Hungarian Administrative Employer–Employee Data: The Admin4 Database

István Boza

Rita Pető

Melinda Tir

KRTK-KTI WP 2026/1

February 2026

INSTITUTE OF ECONOMICS
ELTE CENTRE FOR ECONOMIC AND REGIONAL STUDIES
BUDAPEST, HUNGARY

ABSTRACT
Linked employer–employee administrative databases (LEED/LEE) represent one of the most important data innovations of recent decades, as they enable researchers to follow workers and firms simultaneously over long time horizons. This study first reviews the development of the Hungarian administrative data infrastructure and then presents the most recent Hungarian linked employer–employee dataset---the Linked Administrative Panel Dataset (Admin4)---in detail, with particular emphasis on its supplementary modules providing rich set of information.

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

István Boza
ELTE KRTK
boza.istvan@krtk.hu

Melinda Tir
ELTE KRTK
tir.melinda@krtk.hu

Rita Pető
ELTE KRTK
peto.rita@krtk.hu

Magyar adminisztratív munkáltató–munkavállaló adatok: az Admin4 adatbázis

Boza István

Pető Rita

Tir Melinda

ÖSSZEFOGLALÓ

Az összekapcsolt munkáltató–munkavállaló adminisztratív adatbázisok (LEED/LEE) az elmúlt évtizedek egyik legfontosabb adatinnovációját jelentik, mivel lehetővé teszik a kutatók számára, hogy hosszú időtávon, egyidejűleg kövessék nyomon a munkavállalókat és a vállalatokat. Ez a tanulmány áttekintést nyújt a magyar adminisztratív adatinfrastruktúra fejlődéséről, majd részletesen bemutatja a legfrissebb magyar kapcsolt munkáltató–munkavállaló adatbázist, a Kapcsolt Államigazgatási panel adatot (Admin4), külön hangsúlyozva annak kiegészítő moduljait és a bennük rejlő információkat.

JEL kódok: Q81

Kulcsszavak: kapcsolt munkáltató–munkavállaló adatbázis, adminisztratív adatok

Hungarian Administrative Employer–Employee Data: The Admin4 Database

István Boza (ELTE Centre for Economic and Regional Studies)
Rita Pető (ELTE Centre for Economic and Regional Studies)
Melinda Tir (ELTE Centre for Economic and Regional Studies)

February 4, 2026

Linked employer–employee administrative databases (LEED/LEE) represent one of the most important data innovations of recent decades, as they enable researchers to follow workers and firms simultaneously over long time horizons. This study first reviews the development of the Hungarian administrative data infrastructure and then presents the most recent Hungarian linked employer–employee dataset—the Linked Administrative Panel Dataset (Admin4)—in detail, with particular emphasis on its supplementary modules providing rich set of information.

JEL code: C81

Keywords: linked employer-employee data, administrative data

Introduction

Administrative linked employer–employee data (LEED, or LEE data) have become one of the most important data innovations of recent decades, enabling the observation of individuals and their employers jointly over extended periods. By linking individual employment histories to firm identifiers, such datasets allow the analysis of labor market dynamics, such as worker mobility, wage setting, and firm heterogeneity, within a unified longitudinal framework. Their defining feature is the joint longitudinal observation of workers and firms, which makes it possible to study processes both within firms and across employers.

Historically, linked employer–employee data grew out of employer-based wage surveys, which continue to play an important role in labor market analysis by providing detailed information on earnings components, working hours, and job characteristics. At the same time, survey-based datasets are typically constrained by sampling, limited population coverage, and infrequent observation, and often exclude the unemployed or inactive population. Administrative data sources, such as tax, social security, education, or health registers, address many of these limitations by offering broader population coverage, longer time horizons, and more reliable longitudinal tracking, albeit usually with less detail on working conditions and wage components.

In Hungary, administrative registers have been systematically developed into a large-scale linked employer–employee panel maintained by the Databank of the Centre for Economic and Regional Studies. Successive waves of the Linked Administrative Panel Database (Admin1–Admin3) progressively expanded both the time coverage and the range of observable labor market processes, and have been widely used in applied research on wages, employment dynamics, and inequality. Building on this infrastructure, the most recent wave, Admin4, marks a substantial expansion in both scope and content.

Admin4 integrates individual- and firm-level records from multiple administrative authorities and follows a representative 50% sample of the Hungarian population over a long time horizon (2003-2021). In addition to detailed employment and earnings histories, it also captures unemployment spells, social transfers, education, health care, and firm-level financial characteristics through a modular structure. This paper provides a detailed description of the structure, coverage, and key measurement features of the Admin4 database.

The Linked Administrative Panel Database (Admin4)

The database is constructed by linking individual- and firm-level records from the National Health Insurance Fund Manager (NEAK), the Hungarian State Treasury (MÁK), the Educational Authority (OH), the National Office of Vocational Education and Training and Adult Learning (NSZFH), the Ministry of Technology and Industry (TIM), and the National Tax and Customs Administration (NAV). The sampling frame is provided by NEAK, based on a 50% random sample drawn from the social security number register. This linkage ensures representativeness for the entire Hungarian population. The sample is continuously updated to reflect entries (births, immigration) and exits (deaths).

Anonymized records are processed under strict data protection protocols and stored in a central database by the National Infocommunications Service Company (NISZ). The data is then transferred to the Databank, where they cleaned, harmonized, and converted into a monthly format for research use. The full database follows individuals for up to 228 months. Alongside individuals, every employer of both the public and private sectors that has employed at least one sampled individual is included with their complete history (around 1.7 million firms and institutions).

Structurally, Admin4 consists of a core file and several supplementary modules. While the core file primarily contains essential employment information (e.g., employment and wage data), the supplementary files provide detailed individual- or firm-level information on health care, education, and unemployment, as well as financial data of enterprises. The records are linked through personal identifiers. This structure allows users to tailor the scope of information to the specific requirements of their research questions. It is important to note, however, that most supplementary modules are only available after 2009.

Based on the sensitivity of the information, the Databank classifies accessible data into three categories, as shown in Table 1. While data assigned to the lowest sensitivity level (H1) can be accessed remotely via the Databank's server with prior authorization, data classified at the highest level (H3) are only accessible within a secure on-site data room specifically established for this purpose.

Table 1: Data access levels in the Linked Administrative Panel Database (Admin4)

Level	Data categories	Eligible users	Mode of access
H1	Core data, basic supplementary modules	Applicants for research purposes, subject to Databank approval	KRTK Databank server
H2	More detailed health and education data	KRTK researchers, domain experts, subject to approval	KRTK Databank server
H3	Full-detail supplementary modules (e.g. ICD, medical procedure codes)	KRTK researchers, domain experts, subject to approval of a detailed project plan	KRTK secure data room

I) Core data

The database covers a representative 50% sample of the entire population over the period 2003–2021. Importantly, it includes not only active labour market participants but also inactive individuals—such as students and stay-at-home parents—as well as the unemployed, thereby offering a comprehensive picture beyond the active labour force.[1] For every individual, age and gender are known. Some of the supplementary modules can also be linked to the inactive population.

a) Employment Relationship and Its Characteristics

The available information is primarily based on employment relationships that establish pension entitlements and is derived from employer-submitted declarations. The database includes employees in both the public and private sectors, and the type of employment relationship is also visible (e.g. standard employment contracts, public works, simplified employment, etc.). Each employer, including self-employed businesses, receives a unique anonymized identifier, allowing employers and their associated employees to be tracked over time.

Key features of each job spell are recorded, including contractual weekly hours, insured workdays per month, and all pension-eligible earnings. Occupations are recorded according to the Hungarian Standard Classification of Occupations (FEOR), the national counterpart to the EU's ISCO system. This enables the precise distinction of job positions within a given employer. The database also tracks all jobs held simultaneously by individuals. At any given time, around 10% of employees hold more than one job.

[1] A limited set of information is also available for 2022, which we discuss in more detail later.

The database has certain limitations. For example, employment in older age is only observed when it is subject to social security contributions. Retired employees (and their employers) were exempt from contributions before 2007 and again after 2019. As a result, during these periods, the labor market information of employed pensioners (their employers, earnings, occupations) is not visible. Similarly, before 2012, members of the armed forces (police and military) were recorded in separate registries and thus are not included in the database. Finally, as is unavoidable with administrative sources, undeclared "black" employment cannot be tracked.

Earnings data are available for both employees and the self-employed, but only income forming the basis of pension contributions is recorded. Consequently, self-employed earnings are underestimated in the database. Moreover, until 2012, pension contributions were subject to an upper ceiling, meaning reported wages were censored at this cap for employees as well. This issue affects approximately the top 2% of the income distribution in the relevant years. While workers in the highest earning categories are still observable, their exact wage levels remain unknown.

Until 2012, most employers submitted contribution records annually, with monthly reporting becoming mandatory only from 2013 onwards. As a result, intra-year wage changes cannot be tracked accurately in earlier periods. In those years, the Databank allocated annual earnings across the relevant months. In Hungary, especially at the beginning of the observation period, "grey" employment was also widespread, whereby employers officially reported workers at the minimum wage while paying the remainder in cash. Since no pension contributions were paid on these undeclared amounts, recorded earnings often underestimate the true income—particularly in certain sectors (e.g. construction, retail) and among microenterprises (Köllő and Elek 2012).

While the full database is available for the 2003–2021 period, employment relationship data are also accessible for 2022. However, we should note that the 2022 sample includes only individuals who were active in the labor market in that year. No information is available on the inactive population, nor on individuals entering the Hungarian registers for the first time in 2022. Thus immigrants arriving in 2022 are excluded from the sample.

b) Data on the total population

The database contains information not only on those active in the labor market but also on a representative sample of the entire population. For the total population, we can observe from 2003 onwards whether individuals are engaged in some form of work. In addition, data related to *registered* unemployed persons are available from 2009, including information on registration, benefit payments, program participation, job placement services, as well as participation in public employment schemes (the latter from 2011).[2] The database also includes information on the wage subsidy programs introduced during the COVID-19 pandemic.

Data on pensions and pension-type benefits are available for the entire period. In the case of old-age pensions, the recognized service period is also recorded. Among cash social transfers, data on benefits such as the infant care allowance (TGYÁS/CSED), child care fee (GYED), or sick pay have been available since 2012. For earlier periods, these are only partially accessible or only at the status level, without the exact amounts paid. As the dataset is monthly level, it enables the detailed examination of the temporal dynamics of these benefits.

c) Geographical dimension

According to the legal regulations, the most detailed spatial resolution of the database is the district level. The district corresponding to both the permanent and temporary residence of the persons is known for each month. For Hungarian citizens, the place of birth is also available at the district level, while for those born abroad, it is recorded at the level of broad country-groups (geographical regions).

II) Supplementary modules

d) Information on the employer

Employers are identified through unique anonymized identifiers and can be traced through years in a panel structure. Among the firm-level information, the industry classification

[2]It is important to note, however, that we only observe jobseekers registered at the employment offices, which constitute a (selected) subsample of the total unemployed population. Unfortunately, administrative data do not allow us to distinguish between those who are unemployed but not registered and those who are truly inactive.

(NACE) and certain characteristics from the Wage Survey (available until 2018) are available.[3] In addition, for companies obliged to submit reports, corporate balance sheet data from the tax authority (NAV) is included. The latter includes the main elements of the income statement, such as turnover, material costs, total assets, capital structure, and headcount figures. The information is available on an annual basis.

The industry classification and balance sheet data can be directly linked only for the period between 2011 and 2020. For earlier years, the Databank applied probabilistic matching procedures. Data for the years 2021 and 2022 are not yet included in the current version of the database.

e) Detailed transfer data

For the period from 2003 to 2021, detailed pension data are available to researchers on a monthly basis. The database contains the date of entitlement for pensions and pension-type benefits and allowances, the recognized service period associated with the benefit, the regular monthly payment amounts, and one-off payments. During data processing, the benefits were aggregated; however, the most important types of pensions can also be analyzed separately (old-age pensions, early retirement pensions, disability pensions).

f) Health care

Health care data refer to publicly financed services and are available from 2009 onwards.[4] The main types of services include inpatient and outpatient care, prescription drug use, and diagnostic examinations (e.g., MRI, CT).[5] Services are identified using BNO codes, the Hungarian implementation of ICD (International Classification of Diseases) and OENO (Hungarian National Classification of Medical Procedures) codes. The data are available at the event level. It includes the reimbursed financial costs associated with each service. Based on these codes, aggregated indicators are also available to describe health status.

[3]The Hungarian survey-based linked employer–employee database is called the Wage Survey ("Bértárfia"). The first wave took place in 1986, and it was repeated in 1989. From 1992 and onwards, the survey has been conducted annually.

[4]The Hungarian health care system is strongly centralized, with the National Institute of Health Insurance Fund Management (NEAK) overseeing a single insurance fund that guarantees coverage for almost the entire population.

[5]Information on general practitioner services, inpatient and outpatient care, as well as prescription drug dispensations and CT/MRI examinations is available from 2009. Data on public health subsidy programs are, however, available from 2005, while information on dental care is only available to researchers from 2016 onwards.

g) Education

Educational data are available from 2009 onwards, though in some cases they can be traced back to earlier years. The database includes information on participation in public education and the results of secondary school final examinations (from 2005). From 2008 onward it is also supplemented by the score of a standardized test of mathematics and literacy and detailed background information from a questionnaire provided the National Assessment of Basic Competences. The availability of higher education data is more limited, but institutional participation and degree attainment are known in certain cases. Information on completed educational attainment is available for individuals born after 1991 and was partly reconstructed retrospectively.

h) Family

For a subset of individuals, information on family links is also available. Parent–child relationships are identified from records on applications for and receipt of family allowance and can be observed only when both the parent and the child appear in the sample. Given the 50% random sampling of the population, this implies that roughly one-quarter of all true parent-child links can be reconstructed. The number of children is likewise derived from family allowance records, and observed for sampled individuals who applied for or received family allowance for their child. Data on these applications and disbursements are available from 2008 onwards.

Summary and Outlook

The Hungarian Admin4 database is distinctive internationally in that it is based on a population sample, covering not only employed individuals (in both private and public sectors), but also inactive persons, the unemployed, and the self-employed. This design allows precise tracking of labor market entries, exits, and transitions between employment statuses. Although the sample is limited to 50%, which may constrain analyses relying on very rare events or complete population networks, the monthly and continuous spell-level structure enables detailed analysis of short-term labor market dynamics, such as responses to policy changes, and provides substantially finer temporal resolution than datasets with

only annual observations.[6]

Content-wise, the Admin4 is particularly rich, with modules covering health care, education, financial, social, and demographic information. While many countries operate linked employer–employee databases, supplementary information such as firm financials, social transfers, geographic data, health care usage, education records, or cognitive skill measures is far less common. The Admin4 provides all of these dimensions, although coverage and availability vary across modules.

Shortcomings include the following. Educational attainment is directly available only for younger cohorts with an active student status. Employees are observed at the employer rather than the worksite or establishment level, which limits fine-grained geographic analyses and the identification of workplace-level peer groups. These features should be taken into account when designing empirical applications.

Regarding the future, according to current plans, subsequent waves of the database (Admin5 and beyond) will be updated on an annual basis, extending the longitudinal dimension of the data and enabling more timely observation of labor market adjustments in response to major economic, policy, or demographic shocks.

References

Boza, István and Rita Pető (2025). "The Rise of Linked Employer-Employee Panel Data: Where Are We Now?" In: *KRTK-KTI WORKING PAPERS* (2).

Köllő, János and Péter Elek (2012). "Adóelkerülés, adócsalás, fekete- és szürkefoglalkoztatás". In: *Munkaerőpiaci Tükör 2012*. Ed. by Károly Fazekas, Péter Benczúr, and Álmos Telegdy. Budapest: MTA KRTK–KTI – OFA, pp. 165–190.

[6] A systematic international overview of linked employer–employee datasets, including detailed cross-country comparisons, is provided in Boza and Pető (2025).