ISABEL

MINIMIZE COSTS AND MAXIMIZE BENEFITS IN THE TRANSITIONARY EUROPEAN REGIONAL LABOUR MARKETS

mInimize coSts and mAximize Benefits in the transitionary European regional Labour markets

Deliverable No D3.1

Secondary data collection report

WP3: Data Lake creation and Data Collection on the uneven effects of JCD

Funded by
the European Union

DOCUMENT ID	D3.1
DUE DATE	30/06/2026
DELIVERY DATE	30/06/2026
WORK PACKAGE	WP3
TASK(S)	T3.1
DISSEMINATION LEVEL	R
VERSION	V3
AUTHOR/S	UO, UCPH, US
FILENAME	ISABEL_D3.1_V3

Abstract

This report presents the secondary data collection strategy and estimation methodology developed in Work Package 3 of the ISABEL project to produce a panel database of gross job creation rates (JCR) and job destruction rates (JDR) for European NUTS2 regions over the period 2010–2023, disaggregated by industry, occupation, educational attainment, migrant status, and sex. The construction of such a database is complicated by the absence of any harmonised, geographically complete source of firm-level gross job flows at the sub-national level in Europe. The primary data source — the CompNet database — covers only 18 EU countries and 146 NUTS2 regions, and its estimates are systematically inconsistent with official Eurostat regional employment statistics due to differences in reference populations and statistical methodologies between the two sources.

To address these challenges, WP3 develops a multi-stage estimation pipeline anchored by the constraint that the accounting identity JCR − JDR = NEG must hold exactly at every level of aggregation. The core estimator is a Generalised Cross-Entropy Distributionally Weighted Regression (GCE-DWR) that reconciles CompNet job flow priors with Eurostat net employment growth constraints. A Random Forest regression imputes JCR and JDR for the approximately 28% of EU NUTS2 regions not covered by CompNet, and a cross-entropy matrix balancing step ensures consistency between regional and sectoral estimates. Workforce decompositions by occupation, education, migrant status, and sex are produced using EU Labour Force Survey scientific use microdata, with group-specific priors informed by Location Quotients.

The pipeline is validated through a Monte Carlo simulation — which shows error reductions of 30% or more relative to naive use of CompNet figures — and through an out-of-sample benchmarking exercise against ONS administrative job flow data for UK NUTS2 regions, yielding mean absolute errors of 1.4–1.5 percentage points in non-crisis years. The resulting database reveals substantial cross-regional and cross-sectoral heterogeneity in gross job flows, with average JCR and JDR of approximately 4.7% and 4% of employment per year respectively, and provides a substantially more complete and internally consistent picture of European labour market dynamics than any currently available alternative.

Views and opinions expressed in this document are those of the author(s) and do not necessarily reflect those of the European Union or REA. Neither the European Commission nor the granting authority can be held responsible for them.

Funded by the European Union

Contributors

NAME	ORGANISATION
Esteban Fernández Vázquez	UO
Adrián Marteen Rodríguez del Valle	UO
Nikos Kapitsinis	UCPH
Høgni Kalsø Hansen	UCPH
Maria Heegaard-Poulsen	UCPH
Łukasz Trembaczowski	US

Peer Reviewers

NAME	ORGANISATION
Lefteris Topaloglou	UOWM
Aleksandra Anić	UB

Versions

Version	Date	Author / Organisation	Description
V1	12.6.2026	Esteban Fernández Vázquez/UO	
V2	15.6.2026	Nikos Kapitsinis/UCPH	
V3	25.6.2026	Nikos Kapitsinis/UCPH	

Funded by
the European Union

Executive summary

This report — Deliverable D3.1 of the ISABEL project — presents the secondary data collection strategy and estimation methodology underlying Work Package 3, whose objective is to produce a comprehensive panel database of gross job creation and destruction rates for European NUTS2 regions over the period 2010–2023. The database disaggregates these rates by NACE 2-digit industry, ISCO 2-digit occupation, educational attainment (three tiers), migrant status, and sex, and constitutes the empirical backbone of the ISABEL project's analysis of uneven regional labour market dynamics across Europe in the context of the green and digital transitions.

The starting point for the analysis is the conceptual distinction, formalised by Davis and Haltiwanger (1992), between net employment change — the figure routinely reported in official statistics — and the underlying gross flows of job creation and destruction. Net employment change is the algebraic sum of simultaneous job creation at expanding firms and job destruction at contracting ones, and is therefore profoundly uninformative about the scale of economic restructuring that workers actually experience. Gross job creation and destruction rates in Europe average approximately 4.7% and 4% of employment per year respectively across the NUTS2 regions in the panel, far exceeding the net change they produce, and vary enormously across regions, industries, and workforce groups in ways that aggregate statistics entirely conceal.

Producing these estimates at the NUTS2 level for Europe is a methodological challenge of considerable complexity. The primary firm-level data source — the CompNet database — covers 18 EU countries and 146 NUTS2 regions, leaving approximately 28% of EU NUTS2 units without any direct observation of gross job flows. Even where the database of Competitiveness Research Network (CompNet) provides regional estimates, these are systematically inconsistent with Eurostat regional employment data due to differences in reference populations and statistical methodologies between the two sources: the naive difference between CompNet Job Creation Rate (JCR) and Job Destruction Rate (JDR) does not in general equal the net employment growth rate computed from the Eurostat employment series, with discrepancies exceeding one percentage point in nearly half of all region-year observations.

The estimation pipeline developed in WP3 addresses these challenges in a sequence of internally consistent steps. The core estimator is the Generalised Cross-Entropy Distributionally Weighted Regression (GCE-DWR), an information-theoretic method that replaces the inconsistent CompNet rates with adjusted estimates satisfying the Eurostat net employment growth constraint exactly, while deviating minimally from the CompNet priors in the sense of minimum Kullback-Leibler divergence. A Monte Carlo validation exercise with 100 independent trials demonstrates that the GCE-DWR approach reduces estimation error by 30% or more relative to naive use of CompNet figures, and converges reliably under standard optimisation. For the 28% of NUTS2 regions not covered by CompNet, a Random Forest regression trained on GCE-adjusted estimates from covered regions uses macroeconomic predictors from Eurostat national accounts to impute JCR and JDR; the imputed values are then subjected to the same GCE-DWR adjustment in a second pass. A cross-entropy matrix balancing step reconciles the regional aggregate estimates with independent sectoral estimates to produce a fully balanced region × NACE 2-digit panel.

The workforce decompositions in the second layer of the pipeline use EU Labour Force Survey scientific use microdata — covering 31 countries and approximately 1,200 country-year files for the period 2003–2023 — to obtain employment shares by occupation, education tier, migrant status, and sex for each NUTS2 workplace region and year. These shares enter a GCE decomposition that allocates the regional aggregate JCR and JDR across groups while satisfying the constraint that their employment-weighted average equals the regional total. Group-specific prior rates are informed by Location Quotients, which embed an embeddedness hypothesis: groups structurally concentrated in a region are expected to face higher job

Funded by
the European Union

creation and lower job destruction than the regional average.

Although the results are analysed in detail in the ISABEL D4.1, some initial, key empirical findings are as follows. Average JCR across the sample is approximately 4.7% of employment per year and average JDR approximately 4%, but cross-regional and cross-sectoral dispersion is large. The COVID-19 shock of 2020 produces the only year in the panel in which average JDR exceeds average JCR (5.3% versus 4.1%), consistent with the counter-cyclical spike in job destruction documented in the literature. The post-pandemic recovery is characterised by elevated JCR reaching 5.6% in 2022, followed by a return toward pre-pandemic norms by 2023. Migrant workers — particularly migrant women — display substantially higher gross job flow rates than native workers, reflecting concentration in high-turnover sectors. Low-educated workers face higher destruction rates than their medium-educated counterparts, consistent with dual labour market structures observed across EU economies.

External validity is assessed through a UK benchmarking exercise in which the RF+GCE-DWR pipeline is applied out-of-sample to UK NUTS2 regions — which are absent from CompNet — and compared against administrative job flow data from the Office for National Statistics (ONS). Mean absolute errors average 1.4–1.5 percentage points across non-crisis years, and the GCE-DWR adjustment consistently improves on the Random Forest baseline, confirming that the pipeline extracts genuine information from macroeconomic predictors and uses the net employment growth constraint efficiently.

The WP3 database, delivered as a set of panel datasets at NUTS2 × year × industry and NUTS2 × year × workforce-dimension granularity, provides a substantially more complete and internally consistent picture of European gross job flows than any currently available alternative, and forms the empirical basis for the analyses of uneven job creation and destruction in the transitionary context that constitute the core output of the ISABEL project.

Funded by
the European Union

Table of Contents

Executive summary 4
1. Introduction 7
2. Challenges in the Estimation of Job Creation (JC) and Job Destruction (JD) figures 9
2.1 Net employment growth rate, job creation and job destruction 9
2.2 JC and JD for NUTS2 regions 10
3. Data sources 12
3.1 EUROSTAT: Regional Accounts 12
3.2 EUROSTAT: Labour Force Survey 12
3.3 Compnet database 13
3.4 Other databases 14
4. Methods 15
4.1 A DWR approach 16
4.2 Filling the gaps: a Machine-Learning solution. 17
4.3 Relliability (I): a Monte Carlo experiment 18
4.5. Region × Sector Matrix Balancing 20
4.6. Workforce Disaggregation via GCE 20
5. Results: regional JC/JD 2010-2023 22
5.1 The uneven distribution of JC and JD across regions, industries and occupations 22
5.2 Reliability (II): estimates v “real data” 24
6. Conclusions 27
Bibliography 29
Annexes ..... **Error! Bookmark not defined.**

List of Figures

Figure 1. Overview of the WP3 estimation pipeline. 15
Figure 2. GCE-DWR adjustments to CompNet JCR/JDR. Panel A: scatter of CompNet prior vs GCE-DWR estimate (JCR); Panel B: distribution of adjustments (GCE − CompNet, percentage points) for JCR and JDR; Panel C: share of region-year observations requiring adjustments of different magnitudes. 17
Figure 3. CompNet NUTS2 coverage and Random Forest gap-filling by country (EU27 + Norway). Purple bars: NUTS2 regions with CompNet job flow data; blue bars: regions imputed by the Random Forest. Percentage labels show the share of each country’s NUTS2 regions that are CompNet-covered. 18
Figure 4. Monte Carlo validation results (100 trials, n = 5 simulated regions). Panel A: distribution of MAE (percentage points) across trials for JCR and JDR; Panel B: mean MAPE (±1 SD) for naive CompNet vs GCE-DWR estimator; Panel C: GCE-DWR improvement vs initial CompNet error magnitude (JCR). 19
Figure 5. Average JCR and JDR (left) and Net Employment Growth (right) across NUTS2 regions, 2010–2023. Shaded band marks 2020. 22
Figure 6. Distribution of JCR (left) and JDR (right) across NUTS2 regions by year, 2011–2023. Each box shows interquartile range; whiskers extend to 1.5 IQR. 23
Figure 7. Average JCR and JDR by educational attainment (left) and by migrant status × sex (right), averaged across all available NUTS2 regions and years. 24
Figure 8. UK benchmarking: pipeline estimates versus ONS actual JCR and JDR data. Left column: Random Forest only. Right column: RF + GCE-DWR adjustment. Upper row: JCR; lower row: JDR. The dashed line is the 45-degree line of perfect prediction. Correlation (r) and mean absolute error (MAE) shown in each panel. 25
Figure 9. UK benchmarking: mean absolute error (MAE, percentage points) by year for JCR (left) and JDR (right). Shaded area represents the accuracy gain from the GCE-DWR adjustment over the Random Forest baseline. Grey band marks 2020. 26

Funded by the European Union

1. Introduction

The uneven distribution of job creation and job destruction across space is one of the most consequential and least-understood dimensions of European labour market inequality. While aggregate net employment statistics are routinely published for European regions, they conceal an enormous volume of simultaneous gross flows: at any point in time, some firms are expanding and others contracting, some are entering the market and others exiting, and the net figure observed in official statistics is merely the difference between these opposing forces. Understanding where in Europe jobs are being created and destroyed — and for whom — is essential for designing place-sensitive labour market and industrial policies, for assessing the distributional consequences of structural economic change, and for monitoring the progress of EU cohesion objectives.

This deliverable documents the data collection and estimation strategy developed in Work Package 3 (WP3) of the ISABEL project to meet this challenge. The objective is to construct a comprehensive panel database of gross job creation rates (JCR) and job destruction rates (JDR) for EU NUTS2 regions, covering the period 2010–2023 and disaggregated by industry (NACE 2-digit), occupation (ISCO 2-digit), educational attainment, migrant status, and sex. This database constitutes the empirical foundation upon which the substantive analyses of the ISABEL project are built.

The database described in this deliverable covers the period 2010–2023 as its primary analytical window. However, efforts are currently underway to push the historical coverage back to 2003. This backward expansion is not a minor adjustment: it is intended to bring the job-flow panel into full alignment with the temporal scope of the EU Labour Force Survey scientific-use microdata processed here, thereby enabling a coherent long-run perspective across. The methodological framework, validation procedures, and interpretation guidelines presented in this report remain fully applicable to the extended period, as the estimation strategy has been designed from the outset to operate consistently across heterogeneous labour-market conditions. Extending the series to the pre-2010 years would be particularly valuable because it incorporates the Global Financial Crisis of 2008–2009, a period characterised by unusually sharp and asymmetric shocks to regional labour markets. Including these earlier years will allow the empirical patterns, as initially documented in Section 5, especially those concerning the divergence between job creation and job destruction across EU regions, to be examined under a more turbulent macroeconomic environment.

The construction of such a database is far from straightforward. Unlike the United States, where the Business Employment Dynamics series provides quarterly establishment-level gross job flows for the entire private sector, Europe lacks an equivalent administrative data infrastructure at the sub-national level. No single European data source provides harmonised, internally consistent, and geographically complete estimates of JCR and JDR at NUTS2 resolution across all EU member states. The primary source of firm-level job flow information in Europe — the CompNet database, produced through the collaboration of national central banks and statistical institutes across the EU — provides NUTS2-level estimates for 18 countries and 146 regions, but its coverage is incomplete, its figures are systematically inconsistent with the Eurostat regional employment series due to differences in reference populations and statistical methodologies, and it does not permit simultaneous cross-tabulation by region and industry. Bridging these gaps while maintaining internal consistency of the estimates across all levels of aggregation is the core methodological problem addressed in this report.

The estimation strategy adopted is structured around a multi-stage pipeline designed to ensure that the fundamental accounting identity of the labour market — JCR minus JDR equals net employment growth — holds exactly at every level of aggregation and for every sub-dimension of the workforce. The first layer of the pipeline addresses the regional dimension: a Generalised Cross-Entropy Distributionally Weighted

Funded by
the European Union

Regression (GCE-DWR) estimator reconciles CompNet job flow priors with Eurostat net employment growth constraints for all CompNet-covered regions; a Random Forest regression imputes JCR and JDR for the approximately 28% of EU NUTS2 regions not represented in CompNet; and a cross-entropy matrix balancing step ensures that regional and sectoral estimates are mutually consistent. The second layer decomposes these regional estimates along four workforce dimensions using EU Labour Force Survey microdata as the source of employment composition weights, propagating the aggregate estimates to group-specific rates for occupations, education tiers, migrant status, and sex.

The report is organised as follows. Section 2 sets out the conceptual framework for measuring gross job flows and documents the specific challenges that arise when attempting to produce NUTS2-level estimates in the European data environment. Section 3 describes the three main data sources — Eurostat Regional Accounts, the EU Labour Force Survey scientific use microdata, and the CompNet database — in terms of their coverage, variables, and role in the pipeline. Section 4 presents the estimation methodology in full detail, covering each stage of the pipeline from the baseline GCE-DWR through to the four-dimension workforce decomposition. Section 5 presents the empirical results: the aggregate and cross-sectional patterns of job creation and destruction across EU regions, industries, and workforce groups, and the results of the UK benchmarking exercise that provides an external validity test for the pipeline. Section 6 concludes.

Funded by
the European Union

2. Challenges in the Estimation of Job Creation and Job Destruction figures

2.1 Net employment growth rate, job creation and job destruction

A central insight of modern labour economics is that aggregate net employment change conceals an enormous amount of simultaneous job creation and destruction. At any given moment, some firms are expanding and creating new positions while others are contracting and eliminating them; the *net* change observed in aggregate statistics is merely the difference between these two gross flows. Understanding this distinction is essential for the purpose of the ISABEL project, and it motivates the methodological design of WP3. The foundational framework for measuring these gross flows was established by *Davis and Haltiwanger (1992)*, whose seminal study of U.S. manufacturing formalised the concepts that have since become standard in the literature. In their framework, for any given firm or establishment at time *t*, the **job creation rate** (JCR) is defined as the employment gains at expanding or newly born establishments divided by average employment, and the **job destruction rate** (JDR) is defined as the employment losses at shrinking or dying establishments divided by average employment. The **net employment growth rate** (NEG) then satisfies the accounting identity:

$$NEG = JCR - JDR$$

This identity, which *Davis, Faberman and Haltiwanger (2006)* describe as a *"fundamental accounting identity"* of labour market analysis, is deceptively simple but has profound implications: a region or sector experiencing zero net employment change may nonetheless be undergoing simultaneous substantial job creation and destruction, reflecting a continuous reshuffling of employment positions across firms.

The empirical magnitudes involved are striking. Davis and Haltiwanger (1992) documented that in U.S. manufacturing, gross job creation and destruction rates (JCR+JDR) averaged 20.5% per year respectively—figures that dwarf the absolute net change of −2.1% over their sample period. Davis, Faberman and Haltiwanger (2006) similarly found that quarterly job creation and destruction rates average nearly 8% of employment in the U.S. private sector, meaning that for every dozen filled jobs at a given moment, on average one disappears and a similar number are newly created over the following three months. This phenomenon of excess job reallocation (defined as total job reallocation minus the minimum required to accommodate net employment change) is pervasive across industries and regions.

Three features of gross job flows are particularly important for understanding the limitations of relying exclusively on net employment statistics. First, *simultaneous creation and destruction* occur even in periods of net contraction: Davis and Haltiwanger (1992) show that in 1975, when U.S. manufacturing employment shrank by a dramatic 10%, the gross job creation rate was still 7%. Conversely, in 1973, a year of robust 7% net employment expansion, the gross job destruction rate was 6%. This simultaneity implies that net figures are profoundly uninformative about the labour market churning that workers actually experience.

Second, job flows exhibit *strong cyclical properties* that differ from those of net employment. The gross employment rate change job reallocation rate (JCR + JDR) fluctuates *counter*cyclically, rising during downturns. Davis and Haltiwanger (1992) found a simple correlation of −0.57 between net employment growth and the job reallocation rate in manufacturing, and this pattern has since been documented across many countries. Job destruction is substantially more volatile over the business cycle than job creation; in recessions, destruction spikes sharply while creation falls more moderately. This asymmetry means that during economic crises, the divergence between gross flows and net change is particularly pronounced.

Third, job flows display *substantial cross-sectional heterogeneity*—across industries, firm sizes, firm ages, and geographic regions—that is entirely invisible in net employment aggregates. Davis and Haltiwanger (1992) document annual job reallocation rates ranging from 14% in Tobacco to 28.8% in Lumber and Wood Products across two-digit manufacturing industries. Young and small plants exhibit dramatically higher reallocation rates than older and larger ones. And crucially for the ISABEL project, regional variation is similarly large. This spatial heterogeneity motivates the core research agenda of WP3: understanding *where* in Europe jobs are being created and destroyed, not just *how many* on net.

Funded by
the European Union

It is important to note that, as Davis, Faberman and Haltiwanger (2006) emphasise, JCR and JDR as defined in this tradition are *employer-side* measures reflecting changes in the *number of employment positions*, not the reallocation of individual workers. They differ conceptually from worker flow measures (hires and separations), which additionally capture turnover at stable firms, quits, and temporary layoffs unrelated to net employment changes. Gross job flows represent a lower bound on total worker flows, since workers who separate from shrinking firms and are rehired by expanding ones are counted in both the destruction and creation measures. Davis and Haltiwanger (1992) calculate that job reallocation accounts for 35–56% of total worker reallocation—a substantial fraction, but one that leaves the remainder attributable to worker-side decisions independent of employer-level employment changes.

2.2 JC and JD for NUTS2 regions

Producing reliable estimates of JCR and JDR at the NUTS2 regional level introduces a set of specific methodological challenges that go well beyond those faced in national-level analyses. This section documents these challenges and explains how they motivate the estimation strategy adopted in WP3.

i. **Absence of direct regional job flow data in Europe.** Unlike the United States, where the Business Employment Dynamics (BED) series provides quarterly establishment-level job flows for the entire private sector back to 1990, Europe lacks an equivalent administrative data infrastructure at the regional level. Eurostat's flagship employment data source for NUTS2 regions—the Labour Force Survey (LFS)—provides estimates of total employment (and its changes) by region, but it is a household survey that measures employment from the worker side and does not directly yield the establishment-level expansion and contraction counts needed to compute gross job flows. National administrative data on firm-level employment changes exist in some member states (e.g., the French DADS or the Danish IDA database), but these are not harmonised across the EU, are typically not released at NUTS2 level for research purposes.

ii. **Firms databases: coverage and constraints.** The primary source of company-level job flow information for this project is the *CompNet* database, which aggregates firm-level microdata from national statistical offices and central banks across European countries. CompNet provides JCR and JDR figures (variables *LV04_jcr_sam_N_tot* and *LV14_jdr_sam_N_tot*) at the country level and, for some countries, at the NUTS2 level. However, CompNet coverage is far from complete. Many EU NUTS2 regions are not represented directly in CompNet outputs, and for those that are, the available time series typically does not extend uniformly across all years from 2010 to 2023. Furthermore, a critical structural feature of CompNet is that it *does not permit simultaneous cross-tabulation by region and industry*: country-level CompNet data can be broken down by sector, and region-level data is available in aggregate, but the joint distribution of region × sector job flows is not directly released. This creates a severe data gap when the research objective—as in ISABEL—requires estimates disaggregated by both dimensions.

iii. **Internal inconsistency of the available data.** Even where CompNet provides NUTS2-level JCR and JDR figures, these estimates exhibit a systematic statistical inconsistency with respect to Eurostat regional employment data. The accounting identity NEG = JCR − JDR must hold by construction if JCR, JDR, and NEG are measured from the same data-generating process. In practice, because CompNet's job flow estimates derive from *firm-level* microdata aggregated by national statistical offices, while the NEG computed from Eurostat's regional employment series is based on *household-survey* data, the two are systematically inconsistent: for a typical NUTS2 region and year, the naive difference *JCR—JDR* does not equal the Eurostat-derived NEG. This inconsistency is not a measurement error in the usual sense; it reflects the fact that the two data sources have different coverage rules, reference populations, and statistical methodologies. The main source of inconsistencies is that official Eurostat employment data in the Regional Accounts follow the establishment criterion (where the job is located), while CompNet estimates pay attention to the regional location of the firm (i.e. the company headquarters), regardless of the actual location of the jobs across space. This creates a fundamental problem for any analysis that needs internally consistent regional job flow estimates: naive use of CompNet figures would violate the basic

accounting identity of the labour market and produce misleading results. The magnitude of these inconsistencies is non-trivial. Preliminary diagnostics on the ISABEL dataset revealed that the error exhibits substantial variance across regions and years, with some observations showing discrepancies of several percentage points. Regions with extreme values of this naive error tend to cluster geographically and temporally in ways that suggest systematic differences in the coverage of the underlying microdata rather than random measurement noise.

iv. **Partial coverage and the imputation problem.** Even after reconciling CompNet data with Eurostat employment constraints, a large number of NUTS2 regions—particularly in smaller or newer EU member states—remain entirely uncovered by CompNet at the regional level. These regions have Eurostat employment data and thus an observable NEG, but no prior estimate of how total employment change is decomposed into gross creation and destruction. This coverage gap means that a complete panel of NUTS2 job flows cannot be produced by data reconciliation alone; it requires a *prediction step* that extrapolates from the relationship between observable macroeconomic predictors and job flows in the covered regions to generate estimates for the uncovered ones. This prediction problem is challenging because the determinants of gross job flows (business dynamism, firm-size distributions, sectoral composition) are imperfectly measured at the NUTS2 level for all EU countries.

v. **Sector and workforce disaggregation.** Beyond the baseline regional estimates, the ISABEL project analysis requires disaggregation of job flows along multiple additional dimensions: NACE 2-digit industries, occupational groups (ISCO 2-digit), educational attainment (three tiers), migrant status (native/migrant), and sex. Each of these dimensions introduces further data challenges. No single EU data source provides a direct cross-tabulation of regional job flows by any of these dimensions. The disaggregation must therefore be inferred from partial information: sector-level national accounts data, Labour Force Survey microdata on workforce composition by region and demographic group, and structural assumptions about the relationship between workforce composition and group-specific job creation and destruction rates. The key principle underlying this disaggregation—developed in detail in Section 4—is that the sum of group-specific job flows, weighted by employment shares, must equal the aggregate regional job flow established in the first stage of estimation. This cross-entropy reconciliation approach preserves the internal consistency of the estimates while propagating information from both the top-down (aggregate) and bottom-up (workforce composition) data sources.

vi. **NUTS classification vintage harmonisation.** A further practical challenge for a panel spanning 2010–2023 is that the NUTS classification itself has changed over this period. The NUTS 2010, 2013, 2016, and 2021 vintages differ in the definitions of some NUTS2 boundaries, particularly following administrative reforms in several member states. LFS microdata, CompNet outputs, and Eurostat regional accounts are not all published under the same vintage, and the LFS in particular frequently uses older NUTS codes in its scientific use files. Constructing a consistent panel requires systematic crosswalk operations that map all regional identifiers to the NUTS 2021 classification, with particular care for boundary changes in Germany, France, Poland, and other countries that underwent significant territorial reorganisation during the study period. Failure to harmonise vintages would introduce spurious breaks in the time series and mismatches between the job flow estimates and the demographic data used in the disaggregation stage.

The methodological framework described in Section 4 addresses all of these challenges in a unified and internally consistent way. The Generalised Cross-Entropy Distributionally Weighted Regression (GCE-DWR) approach is designed precisely to handle the situation where a prior estimate of the quantities of interest (from CompNet) exists but is inconsistent with an constraint (the Eurostat NEG fro the Regional Accounts), and where the goal is to recover adjusted estimates that satisfy the constraint while deviating minimally—in an information-theoretic sense—from the prior. As it is described below, a Random Forest imputation extends coverage to unobserved regions. And the sequential GCE decomposition stages propagate these estimates to finer levels of disaggregation using LFS microdata as auxiliary information, while maintaining the accounting identities at every level of aggregation.

Funded by
the European Union

3. Data sources

The construction of the WP3 database draws on three main data sources: (i) the Eurostat Regional Accounts, which supply the employment totals and macroeconomic aggregates that serve both as binding constraints in the estimation procedure and as predictors in the machine-learning imputation step; (ii) the EU Labour Force Survey scientific use microdata, which underpin the workforce composition panel and the disaggregated decompositions; and (iii) the CompNet database, which provides the firm-level job flow information that constitutes the primary input to the GCE-DWR estimation. This section describes each source in terms of its institutional origin, geographic and temporal coverage, the specific variables used, and the role it plays in the pipeline.

3.1 EUROSTAT: Regional Accounts

The *Statistical Office of the European Union* (Eurostat) produces a comprehensive system of harmonised regional statistics covering all EU member states plus several candidate and associated countries. For the purposes of this project, Eurostat data serve two distinct but related functions: they provide the **employment constraint** that the GCE-DWR estimator must satisfy for every region and year, and the **macroeconomic predictors** used by the Random Forest model to impute JCR and JDR figures for regions not covered by CompNet.

The primary dataset for regional employment is extracted from the *Eurostat Regional Accounts* system, specifically the NUTS2-level employment series contained within the national accounts by region tables (*nama_10r_2gdp* and *nama_10r_3gva*). Total employment expressed in thousands of persons is used to calculate the net employment growth rate (NEG) for each NUTS2 region and year as the first difference of employment divided by the employment stock, following the Davis-Haltiwanger symmetric growth rate convention. This NEG figure constitutes the binding information constraint in the GCE problem: the adjusted JCR and JDR estimates must satisfy JCR − JDR = NEG exactly for every region-year cell. The data are available at annual frequency, covering all EU27 NUTS2 regions for the period from the early 1990s to 2023, though the effective panel used in WP3 is restricted to 2010–2023 in order to align with CompNet coverage. Geographic coverage of the Eurostat Regional Accounts extends to all NUTS2 regions in the 27 EU member states as well as Norway, Iceland, and the United Kingdom, amounting to over 240 NUTS2 units under the NUTS 2021 classification. All regional series are harmonised to the NUTS 2021 classification using the official Eurostat correspondence tables, resolving the boundary changes that occurred between the NUTS 2010, 2013, 2016, and 2021 vintages. The data are accessed via the Eurostat API using the *eurostat* R package and stored in a local cache to ensure reproducibility.

3.2 EUROSTAT: Labour Force Survey

The *EU Labour Force Survey* (EU-LFS) is the largest harmonised household survey in Europe, conducted continuously across all EU member states, candidate countries and several associated nations. It is coordinated and published by Eurostat under a common methodological framework designed to ensure comparability across countries and time. The LFS covers the civilian non-institutional population aged 15 and over and collects detailed information on the labour market status of individuals, their sector of employment, occupation, educational attainment, country of birth, and a range of other socio-demographic characteristics. For the purposes of WP3, the LFS is the **sole source** of sub-national workforce composition data and underpins both the structural employment panel and the demographic decompositions of JCR and JDR.

The version of the LFS used in this project is the *scientific use microdata files*, which are distributed by Eurostat to accredited researchers and differ from the publicly available aggregate tables in two important respects. First, the microdata contain *individual-level records* with the full set of survey variables, permitting custom aggregations along any combination of dimensions. Second, and critically for this project, the scientific use files include the workplace region variable (*REGION_2DW* in recent vintages), which identifies

Funded by
the European Union

the NUTS2 region *where the respondent works* rather than where they reside. This distinction is fundamental: a job flows analysis must be anchored to the location of employment positions, not the residential location of workers, and the published Eurostat LFS aggregate tables — which are all coded by residence region — cannot serve this purpose. Access to the workplace region variable therefore makes the scientific use microdata indispensable.

The LFS microdata files used cover 31 countries (all 27 EU member states plus Norway, Iceland, Switzerland, and the United Kingdom) and span the years 2003 to 2023, amounting to approximately 1,200 country-year files in comma-separated format. Each file contains one record per survey respondent, with population weights (*COEFFY*) expressed in thousands of persons, so that weighted sums yield employment totals in thousands directly. From each file, only ILO-employed respondents (*ILOSTAT* = 1) with a known workplace NUTS2 region are retained. The key variables extracted for aggregation are: the two-digit NACE Rev. 2 industry code (*NACE2_2D*), the two-digit ISCO-08 occupation code (*ISCO08_2D*), sex (*SEX*), three-tier educational attainment (*HATLEV1D*, distinguishing Low/ISCED 0–2, Medium/ISCED 3–4, and High/ISCED 5–8), and migrant status derived from the country-of-birth group code (*COUNTRYB*). Migrant status is defined as a binary variable equal to one when the respondent was born outside the country of interview (group code *EU15* or *FOR* in the anonymised scientific use files) and zero for native-born respondents (group code *NAT*). Industry codes are harmonised from NACE Rev. 1 to NACE Rev. 2 for pre-2008 survey years using a standard crosswalk, and ISCO-88 codes are bridged to ISCO-08 at the two-digit level for pre-2011 years.

A significant technical challenge is that the NUTS codes embedded in the LFS microdata do not follow a single vintage. Depending on the country and survey year, regional codes may reflect the NUTS 2003, 2006, 2010, 2013, or 2016 classifications rather than the current NUTS 2021 standard. All codes are harmonised to NUTS 2021 as a post-processing step using Eurostat's official correspondence tables, with proportional allocation applied where a single old region is split across multiple new NUTS2 boundaries. One important exception concerns Germany, for which the LFS scientific use files report only NUTS1 regions (*Länder*) and not NUTS2 sub-regions; Germany is therefore excluded from the workplace-based aggregation and handled separately. The resulting output contains employment totals (in thousands) for each combination of NUTS2 workplace region, NACE2 two-digit industry, ISCO two-digit occupation, sex, education tier, and migrant status, by year.[1] This panel constitutes a key standalone output of WP3 and is also the primary source of employment share weights used in the demographic decompositions of JCR and JDR described in Section 4.

3.3 Compnet database

The *Competitiveness Research Network* (CompNet) database is a micro-aggregated firm-level dataset produced through the joint effort of the European Central Bank, national central banks, national statistical institutes, and other research institutions across the European Union. CompNet aggregates longitudinal firm-level administrative and survey data from each participating country into a set of harmonised distributional statistics on firm performance, employment dynamics, and productivity. Crucially for the purposes of WP3, CompNet implements a job flow methodology closely aligned with the Davis-Haltiwanger-Schuh (DHS) framework: it computes gross job creation and destruction rates by aggregating firm-level employment changes, capturing both the intensive margin (expansion and contraction of continuing firms) and the extensive margin (entry of new firms and exit of closing ones). The job creation rate and job

[1] The database described in this deliverable covers the period 2010-2023 as its primary analytical window, but work is currently underway to extend the estimates back to 2003, which would align the job flow panel with the full temporal coverage of the EU Labour Force Survey scientific use microdata processed in WP3. The methodology, validation results, and interpretation of findings presented in this report apply without modification to the extended period. The empirical patterns documented in Section5 can be expected to carry through to the pre-2010 period, though the earlier years encompass the Global Financial Crisis of 2008-2009, which is likely to generate particularly pronounced divergences between job creation and destruction rates across EU regions.

destruction rate variables used in this project are *LV04_jcr_sam_N_tot* and *LV14_jdr_sam_N_tot* respectively, which represent the DHS-consistent gross flow rates as a proportion of average employment. In terms of geographic coverage, the version of CompNet used provides NUTS2-level job flow estimates for 18 European countries: Belgium, Czechia, Denmark, Spain, Finland, France, Croatia, Hungary, Italy, Lithuania, Latvia, Malta, Poland, Portugal, Romania, Sweden, Slovenia, and Slovakia. Across these countries, the dataset covers 146 distinct NUTS2 regions with at least one valid JCR or JDR observation. Temporal coverage runs from 2000 to 2023 at annual frequency, though the density of coverage varies: some country-region-year cells are missing due to confidentiality suppression in the underlying microdata or gaps in national data delivery to CompNet. In the WP3 pipeline, coverage is set to 2010–2023 for the main panel, with earlier years used only for model training purposes. Notably, several large EU economies — including Germany, Austria, the Netherlands — are not represented in the NUTS2-level CompNet release, which creates the coverage gap that the Random Forest imputation step is designed to fill.

An important structural feature of the CompNet NUTS2 release is that it is aggregated across all sectors: the available JCR and JDR figures represent economy-wide averages for each region, and CompNet does not permit simultaneous cross-tabulation by region *and* industry. The national-level CompNet release provides sectoral breakdowns (by NACE two-digit industry), and these are used in the sectoral extension of the pipeline, but the joint region × sector distribution must be recovered through the matrix balancing step described in Section 4 rather than observed directly. This absence of joint regional-sectoral firm-level data is one of the most significant data constraints facing the project, and it is a key motivation for the multi-step estimation architecture adopted in WP3.

3.4 Other databases

Beyond the job flow rates themselves, the CompNet data file used in WP3 — the *CompNet–Eurostat Merged v1.4* dataset — also contains a set of Eurostat business demography statistics joined at the NUTS2 level, including the number of enterprise births and deaths, employment in new and closing firms, three-year survival rates, and business churn rates. These business demography indicators, drawn from the Eurostat Structural Business Statistics (*bd_r_bdemography*) and the Business Register series, serve as additional contextual variables for the analysis. The file also includes real labour productivity (index, 2015 = 100) derived from national accounts at the NUTS2 level. In total, the merged dataset contains approximately 7,830 region-year observations spanning all available years and countries, forming the primary working dataset for Step 1 of the WP3 estimation pipeline.

Additionally, Eurostat Regional Accounts are also used in a different fashion, since they also provide the *macroeconomic predictors* that feed the Random Forest imputation step. These are drawn from the *nama_10_a64* and *nama_10_a64_e* datasets, which report national accounts aggregates by industry (NACE Rev. 2 two-digit level) at the country level. The variables used as predictors are GVA growth (derived from B1G in current prices), labour productivity (GVA per employee), the growth rate of compensation per employee (derived from D1), and the investment rate (gross fixed capital formation P51G as a share of GVA). These indicators are available at annual frequency at the country level and are matched to the NUTS2 training and prediction datasets via the country code embedded in each NUTS2 identifier. For the regional-level Random Forest, supplementary regional predictors include GDP growth from *nama_10r_2gdp* and the regional unemployment rate from *lfst_r_lfu3rt* (the Eurostat regional labour market statistics derived from the LFS). In combination, these series capture the macroeconomic conditions — structural change, labour cost dynamics, capital intensity, and labour market slack — that the empirical job flows literature[2] identify as the principal drivers of gross job creation and destruction rates.

[2] See Clark (1998) or Haltiwanger et al. (2013) as good examples of papers that explore the empirical relationshipbs between diverse economic indicators and processes of creation and destruction of jobs in the labour market.

Funded by
the European Union

4. Methods

The production of job creation and destruction figures for EU NUTS2 regions involves a multi-stage pipeline that addresses, in sequence, the data limitations described in the preceding section. The architecture is designed around a single organising principle: at every level of aggregation and disaggregation, the accounting identity *JCR − JDR = NEG* must hold exactly. Each stage takes as its input the outputs of the previous one and introduces additional information through constraints or prior distributions while preserving the identities already imposed. Figure 1 provides a schematic of the pipeline.

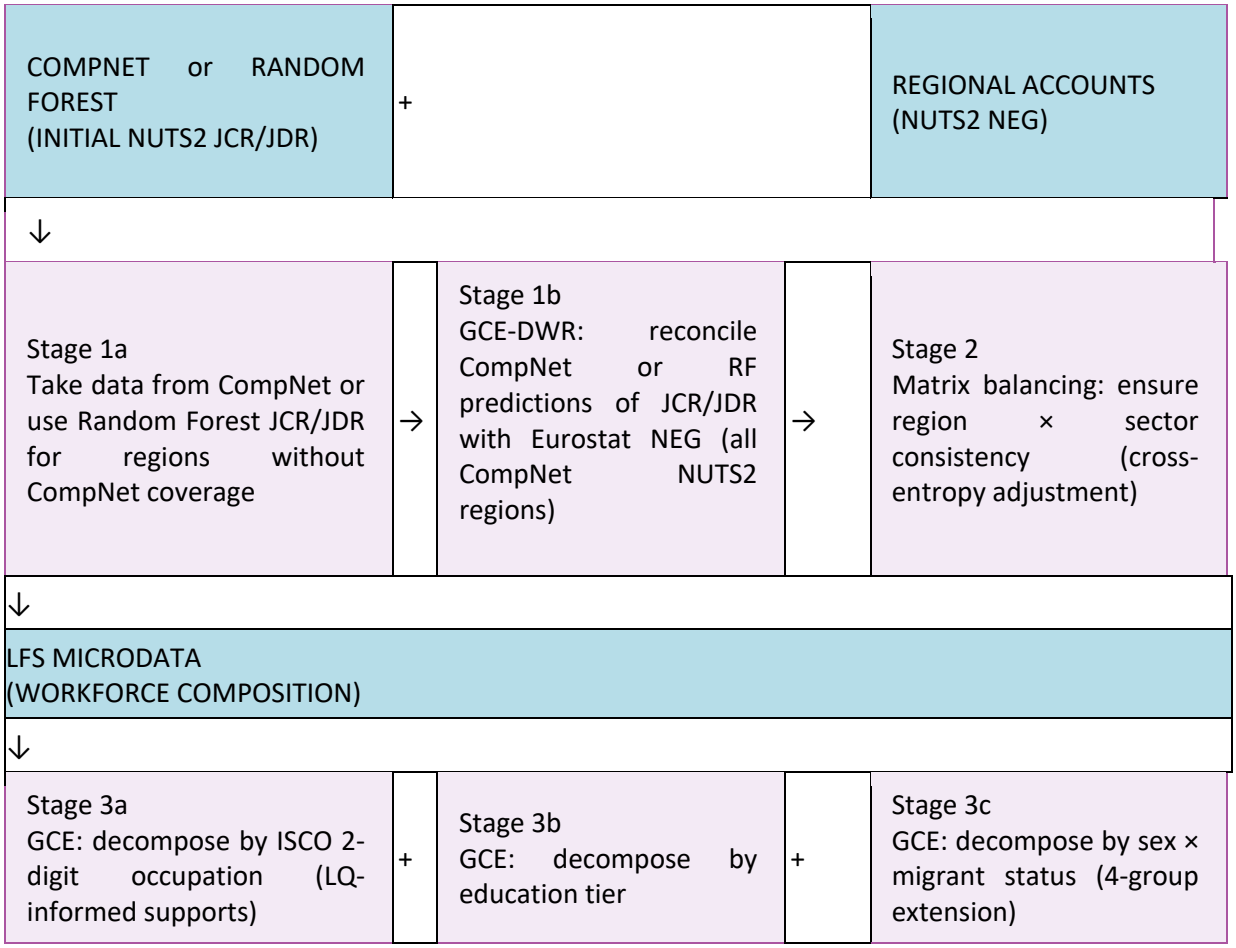


Figure 1. Overview of the WP3 estimation pipeline.

The pipeline shows parts (shaded in pink) the different stages of the process, while other parts (shaded in blue) contain the data that are employed as points of departure or as constraints. The procedure consists of two broad layers. The first layer (Stages 1a–1b) produces consistent regional JCR/JDR estimates using firm-level data from CompNet, Eurostat employment constraints, machine-learning imputation, expanding to a sectoral detail by matrix balancing in the second layer (Stage 2). Finally, the third layer (Stages 3a–3c) decomposes these regional estimates along workforce sub-dimensions — occupation, education, migrant status, and sex — using the EU-LFS microdata as the source of employment composition weights. Each stage is described in detail below.

Funded by the European Union

4.1 A DWR approach

The core estimation challenge in Stage 1a is to produce regional JCR and JDR figures that are internally consistent with official Eurostat employment statistics, given that the CompNet firm-level data provides prior estimates of both rates but these estimates systematically violate the accounting identity when confronted with the Eurostat net employment growth series. The approach adopted is the Generalised Cross-Entropy Distributionally Weighted Regression (GCE-DWR), which belongs to the family of entropy econometrics methods introduced by Golan, Judge and Miller (1996) and applied to labour market data following the estimation strategy developed in the Davis-Haltiwanger-Schuh tradition.

The GCE-DWR problem can be stated as follows. For each region *r* and year *t*, the CompNet database supplies prior estimates of the job creation rate and the job destruction rate. These are treated as signal-bearing but unreliable: they reflect the correct order of magnitude of the underlying gross flows, but they are not consistent with the Eurostat-derived net employment growth rate. The GCE estimator replaces the naive CompNet rates with adjusted estimates JCR_r and JDR_r that satisfy this identity exactly, by solving a minimum Kullback-Leibler divergence problem subject to the constraint. In the DWR parameterization, each adjusted rate is modelled as a weighted expectation over a finite support set $b = \{0, 1, 2\}$: the rate $JCR_{\mathrm{r}} = \beta^{\mathrm{c}}{}_{\mathrm{r}} \cdot jcr_{\mathrm{r}}$, where $\beta^{\mathrm{c}}{}_{\mathrm{r}} \in [0, 2]$ is the adjustment factor whose distribution is estimated by the GCE optimization and a symmetric rationale is adopted for parameterizing job destruction rates. The support set implies that the adjusted rate can range from zero to twice the CompNet prior, which accommodates the empirical distribution of discrepancies observed in the data while ensuring non-negativity.

The information constraint that must be satisfied simultaneously for all *n* regions in a given year is:

$$E[\beta^{\mathrm{c}}{}_{\mathrm{r}}] \cdot jcr_{\mathrm{r}} - E[\beta^{\mathrm{d}}{}_{\mathrm{r}}] \cdot jdr_{\mathrm{r}} = NEG_{\mathrm{r}}$$

where the expectations are taken over the estimated probability distributions on the support set. The statistical structure of this decomposition problem is closely related to the original DWR framework developed by Bidani and Ravallion (1997) for decomposing aggregate social indicators using distributional data. In that original framework, an aggregate indicator for a country is expressed as the population-weighted mean of subgroup-specific latent means, and the subgroup values are then recovered from cross-country variation in the distributional weights. The key insight shared with our approach is to treat the unobserved subgroup quantities as the objects of interest and to exploit an exact accounting identity — linking the aggregate to its subgroup decomposition — as the binding constraint on estimation. Bidani and Ravallion implement this as a random coefficients regression estimated by generalised least squares; our approach instead adopts the entropy econometrics framework of Golan, Judge and Miller (1996), which avoids parametric distributional assumptions on the subgroup quantities and naturally accommodates the bounded, non-negative character of job flow rates. Both approaches share the fundamental principle that the aggregate constraint provides identification while the population shares provide the informational content needed to distribute the aggregate across subgroups.

The joint estimation across all regions for a given year — rather than region by region — is what gives the method its *weighted regression* character: all regions simultaneously inform of their respective employment growth constraints, and these constraints determine the extent to which each region's CompNet rates are adjusted upward or downward. Regions with larger initial inconsistencies receive larger adjustments, while those already close to satisfying the identity receive minimal adjustment, in the information-theoretic sense of minimum deviation from the prior.

The prior distribution over the support is set to the uniform distribution *q = (1/3, 1/3, 1/3)* for both JCR and JDR in Stage 1a, reflecting the absence of additional information distinguishing which of the three support points is more probable for a given region.[3] The problem is solved year-by-year for all regions with valid data in that year, processing the full 2010–2023 panel. A further preparatory step addresses Portugal's data structure: CompNet regional data for Portugal's recently created NUTS2 regions (PT1B, PT1D) is absent in the raw files, and these regions are assigned the rates of the geographically adjacent and structurally similar

[3] The dual problem is solved numerically using the constrOptim.nl function from the *alabama* R package, implementing a BFGS quasi-Newton method with augmented Lagrangian constraint handling.

PT1A before the GCE step is applied.

Figure 2 summarises the magnitude and direction of the GCE-DWR corrections applied to CompNet JCR and JDR rates across all matched region-year observations. Panel A plots the CompNet prior against the GCE-adjusted estimate: most points lie below the 45-degree line for JCR, indicating a systematic downward correction, while the reverse holds for JDR. This pattern is consistent with the 'headquarters bias' present in CompNet: because firm-level employment changes are reported by company headquarters rather than by workplace region, growing multi-plant firms attribute job gains to their head-office location, inflating creation rates in major urban centres while underestimating them in peripheral regions. The GCE-DWR correction redistributes these flows to be consistent with the Eurostat workplace-based NEG. Panel B shows the full distribution of adjustments: 72% of JCR entries are corrected downward (mean adjustment −0.83 percentage points), while 72% of JDR entries are corrected upward (mean +0.20 pp). Panel C documents that non-trivial adjustments (exceeding 1 percentage point) are required in 46% of observations, and adjustments exceeding 2 percentage points in 21% of observations, underlining that naive use of CompNet figures would introduce significant bias into any analysis that requires internal consistency with Eurostat employment totals.

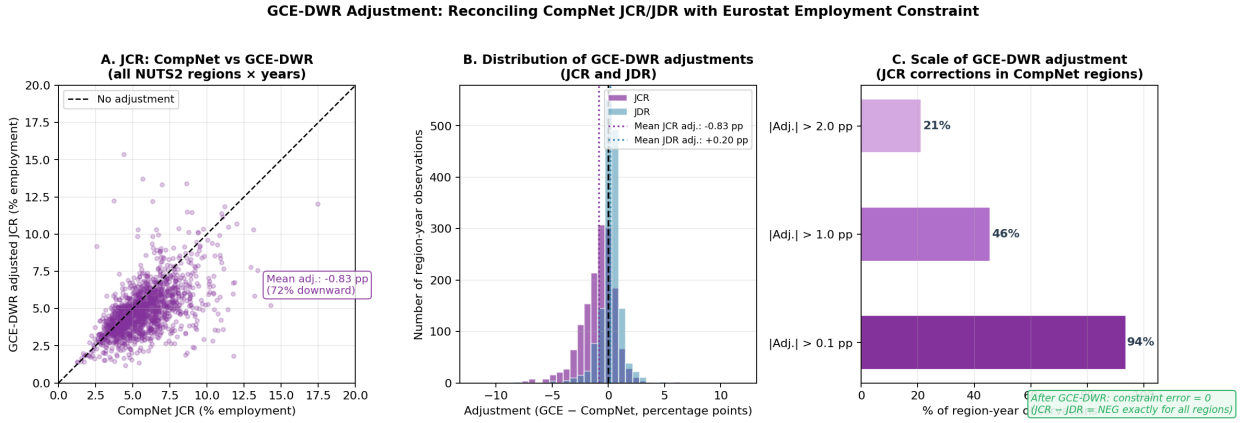


Figure 2. GCE-DWR adjustments to CompNet JCR/JDR. Panel A: scatter of CompNet prior vs GCE-DWR estimate (JCR); Panel B: distribution of adjustments (GCE − CompNet, percentage points) for JCR and JDR; Panel C: share of region-year observations requiring adjustments of different magnitudes.

4.2 Filling the gaps: a Machine-Learning solution.

Stage 1b addresses the coverage gap: 18 EU countries with 146 NUTS2 regions are represented in CompNet, but the full EU27 comprises over 240 NUTS2 units. Regions in countries not covered by CompNet — most notably Germany, Austria, the Netherlands, Greece, Ireland — have Eurostat employment data and thus an observable NEG, but no prior estimate of how net change decomposes into gross creation and destruction. A Random Forest regression is used to impute JCR and JDR for these uncovered regions.

The Random Forest models are trained on the set of CompNet-covered region-year observations where GCE-adjusted JCR and JDR are available. The predictor set is drawn entirely from Eurostat national accounts and regional databases, ensuring that the same predictors are available for both covered and uncovered regions. The variables used are: GVA growth (annual growth rate of gross value added at the country level, from *nama_10r_3gva*); labour productivity (real GVA per employee, 2015 = 100, from the CompNet-Eurostat merged file, used as a level variable reflecting structural economic conditions); regional unemployment rate (from *lfst_r_lfu3rt*); and GDP growth (from *nama_10r_2gdp*). In the sectoral extension of the pipeline (Stage 1b applied at the country × NACE 2-digit level), additional predictors are constructed from *nama_10_a64*: the growth rate of labour productivity, the growth rate of compensation per employee, and the investment rate (GFCF as a share of GVA). These sector-level predictors are available at the country level and are matched to region-sector observations via the country identifier.

Funded by
the European Union

The Random Forest is configured with 500 trees and the standard √predictors split rule at each node. Missing predictor values (which occur in some country-year combinations due to Eurostat data gaps) are imputed using the column median of the training data before prediction, ensuring that the prediction step does not introduce additional missingness. Predicted JCR and JDR rates are bounded to the interval [0.1%, 30%] to exclude implausible extrapolations, and the same GCE-DWR procedure described in Section 4.1 is then applied to the combined dataset (CompNet-covered regions with GCE-adjusted rates, plus RF-predicted regions), enforcing the NEG identity for all regions simultaneously in the second pass. This ensures that the RF predictions, while not themselves guaranteed to satisfy the identity, are corrected to do so in the combined estimation step. The result is a complete panel covering all EU NUTS2 regions for 2010–2023.

Figure 3 documents the CompNet NUTS2 coverage gap and the extent of Random Forest gap-filling across all EU27 countries and Norway. Countries in which CompNet provides full or near-complete NUTS2 coverage — including France, Italy, Spain, Poland, Belgium, and the Scandinavian economies — are shown with their CompNet-covered regions in purple; countries where Random Forest imputation fills the gap are shown in blue. Germany, Austria, the Netherlands, Greece, Ireland, Bulgaria, and Cyprus are among the most important economies not covered by CompNet at the NUTS2 level: together they account for a substantial share of EU employment and regional diversity that would be absent from any analysis relying on CompNet alone. In total, approximately 66 out of 240 NUTS2 regions (28%) in the panel are imputed by the Random Forest step. Crucially, these regions are not randomly distributed across the EU: they tend to be in larger, more centralised economies where administrative data are less likely to have been contributed to CompNet, creating a potential systematic gap in any purely CompNet-based analysis of European regional labour markets.

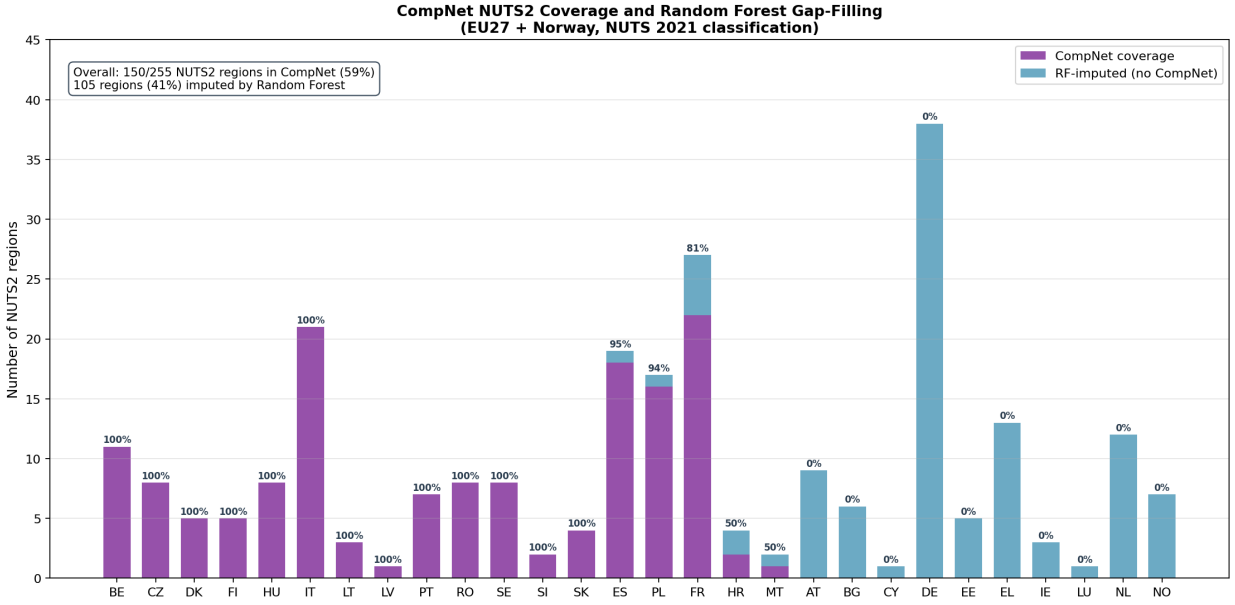


Figure 3. CompNet NUTS2 coverage and Random Forest gap-filling by country (EU27 + Norway). Purple bars: NUTS2 regions with CompNet job flow data; blue bars: regions imputed by the Random Forest. Percentage labels show the share of each country's NUTS2 regions that are CompNet-covered.

4.3 Relliability (I): a Monte Carlo experiment

Before applying the GCE-DWR estimator to the actual data, its finite-sample properties are assessed through a structured Monte Carlo simulation. The simulation design closely mirrors the empirical setting: synthetic job creation and destruction matrices are generated for $n = 5$ regions, with establishment-level JCR and JDR figures drawn from empirically calibrated distributions, and aggregate net employment growth computed from the row sums. The CompNet-style "naive" estimates are the row totals (gross rates), which

by construction do not satisfy the column-sum identity for net employment change. The GCE-DWR estimator is then applied, and its adjusted estimates are compared to the true underlying gross flows.

One hundred independent Monte Carlo trials are run. In each trial, the estimator's performance is measured by three metrics relative to the naive (unadjusted CompNet) baseline: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE), computed separately for JCR and JDR. Across trials, the GCE-DWR method consistently outperforms the naive approach by margins of 30% or more in MAE and RMSE for both the JCR and JDR. Constraint satisfaction is near-perfect in the great majority of trials (maximum violation below 10^{-3} in over 90% of cases), confirming that the GCE algorithm reliably converges to the constrained solution. These results provide a principled justification for preferring the GCE-adjusted estimates over the raw CompNet rates when the goal is to produce internally consistent regional job flow figures.

The simulation also reveals a robustness property of the GCE approach: even when the true adjustment factors β^c and β^d are heterogeneous across regions — as they are in the empirical application — the estimator recovers them accurately, because the cross-regional information constraint system is collectively informative about the distribution of adjustments needed. The uniform prior $q = (1/3, 1/3, 1/3)$ proves adequate as a default: it reflects genuine ignorance about the direction and magnitude of the adjustment for any given region without introducing systematic bias.

Figure 4 presents the results of the Monte Carlo exercise in three complementary panels. Panel A displays the distribution of MAE across 100 independent trials for both JCR and JDR, comparing the naive CompNet estimator against the GCE-DWR approach: the violin plots show that the GCE-DWR estimator consistently achieves lower error, with the improvement visible both in the central tendency and in the upper tail. Panel B reports the mean and standard deviation of the MAPE for both estimators and both rates: the GCE-DWR delivers improvements of approximately 30–35% in MAPE relative to the naive baseline. Panel C reveals a further desirable property: the magnitude of the GCE-DWR improvement is positively correlated with the size of the initial CompNet error — in other words, the estimator corrects more aggressively where correction is most needed, and does so automatically without any tuning, because the objective function penalises larger deviations from the Eurostat constraint more heavily.

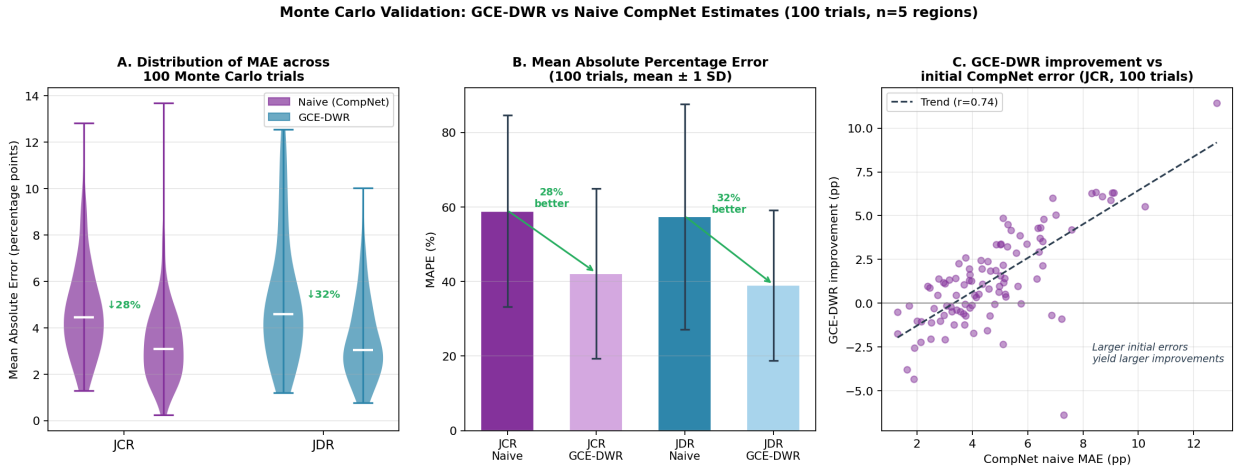


Figure 4. Monte Carlo validation results (100 trials, n = 5 simulated regions). Panel A: distribution of MAE (percentage points) across trials for JCR and JDR; Panel B: mean MAPE (±1 SD) for naive CompNet vs GCE-DWR estimator; Panel C: GCE-DWR improvement vs initial CompNet error magnitude (JCR).

The regional estimates produced in Stages 1a–1b are economy-wide aggregates: they capture total job creation and destruction across all sectors in each NUTS2 region. To produce sectoral estimates, the same GCE-DWR and Random Forest pipeline is replicated at the country × NACE 2-digit × year level, using the national-level CompNet sectoral release as the primary data source for job creation and destruction by industry. CompNet provides JCR and JDR disaggregated by NACE two-digit industry at the country level but, as noted previously, does not simultaneously cross-tabulate by region, so this step necessarily operates at the country level.

A significant practical challenge in this step concerns the NACE classification in Eurostat national accounts.

Funded by the European Union

Eurostat's A*64 industry classification groups several NACE Rev. 2 two-digit codes into combined entries (for example, codes B05 to B09 — all mining sub-sectors — are reported as a single aggregate "B" in some publications). A naive regex-based approach to stripping letter prefixes from Eurostat codes silently drops these sectors, since the letter alone does not map to a single two-digit code. This problem was resolved by constructing a comprehensive crosswalk table of approximately 170 A*64-to-NACE-2d mappings, which handles combined codes explicitly and gives priority to more granular individual-code data when both are present in the same country-year cell.

For country-sector-year cells not covered by CompNet (either because the country is absent from CompNet, or because suppression rules remove specific industry cells), the Random Forest imputation is again applied, this time using the sector-level predictors from *nama_10_a64* described in Section 4.2. The result is a balanced panel of JCR and JDR rates at country × NACE 2-digit × year, serving as the column-sum constraints for the subsequent matrix balancing step.

4.4. Region × Sector Matrix Balancing

After Stage 1a-b produces NUTS2-level aggregate estimates, these results must be reconciled with observed national data by sector: when the regional estimates are aggregated up to the country level, the resulting sector mix does not in general match the sectoral estimates produced independently. The matrix balancing step resolves this inconsistency by adjusting a region × sector matrix of JCR (and separately JDR) rates so that its row sums match the regional totals from Stages 1a-b and its column sums match the sectoral totals available in CompNet while minimising the information loss relative to an initial prior matrix.

The balancing is implemented as a cross-entropy minimisation problem, which is formally equivalent to biproportional (RAS) adjustment but accommodates non-uniform weighting by employment shares and handles structural zeros (cells where a sector is absent in a region) naturally without introducing spurious values. The prior matrix for each country-year is constructed by distributing the regional totals across sectors proportionally to the national sector shares from Stage 1c. The cross-entropy balancer then iterates until both row and column sum constraints are satisfied simultaneously to a pre-specified tolerance. This step is applied country by country and year by year, processing 386 country-year combinations in the empirical application. The output is a fully balanced matrix of JCR and JDR rates at NUTS2 × NACE 2-digit × year, where regional totals, sectoral totals, and the NEG identity all hold simultaneously.

The choice of cross-entropy over simple RAS is motivated by two considerations. First, the employment weights allow larger adjustments in cells that are economically small (where the information content of the prior is lower) and smaller adjustments where employment is large and the prior is likely more reliable. Second, the minimum-information-loss property of the cross-entropy approach ensures that the balancing does not introduce patterns not supported by either the regional or sectoral data: it represents the least assumption-requiring way to reconcile two independently estimated marginals.

4.5. Workforce Disaggregation via GCE

The last layer of the pipeline (Stages 3a–3c) decomposes the NUTS2-level JCR and JDR estimates produced in the first layer along four sub-dimensions of the workforce: occupation (ISCO 2-digit), educational attainment (three tiers), migrant status (binary), and sex (combined with migrant status in a four-group decomposition). All four decompositions share the same structural framework and differ only in the grouping variable, the source of employment shares, and some features of the prior design.

The identifying constraint in each decomposition is that the employment-weighted average of group-specific JCR (and JDR) must equal the regional aggregate estimated in Stages 1a-1b:

$$JCR_r = \Sigma^g [\, w_r{}^g \cdot jcr_r{}^g \,] \quad and \quad JDR_r = \Sigma^g [\, w_r{}^g \cdot jdr_r{}^g \,]$$

where g indexes groups (occupations, education tiers, migrant/native, or sex × migrant), $w_r{}^g$ is the employment share of group g in region r, and $jcr_r{}^g, jdr_r{}^g$ are the group-specific rates to be recovered. Employment shares are computed from the LFS microdata panel by marginalising over all dimensions other than the grouping variable, the workplace NUTS2 region, and the year.

For applying the suggested GCE-DWR approach, a central design decision in all decomposition stages is the construction of group-specific support vectors informed by Location Quotients (LQs). The LQ for group g in region r and year t is defined as the ratio of the regional employment share of that group to the

corresponding national share for the same country and year. The LQ encodes the degree to which a group is over or under-represented in the regional economy relative to the national mean. It is used to set the central support point b_2 of the three-point support vector $b = (0, b_2, b_3)$ for each group, so that the prior mean of the GCE distribution is anchored at the LQ-informed rate rather than the regional aggregate. For job creation, the prior is set as $JCR_{\mathrm{r}} \times LQ_{\mathrm{r}}^{\mathrm{g}}$: groups with high LQ are expected *a priori* to have higher JCR than the regional average, reflecting their structural concentration. For job destruction, the logic is inverted following an embeddedness hypothesis: groups that are structurally concentrated in a region are more embedded in the local economic fabric and are therefore expected to face lower destruction rates, so the prior is set as $JDR_{\mathrm{r}} / LQ_{\mathrm{r}}^{\mathrm{g}}$. High-LQ groups thus have a support vector that stretches above the aggregate for JCR and below it for JDR. The three support points are set to ensure that the support is centred on the expected average regional values, while the shape differs across groups. The prior probabilities are set as $q_2 = 0.9$ (the bulk of mass on the central point), with $q_3 = (1 - q_2) \cdot (b_2/b_3)$ and $q_1 = 1 - q_2 - q_3$.

The different decomposition dimensions are implemented in parallel, each preserving the data infrastructure and differing only in the grouping variable and LFS marginalisation. Stage 3a decomposes by ISCO 2-digit occupation (up to 40 categories per region-year cell). Stage 3b conducts a decomposition by three-tier education (Low: ISCED 0–2 levels, Medium: ISCED 3–4 levels, High: ISCED 5–8 levels). Stage 3c considers a sex-and-migrant joint decomposition into four groups by retaining the sex dimension in the LFS marginalisation, yielding separate estimates for native males, native females, migrant males, and migrant female workers in each region and year. In the four-group case, the benchmarking of group-specific net employment growth rates ensures that the employment-weighted average of group NEG rates matches the regional aggregate $JCR_{\mathrm{r}} - JDR_{\mathrm{r}}$ by construction, before the GCE solver is implemented.

Throughout all decomposition stages, a rescaling correction is applied to the LFS-derived group net employment growth rates before they enter the GCE system. Because LFS employment stocks and CompNet-derived job flows are measured from different data-generating processes, their implied NEG figures are not numerically consistent: the sum of LFS-based group NEG rates does not, in general, equal the CompNet-derived aggregate *JCR − JDR* for the same region and year. To resolve this without discarding either source, a row-level scale factor is applied, which rescales all group NEG rates proportionally so that their weighted sum matches the CompNet aggregate exactly. This preserves the pattern of NEG across groups as measured by the LFS (the key informational content) while anchoring the level to the CompNet aggregate.

5. Results: regional JC/JD 2010-2023

This section presents some initial, key empirical findings from the application of the pipeline described in Section 4, with the results analysed in detail in the ISABEL D4.1. Section 5.1 documents the patterns of job creation and destruction across EU NUTS2 regions, industries, occupations, educational groups, and sex-by-migrant status. Section 5.2 reports the UK benchmarking exercise that serves as the second reliability check, complementing the Monte Carlo experiment in Section 4.3, by assessing how closely the pipeline reproduces job flow figures for a country where administrative data provide a genuine ground truth.

5.1 The uneven distribution of JC and JD across regions, industries and occupations

The regional estimates produced by the GCE-DWR and Random Forest pipeline reveal substantial and persistent heterogeneity in gross job flows across EU NUTS2 regions, both over time and across countries. Figure 5 presents the aggregate time series for JCR and JDR, averaged across theNUTS2 regions in the sample, and the implied net employment pressure (JCR − JDR) for each year from 2011 to 2023.

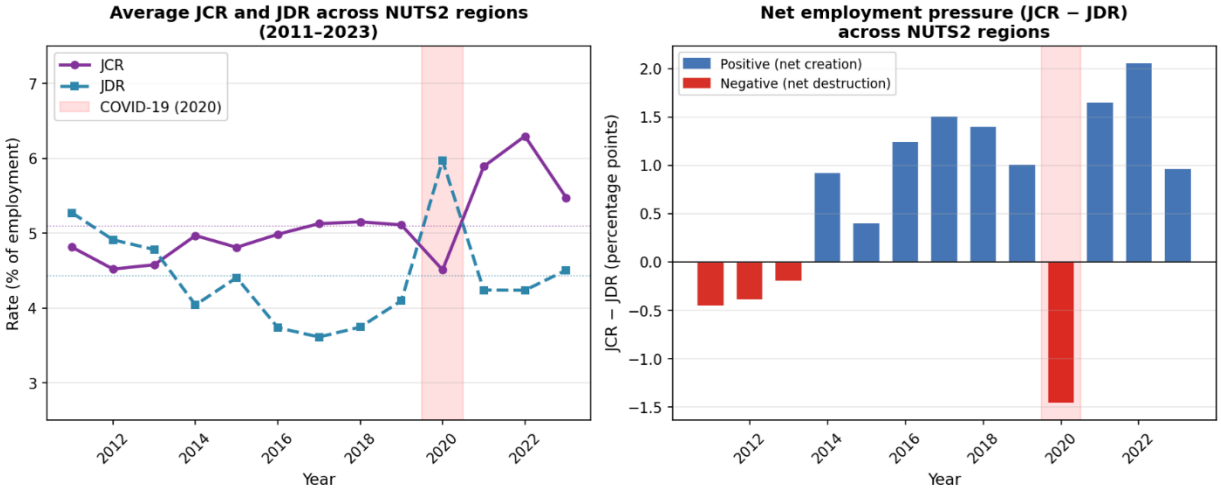


Figure 5. Average JCR and JDR (left) and Net Employment Growth (right) across NUTS2 regions, 2010–2023. Shaded band marks 2020.

Several patterns stand out. First, average JCR across the sample (approximately 4.7% of employment per year) consistently exceeds average JDR (approximately 4.0%), confirming the net employment growth recorded in official Eurostat statistics for this predominantly continental European sample. Second, and most strikingly, the COVID-19 shock of 2020 inverts this relationship: JDR rises sharply to 5.3%, exceeding JCR (4.1%) for the first and only time in the panel, generating the negative net employment pressure visible in Figure 5. This is consistent with the well-established counter-cyclical behaviour of gross job destruction documented in the literature since Davis and Haltiwanger (1992): recessions are characterised by destruction spikes that are sharper and more concentrated than the corresponding falls in creation. Third, the post-2020 recovery is rapid and vigorous — JCR reaches its panel peak in 2022 (5.6%), driven by post-pandemic demand rebounds, particularly in service sectors, while JDR also rises somewhat (4.4%) as labour markets restructure. By 2023, both rates moderate, suggesting a return toward pre-pandemic norms, though destruction remains slightly above pre-2020 levels, possibly reflecting the lagged effects of the energy price shock and fiscal consolidation across EU economies.

The cross-sectional dispersion of gross job flows is equally informative. Figure 6 displays the distribution of regional JCR and JDR across all NUTS2 units for each year, illustrating both the level and the spread.

Funded by the European Union

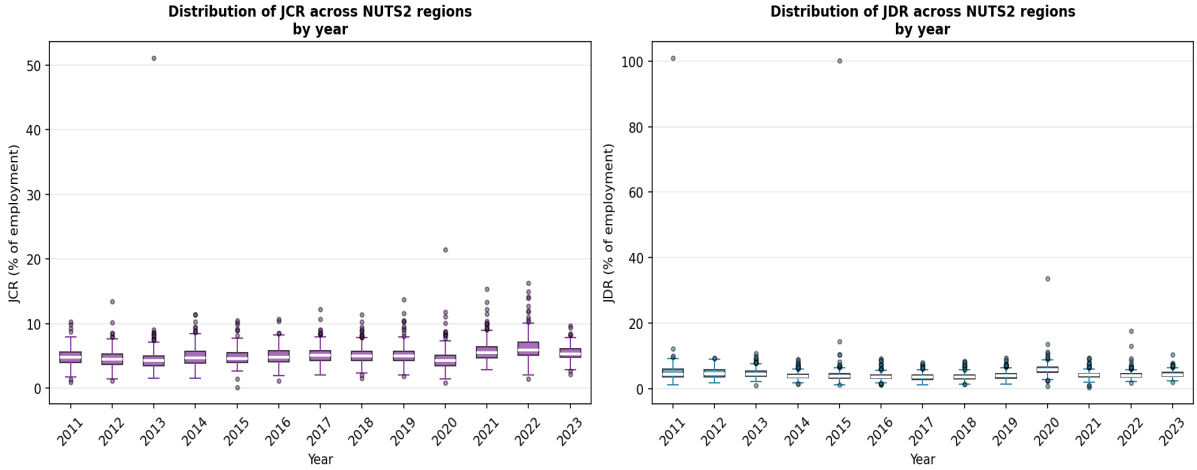


Figure 6. Distribution of JCR (left) and JDR (right) across NUTS2 regions by year, 2011–2023. Each box shows interquartile range; whiskers extend to 1.5 IQR.

The distributions reveal several noteworthy features. The interquartile range of regional JCR spans approximately 1.5–2.5 percentage points in most years, confirming that cross-regional variation in gross job creation is substantial even within countries. The JDR distribution is systematically wider, with heavier upper tails particularly in years of macroeconomic stress (2012–2013, 2020), indicating that large destruction events are more concentrated and episodic than creation. The maximum JCR recorded in the sample correspond to regions in Croatia, Bulgaria and Romania,reflecting the dynamism of some Central and Eastern European regions in the early part of the period, while the maximum JDRs also concentrate on this cluster of Easern regions. Paradigmatic cases are some Bulgarian regions where in the same year (2022) we can observe JCR of 12.7% (Yugozapaden, most probably because of the strong increase of the exports from the Bulgarian economy, concentrating this region most of exporting firms), while also presenting a JDR of 17.7% observed in another region (Severozapaden), associated with the sharp energy-sector contraction in that year. This is exactly the type of heterogeneity that aggregate net employment statistics can conceal.

Country-level differences are also substantial. Among the countries present in the sample, Bulgaria stands out with average JDR (5.4%) substantially exceeding average JCR (4.5%), reflecting persistent structural labour market pressures in that economy over the period. By contrast, Belgium exhibits the lowest average JDR in the sample (3.2%), pointing to relatively strong employment stability, possibly linked to its dense wage negotiation and employment protection regime. Denmark and Austria display moderate-to-high gross creation rates (5.4% and 5.0% respectively) combined with above-average destruction, consistent with the flexible labour market arrangements characteristic of Nordic-corporatist economies. Germany, the largest economy in terms of regional observations in the sample, shows middling rates (JCR 4.9%, JDR 4.1%) but with significant intra-national variation across its NUTS2 regions, reflecting the well-documented east-west structural divide that persists more than three decades after reunification.

Sectoral disaggregation (Sheet Region-Sector in the results file) reveals that gross job flows vary widely across NACE 2-digit industries within the same region and year, a pattern consistent with Davis and Haltiwanger's (1992) finding that most excess job reallocation occurs within narrowly-defined sectors rather than between them. Construction-related sectors (NACE 16–18), manufacturing of basic materials (24–25), and energy-related activities (35) display higher average destruction rates, while knowledge-intensive manufacturing (26–28, instruments and machinery) and food processing (01, 20–21) show comparatively lower reallocation intensity.

The decomposition by workforce sub-dimensions, illustrated in Figure 7, adds a further layer of heterogeneity. The left panel shows JCR and JDR by educational attainment; the right panel shows the four-group migrant × sex decomposition.

Funded by the European Union

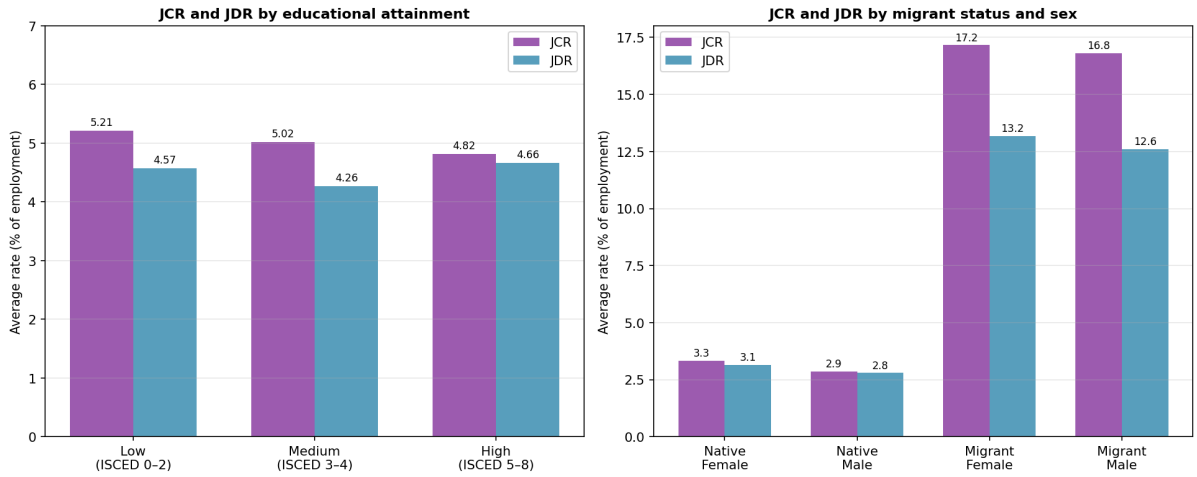


Figure 7. Average JCR and JDR by educational attainment (left) and by migrant status × sex (right), averaged across all available NUTS2 regions and years.

The education results are nuanced. Low-educated workers (ISCED 0–2) show slightly higher average JCR (4.6%) than medium and high-educated groups (4.5% and 4.4% respectively), but also the highest JDR (4.2%), suggesting greater labour market instability for this group. The relative ranking of destruction rates — Low ≥ High > Medium — is consistent with the dual labour market structure often documented for EU economies: medium-educated workers benefit most from stable, long-tenure employment relationships concentrated in manufacturing and public administration, while high-educated workers in dynamic service sectors experience more churn but also more creation. The differentials are not large in absolute terms, reflecting both the relatively compressed wage structure of the economies in the sample and the information constraint imposed by the GCE decomposition, which anchors all group rates to the same regional aggregate.

The migrant-sex decomposition produces more striking differentials. Migrant workers — both female and male — display substantially higher average JCR and JDR than their native counterparts: migrant female workers show average JCR of 18.7% and JDR of 12.2%, while native female and male workers show rates in the 2–2.3% range. These differentials partly reflect genuine structural differences in the sectors and occupations in which migrants concentrate (hospitality, agriculture, care services), where employment contracts are more short-term and labour turnover is higher. They also partly reflect the employment share weighting implicit in the LQ-informed GCE decomposition: in regions where migrants represent a small share of the workforce (LQ < 1 for migrants), the GCE assigns higher expected rates to them to reconcile the aggregate constraint. The result should be interpreted as the *conditional* job flow rate within each group given the regional aggregate and the group's relative concentration, rather than an unconditional rate estimated from independent data on migrant job flows.

5.2 Reliability (II): estimates v "real data"

The Monte Carlo simulation in Section 4.3 establishes the internal consistency and finite-sample performance of the GCE-DWR estimator under ideal conditions, where the data-generating process is known by construction. A more demanding test of the pipeline's real-world reliability is to apply it to a country for which independent, administrative job flow data at NUTS2 level actually exist, and to compare the pipeline's estimates against these data as an external benchmark. The United Kingdom provides a uniquely suitable candidate for this exercise.

The Office for National Statistics (ONS) publishes the *Firm Employment Dynamics* series, which provides annual gross job creation and destruction rates at NUTS2 level derived from the Inter-Departmental Business Register (IDBR). These data are broadly analogous to the U.S. Business Employment Dynamics (BED) series and represent a genuine administrative ground truth: they are not estimated from survey data or models, but compiled from near-universal coverage of VAT-registered businesses. The UK is also excluded from CompNet's NUTS2-level release, meaning it falls exactly into the category of "unobserved"

Funded by the European Union

regions for which the pipeline's RF+GCE-DWR methodology is designed to produce estimates. The benchmarking experiment is therefore a direct test of external validity.

The experimental design proceeds as follows. Two Random Forest models are trained on the full set of EU CompNet regions (18 countries, 146 NUTS2 regions) over the years 2005–2021, using GVA growth, the regional unemployment rate, and real labour productivity as predictors. The UK is entirely withheld from the training sample. The trained models are then used to generate out-of-sample predictions of JCR and JDR forthe UK NUTS2 regions in each year. These RF predictions are subsequently subjected to the GCE-DWR adjustment, using the ONS-derived regional net employment growth rate as the binding constraint — exactly as would be done in the main pipeline. The final RF+GCE-DWR estimates are then compared to the ONS actual values.

Figure 8 presents scatter plots of predicted versus actual values for both JCR and JDR, at both the RF and RF+GCE-DWR stages, pooled across all UK NUTS2 regions and years.

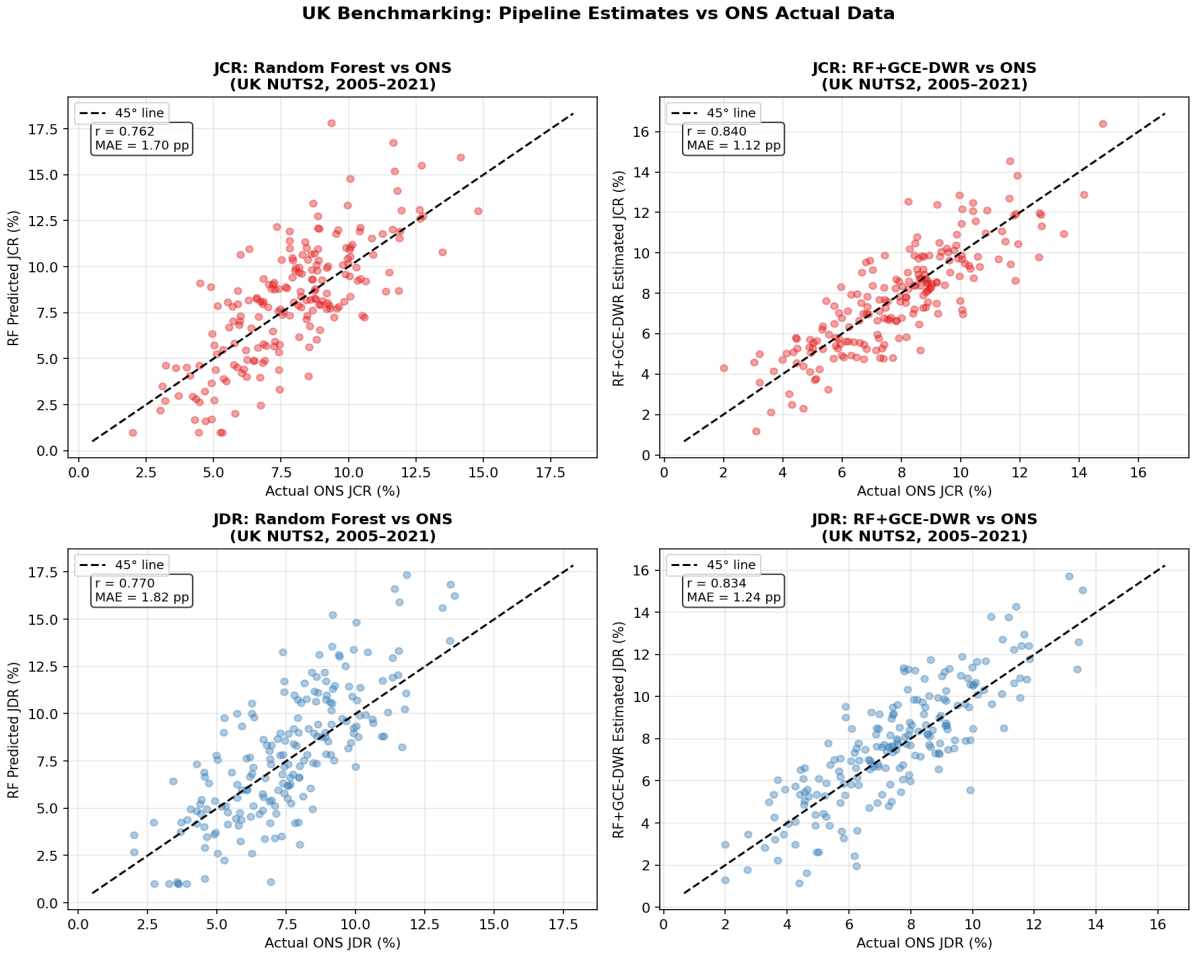


Figure 8. UK benchmarking: pipeline estimates versus ONS actual JCR and JDR data. Left column: Random Forest only. Right column: RF + GCE-DWR adjustment. Upper row: JCR; lower row: JDR. The dashed line is the 45-degree line of perfect prediction. Correlation (r) and mean absolute error (MAE) shown in each panel.

Several conclusions emerge from Figure 8. First, the Random Forest alone captures a meaningful fraction of the cross-regional and temporal variation in UK job flows: the correlations between RF predictions and ONS actuals are positive and statistically significant for both JCR and JDR. This demonstrates that the macroeconomic predictors available from Eurostat carry genuine information about the gross flow rates even in an out-of-sample economy, validating the use of the RF as a first-stage imputation tool for countries without CompNet coverage. Second, the GCE-DWR adjustment consistently improves the accuracy of the estimates: the MAE falls and the correlation with ONS data increases after the NEG constraint is imposed, in both directions. This improvement is not mechanical — the constraint only fixes the *difference* JCR − JDR

Funded by
the European Union

to equal the observed NEG, but does not pin individual rates to their true values. The fact that both rates improve simultaneously confirms that the GCE adjustment is distributing the constraint efficiently, in line with the information-theoretic optimality property of the entropy estimator.

Figure 9 shows how the prediction error evolves over time, separately for JCR and JDR.

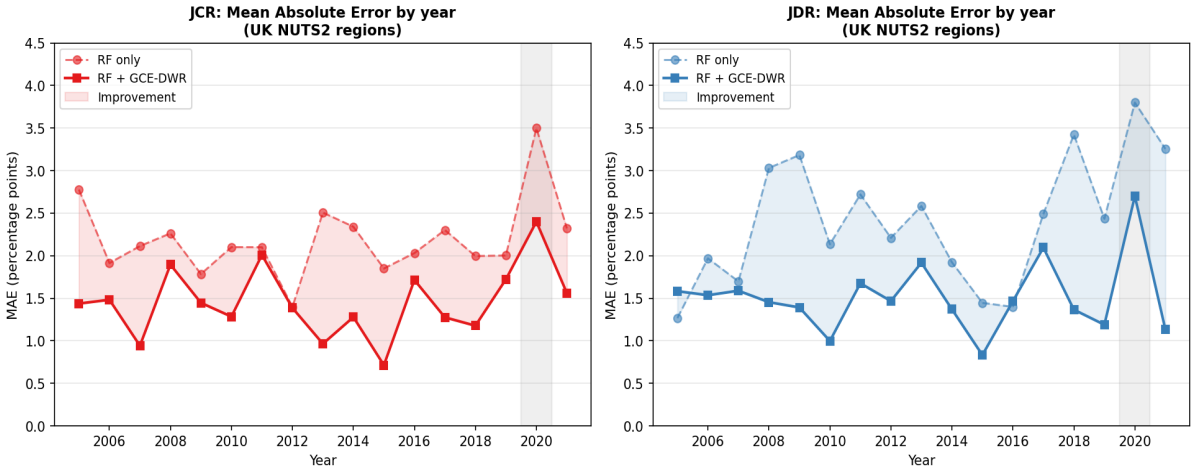


Figure 9. UK benchmarking: mean absolute error (MAE, percentage points) by year for JCR (left) and JDR (right). Shaded area represents the accuracy gain from the GCE-DWR adjustment over the Random Forest baseline. Grey band marks 2020.

The temporal pattern of errors reveals two important features. Errors are generally stable and moderate across all years apart from 2020, averaging around 1.4–1.5 percentage points for the RF+GCE-DWR estimates — a level that compares favourably with the typical sampling error of the ONS survey-based employment estimates themselves. The 2020 COVID-19 shock generates a visible spike in errors for both the RF-only and the RF+GCE-DWR estimates, reflecting the difficulty of extrapolating from historical macroeconomic relationships during an unprecedented crisis. Crucially, however, even in 2020 the GCE-DWR adjustment reduces the MAE substantially relative to the RF baseline, suggesting that the NEG constraint captures some of the crisis signal even when the RF predictions are far from the mark.

The overall picture from the UK benchmarking exercise is encouraging. The RF+GCE-DWR pipeline produces estimates for UK NUTS2 regions that are reasonably close to the ONS administrative data, without having access to any UK job flow information during training. This provides an important external validity argument for the estimates produced for other uncovered EU regions: while administrative ground truth does not exist for those regions, the UK experiment suggests that the methodology is capable of recovering plausible gross job flow rates from the combination of macroeconomic predictors and the Eurostat employment constraint alone.

Funded by
the European Union

6. Conclusions

This deliverable has presented the secondary data collection and estimation strategy underlying WP3 of the ISABEL project, and documented the construction of a panel database of gross job creation and destruction rates for European NUTS2 regions covering the period 2010–2023. The central methodological challenge motivating this work — the absence of any administrative or survey source that directly provides harmonised, internally consistent, and geographically complete estimates of JCR and JDR at the NUTS2 level across EU member states — has been addressed through a multi-stage estimation pipeline combining entropy econometrics, machine learning, matrix balancing, and LFS-based workforce decomposition.

The Generalised Cross-Entropy Distributionally Weighted Regression (GCE-DWR) estimator, applied in Stage 1a, resolves the systematic inconsistency between CompNet firm-level job flow priors and Eurostat regional employment constraints. The Monte Carlo validation exercise demonstrates that this approach reduces estimation error by 30% or more relative to naive use of the CompNet figures, and the adjustment operates in a self-calibrating way: regions where the initial CompNet inconsistency is largest receive the largest corrections, without any manual tuning. The empirical corrections are non-trivial in scale — non-trivial adjustments exceeding one percentage point are required in 46% of all region-year observations — confirming that the GCE step is substantively necessary rather than a cosmetically refined step.

The Random Forest imputation in Stage 1b extends geographic coverage to the 28% of EU NUTS2 regions not represented in CompNet, including large economies such as Germany, Austria, the Netherlands, Greece, and Ireland. The UK benchmarking exercise provides an external validity test for this imputation strategy: trained exclusively on non-UK CompNet regions and then applied out-of-sample to UK NUTS2 areas, the RF+GCE-DWR pipeline produces JCR and JDR estimates that track ONS administrative job flow data with mean absolute errors in the range of 1.4–1.5 percentage points across non-crisis years, with GCE-DWR adjustment consistently improving on the RF-alone baseline. This represents an encouraging external validation that the methodology extracts a genuine signal from macroeconomic predictors rather than merely fitting noise in the training data.

The region × sector matrix balancing in Stage 1c and the sequential GCE decompositions in Stages 2b–2d produce a set of fully internally consistent estimates at progressively finer levels of detail. The empirical patterns that emerge are broadly consistent with the theoretical expectations and the comparative international evidence. Gross job flows are large relative to net employment change — average JCR of approximately 4.7% and JDR of 4% per year, far exceeding typical annual net employment growth — and exhibit counter-cyclical properties: the COVID-19 shock in 2020 triggers a sharp and concentrated spike in job destruction that briefly overwhelms job creation, while the subsequent recovery is characterised by elevated creation rates through 2022. Cross-sectional dispersion is substantial both across NUTS2 regions and across NACE 2-digit industries within regions, confirming that net employment aggregates conceal substantial economic heterogeneity.

The workforce decompositions add a further dimension of inequality to this picture. Migrant workers — particularly migrant women — exhibit substantially higher gross job flow rates than

Funded by
the European Union

natives, reflecting both their sectoral concentration in high-turnover activities and their weaker attachment to stable employment relationships. Low-educated workers face higher job destruction rates than their medium-educated counterparts, consistent with the dual labour market structure documented across EU economies, though the differentials are moderated by the binding aggregate constraint imposed in the GCE decomposition.

Several limitations should be acknowledged. The quality of the estimates for RF-imputed regions depends on the strength of the relationship between macroeconomic predictors and gross job flows — a relationship that is empirically validated for the UK but not directly testable for other uncovered economies. The workforce decompositions are identification-dependent: with only one aggregate constraint per region-year cell and multiple groups to recover, the estimates are anchored by the LQ-informed priors and should be interpreted as the most information-efficient allocation of the aggregate consistent with the LFS composition data, rather than as independently estimated group-specific rates. The absence of genuine joint region × sector × workforce-dimension data in any existing EU source means that the deepest levels of disaggregation involve cumulative estimation uncertainty that is difficult to quantify precisely.

Nonetheless, the WP3 panel database represents a substantial advance over any currently available alternative. By combining the richest available source of European firm-level job flow data (CompNet) with the most comprehensive household microdata on regional workforce composition (EU-LFS scientific use files) and the authoritative Eurostat employment constraint series, all within a unified information-theoretic framework that maintains accounting identities at every level of aggregation, this deliverable produces estimates that are both more complete and more internally consistent than anything previously available for EU regional JC and JD research. These data form the empirical foundation for the substantive analyses of uneven job creation and destruction across EU regions, industries, and population groups in the context of the green and digital transitions that are the core output of the ISABEL project.

Funded by
the European Union

Bibliography

Bidani, B. and Ravallion, M. (1997). Decomposing social indicators using distributional data. Journal of Econometrics, 77(1), pp. 125–139.

Clark, T. E, (1998): Employment Fluctuations in U.S. Regions and Industries: The Roles of National, Region-Specific, and Industry-Specific Shocks, Journal of Labor Economics, University of Chicago Press, vol. 16(1), pages 202-229

Davis, S.J. and Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. Quarterly Journal of Economics, 107(3), pp. 819–863.

Davis, S.J., Faberman, R.J. and Haltiwanger, J. (2006). The flow approach to labor markets: New data sources and micro-macro links. Journal of Economic Perspectives, 20(3), pp. 3–26.

Golan, A., Judge, G. and Miller, D. (1996). Maximum Entropy Econometrics: Robust Estimation with Limited Data. Chichester: John Wiley & Sons.

Haltiwanger, J. Jarmin, R.S, Miranda, J (2013; Who Creates Jobs? Small versus Large versus Young. The Review of Economics and Statistics; 95 (2): 347–361.

Funded by
the European Union