111/jors.12366

Di Cintio, M., & Grassi, E. (2013). Internal migration and wages of Italian university graduates. Papers in Regional Science, 92(1), 119-141.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica, 64(5), 1001–1044. https://doi.org/10.2307/2171954

Faggian, A., McCann, P., & Sheppard, S. (2006). An analysis of ethnic differences in UK graduate migration behaviour. The Annals of Regional Science, 40(2), 461–471. https://doi.org/10.1007/s00168-006-0061-y

Faggian, A., & McCann, P. (2008). Human capital, graduate migration and innovation in British regions. Cambridge Journal of Economics, 33(2), 317–333. https://doi.org/10.1093/cje/ben042

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis, 20(1), 25–46. https://doi.org/10.1093/pan/mpr025

Jauhiainen, S. (2011). Overeducation in the Finnish regional labour markets. Papers in Regional Science, 90(3), 573–589. https://doi.org/10.1111/j.1435-5957.2010.00334.x

Jewell, S., & Faggian, A. (2013). Interregional migration ‘Wage Premia’: the case of creative and science and technology graduates in the UK. In Applied regional growth and innovation models (pp. 197-214). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kazakis, P., & Faggian, A. (2017). Mobility, education and labor market outcomes for US graduates: Is selectivity important?. The Annals of Regional Science, 59, 731-758.

Lehmer, F., & Ludsteck, J. (2011). The returns to job mobility and inter-regional migration: Evidence from Germany. Papers in Regional Science, 90(3), 549-572.

Mitze, T., & Javakhishvili-Larsen, N. (2020). Graduate Migration and Early-career Labor Market Outcomes: Do Education Programs and Qualification Levels Matter?. Labour, 34(4), 477-503.

Nakosteen, R. A., Westerlund, O., & Zimmer, M. (2008). Migration and Self-Selection: Measured Earnings and Latent Characteristics. Journal of Regional Science, 48(4), 769–788. https://doi.org/10.1111/j.1467-9787.2008.00576.x

Rehák, Š., & Eriksson, R. (2020). Migration of university graduates and structural aspects of regional higher education. European Planning Studies, 28(10), 1941–1959. https://doi.org/10.1080/09654313.2019.170048

Rodrigues, N., Oliveira, F., Biscaia, R., Silva, P., Rudakov, V., Inácio, A. (2024). *Eurograduate 2022: Relatório Final*. Available at https://www.dgeec.medu.pt/api/ficheiros/65fb21c70c7bc47914575919

Rodríguez-Pose, A., & Tselios, V. (2010). Returns to migration, education and externalities in the European Union. Papers in Regional Science, 89(2), 411-434.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. Journal of Political Economy, 70(5), 80–93.

Zhao, M., & Hu, Y. (2019). Migration premium? The economic returns to youth inter-province migration in post-reform China. Journal of Youth Studies, 22(10), 1409-1427.

Appendix

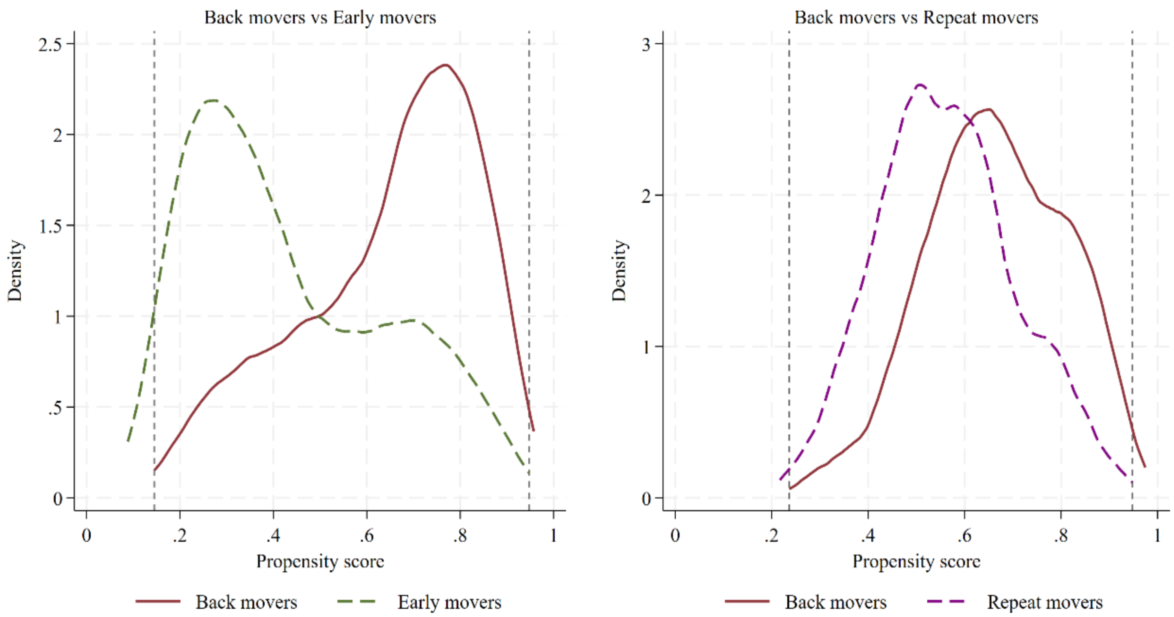

Figure A1. *Propensity-score overlap and common-support bounds, 2016-2017 cohort.*

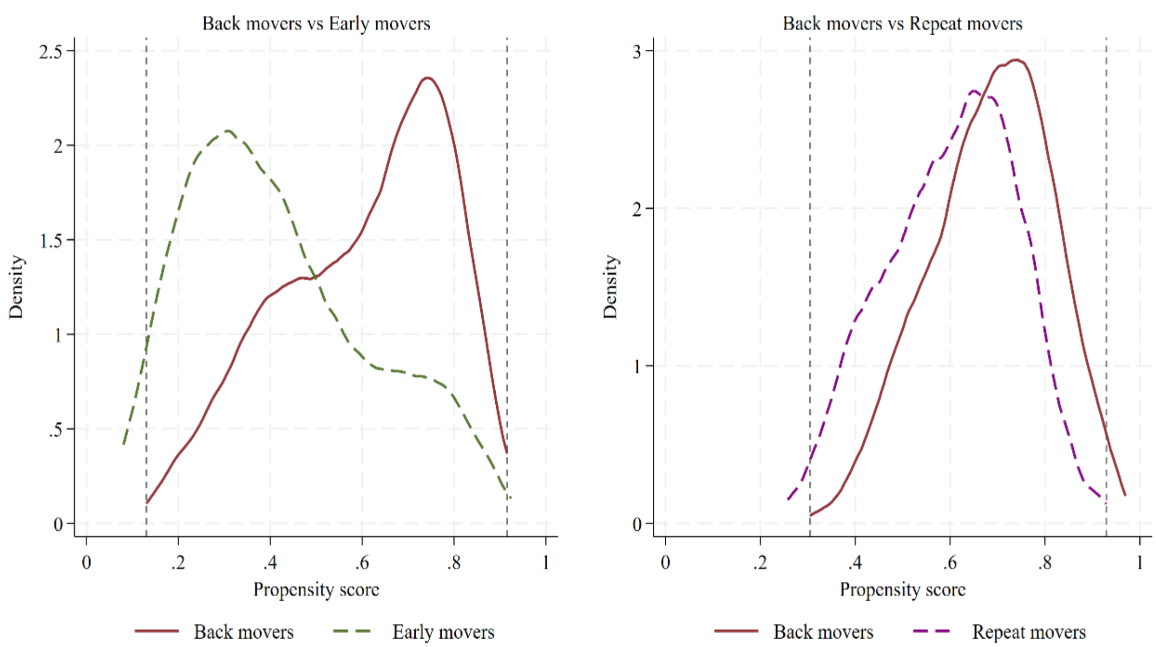

Figure A2*. Propensity-score overlap and common-support bounds, 2020-2021 cohort.*

	2016-2017		2020-2021	
Sample restriction	Observations	% previous	Observations	% previous
Valid survey responses	7,453	100.00	10,827	100.00
Employed respondents	6,514	87.40	7,186	66.37
Valid monthly earnings	6,214	95.39	6,745	93.86
Valid hourly-wage information	6,176	99.39	6,705	99.41
Valid domestic geography and mobility type	5,221	84.54	5,697	84.97
Final baseline estimation sample	4,847	92.84	5,273	92.56

Table A1 - Sample selection by graduation cohort

Notes: Each column pair is calculated independently within the indicated cohort; no pooled sample count is reported. Percentages report retention relative to the preceding row. The final analytical sample requires employment, a valid hourly wage, domestic home-HEI-job region, a mobility classification, valid full-time/part-time status, complete demographic, family-background, education, and institution controls, and non-singleton home and study cells for the fixed-effects specification. This same sample is used throughout the descriptive figures, OLS benchmarks, common-support diagnostics, and entropy balancing. Among final-sample respondents with valid citizenship information, 99.62% in the 2016–2017 cohort and 99.48% in the 2020–2021 cohort report Portuguese citizenship. Source: Own computations using the EG22 survey.

Variable	Description
Log hourly wage	Log of wage/hour winsorized at 1% and 99%. Hours = actual hours worked; when missing, contractual hours are used.
Home region (H)	NUTS III region of residence during high school (Origin proxy)
Study region (S)	NUTS III region of the higher education institution
Workplace region (W)	NUTS III region of the workplace
Studied-away indicator	=1{S≠H}; defines the subsamples for the decomposition
Workplace path	Conditional in A_i=1: h/s/o (home / study-region / third region)
Metropolitan origin	Origin in Lisboa or Porto NUTS III
Metropolitan study region	Higher Education Institution in Lisboa/Porto NUTS III
Nearby HEI access	Number of higher education institutions less than 50km from the NUTS IV/LAU I of the graduate's residence
Home-to-study distance (km)	Straight line distance between NUTS IV/LAU I of residence to NUTSIV/LAU I of HEI
Secondary-school final grade	Average grade (secondary-school)
Female	=1 female, =0 male
Mother's education	4 categories: less than secondary education; secondary education; vocational HE; academic HE
Father's education	4 categories (same)
Age at survey	Age ≈ 2022 - year of birth
Degree level	CTeSP (ISCED 5); bachelor (ISCED 6); master (ISCED 7)
Tertiary final grade	Final grade at higher education
University status	=1 if university
Public HEI status	=1 if public
Field of study	Education/humanities; social sciences/law (ref.); business/services; STEM; health

Table A2 – Variable definition

Covariate	Back-Early raw	Back-Early weighted	Back-Repeat raw	Back-Repeat weighted
Secondary-school final grade	-0.1863	0.0000	0.0065	0.0000
Female	0.1150	0.0000	0.1541	0.0000
Mother: lower secondary or below	0.1872	0.0000	0.0728	0.0000
Mother: upper/post-secondary	0.0171	0.0000	0.0496	0.0000
Mother: tertiary vocational	-0.0249	0.0000	-0.0312	0.0000
Mother: higher education	-0.2460	0.0000	-0.1389	0.0000
Father: lower secondary or below	0.2016	0.0000	0.0035	0.0000
Father: upper/post-secondary	-0.0672	0.0000	0.0316	0.0000
Father: tertiary vocational	-0.0269	0.0000	0.0081	0.0000
Father: higher education	-0.1826	0.0000	-0.0490	0.0000
Age at survey	0.0818	0.0000	0.1396	0.0000
Age at survey squared	0.0988	0.0000	0.1568	0.0001
HEIs within 50 km of home	0.4352	0.0000	0.1526	0.0000
Home: Lisbon/Porto metropolitan area	0.3477	0.0000	0.3118	0.0000
Study region: Lisbon/Porto metropolitan area	-0.7630	0.0000	0.1815	0.0000
Home-to-study distance (km)	0.0866	0.0000	0.2648	0.0000
Degree: short cycle	0.1639	0.0000	0.0861	0.0000
Degree: bachelor’s	0.1426	0.0000	0.1027	0.0000
Degree: master’s or doctorate	-0.1938	0.0000	-0.1339	0.0000
Tertiary final grade	-0.0635	0.0000	0.1059	0.0000
University	-0.2764	0.0000	-0.1016	0.0000
Public HEI	-0.0439	0.0000	-0.1450	0.0000
Field: education and humanities	0.1007	0.0000	0.0607	0.0000
Field: social sciences and law	0.0071	0.0000	0.0334	0.0000
Field: business and services	0.0029	0.0000	0.1551	0.0000
Field: STEM	-0.1982	0.0000	-0.2401	0.0000
Field: health	0.1295	0.0000	0.0370	0.0000

Table A3. Covariate balance, 2016-2017 cohort.

Covariate	Back-Early raw	Back-Early weighted	Back-Repeat raw	Back-Repeat weighted
Secondary-school final grade	-0.2696	0.0000	-0.0305	0.0000
Female	0.1228	0.0000	0.2065	0.0000
Mother: lower secondary or below	0.2220	0.0000	0.0633	0.0000
Mother: upper/post-secondary	-0.0685	0.0000	-0.0829	0.0000
Mother: tertiary vocational	-0.0295	0.0000	-0.0187	0.0000
Mother: higher education	-0.1775	0.0000	0.0227	0.0000
Father: lower secondary or below	0.2434	0.0000	0.0072	0.0000
Father: upper/post-secondary	-0.0944	0.0000	-0.0124	0.0000
Father: tertiary vocational	-0.0413	0.0000	-0.0167	0.0000
Father: higher education	-0.1982	0.0000	0.0151	0.0000
Age at survey	0.1248	0.0000	0.0580	0.0000
Age at survey squared	0.1251	0.0000	0.0623	0.0001
HEIs within 50 km of home	0.2779	0.0000	0.0540	0.0000
Home: Lisbon/Porto metropolitan area	0.2634	0.0000	0.1777	0.0000
Study region: Lisbon/Porto metropolitan area	-0.6614	0.0000	0.2362	0.0000
Home-to-study distance (km)	0.1171	0.0000	0.2336	0.0000
Degree: short cycle	0.1303	0.0000	0.1421	0.0000
Degree: bachelor’s	0.1774	0.0000	0.1198	0.0000
Degree: master’s or doctorate	-0.2245	0.0000	-0.1619	0.0000
Tertiary final grade	-0.1764	0.0000	-0.0587	0.0000
University	-0.2972	0.0000	-0.1357	0.0000

Covariate	Back-Early raw	Back-Early weighted	Back-Repeat raw	Back-Repeat weighted
Public HEI	0.0450	0.0000	-0.0856	0.0000
Field: education and humanities	0.1682	0.0000	0.0225	0.0000
Field: social sciences and law	-0.1029	0.0000	-0.0769	0.0000
Field: business and services	0.0690	0.0000	0.1542	0.0000
Field: STEM	-0.2294	0.0000	-0.1909	0.0000
Field: health	0.1192	0.0000	0.0935	0.0000

Table A4. Covariate balance, 2020-2021 cohort.