INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

HUN-REN
Hungarian Research Network

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

The Rise of Linked Employer-Employee Panel Data: Where Are We Now?

ISTVÁN BOZA – RITA PETŐ

KRTK-KTI WP – 2025/2

February 2025

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
HUNGARIAN RESEARCH NETWORK (HUN-REN)
BUDAPEST

ABSTRACT
In recent decades, one of the most significant advancements in empirical labor economics was the emergence of longitudinal linked employer-employee datasets. This paper aims to provide a snapshot of this data revolution. With LEE panels now available in over 30 countries, we survey their general availability and key characteristics. Beyond common features, we highlight the more complex aspects of these datasets, which enable rigorous, large-scale research across diverse subfields. Finally, we explore emerging directions in LEE-based research, with the goal of engaging researchers, policymakers, and data providers.

JEL codes: C81
Keywords: linked employer-employee data, administrative data

István Boza
HUN-REN CERS
boza.istvan@krtk.hun-ren.hu

Rita Pető
HUN-REN CERS
peto.rita@krtk.hun-ren.hu

A kapcsolt munkáltató-munkavállaló paneladatok térhódítása – hol tartunk most?

BOZA ISTVÁN – PETŐ RITA

ÖSSZEFOGLALÓ
Az empirikus közgazdaságtan, különösen a munkagazdaságtan területén, az elmúlt évtizedek egyik legjelentősebb fejleménye a kapcsolt munkáltató-munkavállaló paneladatbázisok megjelenése és elterjedése volt. Világszerte több, mint 30 olyan ország van már, amely rendelkezik ilyen panel adatbázissal. Tanulmányunk legfontosabb célja, hogy egy átfogó képet adjunk ezen adatbázisok általános jellemzőiről. Külön kitérünk számos (adminisztratív forrásokból építkező) adatbázis esetén megjelenő sajátosságra is, amelyek révén korábban nem vizsgálható jelenségek kutatása vált lehetővé az elmúlt évtizedekben. Végül ismertetjük az elmúlt évek legfontosabb fejleményeit az ilyen adatokkal kapcsolatban, valamint azokat az új kutatási és együttműködési irányokat, amelyek nemcsak a kutatók, hanem a döntéshozók és az adatgazdák érdeklődésére is számot tarthatnak.

JEL kódok: C81
Kulcsszavak: kapcsolt munkáltató-munkavállaló adat, adminisztratív adat

The Rise of Linked Employer-Employee Panel Data: Where Are We Now?*

István Boza (HUN-REN Centre for Economic and Regional Studies)

Rita Pető (HUN-REN Centre for Economic and Regional Studies)

February 17, 2025

Abstract

In recent decades, one of the most significant advancements in empirical labor economics was the emergence of longitudinal linked employer-employee datasets. This paper aims to provide a snapshot of this data revolution. With LEE panels now available in over 30 countries, we survey their general availability and key characteristics. Beyond common features, we highlight the more complex aspects of these datasets, which enable rigorous, large-scale research across diverse subfields. Finally, we explore emerging directions in LEE-based research, with the goal of engaging researchers, policymakers, and data providers.

JEL code: C81

Keywords: linked employer-employee data, administrative data

*The authors are grateful for the contributions of Beatrice Allamand Turner, Antoine Bertheau, Bernardo Fanfani, Olivier Godechot, Martin Hallsten, Are Skeie Hermansen, Miklós Koren, Alena Křížková, Zoltán Lippényi, Vladimír Peciar, Andrew Penner, Malte Reichelt, Vera Rocha, Halil Sabanci, Marta Silva, and Andrea Weber. There is a preliminary accompanying website for this paper at leedata.eu. The aim is/was to keep the main tables of this study up to date, and store them on this site.

Introduction

In recent decades, one of the most significant developments in the economic data revolution was the emergence and spread of administrative linked employer-employee data (LEED or LEE data). Looking at the program of any labor economics conference reveals that most new research is not only empirical but increasingly relies on employer-employee panels or other administrative datasets. The study by Currie et al. (2020) confirms this trend by analyzing the content of the NBER Working Papers and the articles published in the top five economics journals over several decades. The authors find that the share of Top 5 journal articles using administrative data rose from around 8% in 2005 to over 20% by 2019. In NBER working papers, this ratio increased from below 10% to nearly 30% over the same period. The upward trend has likely continued in the past 5 years.

A key advantage of this type of data is that it provides a comprehensive view of both individual labor market outcomes and employer characteristics. Such data allow researchers to investigate firm-level dynamics and differences across firms in greater depth. While researchers have long had access to plant-level surveys and firm-level workforce data, these earlier sources were often limited in scope.[1] Similarly, longitudinal survey studies, such as the Panel Study of Income Dynamics (PSID) in the United States, and the British Household Panel Survey (BHPS) in the UK, have tracked individuals over time.

However, when a dataset follows both individuals and firms over several years – forming a linked employer-employee panel (LEEP) – researchers gain new research opportunities. The linked and longitudinal structure enables the study of worker mobility between firms and the consequences of within-firm changes (e.g., ownership changes) or broad policy shifts, offering deeper insight into the labor market dynamics.

In this paper, we provide an overview of the different types of LEE datasets and, as a key contribution, we examine their growing availability in 33 countries. We discuss the main data features of these state-of-the-art datasets, highlighting their strengths and limitations. In the final section, we present a broad range of features that can elevate the main body of these datasets and discuss the possibilities emerging through additional data linkage.

[1] See the summary of Abowd and Kramarz (1999) for a review of such datasets and early examples of administrative linked employer-employee data

Sources of LEE data

There are two primary types of data sources used to construct linked employer-employee (LEE) data. Historically, the earliest LEE datasets were based on wage surveys regularly completed by employers. The main purpose of these surveys was statistical, providing valuable insights for statisticians, researchers, and policymakers. One of the earliest examples is the 1966 French Enquête sur la Structure des Salaires (ESS), but many similar datasets emerged across Europe in the late 1970s (Abowd & Kramarz, 1999).

Survey-based data collection provides highly detailed information about employees' working conditions. This may include not only job-related information such as hours worked, overtime, and bonuses but also details on employees' qualifications, skills, and the equipment they use. Additionally, surveys allow for a broader range of questions, offering better insights into economic trends.

However, surveys are typically not exhaustive and are often restricted to a sample of firms or a subset of employees within a firm. Additionally, collecting and organizing data is costly. Furthermore, surveys only cover employees working at the time of data collection, excluding the unemployed, inactive, or self-employed population. It is also worth noting that willingness to participate in surveys has been decreasing significantly (Fujita et al., 2024).

To address these challenges, researchers have increasingly turned to databases constructed from administrative data sources. These databases are typically constructed by linking information from public institutions such as tax authorities, pension systems, or central banks. While the primary purpose of collecting this data is to support government operations, it has also become a valuable resource for research. Some of the earliest examples of administrative data usage in research are from the United States, including the works of Topel and Ward (1992) and Jacobson et al. (1993), which utilized social security records. Since then, researchers in many countries have developed databases based on similar administrative records.

Administrative databases are cost-efficient, requiring only a one-time cost for data acquisition. Updating these databases is cheaper than conducting new survey waves, although not without costs, given the expenses associated with data collection, storage, and hardware maintenance. Moreover, they enable monitoring of the entire population (employed, inactive, and unemployed) over time, facilitating research on topics such as

labor market dynamics and unemployment. Data from public entities are also thought to be more reliable than survey responses.

However, administrative datasets contain only the information generated by public administrations. For example, income data may reflect only taxable components, omitting details on overtime pay or performance-related bonuses. These datasets often lack comprehensive information on working conditions (including actual working hours), family and social backgrounds, and education history—unless such data were previously collected by public administrations. Additionally, administrative datasets do not cover the informal labor market, which constitutes a substantial part of the economy in some countries.

Two important points should be noted. First, with the widespread adoption of digital technologies, firms increasingly use the same data for both wage surveys and administrative reporting, mitigating concerns about data validity in survey-based datasets. Second, some countries conduct economy-wide surveys where all firms report workforce data. While technically classified as survey data, these datasets (such as Portugal's Quadros de Pessoal) do share key characteristics with administrative datasets.

Given these considerations, it is more useful to distinguish LEE datasets not by their primary data source but by whether they allow tracking individuals over time. Datasets that facilitate such tracking are often called longitudinal LEE data or LEE panel data (LEEP). In contrast, those that do not link individual observations over time are classified as cross-sectional LEE (Bryson et al., 2006). The latter primarily consists of survey datasets, whereas longitudinal LEE datasets include both survey-based and administrative data sources. Figure 1 illustrates these classifications.

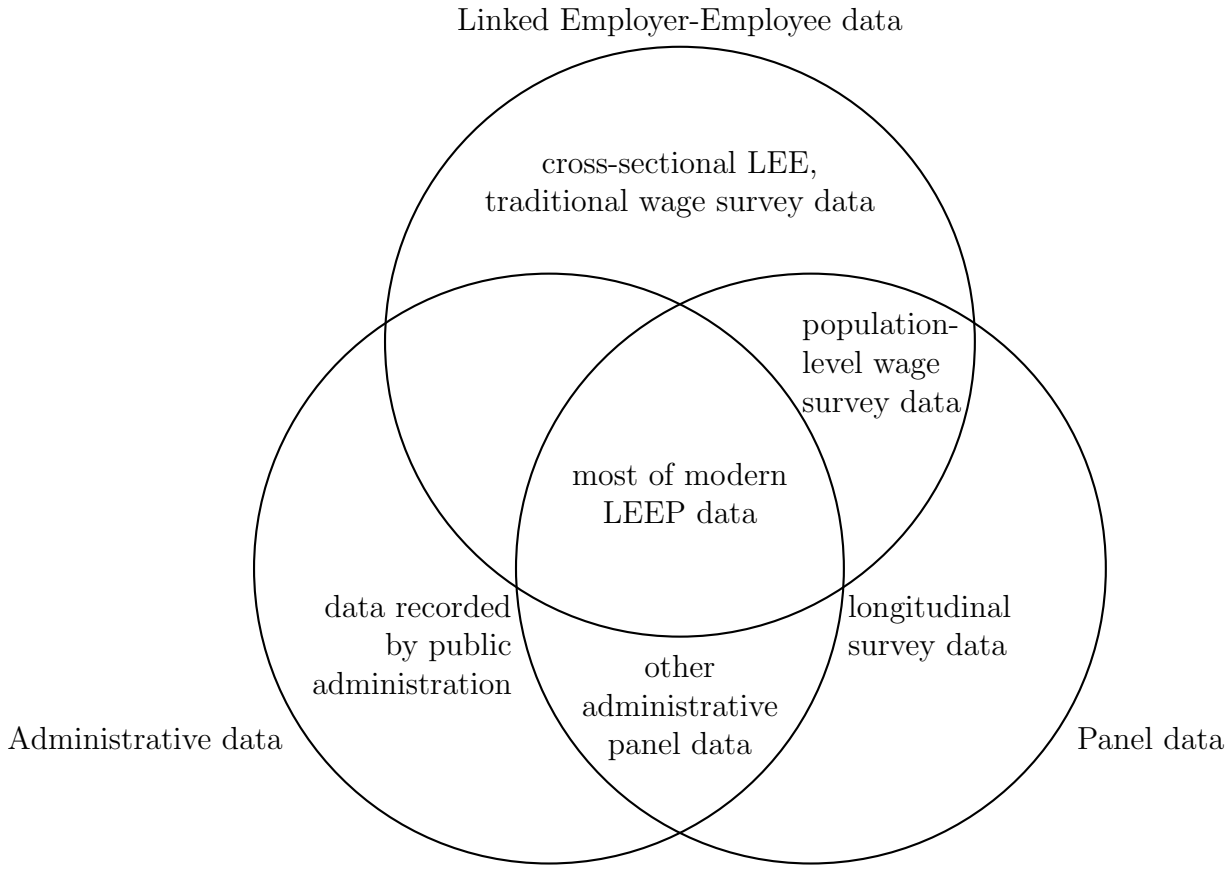

Figure 1: Classification of LEE data

In the following, we focus on linked employer-employee (LEE) data that are longitudinal, meaning they have a panel structure at the individual level. This category includes both administrative datasets and population-level wage surveys in certain countries that generate comparable longitudinal data, enabling similar analyses. We do not consider cross-sectional LEE datasets, which are often available alongside administrative LEE data or, in some cases, serve as the only form of LEE data – such as in the Czech Republic and Belgium in Europe, and Japan and South Korea in Asia. Additionally, non-labor administrative data collected over time in a panel structure are often merged with LEEP datasets, significantly expanding research possibilities.[2]

Main features of LEE data

Tables 1 and 2 summarize the countries known to have longitudinal LEE datasets. Our data collection process relied on input from co-authors and acquaintances familiar with these

[2]In some cases, repeated waves of wage surveys have been linked to LEEP data, providing richer insights into working conditions and wage structures.

datasets. In cases Where direct contacts were unavailable, we consulted publicly available comparative studies (such as OECD (2021)), relevant research papers, or colleagues. We include all such datasets, even if they are highly confidential and can be accessed by only a very limited number of researchers. Overall, we identified 33 countries where large-scale LEE data are available. For approximately half of these, scholars with direct access to the datasets—or the data owners themselves—have validated the information. The table presents key features of the datasets alongside their primary dimensions.[3] Figures 2 and 3 illustrate the geographical distribution of longitudinal LEE data. Please note that Israel, Taiwan, Australia, and New Zealand are not included on the maps.

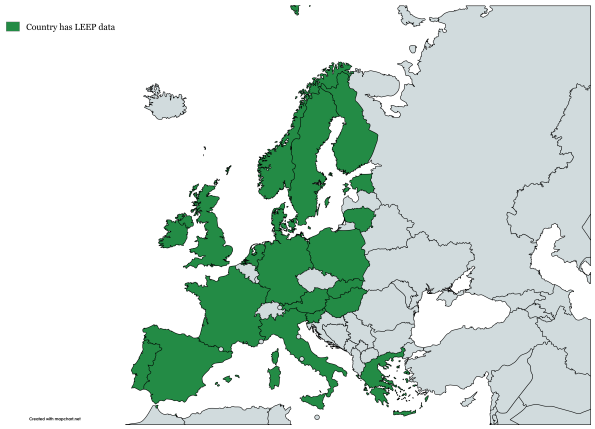

Figure 2: Availability of longitudinal LEE data in Europe

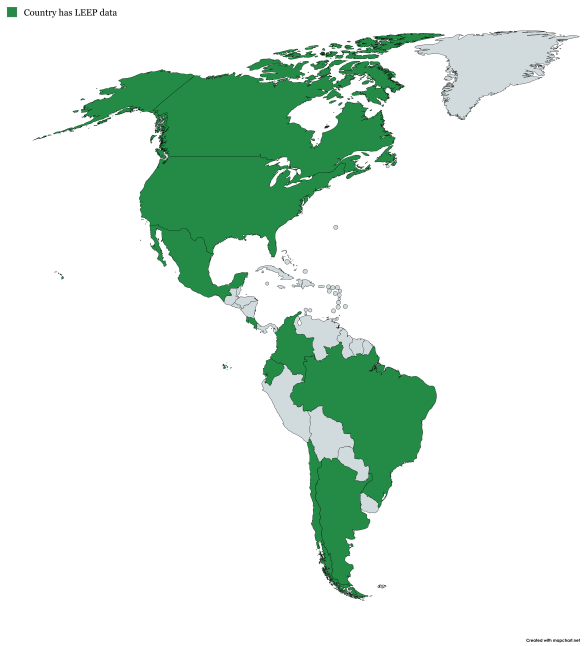

Figure 3: Availability of longitudinal LEE data in the Americas

[3]Nonetheless, the authors take responsibility for any potential errors.

The scope of LEE datasets is shaped by two key factors that define population coverage. First, the base population for data collection may be limited to a specific sector, such as the private sector, or it may encompass all formal employment within a given region or country. A common limitation is that many datasets cover only the private sector, excluding the public sector. Second, administrative datasets are typically designed to capture formal employment, which may represent only a fraction of the overall labor force – an especially relevant distinction in certain Latin American countries. Some administrative LEE datasets go beyond employment records, incorporating information on unemployed individuals. Many also integrate additional administrative data outside the traditional LEED framework, such as health expenditures and social policies for all individuals not just those employed. In such cases, these datasets are categorized as *all individuals* in Tables 1 and 2.

Table 1: Linked Employer-employee Data availability – Europe

Country	**Population**[a]	**Sampling**[b]	**Resolution**[c]	**Time window**[d]
Austria	all employed	universe	spell	1972 - 2020
Denmark	all individuals	universe	yearly	1980 - 2024
Estonia	all employed	universe	yearly	2001 - 2017
Finland	all employed	universe	yearly	2004 - 2018
France	all employed	universe	yearly	1994 - 2020
Germany	all individuals	universe	spell	1993 - 2019
Greece	private sector	universe	monthly	2002 - 2023
Hungary	all individuals	sample (50)	spell	2003 - 2017
Ireland	all employed	sample (10)	yearly	2011 - 2022
Italy	private sector	universe	monthly	1974 - 2021
Lithuania	all individuals	sample (25)	monthly	2000 - 2020
Netherlands	all employed	universe	monthly	2006 - 2022
Norway	all employed	universe	monthly	1995 - 2021
Poland	all employed	universe	yearly	2000 - 2019
Portugal	private sector	universe	yearly	1985 - 2019
Slovakia	all employed	universe	monthly	2014 - 2021
Slovenia	all employed	universe	yearly	1992 - 2014
Spain	all individuals	sample (5)	yearly	2006 - 2018
Sweden	all individuals	universe	yearly	1990 - 2021
United Kingdom	all employed	sample (1)	yearly	1975 - 2019

[a] *All employed* always refers to all formal employment, excluding informal or illegal forms of employment. Country-specific minor exceptions (e.g., small or large companies, certain sectors or occupations) are not listed here."

[b] Sampling is usually simple random sampling of individuals. The number in parentheses refers to data available to researchers in percentage of the total population.

[c] Lowest achievable resolution. This resolution may only be available in later years in the data, not for the whole period. *Spell* indicates countries that have access to spell-level (daily) data.

[d] Latest years refers to the most recent year appearing in a publication we are aware of or the latest year available, as confirmed by a scholar with data access

Table 2: Linked Employer-employee Data availability – Americas and Rest of the World

Country	**Population**[a]	**Sampling**[b]	**Resolution**[c]	**Time window**[d]
Argentina	private sector	sample (3)	monthly	1996 - 2015
Australia	all employed	universe	spell	2011 - 2020
Brazil	all employed	universe	yearly	1985 - 2018
Canada	all employed	universe	yearly	1983 - 2019
Chile	all employed	universe	monthly	2002 - 2022
Columbia	all employed	universe	monthly	2008 - 2014
Costa Rica	all employed	universe	monthly	2006 - 2017
Ecuador	all employed	universe	monthly	2006 - 2012
Israel	all employed	universe	monthly	1983 - 2015
Mexico	private sector	universe	monthly	2005 - 2019
New Zealand	all employed	universe	monthly	2000 - 2017
Taiwan	private sector	universe	monthly	1998 - 2004
United States	all employed	universe	yearly	1993 - 2021

[a] *All employed* always refers to all formal employment, excluding informal or illegal forms of employment. Country-specific minor exceptions (e.g., small or large companies, certain sectors or occupations) are not listed here."
[b] Sampling is usually simple random sampling of individuals. The number in parentheses refers to data available to researchers in percentage of the total population.
[c] Lowest achievable resolution. This resolution may only be available in later years in the data, not for the whole period. *Spell* indicates countries that have access to spell-level (daily) data.
[d] Latest years refers to the most recent year appearing in a publication we are aware of or the latest year available, as confirmed by a scholar with data access

To protect privacy and ensure data confidentiality, LEE datasets often incorporate sampling in the versions available to researchers. This approach may also be necessary due to data size constraints. Sampling can be applied at different levels - individuals, firms, or employment spells - with common sample sizes including 1%, 5%, 20%, and 50%. This method ensures both usability and security, by allowing researchers to analyze a representative subset of the population while upholding data protection standards. However, in certain research fields, such as network-related studies, this practice can significantly limit potential applications.

Regarding the temporal resolution of data, our table specifies the frequency at which an individual's employer is recorded. Observations may occur annually (on a specific date), monthly, or even continuously, providing/offering different levels of detail on employment relationships. The frequency of observation directly affects the extent to which researchers can analyze short-term trends and labor market fluctuations. Additionally, it is crucial/important to consider the reference period for wage measurements, as this may differ from the recorded period of the employer-employee relationship.[4]

[4]For instance, the Sweden LISA dataset records an individual's employer as of November but reports total annual earnings from all sources and earnings from the three most significant sources for the given year without exact employment dates.

These tables do not cover data access, as practices vary significantly across countries. Even within a single country, access methods and restrictions may differ depending on the institution providing the data. However, we summarize the available information in Appendix Tables A1 and A2. Most institutions restrict data access to individuals with domestic affiliation or even based on nationality. Some allow access only on-site within secure working environments, often supplemented by data disclosure or output-checking procedures to prevent researchers from recording individual-level data. Other countries permit remote access under strict security protocols, such as multi-factor authentication.

Main research topics using LEE data

Linked Employer-Employee (LEE) datasets share a core set of information, including an employer ID, which identifies the firm or institution where an employee works, a worker ID, calendar time referring for the time of observation, and a measure of work income, such as total monthly earnings or hourly wages. In some cases, an establishment-level identifier is also available, distinguishing between different sites of the same company. These datasets typically include basic demographic information, such as age, gender (sex), and country of birth. Additionally, many countries provide occupational classifications, with most European countries using the International Standard Classification of Occupations (ISCO) or a national variation.

Even with this minimal set of variables, the emergence of administrative LEE data has allowed researchers to explore a wide range of new questions (Abowd & Kramarz, 1999). In this study, we first highlight key strands of research that rely solely on this basic set of information before discussing the advanced features found in more recent LEE datasets.[5]

Linked Employer-Employee (LEE) data is crucial for analyzing wage disparities in modern labor economics research. This data provides a detailed and comprehensive perspective on the various factors contributing to wage inequality. By linking employers and employees, researchers can conduct sophisticated analyses to isolate the key determinants of wage differences, such as firm-specific wage-setting strategies, workforce composition, and occupational segregation (Abowd et al., 1999; Torres et al., 2018). The role of

[5]We acknowledge that comprehensively summarizing all relevant works within the research areas discussed is an immense task – one we did not undertake. In selecting a limited set of references, we aimed to include seminal works in each field alongside some recent studies. We have occasionally cited works by researchers who played a significant role in bringing this project to fruition.

these wage components in overall wage variation (and its evolution over time) has been extensively explored/analyzed in a branch of studies building on the works of Card et al. (2013) and Song et al. (2019).

Researchers utilize LEE datasets and wage decompositions to investigate wage disparities across multiple dimensions, including gender, race, and skill levels (Card et al., 2016; Cardoso et al., 2018; Gerard et al., 2021). Employer identifiers enable the distinction between within-firm and between-firm sources of wage differences, facilitating cross-country comparative analysis (Penner et al., 2022; Tomaskovic-Devey et al., 2020). Furthermore, intra-firm and inter-firm wage dynamics have become prominent research fields, encompassing topics such as job ladders, job hierarchies, and internal labor markets Cestone et al. (2023) and Huitfeldt et al. (2023).

In countries where administrative LEE data comprehensively captures employment histories/spells - including accurate start and end dates - researchers can reliably study the duration of unemployment and assess the effects of changes in unemployment insurance policy (Lindner & Reizer, 2020; Nekoei & Weber, 2017). Such data also enables the estimation of job-search models by focusing on the unemployed population (Dellavigna et al., 2017). Additionally, panel data allows the identification of individuals who exit the workforce entirely, facilitating an in-depth analysis of external factors affecting employment levels (Bertheau, Acabbi, et al., 2023; Fanfani, 2023). When the dataset includes information on entrepreneurs, researchers can further investigate the returns of entrepreneurship over standard employment (Merida & Rocha, 2021).

The ability to track workers in firms over multiple years allows researchers to approximate professional networks based on former co-workers. A growing body of literature employs this approach to assess the effects of professional contacts (weak ties) on job-finding prospects/chances and wage outcomes (Boza & Ilyés, 2020; Eliason et al., 2023; Glitz & Vejlin, 2021; Saygin et al., 2021). Tracking workers across employers also enables the study of knowledge spillover effects (Köllő et al., 2021; Poole, 2013). Additionally, long administrative LEE panels provide valuable insight into career trajectories, including the long-run scarring effects of job loss or employment in lower-quality/inferior firms (Arellano-Bover, 2024; Liu et al., 2016; Oreopoulos et al., 2012).

Advanced features in LEE data

Table 3 summarizes the availability of specific topics across various LEE datasets. While these topics are not universally present in all datasets, their inclusion offers valuable research opportunities. The scope of this table is currently limited by the number of researchers we were able to directly contact. Detailed, country-specific responses are available in Appendix Tables A3 and A4. We are actively working to gather additional responses to expand this information.[6]

Table 3: Features in LEE datasets in selected countries.

	Has any	Has none	Unknown
Location of workplace	12	3	18
Residence	12	3	18
Work conditions (hours, overtime, bonuses)	10	5	18
Firm financial records/ ownership	11	4	18
Health data	7	8	18
Social benefits/transfers	10	5	18
Number of kids	9	6	18
Family relations	7	8	18
Immigration status	12	3	18
Observe educational institutions	8	7	18
Cognitives skills	6	9	18
Mergeable to other data sources	11	4	18

Detailed information is currently available from 15 countries, but have not yet been collected from 18. Country-specific information on availability in these general categories can be found in Appendix Tables A3 and A4.

We discuss these common (or rather uncommon) features, highlighting potential applications in the following section. Again, our goal is not to provide an exhaustive review of these research strands but to highlight some seminal works and innovative studies from recent years.

[6]The most up-to-date version of these tables is available on our website, leedata.eu. We welcome feedback and contributions, particularly from countries we have not yet had the opportunity to survey in detail.

Firm characteristics and work conditions

Local geographic conditions play a crucial role in shaping market dynamics. In many countries, researchers can access basic geographic data in LEE datasets, including individuals' home and workplace locations. This data supports analyses of labor market determinants influencing residential mobility (Eriksson & Lengyel, 2019; Ilyés et al., 2023; Schmutz & Sidibé, 2019) and enables the estimation of regional wage differentials between commuting zones, as worker mobility across regions can be monitored over time (Card et al., 2024). Additionally, this data facilitates investigations into the role of cities in wage inequalities (Dauth et al., 2022).

A detailed understanding of workers' working conditions is crucial for researchers. Administrative data, typically collected by public authorities, often provide limited insights into working conditions, usually restricted to contracted hours.[7] However, in some countries, this data can be supplemented with survey-based information, such as wage components (e.g., bonuses or overtime hours) from the Structure of Earnings Survey. This integration results in a more comprehensive dataset on working conditions enabling researchers to analyze the factors contributing to the gender wage gap (among other topics) (Boza & Reizer, 2023; Burbano et al., 2023).

Furthermore, administrative databases often contain enterprise-related characteristics, enabling researchers to explore firm-level factors, such as productivity, ownership, export activities, and other trade-related measures, alongside detailed workforce data. These datasets have fueled renewed interest in estimating productivity-wage passthrough parameters in rent-sharing literature (Card et al., 2018; Criscuolo et al., 2021; Torres et al., 2018). Moreover, firm-level exporting activity data allows assessments of its effects on labor demand (Lichter et al., 2017), the firm-level gender wage gap (Bøler et al., 2018), and even spillovers across labor mobility (Choquette & Meinen, 2015). Additionally, ownership information provides insight into wage effects associated with foreign ownership (Köllő et al., 2021).

Health and society

A key advantage of certain LEE datasets that integrate multiple national data sources—rather than relying solely on employment reports—is the availability of highly detailed informa-

[7]The presence of such information is indicated in the third row of Table 3.

tion on social transfers, labor market programs, individual health status, and healthcare and pharmaceutical costs. Governments often maintain comprehensive records on the transfers they distribute, making this data readily accessible through various datasets. However, obtaining information on individuals' health status can be more challenging. Even when health data is available, coverage may be uneven, particularly in countries where the private sector plays an increasing role in healthcare provision, limiting the government's direct access to this information.

Nevertheless, several studies have successfully linked employment and health data. One line of research investigates the impact of employment shocks on health outcomes (Bíró & Elek, 2020) and conversely, the effect of health shocks on employment outcomes (Bíró et al., 2024). Additionally, the role of firms in medical spending variation can be examined (Ahammer et al., 2024). Access to medical records also enables researchers to identify unexpected death events in the data, providing an exogenous source of variation for studies on worker substitutability (Bertheau et al., 2022; Jäger & Heining, 2022). Furthermore, administrative mortality records serve as valuable outcome variables, as demonstrated by Sullivan and Wachter (2009), who estimate the effects of job displacement on mortality rates.

In some countries, available data also enables the assessment of social benefit uptake based on individual or firm characteristics (Aizawa et al., 2022; Bana et al., 2023; Bíró et al., 2024; Lachowska et al., 2022). Similarly, the effectiveness of active labor market policies, such as training programs, can be evaluated through participation data (Kauhanen, 2021).

Certain datasets also make it possible to identify incarcerated individuals based on specific transfer records, enabling research on long-term labor market effects of imprisonment (Köllő et al., 2023). In other cases, researchers have access to direct crime data, allowing the estimation of the effects of job loss on criminal activity (Britto et al., 2022; Rose, 2018).

Family and background

Differences in database creation methods and legal frameworks that govern data linking lead to considerable variation in the availability of family and social background data across countries. Public institutions often collect this information for purposes such as family taxation or social support. For example, in many cases, information related to

children is recorded primarily for mothers, and family-related data is often indirectly inferred through social transfers. However, in some countries, the core data can be supplemented with survey-based data, offering more extensive and detailed information on family backgrounds. Despite these challenges, the labor market effects of parenthood remain a prominent area of research (Kleven et al., 2019; Lundborg et al., 2017).[8]

Nevertheless, gender is consistently recorded across all LEE datasets. This facilitated a vast body of research aimed at understanding the sources and dimensions of gender-based employment and wage differences. Notable examples include seminal papers by Cardoso et al. (2016), Card et al. (2016), and more recent contributions such as Casarico and Lattanzio (2024), Gallen et al. (2019), and Boza and Reizer (2024), as well as comparative studies like Penner et al. (2022).

In several countries, particularly in the Nordic region, detailed family background information extends beyond the presence and number of children, including household or other familial relationships. This level of detail allows researchers to map family networks, enabling studies on various topics, such as the firm-level dimensions of intergenerational mobility (Bennett et al., 2023; Dobbin & Zohar, 2023; Hällsten & Kolk, 2023), family effects of job-loss (Jensen et al., 2023), the effects of joining a family member's firm (Corak & Piraino, 2011; Staiger et al., 2023), and the impact of family contacts on labor market and other outcomes, such as job-finding (San, 2021) or residential mobility (Ilyés et al., 2023).

The labor market performance of immigrants and the role of social connections have been extensively analyzed in numerous Western and Scandinavian studies, facilitated by the availability of detailed immigration data, including birthplace, nationality, and immigration status. In some countries, records on parents' immigration status also enable the analysis of second-generation immigrants. This information supports research on labor market differences between immigrants and natives, including aspects of segregation (Ansala et al., 2022; Arellano-Bover & San, 2020; Hermansen, 2013), and wage disparities (Dostie et al., 2023; Drange et al., 2024; Hermansen et al., 2023). Moreover, immigrant status data also facilitates researchers to estimate the heterogeneous effects of other phenomena, such as the motherhood penalty (Nieto, 2021) or the effects of business cycles (Dustmann et al., 2010), showing the importance of intersectional inequalities.

[8]This literature also includes family composition, when available. For instance, Andresen and Nix (2022) focuses on parenthood in same-sex couples.

Education and skills

Administrative data do not always include information on individuals' educational attainment, unlike wage surveys, which typically capture this detail. However, student status data is available in some countries, allowing researchers to infer educational levels. Additionally, tracking individuals' past educational institutions facilitates studies on the correlations between school peers' outcomes (Boza & Horn, 2023; Hermansen et al., 2020) and the influence of peers network on various labor market outcomes (Ilyés & Sebők, 2023; Ilyés et al., 2023).

In some countries, data extends beyond basic educational institution records, including individual-level school performance metrics or standardized assessments, such as military entrance exams. These variables are highly valuable as they provide reliable proxies for workers' true, often unobserved, cognitive skills. The economic returns to such skills can then be assessed concerning future labor market outcomes (Böhm et al., 2023; Hermann et al., 2022) or used as indicators of worker quality (Hensvik & Skans, 2016).

Discussion – The research frontier and the way forward

For the final part of our manuscript, we present recent innovations and highlight potential directions for future research using LEE data.

International collaborations

In addition to studies using data from a single country, comparative studies can provide strong evidence for behavioral patterns or market processes that are common across modern economies. Due to data access restrictions and the considerable effort required to master these datasets, it is unlikely that any one scholar will work with a large set of different LEE datasets. However, cross-country collaborations, utilizing a distributed coding (shared computation) approach, have emerged to address this gap. We are aware of three major efforts in this area, alongside several research papers.

The **Comparative Organizational Inequality Network (COIN)** was established in 2015 by a collaboration of researchers from 7 countries interested in studying inequality. Its primary objective is to examine the role of workplaces as institutions in shaping (income) inequality. Over time, COIN has expanded to include sociologists, economists, labor,

and management scholars from more than 16 countries. Research within this network focuses on earnings disparities, immigration, gender differences, mobility networks, and other sources of organizational inequalities (Godechot et al., 2024; Penner et al., 2022; Tomaskovic-Devey et al., 2020).

The **LINKEED research group** of the OECD leverages harmonized linked employer-employee data across 20 OECD countries, making it one of the most ambitious efforts to apply administrative data in a cross-country context. After publishing a volume in 2021 on the role of firms in wage inequality (OECD, 2021), the research group (now known as LINKEED V2) has expanded its research agenda. Key topics under investigation include the implications of firm-level wage-setting practices for job mobility, the role of job transition in career progression, and labor market efficiency, and comparative analysis of gender gaps in firm-level wage premia and immigrant-native wage disparities.

The **Global Repository of Income Dynamics (GRID)** project provides an open-access international database containing harmonized microstatistics on income inequality and income dynamics at the individual level. All statistics in the database are derived/computed from administrative earnings records data and are designed to be comparable across countries. Currently, 13 countries participate in GRID, but this number is just about to expand over 25 countries in GRID V2.0 in the near-future. In addition to making the dataset publicly available, accompanying research papers have highlighted key trends in income inequality and income dynamics for each country. Furthermore, Guvenen et al. (2022) identifies global trends based on data from the original 13 GRID countries.[9]

Beyond these major initiatives, smaller-scale cross-country studies studies using LEE data have also emerged. For example, Bertheau, Acabbi, et al. (2023) utilize LEE data from seven countries to assess the causal effects of job loss on workers' labor outcomes in a comparative framework. Another distinctive study by Bütikofer et al. (2024) leverages cross-country matched registry data to analyze the labor market effects of the Öresund Bridge's construction. This approach enabled the authors to track employment and wage changes for individuals on both sides of the Sweden-Denmark border.

[9]For more details, see the project's website at grid-database.org.

Extending the scope of LEE data

As discussed earlier, a key limitation of administrative databases is that they typically contain only information generated by public institutions. However, this drawback can be easily addressed by integrating additional data sources - a common practice in many international databases. Wage, household budget, and time-use surveys are the most frequently incorporated sources. To link these datasets with the master LEE data, researchers must obtain authorization from data holders and adhere to strict anonymization protocols. Once permission is granted and anonymization is completed, existing or newly collected data can be successfully integrated into an established LEE database.[10]

One way for researchers to distinguish their work in the increasingly crowded field of administrative data studies is by incorporating entirely new data features. Recognizing this potential, academic institutions, research centers, and statistical agencies in various countries are expanding the scope of linked datasets. Increasingly, they are integrating additional information on education, health, or business networks. Countries such as Germany, Denmark, and France have pioneered linking individual or firm-level questionnaires, and even data from economic lab experiments to administrative records with work histories. These linkages are conducted with the consent of participants and in compliance with stringent privacy regulations.

The German Institute for Employment Research (IAB) has been at the forefront of this effort, developing innovative datasets that merge firm-level surveys with administrative registers. One such dataset (Integrated Employment Biographies - IEB) builds on a detailed survey of vacancies within firms, probabilistically linking them to employers who later appear in the administrative registers as entrants of the firm, filling these positions (Lochner, 2023).[11] Another example, the Linked Personnel Panel, integrates two firm-level surveys (Mackeben et al., 2023). Jäger et al. (2024) link a new survey, embedded in the German Socio-Economic Panel, to the IAB data to assess workers' beliefs about outside options from a given firm. Meanwhile, Altmann et al. (2018) uses experimental data linked to the employer-employee records.[12]

[10] As discussed earlier, certain administrative data sources, such as health or education data, are also available (directly) alongside the main LEE datasets.

[11] A similar data linkage is employed by Mueller et al. (2024), who used linked vacancy-employer-employee data from Austria to estimate wage elasticities of vacancy filling rates.

[12] Similar initiative exists in Denmark, and France. In Hungary, Boza and Reizer (2024) links wage survey data to an administrative LEEP dataset using probabilistic matching.

Other notable research projects have expanded LEE data in creative ways. For example, Heining et al. (n.d.) conducted a firm-level survey collecting on hiring practices (such as whether firms post wages publicly or engage in multi-step bargaining during hiring) and linked this dataset with LEE data to examine the connection between hiring practices and firm-level gender pay gaps. Meanwhile, Bertheau and Hoeck (2023) compares a firm's self-assessed position in the wage ladder to its actual standing in the wage distribution, using a linked survey-administrative dataset from Denmark.[13] Another example is Cirillo et al. (2024), who linked detailed firm-level information on automation to LEE data in order to assess the effect of technological change on labor market flows in Italy.

Beyond direct data integration, innovative approaches can further enhance existing LEE datasets. For instance, Reichelt and Müller (n.d.), in cooperation with the IAB, applied a name classification procedure to the raw adminsitrative records to approximate workers' country or region of origin. This information could be then used in the otherwise strictly anonymized dataset for research. The spread of modern classification and language-based algorithms may open up additional possibilities, as long as security protocols are maintained and data providers are willing to cooperate.

Policy relevance of LEE data

It is important to recognize that LEE data can offer significant value not only to academic researchers but also to decision-makers, policymakers, and public policy program planners. Given this broad relevance, there is a shared interest in ensuring that data can be analyzed and retrieved efficiently. Currently, large administrative databases are primarily used to study general behavioral patterns, economic regularities, and ex-post evaluations of major regulatory changes. However, increasing the frequency of data delivery could significantly expand their applications.

In many countries, LEE datasets are updated annually, typically incorporating data from the previous year. However, this is not a universal practice, and some countries construct their LEE datasets only on a non-regular basis. Since many administrative databases sources, such as employer return forms, are generated monthly, there is a potential to make these datasets available with only a brief delay. Establishing a continuous data provider protocol could allow researchers and policymakers to analyze recent trends

[13]The same survey is also used to assess the firm's decision to hire (Bertheau et al., 2024) and its decision to lay off employees (Bertheau, Kudlyak, et al., 2023).

with minimal delay.

This capability would be particularly valuable in assessing the labor market effects of sudden economic shocks - such as the COVID-19 pandemic - or tracking the consequences of large-scale immigration following armed conflict. Faster access to high-quality administrative data would provide timely, evidence-based insight for public policy decisions while also enhancing scientific competition by allowing researchers to respond to emerging issues with great speed and accuracy.

References

Abowd, J. M., & Kramarz, F. (1999). Chapter 40 the analysis of labor markets using matched employer-employee data. https://doi.org/10.1016/S1573-4463(99)30026-2

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*, 251–333. https://doi.org/10.1111/1468-0262.00020

Ahammer, A., Packham, A., & Smith, J. (2024). Firms and worker health. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4685812

Aizawa, N., Mommaerts, C., & Rennane, S. (2022). Explaining heterogeneity in use of non-wage benefits: The role of worker and firm characteristics in disability accommodations. *AEA Papers and Proceedings*, *112*, 376–80. https://doi.org/10.1257/PANDP.20221124

Altmann, S., Falk, A., Jäger, S., & Zimmermann, F. (2018). Learning about job search: A field experiment with job seekers in germany. *Journal of Public Economics*, *164*. https://doi.org/10.1016/j.jpubeco.2018.05.003

Andresen, M. E., & Nix, E. (2022). What causes the child penalty? evidence from adopting and same-sex couples. *https://doi.org/10.1086/718565*, *40*, 971–1004. https://doi.org/10.1086/718565

Ansala, L., Åslund, O., & Sarvimäki, M. (2022). Immigration history, entry jobs and the labor market integration of immigrants. *Journal of Economic Geography*, *22*, 581–604. https://doi.org/10.1093/jeg/lbaa038

Arellano-Bover, J. (2024). Career consequences of firm heterogeneity for young workers: First job and firm size. *Journal of Labor Economics*, *42*. https://doi.org/10.1086/723500

Arellano-Bover, J., & San, S. (2020). The role of firms in the assimilation of immigrants. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3594778

Bana, S., Bedard, K., Rossin-Slater, M., & Stearns, J. (2023). Unequal use of social insurance benefits: The role of employers. *Journal of Econometrics*, *233*, 633–660. https://doi.org/10.1016/J.JECONOM.2022.02.008

Bennett, P., Liu, K., & Salvanes, K. G. (2023). The decline of routine tasks, education investments, and intergenerational mobility. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4404498

Bertheau, A., Acabbi, E. M., Barceló, C., Gulyas, A., Lombardi, S., & Saggio, R. (2023). The unequal consequences of job loss across countries. *American Economic Review: Insights*, *5*, 393–408. https://doi.org/10.1257/AERI.20220006

Bertheau, A., Cahuc, P., Jäger, S., & Vejlin, R. (2022). Turnover costs: Evidence from unexpected worker separations.

Bertheau, A., & Hoeck, C. P. (2023). Firm beliefs about wage setting. www.antoinebertheau.com

Bertheau, A., Kudlyak, M., Larsen, B., & Bennedsen, M. (2023). Why firms lay off workers instead of cutting wages: Evidence from linked survey-administrative data.

Bertheau, A., Larsen, B., & Zhao, Z. (2024). What makes hiring difficult? evidence from linked survey-administrative data.

Bíró, A., Boza, I., Gyetvai, A., & Prinz, D. (2024). Health shocks, social insurance, and firms. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4868760

Bíró, A., & Elek, P. (2020). Job loss, disability insurance and health expenditure. *Labour Economics*, *65*. https://doi.org/10.1016/j.labeco.2020.101856

Böhm, M. J., Esmkhani, K., & Gallipoli, G. (2023). Firm heterogeneity in skill returns. https://ssrn.com/abstract=4653665

Bøler, E. A., Javorcik, B., & Ulltveit-Moe, K. H. (2018). Working across time zones: Exporters and the gender wage gap. *Journal of International Economics*, *111*. https://doi.org/10.1016/j.jinteco.2017.12.008

Boza, I., & Horn, D. (2023). *Contribution of high school heterogeneity to the wage variation of young workers*.

Boza, I., & Ilyés, V. (2020). Decomposition of co-worker wage gains. *IZA Journal of Labor Economics*, *9*. https://doi.org/10.2478/izajole-2020-0008

Boza, I., & Reizer, B. (2023). The role of flexible wage components in gender wage differences.

Boza, I., & Reizer, B. (2024). The role of flexible wage components in gender wage difference. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4889211

Britto, D. G. C., Pinotti, P., & Sampaio, B. (2022). The effect of job loss and unemployment insurance on crime in brazil. *Econometrica*, *90*. https://doi.org/10.3982/ecta18984

Bryson, A., Forth, J., Barber, C., of Trade, G. B. D., & Industry. (2006). Making linked employer-employee data relevant to policy. *DTI occasional paper*.

Burbano, V. C., Folke, O., Meier, S., & Rickne, J. (2023). The gender gap in meaningful work. *Management Science*. https://doi.org/10.1287/mnsc.2022.01807

Bütikofer, A., Løken, K. V., & Willén, A. (2024). Building bridges and widening gaps. *The Review of Economics and Statistics*, *106*, 681–697. https://doi.org/10.1162/rest_a_01183

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, *36*, S13–S70. https://doi.org/10.1086/694153

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, *131*, 633–686. https://doi.org/10.1093/qje/qjv038

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, *128*, 967–1015. https://doi.org/10.1093/qje/qjt006

Card, D., Rothstein, J., Yi, M., Albouy, D., Dauth, W., Davis, D., Diamond, R., Moretti, E., Kahn, M., Kline, P., Roca, J. D. L., & Wu, A. (2024). Location, location, location. *American Economic Journal: Applied Economics (forthcoming)*.

Cardoso, A. R., Guimarães, P., & Portugal, P. (2016). What drives the gender wage gap? a look at the role of firm and job-title heterogeneity. *Oxford Economic Papers*, *68*, 506–524. https://doi.org/10.1093/oep/gpv069

Cardoso, A. R., Guimarães, P., Portugal, P., & Reis, H. (2018). The returns to schooling unveiled. *IZA Discussion Papers*, 1–45. https://ideas.repec.org/p/iza/izadps/dp11419.html

Casarico, A., & Lattanzio, S. (2024). What firms do: Gender inequality in linked employer-employee data. *https://doi.org/10.1086/723177*, *42*, 325–355. https://doi.org/10.1086/723177

Cestone, G., Fumagalli, C., Kramarz, F., & Pica, G. (2023). Exploiting growth opportunities: The role of internal labour markets. *Review of Economic Studies*. https://doi.org/10.1093/restud/rdad094

Choquette, E., & Meinen, P. (2015). Export spillovers: Opening the black box. *World Economy*, *38*. https://doi.org/10.1111/twec.12225

Cirillo, V., Mina, A., & Ricci, A. (2024). Digital technologies, labor market flows and training: Evidence from italian employer-employee data. *Technological Forecasting and Social Change*, *209*, 123735. https://doi.org/10.1016/J.TECHFORE.2024.123735

Corak, M., & Piraino, P. (2011). The intergenerational transmission of employers. *Journal of Labor Economics*, *29*. https://doi.org/10.1086/656371

Criscuolo, C., Hijzen, A., Koelle, M., Schwellnus, C., Barth, E., Chen, W.-h., Fabling, R., Fialho, P., Garloff, A., Grabska, K., Kambayashi, R., Lankester, V., Stadler, B., Skans, O. N., & Murakozy, B. (2021). The firm-level link between productivity dispersion and wage inequality : A symptom of low job mobility ? *OECD Economics Department Working Papers*.

Currie, J., Kleven, H., & Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings*, *110*, 42–48. https://doi.org/10.1257/pandp.20201058

Dauth, W., Findeisen, S., Moretti, E., & Suedekum, J. (2022). Matching in cities. *Journal of the European Economic Association*, *20*. https://doi.org/10.1093/jeea/jvac004

Dellavigna, S., Lindner, A., Reizer, B., & Schmieder, J. F. (2017). Reference-dependent job search: Evidence from hungary. *Quarterly Journal of Economics*, *132*. https://doi.org/10.1093/qje/qjx015

Dobbin, C., & Zohar, T. (2023). Quantifying the role of firms in intergenerational mobility. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4636017

Dostie, B., Li, J., Card, D., & Parent, D. (2023). Employer policies and the immigrant–native earnings gap. *Journal of Econometrics*, *233*, 544–567. https://doi.org/10.1016/j.jeconom.2021.07.012

Drange, I., Helland, H., & Hermansen, A. S. (2024). Unionization, licensure and workplace variation in pay inequality between immigrants and natives. *Socio-Economic Review*, *00*, 1–36. https://doi.org/10.1093/SER/MWAD075

Dustmann, C., Glitz, A., & Vogel, T. (2010). Employment, wages, and the economic cycle: Differences between immigrants and natives. *European Economic Review*, *54*. https://doi.org/10.1016/j.euroecorev.2009.04.004

Eliason, M., Hensvik, L., Kramarz, F., & Skans, O. N. (2023). Social connections and the sorting of workers to firms. *Journal of Econometrics*, *233*. https://doi.org/10.1016/j.jeconom.2021.12.018

Eriksson, R. H., & Lengyel, B. (2019). Co-worker networks and agglomeration externalities. *Economic Geography*, *95*. https://doi.org/10.1080/00130095.2018.1498741

Fanfani, B. (2023). The employment effects of collective wage bargaining. *Journal of Public Economics*, *227*, 105006. https://doi.org/10.1016/J.JPUBECO.2023.105006

Fujita, S., Moscarini, G., & Postel-Vinay, F. (2024). Measuring employer-to-employer reallocation. *American Economic Journal: Macroeconomics*, *16*, 1–51. https://doi.org/10.1257/MAC.20210076

Gallen, Y., Lesner, R. V., & Vejlin, R. (2019). The labor market gender gap in denmark: Sorting out the past 30 years. *Labour Economics*, *56*, 58–67. https://doi.org/10.1016/J.LABECO.2018.11.003

Gerard, F., Lagos, L., Severnini, E., & Card, D. (2021). Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil. *American Economic Review*. https://doi.org/10.1257/AER.20181596

Glitz, A., & Vejlin, R. (2021). Learning through coworker referrals. *Review of Economic Dynamics*, *42*. https://doi.org/10.1016/j.red.2020.10.007

Godechot, O., Tomaskovic-Devey, D., Boza, I., Henriksen, L. F., Hermansen, A. S., Hou, F., Jung, J., Kodama, N., Křížková, A., Lippényi, Z., Melzer, S. M., Mun, E., Sabanci, H., Thaning, M., Apascaritei, P., Avent-Holt, D., Bandelj, N., Baudour, A., Cort, D., ... Soener, M. (2024). The great separation: Top earner segregation at work in advanced capitalist economies1. *https://doi.org/10.1086/731603*, *130*, 439–495. https://doi.org/10.1086/731603

Guvenen, F., Pistaferri, L., & Violante, G. L. (2022). Global trends in income inequality and income dynamics: New insights from grid. *Quantitative Economics*, *13*, 1321–1360. https://doi.org/10.3982/qe2260

Hällsten, M., & Kolk, M. (2023). The shadow of peasant past: Seven generations of inequality persistence in northern sweden. *American Journal of Sociology*, *128*. https://doi.org/10.1086/724835

Heining, J., Caldwell, S., & Haegele, I. (n.d.). *Bargaining and the gender pay gap*.

Hensvik, L., & Skans, O. N. (2016). Social networks, employee selection, and labor market outcomes. *Journal of Labor Economics*, *34*, 825–867. https://doi.org/10.1086/686253

Hermann, Z., Horn, D., Köllő, J., Sebők, A., Semjén, A., & Varga, J. (2022). A kompetenciaeredmények hatása a munkaerőpiaci sikerességre. *Közgazdasági Szemle*, *69*, 177–198. https://doi.org/10.18414/KSZ.2022.2.177

Hermansen, A. S. (2013). Occupational attainment among children of immigrants in norway: Bottlenecks into employment - equal access to advantaged positions? *European Sociological Review*, *29*. https://doi.org/10.1093/esr/jcr094

Hermansen, A. S., Borgen, N. T., & Mastekaasa, A. (2020). Long-term trends in adult socio-economic resemblance between former schoolmates and neighbouring children. *European Sociological Review*, *36*, 366–380. https://doi.org/10.1093/esr/jcz066

Hermansen, A. S., Penner, A., Elvira, M., Godechot, O., Hällsten, M., Henriksen, L. F., Hou, F., Lippényi, Z., Petersen, T., Reichelt, M., Sabanci, H. I., Safi, M., Tomaskovic-Devey, D., & Vickstrom, E. (2023, December). *Immigrant–native pay gap driven by lack of access to high-paying jobs* [working paper or preprint]. https://doi.org/10.31235/osf.io/2p4vw

Huitfeldt, I., Kostøl, A. R., Nimczik, J., & Weber, A. (2023). Internal labor markets: A worker flow approach. *Journal of Econometrics*, *233*, 661–688. https://doi.org/10.1016/J.JECONOM.2021.12.016

Ilyés, V., Boza, I., Lőrincz, L., & Eriksson, R. H. (2023). How to enter high-opportunity places? the role of social contacts for residential mobility. *Journal of Economic Geography*, *23*, 371–395. https://doi.org/10.1093/JEG/LBAC019

Ilyés, V., & Sebők, A. (2023). University peers and career prospects: The impact of university ties on early labor market outcomes. *Economics of Education Review*, *96*. https://doi.org/10.1016/j.econedurev.2023.102456

Jacobson, L. S., Lalonde, R. J., & Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review*, *83*.

Jäger, S., & Heining, J. (2022). How substitutable are workers? evidence from worker deaths. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4301638

Jäger, S., Roth, C., Roussille, N., & Schoefer, B. (2024). Worker beliefs about outside options. *Quarterly Journal of Economics*, *139*. https://doi.org/10.1093/qje/qjae001

Jensen, S. S., Lindemann, K., & Weiss, F. (2023). Parental job loss and the role of unemployment duration and income changes for children's education. *European Sociological Review*. https://doi.org/10.1093/esr/jcad068

Kauhanen, A. (2021). The effects of an education-leave program on educational attainment and labor-market outcomes. *Education Economics*, *29*. https://doi.org/10.1080/09645292.2021.1929849

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*, 181–209. https://doi.org/10.1257/APP.20180010

Köllő, J., Boza, I., & Balázsi, L. (2021). Wage gains from foreign ownership: Evidence from linked employer–employee data. *Journal for Labour Market Research*, *55*, 3. https://doi.org/10.1186/s12651-021-00286-0

Köllő, J., Boza, I., Ilyés, V., Kőműves, Z., & Mark, L. K. (2023). How do firms deal with the risks of employing ex-prisoners? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4653666

Lachowska, M., Sorkin, I., & Woodbury, S. (2022). Firms and unemployment insurance take-up. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4165338

Lichter, A., Peichl, A., & Siegloch, S. (2017). Exporting and labour demand: Micro-level evidence from germany. *Canadian Journal of Economics*, *50*. https://doi.org/10.1111/caje.12290

Lindner, A., & Reizer, B. (2020). Front-loading the unemployment benefit: An empirical assessment. *American Economic Journal: Applied Economics*, *12*, 140–74. https://doi.org/10.1257/APP.20180138

Liu, K., Salvanes, K. G., & Sørensen, E. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, *84*, 3–17. https://doi.org/10.1016/j.euroecorev.2015.08.015

Lochner, B. (2023). Iab job vacancy survey data linked to administrative data of the iab (iabse-adiab) 1975-2020.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, *107*, 1611–1637. https://doi.org/10.1257/aer.20141467

Mackeben, J., Ruf, K., Wolter, S., & Grunau, P. (2023). Lpp survey data linked with administrative data of the iab (lpp-adiab) 1975-2021.

Merida, A. L., & Rocha, V. (2021). It's about time: The timing of entrepreneurial experience and the career dynamics of university graduates. *Research Policy*, *50*. https://doi.org/10.1016/j.respol.2020.104135

Mueller, A. I., Osterwalder, D., Zweimüller, J., & Kettemann, A. (2024). Vacancy durations and entry wages: Evidence from linked vacancy–employer–employee data. *The Review of Economic Studies*, *91*, 1807–1841. https://doi.org/10.1093/RESTUD/RDAD051

Nekoei, A., & Weber, A. (2017). Does extending unemployment benefits improve job quality? *American Economic Review*, *107*. https://doi.org/10.1257/aer.20150528

Nieto, A. (2021). Native-immigrant differences in the effect of children on the gender pay gap. *Journal of Economic Behavior and Organization*, *183*. https://doi.org/10.1016/j.jebo.2021.01.015

OECD. (2021). *The role of firms in wage inequality*. https://doi.org/https://doi.org/https://doi.org/10.1787/7d9b2208-en

Oreopoulos, P., von Wachter, T., & Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, *4*, 1–29. https://doi.org/10.1257/app.4.1.1

Penner, A. M., Petersen, T., Hermansen, A. S., Rainey, A., Boza, I., Elvira, M. M., Godechot, O., Hällsten, M., Henriksen, L. F., Hou, F., Mrčela, A. K., King, J., Kodama, N., Kristal, T., Křížková, A., Lippényi, Z., Melzer, S. M., Mun, E., Apascaritei, P., ... Tufail, Z. (2022). Within-job gender pay inequality in 15 countries. *Nature Human Behaviour 2022 7:2*, *7*, 184–189. https://doi.org/10.1038/s41562-022-01470-z

Poole, J. P. (2013). Knowledge transfers from multinational to domestic firms: Evidence from worker mobility. *Rev. Econ. Stat.*, *95*, 393–406. https://doi.org/10.1162/rest_a_00258

Reichelt, M., & Müller, C. (n.d.). *Islamist terrorism and the employment of middle eastern men: A relational approach to event-based labor market effects*.

Rose, E. K. (2018). The effects of job loss on crime: Evidence from administrative data. *Working Paper*.

San, S. (2021). Who works where and why? parental networks and the labor market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3726993

Saygin, P. O., Weber, A., & Weynandt, M. A. (2021). Coworkers, networks, and job-search outcomes among displaced workers. *ILR Review*, *74*. https://doi.org/10.1177/0019793919881988

Schmutz, B., & Sidibé, M. (2019). Frictional labour mobility. *Review of Economic Studies*, *86*. https://doi.org/10.1093/restud/rdy056

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Wachter, T. V. (2019). Firming up inequality. *The Quarterly Journal of Economics*, *134*, 1–50. https://doi.org/10.1093/qje/qjy025

Staiger, M., Abra-Ham, K., Agarwal, R., Black, S., Caldwell, S., Chetty, R., Coglianese, J., Friedman, J., Fogel, J., Hendren, N., Kearney, M., Murray, S., Palloni, G., Paserman, D., Pope, N., Ricks, M., Sabelhaus, J., Shea, J., Sprung-Keyser, B., ... Wiemers, E. (2023). The intergenerational transmission of employers and the earnings of young workers.

Sullivan, D., & Wachter, T. V. (2009). Job displacement and mortality: An analysis using administrative data. *The Quarterly Journal of Economics*, *124*, 1265–1306. https://doi.org/10.1162/QJEC.2009.124.3.1265

Tomaskovic-Devey, D., Rainey, A., Avent-Holt, D., Bandelj, N., Boza, I., Cort, D., Godechot, O., Hajdu, G., Hällsten, M., Henriksen, L. F., Hermansen, A. S., Hou, F., Jung, J., Kanjuo-Mrčela, A., King, J., Kodama, N., Kristal, T., Krízková, A., Lippényi, Z., ... Tufail, Z. (2020). Rising between-workplace inequalities in high-income countries. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9277–9283. https://doi.org/10.1073/pnas.1918249117

Topel, R. H., & Ward, M. P. (1992). Job mobility and the careers of young men. *Quarterly Journal of Economics*, *107*. https://doi.org/10.2307/2118478

Torres, S., Portugal, P., Addison, J. T., & Guimarães, P. (2018). The sources of wage variation and the direction of assortative matching: Evidence from a three-way high-dimensional fixed effects regression model. *Labour Economics*, *54*, 47–60. https://doi.org/10.1016/j.labeco.2018.06.004

Appendix A

Table A1: Access details of LEE datasets – Europe

Country	Name	Availability	Affiliation req.	Note
Austria	Austrian Social Security Database (ASSD)	remote	EU	?
Denmark	Integrated Database for Labour Market Research (IDA)	remote	national	?
Estonia	?	?	?	?
Finland	FLEED	?	?	?
France	Base tous salariés (BTS); also known Déclaration Annuelle de Données Sociales (DADS)	remote	national?	?
Germany	Linked-Employer-Employee-Data of the IAB (LIAB)	local*	institutional**	*remote after an on-site period; **access for sample is less restricted
Greece	?	?	?	?
Hungary	Kapcsolt Államigazgatási Adatgyűjtés (Admin3)	remote	national*	*or institutional co-author
Ireland	?	?	?	?
Italy	Visitinps	secure	none*	*competitive calls for projects
Lithuania	Administrative data from the State Social Insurance Fund Board (SoDra)	?	?	?
Netherlands	Stelsel van Sociaal-statistische Bestanden (SSB)	remote	EU	?
Norway	Statistics Norway	remote	national?	?
Poland	PIT Database	?	?	?
Portugal	Quadros de Pessoal	remote	national*	*and specific international organizations
Slovakia	?	delegated	national?	?
Slovenia	?	?	?	?
Spain	Muestra Continua de Vidas Laborales	remote	?	?
Sweden	Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA)	remote*	national**	*in EU, ** projects hosted in Sweden
United Kingdom	?	?	?	?

Notes: Secure access occurs in controlled environments like data rooms. local access is on institutional computers, while remote access includes secure server connections or data downloads to personal devices. Delegated access means that researchers submit scripts and receive results without handling the data directly. Both secure and remote access may involve output-checking procedures. For affiliation requirements institutional refers to one (or few) specific institutions (the data handlers), and national covers mostly researchers scientific institutions in the given countries. We do not include whether access is provided on a one-on-one basis, through competitive calls, or other ways. Also, we do not include whether data access incurs costs.

Table A2: Access details of LEE datasets – Americas and Rest of the World

Country	**Name**	**Availability**	**Affiliation req.**	**Note**
Argentina	?	?	?	?
Australia	?	?	?	?
Brazil	Relacao Anual de Informacoes Sociais (RAIS)	?	?	?
Canada	Canadian Employer-Employee Dynamics Database (CEEDD)	local*	insitutional	*minor queries as remote (paid)
Chile	Muestra Afiliados al Seguro de Cesantía	remote	none	?
Columbia	Planilla Integrada de Liquidaciön de Paortes (PILA)	?	?	?
Costa Rica	Caja Costarricense de Seguro Social (CCSS)	?	?	?
Ecuador	Instituto Ecuatoriano de Seguridad Social	?	?	?
Israel	?	?	?	?
Mexico	?	?	?	?
New Zealand	?	?	?	?
Taiwan	?	?	?	?
United States	IRS tax data	remote*	national*	*US access ** residency requirement

Notes: Secure access occurs in controlled environments like data rooms. local access is on institutional computers, while remote access includes secure server connections or data downloads to personal devices. Delegated access means that researchers submit scripts and receive results without handling the data directly. Both secure and remote access may involve output-checking procedures. For affiliation requirements institutional refers to one (or few) specific institutions (the data handlers), and national covers mostly researchers scientific institutions in the given countries. We do not include whether access is provided on a one-on-one basis, through competitive calls, or other ways. Also, we do not include whether data access incurs costs.

Table A3: Features in LEE datasets by countries – Europe

	Austria	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Lithuania	Netherlands	Norway	Poland	Portugal	Slovakia	Slovenia	Spain	Sweden	United Kingdom
Location of workplace	✓	✓	?	?	✓	✓	?	×	?	✓	?	✓	✓	?	✓	×	?	✓	✓	?
Residence	×	✓	?	?	✓	✓	?	×	?	✓	?	✓	✓	?	×	✓	?	✓	✓	?
Work conditions (hours, overtime, bonuses)	×	✓	?	?	✓	✓	?	✓	?	✓	?	✓	✓	?	✓	×	?	×	✓	?
Firm financial records/ ownership	×	✓	?	?	✓	×	?	✓	?	✓	?	✓	✓	?	✓	✓	?	×	✓	?
Social benefits/transfers	✓	✓	?	?	×	×	?	✓	?	✓	?	✓	✓	?	×	×	?	×	✓	?
Health data	✓	✓	?	?	×	✓	?	✓	?	✓	?	✓	✓	?	×	✓	?	×	✓	?
Number of kids	✓	✓	?	?	✓	×	?	×	?	✓	?	✓	✓	?	×	×	?	×	✓	?
Family relations	×	✓	?	?	✓	×	?	×	?	×	?	✓	✓	?	×	×	?	×	✓	?
Immigration info	✓	✓	?	?	✓	×	?	×	?	✓	?	✓	✓	?	✓	✓	?	✓	✓	?
Observe educational institutions	×	✓	?	?	✓	×	?	✓	?	✓	?	✓	✓	?	×	×	?	×	✓	?
Individual (cogn.) skill data could be linked	×	✓	?	?	×	×	?	✓	?	×	?	×	✓	?	×	×	?	×	✓	?
Can be merged to any other data source	✓	✓	?	?	✓	✓	?	✓	?	×	?	✓	✓	?	×	✓	?	×	✓	?

Notes: The responses marked by a tick (✓) indicate the presence of any kind of related data in the current category (not necessary everything), while crosses (×) indicate the lack of any data on the given topic. Question marks (?) indicate that we have no or unclear information at the moment about the given item for the given country. A more detailed table on the availability of features (e.g. separate info on working hours, overtime and bonuses) is available at leedata.eu. If anything is missing / not correct / outdated, contact us at bozaistvan@gmail.com.

Table A4: Features in LEE datasets by countries – Americas and Rest of the World

	Argentina	Australia	Brazil	Canada	Chile	Columbia	Costa Rica	Ecuador	Israel	Mexico	New Zealand	Taiwan	United States
Location of workplace	?	?	?	✓	✓	?	?	?	?	?	?	?	×
Residence	?	?	?	✓	✓	?	?	?	?	?	?	?	✓
Work conditions (hours, overtime, bonuses)	?	?	?	×	×	?	?	?	?	?	?	?	✓
Firm financial records/ ownership	?	?	?	✓	✓	?	?	?	?	?	?	?	×
Social benefits/transfers	?	?	?	×	×	?	?	?	?	?	?	?	×
Health data	?	?	?	✓	✓	?	?	?	?	?	?	?	×
Number of kids	?	?	?	✓	×	?	?	?	?	?	?	?	✓
Family relations	?	?	?	✓	×	?	?	?	?	?	?	?	✓
Immigration info	?	?	?	✓	✓	?	?	?	?	?	?	?	✓
Observe educational institutions	?	?	?	×	✓	?	?	?	?	?	?	?	×
Individual (cogn.) skill data could be linked	?	?	?	✓	×	?	?	?	?	?	?	?	✓
Can be merged to any other data source	?	?	?	✓	×	?	?	?	?	?	?	?	✓

Notes: The responses marked by a tick (✓) indicate the presence of any kind of related data in the current category (not necessary everything), while crosses (×) indicate the lack of any data on the given topic. Question marks (?) indicate that we have no or unclear information at the moment about the given item for the given country. A more detailed table on the availability of features (e.g. separate info on working hours, overtime and bonuses) is available at leedata.eu. If anything is missing / not correct / outdated, contact us at bozaistvan@gmail.com.