Improving Official Employment Statistics: New Insights from Private-sector Data*

André Kurmann† Etienne Lalé‡

April 15, 2026

PRELIMINARY AND INCOMPLETE

Abstract

This paper develops new methodology to improve real-time employment measurement for small business establishments in high-turnover service sectors, where official estimates such as the Current Employment Statistics (CES) face their greatest challenges. Building on Kurmann et al. (2025), we make two methodological advances. First, we improve the identification of true business openings and closings in private-sector payroll data by exploiting self-reported company creation dates and exit reasons together with independent information on business-level foot-traffic and transactions, thus resolving a fundamental timeliness–accuracy trade-off in prior work. Second, we extend employment estimates beyond the payroll sample to a broader population of small businesses using a supervised prediction framework that maps observable business activity signals to payroll outcomes, validated against administrative population data. Applying the methodology to Homebase scheduling and payroll data from 2018 through early 2026, we produce weekly estimates of employment, hours, wages, and job flows. We then evaluate the performance of our method by comparing the resulting estimates against administrative population data.

*We thank Madison Perry for outstanding research support. Portions of this manuscript were drafted with the assistance of Claude AI. The output was reviewed, verified, and edited by the authors.

†Drexel University. Email: andre.kurmann@drexel.edu

‡York University. Email: elale@yorku.ca

1 Introduction

The accuracy of official U.S. employment statistics has come under renewed scrutiny following historically large benchmark revisions to the Current Employment Statistics (CES) survey. The preliminary benchmark revision for March 2025 was −911,000 jobs, the largest in absolute terms on record and the largest in percentage terms since the Great Recession. The previous year's preliminary revision of −818,000 (finalized at −598,000) was itself the second largest. While representing only about 0.5% relative to total non-farm employment, these magnitudes exceed a threshold that matters for policymaking: the Federal Reserve, financial markets, and private businesses all condition decisions on CES payroll growth estimates that, as it turns out, may overstate—or in other instances understate—actual employment growth by hundreds of thousands of jobs.

The CES is the product of decades of methodological refinement and rigorous statistical practice. As such it remains the benchmark for real-time employment measurement by design. Yet its inherent constraints – a survey-based design with increasing non-response issues, estimates that are available only for relatively coarse industry-geography cells, and publication at monthly frequency with considerable time-lag – limit its usefulness and reliability, especially during periods of sharp disruptions such as the COVID-19 pandemic and around business cycle turning points. In response, a substantial literature has developed that uses novel private-sector datasets to provide more timely, higher-frequency, and more granular employment estimates (e.g., Bartik et al. 2020; Cajner et al. 2020; Chetty et al. 2024; Kurmann et al. 2025), which received widespread attention in recent years.

While very useful, estimates from these private-sector datasets are subject to many of the same fundamental measurement challenges that affect the CES and other official statistics. The two central difficulties are identifying true business openings and closings, and ensuring the representativeness of the resulting estimates. Private-sector opportunity samples typically grow over time as platform adoption expands, so new entrants to the data are a mix of genuinely new businesses and existing businesses that recently adopted the platform. Without direct information on which is which, naive employment counts confound true economic dynamics with platform growth. Furthermore, because adoption of a particular platform is unlikely to be random, these samples may attract particular types of establishments, implying compositional biases that may evolve over time as sample frame expands. Indeed, these problems are potentially more severe than in the CES, where the sampling frame is anchored to the near-universe Quarterly Census of Employment and Wages (QCEW) and decades of methodological investment underpin the birth-death and nonresponse corrections behind the

offical estimates.[1]

This paper develops new methodology to directly address these challenges, with a focus on the segment of the labor market that is both hardest to measure and likely most consequential for measurement error: small establishments in high-turnover service industries. Indeed, small businesses account for a disproportionate share of employment dynamics through births and deaths, and as we document in Section 2, the sectoral concentration of recent CES benchmark revisions falls squarely on the industries where the employment share of small establishments is among the largest.

We combine private-sector payroll data with independent information on business activity from geo-located cellphone visits (and, in future iterations, establishment-level transactions data) to construct high-frequency estimates of employment, hours, wages, and job flows that distinguish true business openings and closings from sample churn (i.e., entries of already operating businesses and exits of continuing businesses). In a second step, we then exploit the much larger coverage of the independent data sources on business activity to extrapolate these estimates beyond the available payroll sample to a broader universe of small establishments using a supervised prediction framework. This yields more granular and more representative estimates than either data source could provide alone. As we discuss throughout the paper, this task is subject to several potential caveats, and we do not claim to resolve all measurement challenges. Rather, we aim to make targeted progress on the margins where the potential gains are largest.

The new methodology is motivated by a growing body of research that traces measurement difficulties in the CES to its "net birth-death model", which estimates the net contribution of employment from newly created and recently closed establishments based on imputation and historical aggregate data, and to declining survey response rates (Mueller, 2006; Decker, 2026). Both of these problems are disproportionately concentrated in small establishments in high-turnover service sectors, precisely where employment dynamics are most volatile and hardest to capture in real time. These observations point to a clear set of desiderata for improved measurement, particularly for this segment of the economy: (i) automatic, administrative-style data collection; (ii) direct information on business openings and closings; (iii) high-frequency real-time availability; and (iv) large, representative samples. No single existing source satisfies all four. Administrative population counts from the QCEW by definition work well on accuracy and coverage but are published only with a multi-quarter lag. The CES is timely but suffers from the birth-death and non-response problems

[1] In addition, there are other potential caveats about relying on private-sector data that are non-statistical in nature; in particular continuity of data access and business-driven changes in data collection and reporting. While very important, the present paper does not address these issues.

noted above. Private-sector microdata can in principle deliver on all four dimensions, but only if the twin challenges of sample churn and representativeness are properly addressed, as Kurmann et al. (2025) demonstrated in the context of the COVID-19 pandemic. The present paper develops that approach systematically in two directions.

Our **first contribution** addresses desiderata (i) and (ii), automatic data collection and direct information on business activity, through a significant advance in how business openings and closings are identified in private-sector data. In Kurmann et al. (2025), we addressed this by matching payroll establishment records to independent signals of business activity from Google Places and Facebook posts. While effective, this approach had important limitations: it was essentially limited to businesses with physical store locations, and it faced a fundamental trade-off between timeliness and accuracy since identifying closings required independent signals for several weeks beyond the exit of an establishment from the payroll sample frame.

In the present paper, we exploit two new complementary approaches that resolve these limitations. The first relies on business activity information that is internal to the private-sector payroll data itself: self-reported company creation dates, which indicate whether a sample entrant is a genuinely new business or an existing one that recently adopted the software, and self-reported reasons for leaving the platform—such as permanent closure, switching to a competing provider, or seasonal shutdown—recorded for a subset of exiting establishments. The second approach uses foot-traffic data geo-located cell-phones as well as transaction data from payment processors as an independent, high-frequency measure of business activity. Together, the two sources of information on business activity in real- or near-real time, and can be combined to validate and extend the identification of business openings and closings in the payroll sample.

We apply the new method to data from Homebase, a scheduling and payroll software provider used primarily by small businesses. Preliminary results with only Homebase-internal information on business creation and closure indicate that the proposed methodology works surprisingly well: small business estimates for three prominent service sectors, retail trade, professional and business services, and leisure & hospitality match annual ex-post population data from the QCEW surprisingly well, implying relatively modest revisions.

In addition, the preliminary estimates reveal interesting seasonal patterns and real-time developments about small business dynamics and job flows not currently available in the official statistics. In particular, we find that small business openings for the second half of 2025 and the beginning of 2026 are unusually low, below pandemic levels. Furthermore, small business job growth from new hires is at declining rate while average hourly wages have increased by less than for larger businesses since the end of the pandemic.

Our **second contribution** addresses the desideratum (iv), coverage and representativeness, by extending the estimates beyond the available payroll sample to a broader population of (small business) establishments. For instance, in the Homebase sample used in Kurmann et al. (2025) was broadly representative of small businesses in the covered sectors, but its coverage is limited to approximately 130,000 establishments per year. While substantial, this represents only a fraction of the total small business population in those sectors. To improve representativeness and geographic granularity, we propose to use the establishments in the payroll sample merged with visits and payments data as a training sample for a predictive framework that maps observable activity signals (visit counts, dwell times, transaction volumes) to payroll outcomes (employment, hours, wages, opening/closing status). The predictive framework uses K-fold cross validation as applied in an economics context by Kleinberg et al. (2015) and Mullainathan and Spiess (2017). Furthermore, the temporal stability of the resulting activity-to-payroll prediction can be assessed with out-of-sample validation as proposed for instance in Ginsberg et al. (2009) as well as external benchmarking against QCEW and BED population statistics.

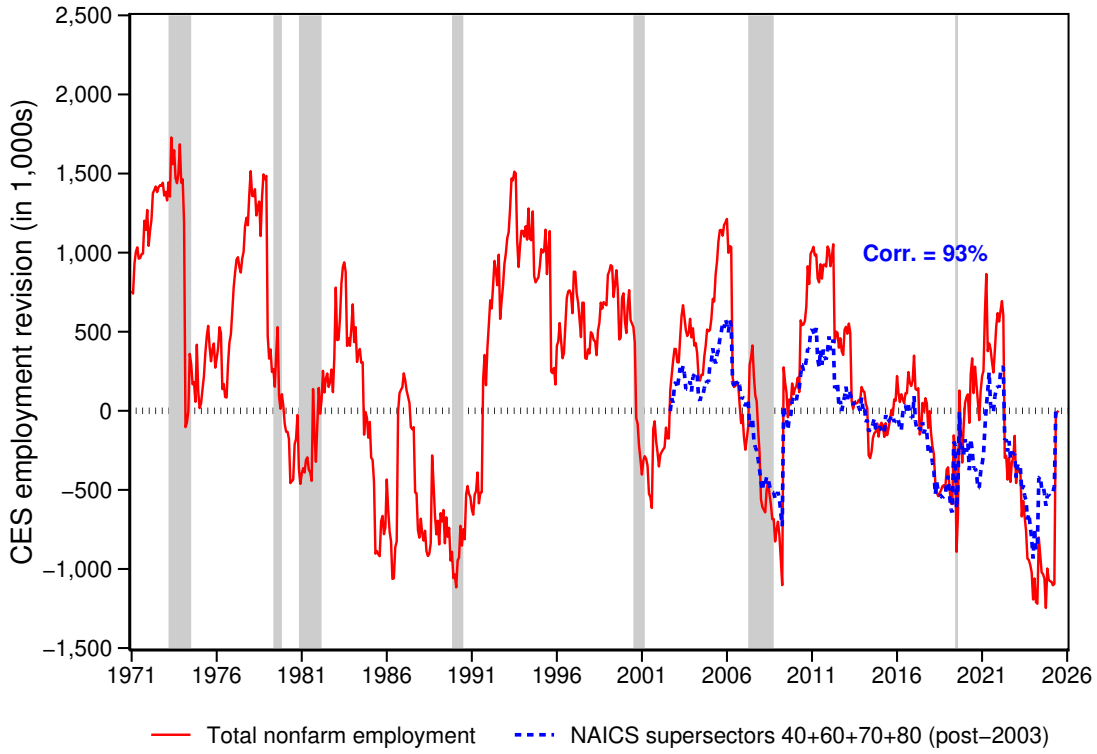
Our work contributes to an active policy and research agenda on improving official employment measurement. Decker (2026) surveys potential improvements to the CES, including expanding the data sample through private payroll providers, more frequent benchmarking, and improved birth-death estimation using contemporaneous external data. Cajner et al. (2022) demonstrate the value of combining multiple private-sector data sources to improve payroll employment measurement. Crane et al. (2022) use non-traditional indicators to track business exit during the pandemic. Our paper provides a concrete methodological framework and empirical demonstration of how private-sector microdata, combined with supplementary information on business activity, can address all four measurement desiderata for the segment of the economy that is hardest for official statistics to measure. The approach is transparent, produces standard errors, and can be readily benchmarked against population statistics—addressing Decker's (2026) emphasis that the most useful approach is access to underlying microdata so that methodological choices remain with trained statistical staff.

The remainder of the paper is organized as follows. Section 2 presents the empirical case for focusing on small establishments in high-turnover service sectors. Section 3 develops the methodology for estimating employment, job flows, hours, and wages from private-sector data, and describes the prediction-based coverage extension. Sections 4 and 5 apply the methodology to small business payroll data from Homebase and present results. Section 6 provides further details of the prediction-based coverage extension. Section 7 concludes.

2 The Case for Focusing on Small Establishments

To provide context, we begin with Figure 1 providing a long-run perspective on CES employment revisions. The solid line denotes the difference between the initial and final release of CES estimates for total non-farm employment. As previous studies have documented (Haltom, 2005; Phillips and Nordlund, 2012; Quinlan and Pinheiro, 2026), these revisions are typically positive during expansion periods and negative during economic downturns. The historical significance of the March 2025 revision is directly visible by comparing the most recent data points of the series with the revisions that followed the 1990s recession as well as the Great Recession of 2007-09.

Figure 1: CES employment revisions



NOTES: The figure shows CES revisions (defined as the difference between first and final release of data) for total nonfarm employment (solid line) and employment in NAICS supersectors 40 (Trade, transportation, and utilities), 60 (Professional and business services), 70 (Leisure and hospitality), and 80 (Other services) combined (dotted line). NAICS supersector data revisions are available from May 2003 onward only. Data are seasonally adjusted. Gray-shaded areas indicate NBER recession periods. Source: St. Louis Fed: Archival FRED and Bureau of Labor Statistics.

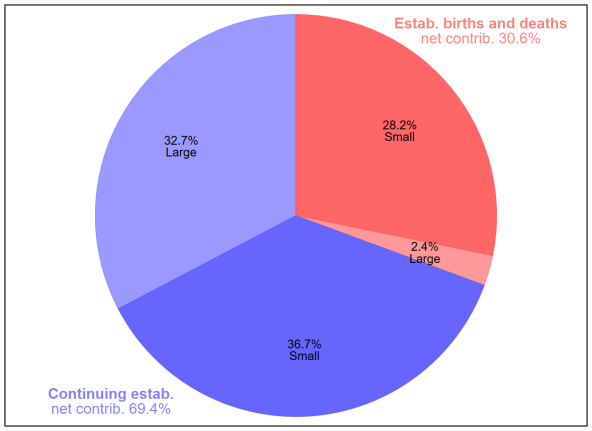
Next to the solid line in Figure 1, the dotted line shows CES employment revisions for four of of the service supersectors combined: Trade, transportation, and utilities (40),

Professional and business services (60), Leisure and hospitality (70), and Other services (80). While they represent 38% of total nonfarm employment on average, these four supersectors explain the lion's share of CES total employment revisions, as demonstrated by the close comovements between the two lines.

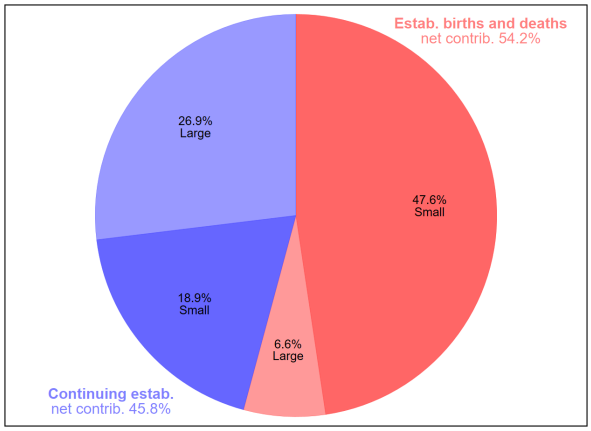
In Figure 2, we turn to annual data from the Business Dynamics Statistics (BDS) to assess the relative role of births/deaths by establishment size in driving aggregate employment growth. As shown in Panel (a), net employment from births/deaths accounts for almost a third of net employment growth at the yearly frequency.[2] Small establishments account for the bulk of net employment from births/deaths, and they also drive about half of the employment growth that is coming from continuing establishments.[3]

Figure 2: Contribution of births/deaths and continuing establishments, by size

(a) Total nonfarm employment



(b) NAICS supersectors 40+60+70+80



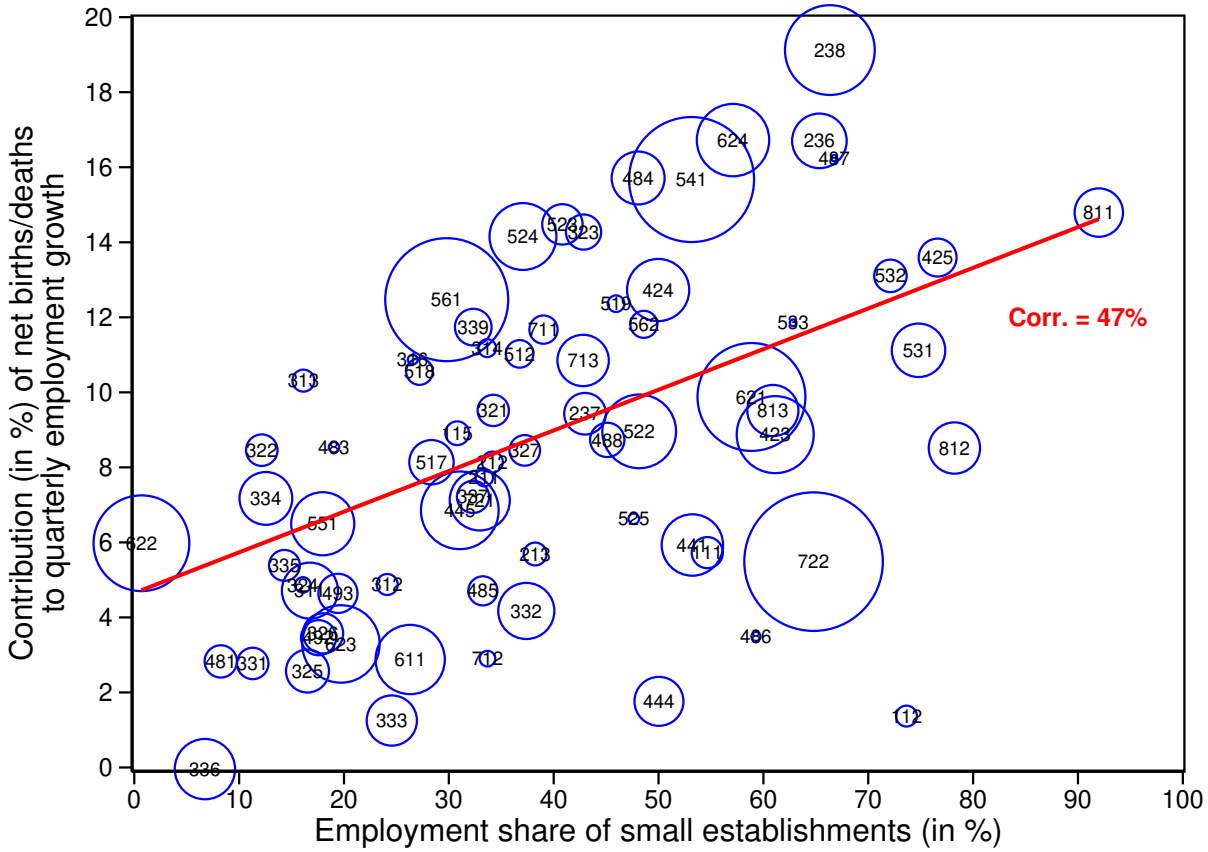
NOTES: The figure shows the contribution of net employment from births/deaths and net employment from continuing establishments to yearly employment growth, on average over the years 1979-2023 in data from the Business Dynamics Statistics. Small establishments are defined as establishments with 1 to 99 employees.

Next, in panel (b) of Figure 2, we focus on the four sectors that we previously highlighted. For those, employment growth through births/deaths explain almost 55% of total employment growth, with much of this margin driven by small establishments.

[2]The frequency of measurement is important here: for instance, any employment that is added by an establishment, say, one or two quarters after it is born is still counted as job creation from a newly-born establishment, whereas in quarterly data this would be counted as job creation from a continuing establishment.

[3]In the BDS, we can either define small establishments as those with 1-19 employees or 1-99 employees. Using the former definition, we find that small establishments account for 18.4% of employment growth through births/deaths and 13.4% through continuing establishments. Hence, while the cutoff at 99 employees clearly inflates the role of small establishments regarding the latter margin, it remains that small establishments drive the largest part of employment growth through births/deaths.

Figure 3: Net births/deaths contribution to employment growth and estab. size



NOTES: The figure shows the contribution of net employment from births/deaths to quarterly employment growth within each 3-digit NAICS industry, plotted against the share of employment within each industry accounted for by small establishments (fewer than 50 employees). Data is weighted by employment. The solid line is the OLS regression line ($N = 72$ 3-digit NAICS industries). Source: Business Employment Dynamics and Quarterly Census of Employment and Wages.

Finally, in Figure 3, we document a striking correlation between the importance of establishment births and deaths for industry-specific employment dynamics and the size distribution of establishments. Using quarterly data from the Business Employment Dynamics (BED), we first measure the contribution of net employment growth coming from births/deaths to employment growth within each available 3-digit NAICS industry. Note that, due to the higher frequency of measurement, these contributions are lower than in our analysis of BDS data. We then correlate these with the employment share of small establishments (those with fewer than 50 employees) for each industry, which we obtain from the Quarterly Census of Employment and Wages (QCEW). The strong correlation reported in Figure 3 indicates that industries with mostly small establishments tend to be industries where births/deaths play a relatively more important role in employment growth.

3 Methodology

This section presents the methodology at a conceptual level. Section 3.1 describes the recursive employment estimator, building on Kurmann et al. (2025), including the central problem of distinguishing true business openings and closings from sample churn. Section 3.2 discusses the additional labor market statistics we construct—hours, wages, job flows, and gross worker flow rates including recall—and explains why sample churn is a less direct concern for these rate variables. Section 3.3 introduces the idea of prediction-based coverage extension based on business activity signals from a larger sample, with the full formalization and implementation deferred to Section 6.

3.1 Employment Estimator

Consider estimating employment based on a sample of establishments from a private-sector provider or a survey. Starting with a reference employment level $\hat{E}_0$ taken from administrative population data, the estimate for employment in period t is computed recursively as

$$\hat{E}_t = \hat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} \right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} \right)} \tag{1}$$

where ω_i are sampling weights for cell i (e.g., industry $\times$ size class $\times$ region) constructed to make the cells representative of the population; $\hat{e}_{i,t}^{\mathcal{A}_{i,t}}$ is employment of establishments active in both t and $t-1$; $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ is employment gains from establishments newly opening or reopening in t; and $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}$ is employment losses from establishments closing in t.

The central challenge is to distinguish true business openings and closings from sample churn—establishments entering or exiting the data provider's platform without any change in underlying business activity. As shown in Kurmann et al. (2025), naive treatments of sample entry and exit (either treating all entries and exits as openings and closings, or ignoring them entirely) can generate employment estimates that differ dramatically from properly adjusted estimates if the underlying sample frame is expanding or contracting. Following Kurmann et al. (2025), we adopt the following framework to estimate employment gains from openings and losses from closings as

$$\hat{e}_{i,t}^{\mathcal{O}_{i,t}} = \sum_{\ell \in i} \hat{p}(\mathcal{O}_{\ell,t} \mid \text{entry}_{\ell,t}) \times \theta_{i,t}^{\mathcal{O}} \times e_{\ell,t}, \tag{2}$$

$$\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} = \sum_{\ell \in i} \hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t}) \times \theta_{i,t}^{\mathcal{C}} \times e_{\ell,t-1}, \tag{3}$$

where $\hat{p}(\mathcal{O}_{\ell,t} \mid \text{entry}_{\ell,t})$ and $\hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t})$ are the estimated probabilities that establishment ℓ is a true opening or closing conditional on entering or exiting the sample; and $\theta^{\mathcal{O}}_{i,t}$ and $\theta^{\mathcal{C}}_{i,t}$ are adjustment factors to correct for possible selection and measurement issues that could arise if (i) establishments in the sample have systematically different propensities to be true openings or closings than in the population; (ii) the independent sources used to estimate opening and closing probabilities are not fully representative; or (iii) the information from these sources contains systematic measurement error.

Estimating opening and closing probabilities

In Kurmann et al. (2025), opening and closing probabilities were estimated by matching Homebase establishments to independent information on business activity from Google Places and Facebook postings. For entrants with no evidence of prior business activity in these sources, $\hat{p}(\mathcal{O}_{\ell,t} \mid \text{entry}_{\ell,t}) = 1$ was assigned; for entrants with documented prior activity, $\hat{p}(\mathcal{O}_{\ell,t} \mid \text{entry}_{\ell,t}) = 0$; and unmatched entrants received the cell-average probability. For exits, establishments that re-entered the sample within the estimation window were classified as temporary closings ($\hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t}) = 1$); remaining exits were assigned $\hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t}) = 1$ if Google Places recorded a closure or Facebook showed no posting activity, and $\hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t}) = 0$ otherwise; unmatched exits received the cell-average probability.

This approach had two key limitations. First, it faced a fundamental trade-off between timeliness and accuracy: real-time estimates, which could only use information available up to period $t + s$ for small s, were prone to false positives and false negatives because exits that would eventually re-enter the sample had not yet done so. Retrospective estimates were more accurate but available only with a delay. Second, the approach was essentially limited to businesses with physical store locations.

In this paper, we exploit two new, complementary approaches to estimating opening and closing probabilities. The first relies on business activity information that is internal to the payroll or scheduling platform itself: self-reported company creation dates, which indicate whether a sample entrant is a genuinely new business or an existing one that recently adopted the software, and self-reported reasons for leaving the platform—such as permanent closure, switching to a competing provider, or seasonal shutdown—recorded for a subset of exiting establishments. The second approach uses transaction data from payment processors as an independent, high-frequency measure of business activity. Transaction data serve two complementary roles: extending coverage of the probability estimates to establishments for which platform-internal information are unavailable, and cross-validating those estimates against an independent source. Section 4.2 describes the implementation and validation of both approaches.

Both approaches are likely observed at the moment or near sample entry or exit, thus addressing the timeliness–accuracy trade-off that required the two-stage estimation in Kurmann et al. (2025). Furthermore, both approaches are applicable to business with and without in-person customers.

Adjustment factors

Following Kurmann et al. (2025), the adjustment factors are calibrated over a reference period $\mathbf{0}$ such that the implied birth and death rates match their population counterparts from administrative data (BED/BDS) for the same period. The approach implicitly assumes that selection and measurement issues remain stable between the reference period and the estimation period.

For openings, the adjustment factor for cell i is allowed to vary over time as:

$$\theta^{\mathcal{O}}_{i,t} = \theta^{\mathcal{O}}_{i,\mathbf{0}} \times \left(\frac{n^{\text{entry}}_{i,t} / n^{\mathcal{A}}_{i,t}}{n^{\text{entry}}_{i,\mathbf{0}} / n^{\mathcal{A}}_{i,\mathbf{0}}} \right)^{-1} \tag{4}$$

where $n^{\text{entry}}_{i,t}$ is the number of new entrants and $n^{\mathcal{A}}_{i,t}$ is total sample size in cell i at time t. This time variation, which is akin to inverse probability weighting, accounts for the fact that entry rates can change significantly with sample size. For instance, if a sample expansion doubles the rate of entry while average opening probabilities remain unchanged, employment gains from openings would spuriously double as well. The adjustment factor in (4) controls for this by scaling inversely with the relative entry rate.

For closings, the adjustment factor remains fixed at its reference-period value:

$$\theta^{\mathcal{C}}_{i,t} = \theta^{\mathcal{C}}_{i,\mathbf{0}} \tag{5}$$

since exit rates are naturally bounded by the available sample size and do not exhibit the same scale dependence as entry rates. The closing probabilities $\hat{p}(\mathcal{C}_{\ell,t} \mid \text{exit}_{\ell,t})$ themselves vary over time, but the adjustment factor does not need to be updated.

These adjustment factors can be updated whenever new administrative population data (e.g., from the QCEW) become available. Conceptually, this updating procedure is analogous to the benchmark revisions used in the CES, but applied at the much deeper industry–size cell level as part of the opening and closing probability estimator.

3.2 Other Labor Market Statistics

Another advantage of using longitudinally linked worker-firm level data from scheduling and payroll processing providers is that it often allows one to construct high-frequency labor market statistics of interest beyond employment such as gross hiring and separation rates (including recall) as well as average hours worked and average wages. Importantly, these statistics are rates or unit measures rather than levels. Sample churn therefore does not pose a direct challenge as it does for total employment as long as the sample is representative within cells. Furthermore, compositional change, which can be important during times of disruptions, can be addressed by computing the statistics on a balanced panel of continuously active establishments, respectively continuously employed workers. (Kurmann et al., 2025). Section 5 provides details.

3.3 Prediction-Based Coverage Extension

A natural concern with any private-sector data source is coverage. For instance, the Homebase payroll data we use in our application covers approximately 130,000 mostly small establishments per year across four service sectors. While substantial, this represents only a fraction of the total small business population in those sectors.[4] This raises questions about the representativeness of the resulting estimates and limits the geographical granularity of estimates that one can reliably produce.

To address these limitations, we develop a supervised prediction framework that exploits other signals of business activity (e.g., geo-located visits based on cell-phone data or transactions from payments processors) for a much larger universe of establishments of which the available payroll data is a subset. The overlap between the two datasets allows us to estimate a predictive mapping from activity signals to employment outcomes, which can then be applied to the much larger universe of establishments for which only the business activity signals are available. Aggregating the establishment-level predictions then yields employment estimates with substantially greater coverage that should be more representative and can be analyzed at greater geographic granularity than the payroll data alone.

Section 6 contains a formal description of the prediction framework including a discussion of potential caveats. Here, we provide a conceptual description of the coverage extension.

To formalize the description, it is useful to rewrite the recursive employment estimator in equation (1) in terms of individual establishment contributions. For any establishment ℓ

[4]It should be noted that CES coverage of small establishments in these industries is likely considerably smaller. See Kurmann et al. (2025).

in cell i at time t, let

$$n_{\ell,t} \;=\; e_{\ell,t} \cdot \mathbf{1}_{A_{\ell,t} \cup O_{\ell,t}}, \qquad d_{\ell,t} \;=\; e_{\ell,t-1} \cdot \mathbf{1}_{A_{\ell,t} \cup C_{\ell,t}}, \tag{6}$$

where $A_{\ell,t}$, $O_{\ell,t}$, and $C_{\ell,t}$ denote the events that establishment ℓ is active in both $t-1$ and t, opens in t, or closes in t, respectively. The composite $n_{\ell,t}$ equals current employment if ℓ contributes to the numerator of (1) and zero otherwise; $d_{\ell,t}$ equals lagged employment if ℓ contributes to the denominator and zero otherwise. For establishments in the payroll sample, these composites are computed using the opening and closing probabilities from Section 3.1. Defining $N_{i,t} = \sum_{\ell \in i} n_{\ell,t}$ and $D_{i,t} = \sum_{\ell \in i} d_{\ell,t}$, equation (1) becomes

$$\hat{E}_t \;=\; \hat{E}_{t-1} \;\times\; \frac{\sum_i \omega_i \, N_{i,t}}{\sum_i \omega_i \, D_{i,t}}. \tag{7}$$

The extrapolation problem is to estimate $N_{i,t}$ and $D_{i,t}$ for establishments not in the payroll sample using observable signals of business activity from other sources.

We formalize this as a *mass imputation* problem (Kim and Rao, 2012; Yang et al., 2021). Let $\mathcal{L}$ denote the labeled sample of establishments for which both the composite outcomes $(n_{\ell,t}, d_{\ell,t})$ and a vector of activity signals $x_{\ell,t}$ are observed, and let $\bar{\mathcal{L}}$ denote the unlabeled sample of establishments for which only $x_{\bar{\ell},t}$ is observed, with $|\bar{\mathcal{L}}| \gg |\mathcal{L}|$. We then train prediction functions $\hat{f}_n : x \mapsto \hat{n}$ and $\hat{f}_d : x \mapsto \hat{d}$ on the labeled sample by minimizing regularized empirical loss functions, with the regularization parameter and function class selected by K-fold cross-validation (Kleinberg et al., 2015; Mullainathan and Spiess, 2017). Their out-of-sample performance can be further assessed through out-of-time validation (i.e., training on an earlier period and forecasting a later one), which directly tests whether the activity-to-payroll relationship is stable enough to support real-time estimation (Ginsberg et al., 2009). Details of the estimation procedure, including the choice of $\hat{f}_n$ and $\hat{f}_d$, are deferred to Section 6.

Applying these prediction functions to the unlabeled sample yields imputed contributions for each $\bar{\ell} \in \bar{\mathcal{L}}$, which aggregate within cells to

$$\hat{N}^{\bar{\mathcal{L}}}_{i,t} \;=\; \sum_{\bar{\ell} \in i} \hat{f}_n(x_{\bar{\ell},t}), \qquad \hat{D}^{\bar{\mathcal{L}}}_{i,t} \;=\; \sum_{\bar{\ell} \in i} \hat{f}_d(x_{\bar{\ell},t-1}). \tag{8}$$

The *prediction-based employment estimator* then extends (7) to the combined labeled and

unlabeled sample:

$$\hat{E}_t^{\text{extended}} = \hat{E}_{t-1}^{\text{extended}} \times \frac{\sum_i \omega_i \left(N_{i,t}^{\mathcal{L}} + \hat{N}_{i,t}^{\bar{\mathcal{L}}} \right)}{\sum_i \omega_i \left(D_{i,t}^{\mathcal{L}} + \hat{D}_{i,t}^{\bar{\mathcal{L}}} \right)}, \tag{9}$$

where $N_{i,t}^{\mathcal{L}} = \sum_{\ell \in \mathcal{L} \cap i} n_{\ell,t}$ and $D_{i,t}^{\mathcal{L}} = \sum_{\ell \in \mathcal{L} \cap i} d_{\ell,t}$ are the directly observed labeled contributions. As for the original estimator, the reference level $\hat{E}_0^{\text{extended}}$ is initialized from administrative population data covering the target population.

Two features about (9) are worth noting. First, it preserves the recursive ratio structure of (1): unlabeled establishments enter symmetrically alongside labeled ones within each cell, and the correction for sample churn embedded in Section 3.1 implicitly carries over to $N_{i,t}^{\mathcal{L}}$ and $D_{i,t}^{\mathcal{L}}$. This carry-over, however, applies cleanly only to unlabeled establishments continuously present in the activity data across $t-1$ and t. For unlabeled entrants and exits, the activity data's own sample frame dynamics create a potential identification problem that is analogous to the one that we address in Section 3.1 for the payroll sample: it is not possible to cleanly determine from activity data alone whether an establishment's appearance or disappearance reflects a genuine business opening or closing, or merely a change in the data provider's client base. The coverage extension is therefore on stronger footing for the intensive margin—employment levels, hours, wages, and worker flows of continuously present establishments—than for the extensive margin of openings and closings. Second, consistency of the extrapolated estimator requires that the conditional expectations $\mathbb{E}[n_{\ell,t} \mid x_{\ell,t}]$ and $\mathbb{E}[d_{\ell,t} \mid x_{\ell,t}]$ be the same in the labeled and unlabeled populations, which is an assumption that may fail if selection into the payroll sample is correlated with payroll outcomes conditional on observables (Rao and Molina, 2015; Shimodaira, 2000). Section 6 discusses both of these issues and possible remedies in detail.

4 Application with Homebase Small Business Data

To illustrate how the proposed methodology performs in practice, we use data from Homebase, a scheduling and payroll software provider used primarily by small businesses. The majority of these businesses operate in service sectors with a high propensity for in-person interaction that are often most exposed to business cycles, including the disruptions and stay-at-home orders at the beginning of the COVID-19 pandemic. As such, the application with HB data represents an ideal testing ground.

We begin by describing the data and the construction of the estimation sample (Section 4.1) and then turn to the estimation of opening and closing probabilities using new

self-reported information on company creation dates and exit reasons (Section 4.2). We validate the resulting estimates against administrative birth and death rates (Section 4.3) and finally, we present employment estimates and compare against population counterparts from the QCEW (Section 4.4).

4.1 Homebase Data

The Homebase data consist of anonymized daily records of individual hours worked and wages of employees, linked longitudinally to the establishment where they work and the firm that owns the establishment. The data is recorded in real-time through HB's proprietary software and is used by many of the businesses for payroll processing. HB provides free data access to researchers and updates the data regularly with the latest observations.

The raw HB data contain between 200,000 and 250,000 distinct establishments per year; however, many do not use HB regularly. We therefore retain establishments that appear in HB for at least three consecutive weeks with at least 40 weekly hours worked across their employees. Furthermore, we restrict the analysis to establishments with fewer than 50 employees, which comprises most of HB's customers, in the following sectors where we have the most coverage: Retail Trade (NAICS 44–45), Professional and Business Services (NAICS 54-56), Leisure & Hospitality (NAICS 71 and 72), and Other Services (NAICS 81).[5] According to official statistics, establishments with fewer than 50 employees accounted for about half of all jobs across these four sectors and about 19% of all private-sector jobs prior to the pandemic. Hence, our sample represents a sizable share of total employment.

An establishment in our sample is active in a given week if any of the employees register positive hours. Weekly employment equals the sum of individuals with positive hours.[6]

[5]We use HB's internal NAICS classification, which is available starting in December 2020. The HB data also include a separate, less detailed industry classification that extends back to January 2018 (when the HB data begin) and continues past December 2020. We use this classification to impute NAICS codes prior to December 2020 by assigning, within each industry-geography cell, the most probable NAICS code as revealed by the post-December 2020 data.

[6]While in Kurmann et al. (2025) our employment counts included active HB users with untracked hours — presumably the owner or manager of an establishment — this information is not currently available in the real-time HB data, and our measure of employment is therefore restricted to hourly employees.

Table 1: Homebase sample counts

	2022		2023		2024		2025	
Mid-February base sample	81,817	(100%)	96,774	(100%)	108,493	(100%)	113,909	(100%)
Active in mid-Feb.	74,586	(91%)	88,444	(91%)	99,158	(91%)	104,170	(91%)
Temp. inactive in mid-Feb.	7,231	(9%)	8,330	(9%)	9,335	(9%)	9,739	(9%)
Exits without return	24,802	(30%)	28,766	(30%)	29,528	(27%)	31,898	(28%)
New entrants	45,046	(55%)	45,507	(47%)	38,819	(36%)	29,743	(26%)

Notes: The table shows counts of establishments from mid-February of each year to mid-February of the following year that belong to the service-sector industries studied and have fewer than 50 workers when active in mid-February or when entering Homebase.

Table 1 presents the sample used in our estimations for each year starting with the 2022 data. The sample has been growing over time. For instance, the total sample for 2022 consists of 81,817 establishments from the mid-February base sample together with the 45,046 newly entering establishments for that year, yielding a combined sample size of 126,863. In 2025, the resulting total sample size is 143,652 establishments. The underlying composition of the mid-February samples is very stable over time, with about 9% of establishments temporarily inactive in mid-February but active before and after this period.

Consistent with Kurmann et al. (2025), Table 1 highlights the substantial sample churn that characterizes the HB data. Each year, almost 30% of the sample exits HB permanently, which is far larger than yearly establishment exit rates in official population data. Similarly, entry rates in HB samples are much larger than establishment entry rates in official data.

[A few words about representativeness.]

4.2 Estimating Opening and Closing Probabilities

A novelty of our empirical analysis in this paper is that we now have access to self-reported company creation dates – indicating whether a sample entrant is a genuinely new business or an existing one that recently adopted the software – and self-reported reasons for leaving the platform – and specifically whether this is due to a permanent closure – directly within Homebase. This allows us to estimate $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ and $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ in real-time directly without relying on external source of information.

Table 2 reports the estimated probabilities averaged over the sample period. First, consistent with our earlier analysis in Kurmann et al. (2025), true openings and closings conditional on entering or exiting Homebase account for only a small share of all entries and exits. Second, there is substantial heterogeneity across sectors. For example, in Professional and Business Services, fewer than 7% of entering businesses are truly new, compared with nearly 19% in Leisure and Hospitality. Third, in Leisure and Hospitality, the probability of closing

Table 2: Opening and closing probabilities

	$\hat{p}(\mathcal{O}_{\ell,t}\|\text{entry}_{\ell,t})$	$\hat{p}(\mathcal{C}_{\ell,t}\|\text{exit}_{\ell,t})$
Retail Trade	12.2%	13.3%
Professional and Business Services	6.9%	7.3%
Leisure and Hospitality	18.5%	11.1%
Other Services	8.9%	10.3%

Notes: The table shows the average estimated probability of true opening or closing conditional on entering or exiting the Homebase sample.

conditional on exit differs markedly from the probability of opening conditional on entry. As a result, treating all entries and exits as openings and closings under the assumption that these effects offset each other would introduce substantial bias, at least in this sector.

Table 3: Birth and death adjustment factors

	$\theta^{\mathcal{O}}_{i,\mathbf{0}}$		$\theta^{\mathcal{C}}_{i,\mathbf{0}}$	
	Small	Larger	Small	Larger
Retail Trade	2.78	0.44	4.64	1.03
Professional and Business Service	7.83	1.50	9.71	3.33
Leisure and Hospitality	2.65	1.36	6.61	3.80
Other Services	4.59	0.59	9.05	1.72

Notes: The table shows the adjustment factors calibrated to match BED/BDS establishment birth and death rates, separately for smaller establishment (1-4 employees) and larger establishments combined (5-49 employees).

Next to opening and closing probabilities, we report in Table 3 the adjustment factors for births and deaths (see Section 3 for details) calibrated to replicate the BED/BDS counterparts. For simplicity, in this table we group establishments into two groups: small, with 1-4 employees, and larger establishments. The key point is that smaller establishments are typically assigned a larger adjustment factor to reproduce the much higher births and deaths rate for this category.

4.3 Validation Against Administrative Birth and Death Rates

Figure 4, panel (a), compares the new opening rates from our HB data, together with the corresponding opening rates from the combined BED/BDS data.[7] To facilitate comparisons, each weekly Homebase rate is converted to the quarterly frequency. Recall, more-

[7]Since the BED does not report statistics by establishment size class, we use data from the BDS to construct conversion factors by industry and size class that we apply to BED opening and closing rates.

over, that our adjustment factors are calibrated to make the quarterly HB rates match the BED/BDS counterparts – which is visible in the comovements of the two lines. Not surprisingly, the Homebase estimates show larger movements that are somewhat muted in the quarterly BED/BDS rates.

Figure 4, panel (a), compares the new opening rates from our HB data with the corresponding opening rates from the combined BED/BDS data.[8] To facilitate comparisons, each weekly Homebase rate is converted to the quarterly frequency. Recall, moreover, that our adjustment factors are calibrated to make the quarterly HB rates match the BED/BDS counterparts — which is visible in the comovements of the two lines. Not surprisingly, the Homebase estimates display larger movements than the more muted quarterly BED/BDS rates.

Likewise, panel (b) compares the closing rates from HB with the combined BED/BDS data. Here, the higher frequency movements from HB become very pronounced: while on average the HB line matches the BED/BDS counterpart, it appears to be below in most weeks. What is equally striking is the spike in business closings right at the onset of the COVID pandemic very well captured by Homebase. Importantly, as discussed momentarily, such closures include both permanent as temporary business closures. This feature explains why, according to the Homebase estimate, we also observe an uptick in business closures at the end of the sample period.

Figure 5 complements Figure 4 by reporting *cumulative* new opening and closing rates for selected years in our data – the most recent years, and the 2020 COVID period for comparison.

Beginning with cumulative new opening rates (panel (a)), it appears clearly that 2025 shows reduced rates of new opening, closer to 2020 rates than the higher rates of new opening that have prevailed after the COVID recession. Note that this slowdown is only partially visible in the BED/BDS data at the end of 2024 (Figure 4).

For total closing rates (panel (b) in Figure 5), by contrast, the series appears stable and broadly comparable to the rates observed in the years following the COVID recession. The figure also provides an important qualification to the spike in business closings at the onset of the pandemic reported in Figure 4: most of these closures were temporary,[9] such that the total closing rate by the end of 2020 was not significantly different from that at the end of other years in our sample.

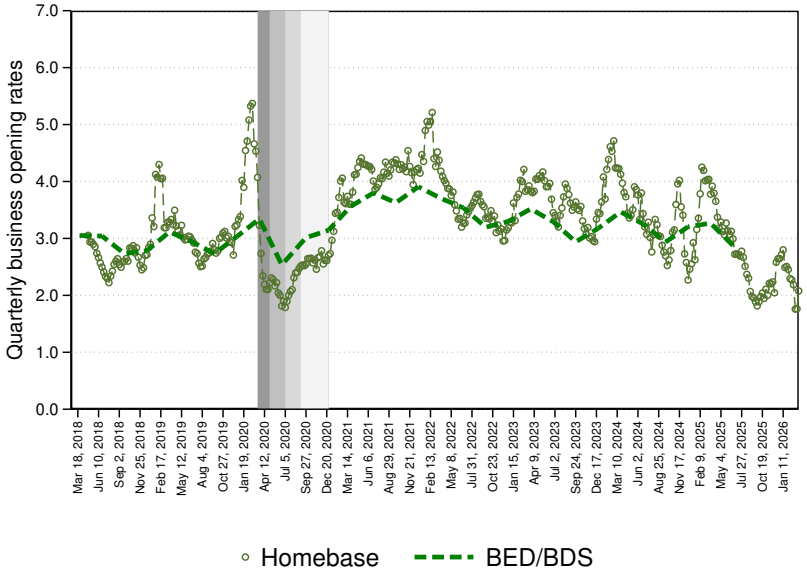
In sum, if correct, the HB estimates in Figure 5 give an early sign of a slowdown in small

[8]Since the BED does not report statistics by establishment size class, we use data from the BDS to construct conversion factors by industry and size class that we apply to BED opening and closing rates.

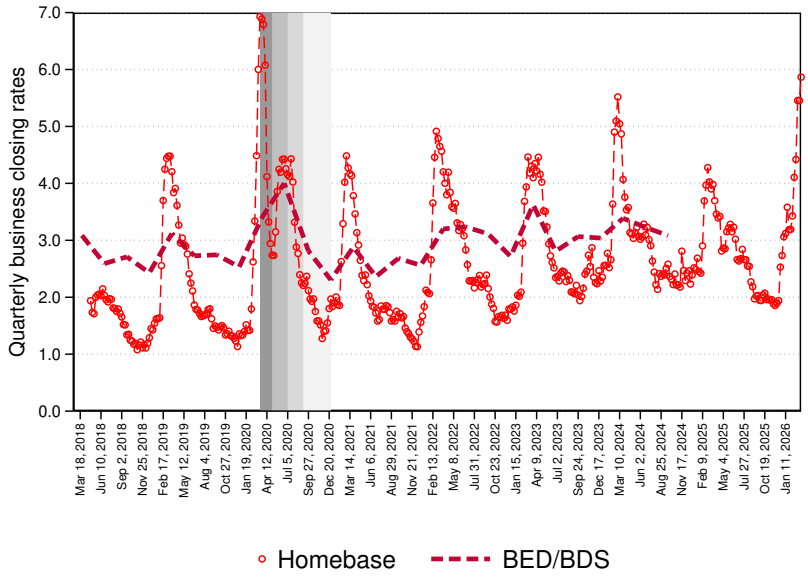
[9]We calculate that only about one third of all closings in mid-March were permanent.

Figure 4: New business opening and closing rates: Homebase vs. BED/BDS data

(a) New opening rates



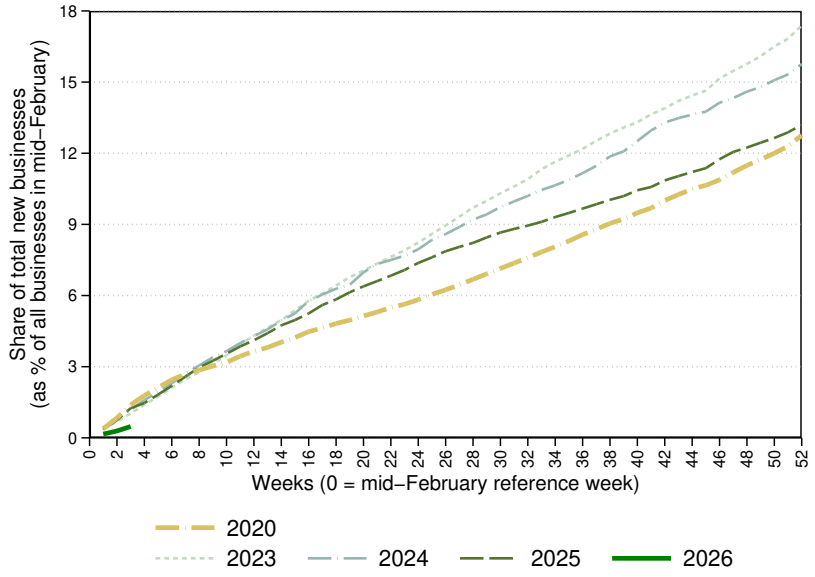
(b) Closing rates



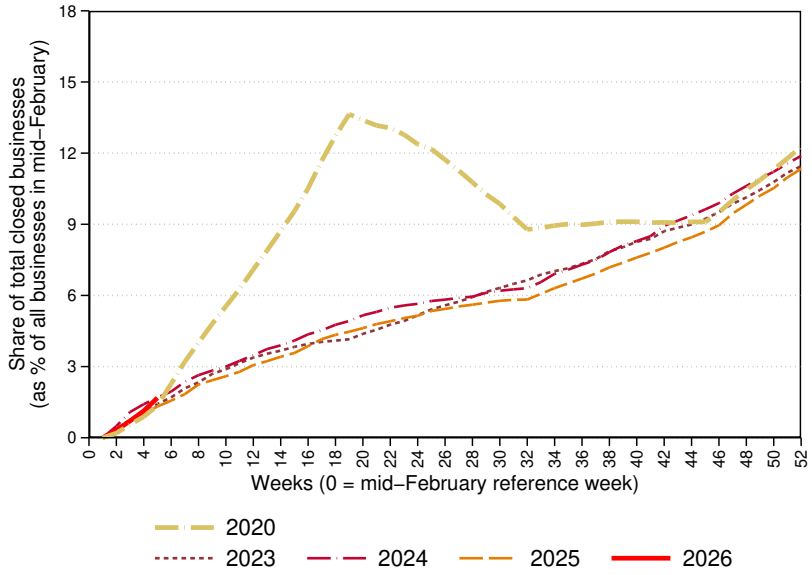
NOTES: The figure shows new opening and closing rates for small businesses for in the four sectors combined (Retail Trade, Professional and Business Services, Leisure and Hospitality, and Other Services) in Homebase and combined BED/BDS data. The Homebase weekly estimates are converted to the quarterly frequency and smoothed using a two-period, two-sided MA filter. The gray-shaded areas indicate the COVID pandemic period.

Figure 5: Cumulative new opening and closing rates

(a) Cumulative new opening rates



(b) Cumulative closing rates



NOTES: The figure shows Homebase new opening and closing rates for small businesses in the four sectors combined (Retail Trade, Professional and Business Services, Leisure and Hospitality, and Other Services). All rates are computed as a % of the total count of active businesses in mid-February (week 0) of each year.

business dynamics along the entry margin – which at the time of this writing cannot be detected in administrative data.

4.4 Small Business Employment

Figure 6 presents our weekly employment estimates separately for each four sectors – Retail Trade, Professional and Business Services, Leisure and Hospitality, and Other Services. In each plot, the red square in mid-February of each year is the QCEW employment count for small establishments. By construction of our recursive employment estimator (see equation (1)), the HB estimate for that year onward starts off from the QCEW data point.

Consider first our employment estimates for Retail Trade, starting with year 2022. Employment increases during the first two quarter, then stabilizes before dropping around the Christmas period, and remaining lower in January 2023. By the time of rebasing, i.e. mid-February 2023, the estimates comes very close to the new QCEW employment count. The same holds true for Retail Trade employment estimates in the following years.

Turning to the other sectors in Figure 6, we note that the fit to QCEW employment counts for Professional and Business Services is perfect for 2023, while there is a slight undercount for 2022 and a slight overcount for 2024. Similarly, small discrepancies arise for Leisure and Hospitality in some years. Performance is less satisfactory for Other Services, where the employment estimate overstates the next QCEW data point in each year, generating a clear discrepancy at the time of rebasing.

Overall, however, given the very large within-year variation in employment – presumably reflecting strong seasonal patterns for small establishments in these industries – the fit to the QCEW data is remarkable.

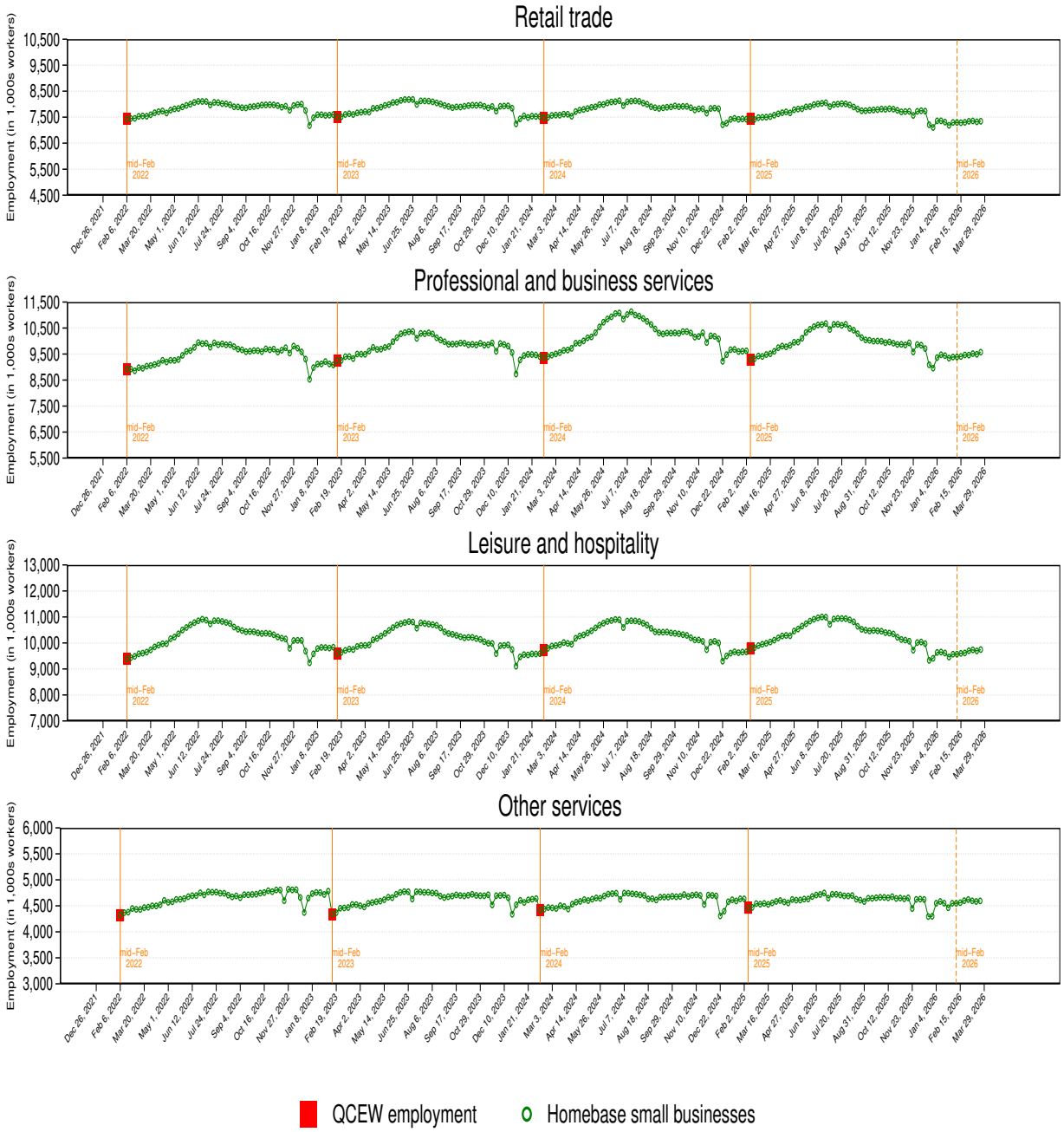
5 Other Labor Market Statistics

This section presents results on the additional labor market statistics introduced conceptually in Section 3.2: average hours per worker and hourly wages (Section 5.1), gross job creation and destruction rates (Section 5.2), and gross hiring and separation rates including recall (Section 5.3).

5.1 Hours and Wages

To estimate average hours per worker and average hourly wages, we follow Section 2.3 of Kurmann et al. (2025) and construct the measures on a balanced panel of continuously

Figure 6: Weekly employment estimates



NOTES: The figure shows weekly employment estimates for Retail Trade (NAICS 44-45), Professional and Business Services (NAICS 54-56), Leisure and Hospitality (NAICS 71-72), and Other Services (NAICS 81) from Homebase, and the corresponding annual employment estimates for small establishments from the QCEW. None of the estimates are seasonally adjusted.

active establishments. By restricting to continuously active establishments, we abstract from compositional change due to sample entry and exit and isolate the intensive-margin adjustment. Specifically, let $\mathcal{B}_{t_0,t_1}$ denote the set of establishments continuously active from period t_0 through t_1. Average hours per worker in period $t \in [t_0, t_1]$ is:

$$\hat{h}_t = \frac{\sum_i \omega_i \sum_{\ell \in \mathcal{B}_{t_0,t_1} \cap i} h_{\ell,t}}{\sum_i \omega_i \sum_{\ell \in \mathcal{B}_{t_0,t_1} \cap i} n_{\ell,t}} \tag{10}$$

where $h_{\ell,t}$ denotes total hours worked at establishment ℓ, $n_{\ell,t}$ is employment, and ω_i are sampling weights as before. Average hourly wages are computed analogously by replacing hours in the numerator with total wages paid. The length of the balanced-panel window $[t_0, t_1]$ involves a trade-off: longer windows produce a smaller balanced panel, while shorter windows limit the horizon over which dynamics can be tracked.

Figure 7 presents our Homebase estimates along with the CES estimates for comparison, where the latter comprise businesses of all sizes in the corresponding sectors. Note that the CES estimates at the beginning of the sample period serve as the reference level for the recursive formula in (10) — which effectively assumes that average weekly hours and hourly wages are independent of establishment size.

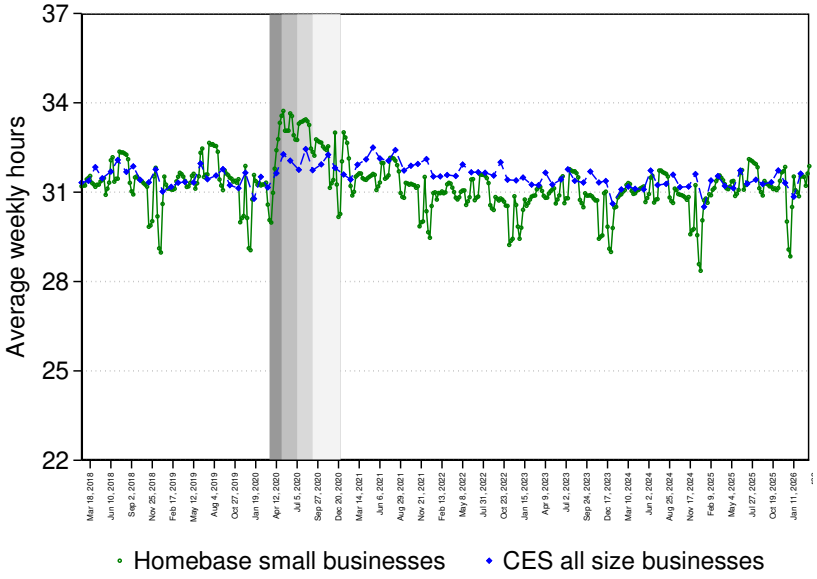
Panel (a) reveals significant week-to-week variation in hours per worker in small businesses, as well as a substantial increase in hours per worker at the beginning of the COVID pandemic period. It is worth recalling that, unlike the CES estimates (which show little change during this period), our HB estimates control for compositional changes in active establishments and in workers employed within those establishments. Turning to Panel (b), we observe something of a divergence between the HB small-business hourly wage and the CES all-size estimate after the pandemic. The trend in the two series, however, is similar, indicating that small businesses slowed wage growth at the end of the pandemic but eventually kept up with overall wage growth.

5.2 Job Creation and Destruction

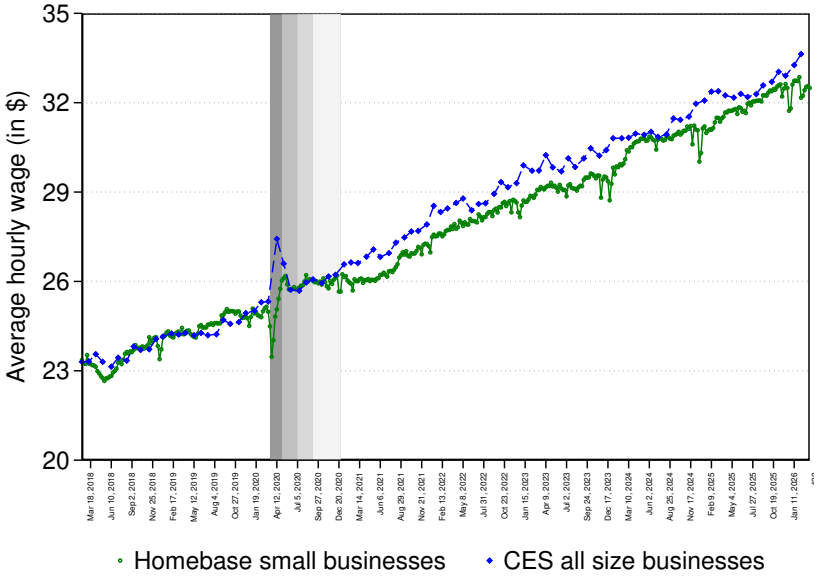
Job creation and destruction rates decompose net employment change into gross flows. For establishments active in the sample, gross job creation in period t in cell i is $JC_{i,t} = \sum_{\ell \in i} \max(\Delta e_{\ell,t}, 0)$ and gross job destruction is $JD_{i,t} = \sum_{\ell \in i} |\min(\Delta e_{\ell,t}, 0)|$, where $\Delta e_{\ell,t} = e_{\ell,t} - e_{\ell,t-1}$. Following Davis et al. (1996), we express these as rates by normalizing with the average of cell employment in t and $t-1$, which accommodates both entry and exit. The flows can be further decomposed into intensive-margin contributions (from continuing establishments) and extensive-margin contributions (from openings and closings, using the

Figure 7: Average weekly hours and hourly wages

(a) Average weekly hours



(b) Hourly wages



NOTES: The figure shows average weekly hours (panel (a)) and hourly wages (panel (b)) of employees in the four sectors combined (Retail Trade, Professional and Business Services, Leisure and Hospitality, and Other Services). The small business estimates are constructed from Homebase for employees who are continuously employed. The all size business estimates are from the CES and average weekly hours and hourly wages include all employees. None of the estimates are seasonally adjusted. The gray-shaded areas indicate the COVID pandemic period.

methodology of Section 3.1).

In Figure 5, we have documented a slowdown new establishment opening rates in 2025, and in Figure 9 below we further document a slight decline in new hiring rates that predates 2025. We find that this affects mostly employment in Leisure and hospitality. This pattern seems to line up well with data from the Job Openings and Labor Turnover Survey (JOLTS). As shown in Figure 8 (looking at net job creation rates, i.e. the difference between the new hiring and total separation rates, for which the frequency of measurement – weekly vs. monthly – does not affect the comparison), this industry has been experiencing a slight slowdown since the pandemic, while the other three industries in this plot have experienced stable rates of net job creation. Another interesting pattern in Figure 8 is that the Homebase estimates seem to exhibit a positive lead with respect to the JOLTS net job creation series.

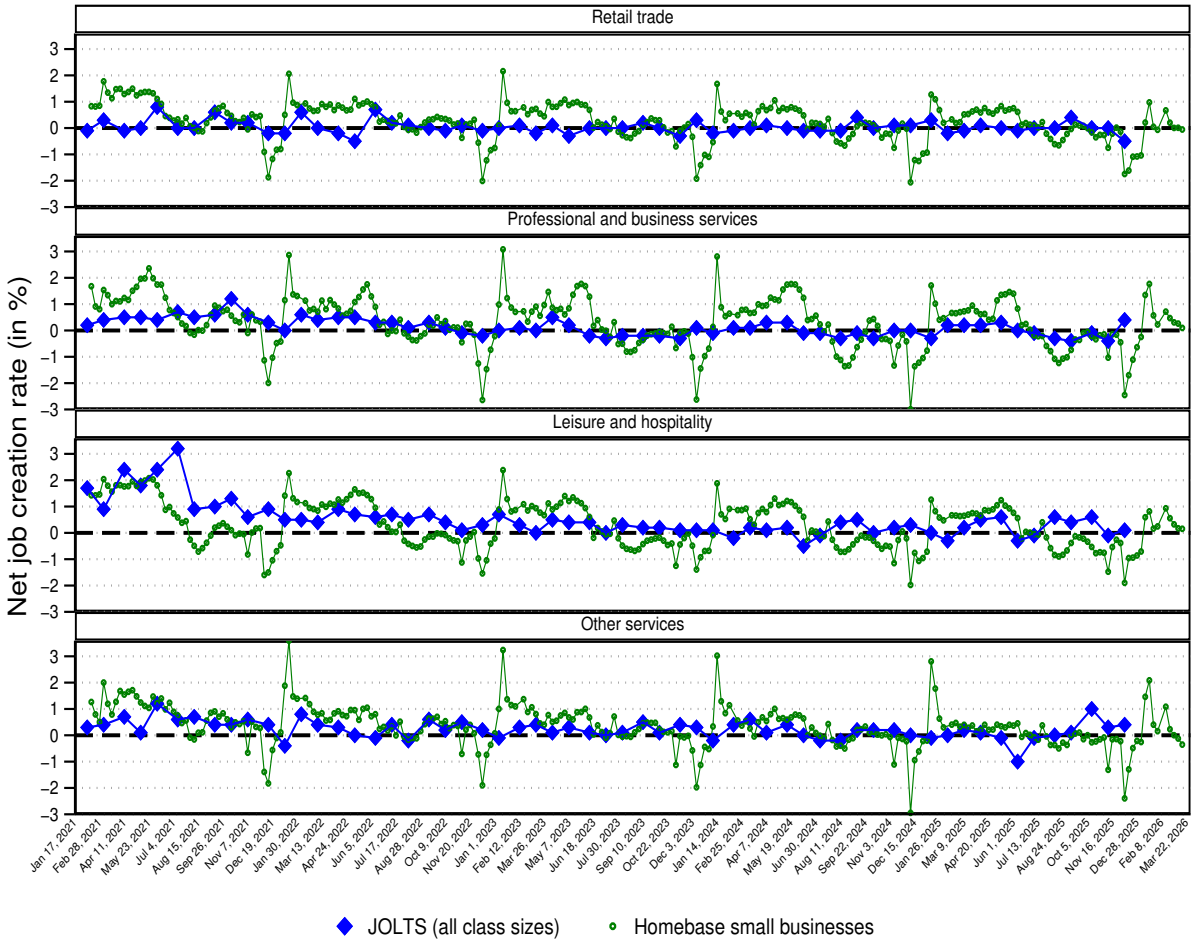
5.3 Gross Hiring, Separation, and Recall Rates

The Homebase data link individual workers longitudinally to establishments, which allows us to compute gross hiring and separation rates that are conceptually distinct from job creation and destruction. A hire occurs when a worker first appears at an establishment in period t after being absent in $t-1$; a separation when a worker present in $t-1$ no longer appears in t. The difference between aggregate gross hires and gross separations equals net employment change, but the gross flows are typically much larger because of constant within-establishment churn even in the absence of net employment change.

The worker-level panel structure also allows us to identify *recalls*: separations followed by a re-hire of the same worker at the same establishment within a specified window. Letting $H_{i,t}$ denote total hires and $H_{i,t}^{\text{recall}}$ denote hires that are recalls of previously separated workers at the same establishment, the recall rate is $H_{i,t}^{\text{recall}}/H_{i,t}$. Recalls are particularly relevant in the sectors covered by Homebase, where temporary layoffs and seasonal employment are common; treating recalled workers as new hires would overstate true worker reallocation and obscure the distinction between adjustment along an existing worker-establishment match and the formation of new matches.

Figure 9 presents new hires, recalls, job separations, and excess turnover rates in small businesses for the four sectors combined. Panel (a) shows a slight downward trend in weekly hiring rates. At the same time, recalled workers account for an increasing share of total hires since the pandemic (panel (c)), implying that rates of *new* hires have declined substantially in recent years. Total job separation rates, by contrast, are stable or only mildly decreasing, but they include separations of workers who are subsequently recalled. In light of the rise in recalls, this suggests that permanent separation rates have also declined after the pandemic.

Figure 8: Net job creation rates



NOTES: The figure shows net job creation rates for Retail trade (NAICS 44-45), Professional and business Services (NAICS 54-56), and Leisure and hospitality (NAICS 71-72) from Homebase, and the corresponding estimates from the JOLTS. The weekly Homebase estimates are smoothed using a two-period, two-sided MA filter. The JOLTS estimates are seasonally adjusted.

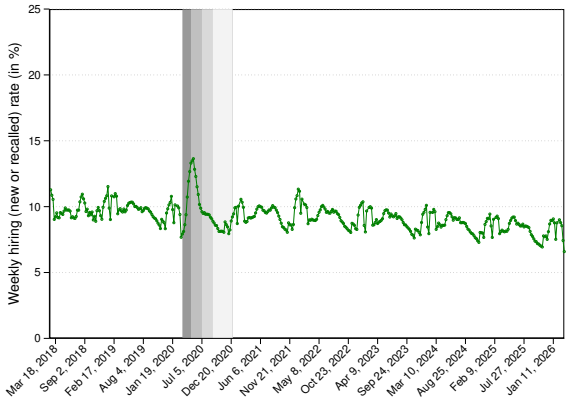
Taken together, these patterns point to a shift in worker turnover toward a greater role for recalls relative to new hires and permanent separations.

6 Prediction-based Coverage Extension

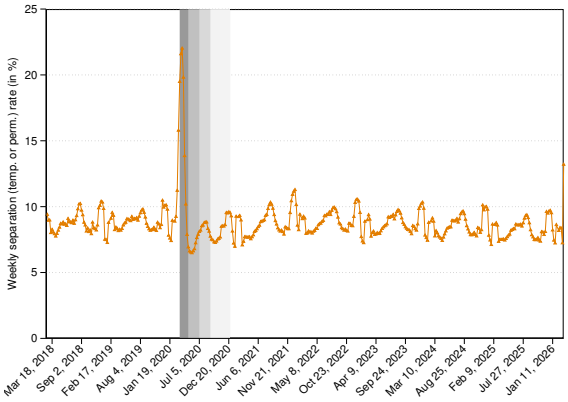
This section describes the details of how to implement the supervised prediction framework introduced conceptually in Section 3.3 and presents the resulting prediction-based estimates. Section 6.1 introduces the notation and describes the labeled and unlabeled samples. Sec-

Figure 9: New hires, recalls, job separations, and excess turnover in small businesses

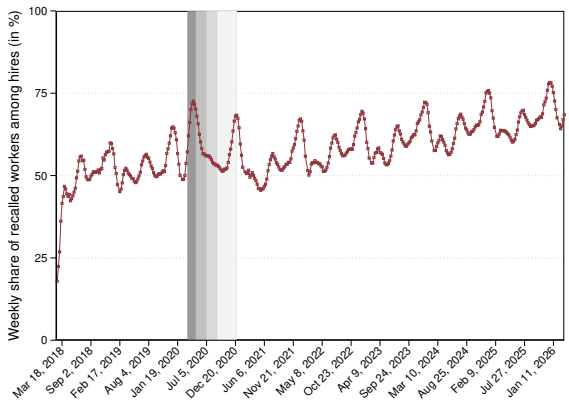
(a) Weekly rate of hires and recalls



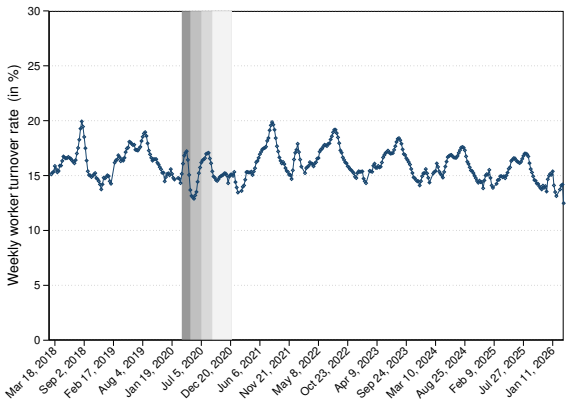
(b) Weekly rate of total job separation



(c) Share of recalls in weekly hires



(d) Weekly rate of excess turnover



NOTES: The figure shows weekly rates of new hires and recalls, job separation, share of recalls in weekly hires, and weekly excess turnover rates in the four sectors combined (Retail Trade, Professional and Business Services, Leisure and Hospitality, and Other Services). The weekly estimates are smoothed using a two-period, two-sided MA filter. The gray-shaded areas indicate the COVID pandemic period.

tion 6.2 formalizes the prediction objective. Section **??** addresses the covariate shift problem that arises when extending estimates from a selected sample to a broader population. Section **??** describes the construction of aggregate estimates from establishment-level predictions, and Section 6.5 presents the implementation, validation, and extrapolated estimates.

6.1 Setup

Section 3.3 conceptualized the prediction-based coverage extension and introduced the extended employment estimator (9), which is reproduced here for convenience:

$$\hat{E}_t^{\text{extrap}} = \hat{E}_{t-1}^{\text{extrap}} \times \frac{\sum_i \omega_i \left(N_{i,t}^{\mathcal{L}} + \hat{N}_{i,t}^{\bar{\mathcal{L}}} \right)}{\sum_i \omega_i \left(D_{i,t}^{\mathcal{L}} + \hat{D}_{i,t}^{\bar{\mathcal{L}}} \right)}, \tag{11}$$

where $N_{i,t} = \sum_{\ell \in i} n_{\ell,t}$ and $D_{i,t} = \sum_{\ell \in i} d_{\ell,t}$ denote the contributions from the labeled (observed) sample $\mathcal{L}$ of payroll establishments matched to business activity data $x_{\ell,t}$; and $\hat{N}_{i,t}^{\bar{\mathcal{L}}} = \sum_{\bar{\ell} \in i} \hat{f}_n(x_{\bar{\ell},t})$ and $\hat{D}_{i,t}^{\bar{\mathcal{L}}} = \sum_{\bar{\ell} \in i} \hat{f}_d(x_{\bar{\ell},t-1})$ denote the contributions from the unlabeled sample $\bar{\mathcal{L}}$ of establishments for which only the business activity data $x_{\bar{\ell},t}$ is available.

The goal is to train prediction functions $\hat{f}_n : x \mapsto \hat{n}$ and $\hat{f}_d : x \mapsto \hat{d}$ from the labeled sample and apply them generate predicted outcomes $\hat{f}_n(x_{\bar{\ell},t})$ and $\hat{f}_d(x_{\bar{\ell},t})$ for all $\bar{\ell} = 1, \ldots, \bar{L}$. This is a standard supervised learning prediction problem as formalized in the machine learning literature and discussed in the economics context by Kleinberg et al. (2015) and Mullainathan and Spiess (2017).

6.2 Prediction Objective and Model Selection

Each prediction function $\hat{f}_n$ and $\hat{f}_d$ is chosen to minimize expected out-of-sample prediction error. Formally, let $\mathcal{F}$ denote a class of candidate functions and let $\mathcal{L}(y, \hat{y})$ denote a loss function measuring the discrepancy between the true outcome and the prediction. For each $j \in [n, d]$, the objective is:

$$\hat{f}_j = \arg\min_{f_j \in \mathcal{F}} \mathbb{E}\big[\mathcal{L}\big(y, f_j(\mathbf{x})\big)\big], \tag{12}$$

where the expectation is over the joint distribution of $(y, \mathbf{x})$ in the target population. For continuous outcomes such as employment, the natural loss function is squared error, $\mathcal{L}(y, \hat{y}) = (y - \hat{y})^2$, and the optimal predictor is the conditional expectation $f^*(\mathbf{x}) = \mathbb{E}[y \mid \mathbf{x}]$. The choice of function class $\mathcal{F}$ determines the flexibility of the prediction model. We consider several candidates: penalized linear regression (LASSO, ridge), random forests (Athey and Imbens, 2019), and gradient-boosted regression trees.

Since the true joint distribution is unknown, $\hat{f}$ is estimated by minimizing a regularized empirical loss over the labeled sample:

$$\hat{f}_j = \arg\min_{f_j \in \mathcal{F}} \frac{1}{L} \sum_{\ell=1}^{L} \mathcal{L}\big(y_\ell, f_j(\mathbf{x}_\ell)\big) + \lambda \cdot R(f_j) \tag{13}$$

where $R(f_j)$ is a regularization penalty that controls the complexity of f_j and $\lambda > 0$ is a tuning parameter selected by K-fold cross-validation. As emphasized by Mullainathan and Spiess (2017), the regularization term distinguishes the prediction objective from standard econometric estimation: it introduces deliberate bias in exchange for reduced variance, thereby improving out-of-sample performance.

The regularization parameter λ and the choice among function classes are selected by K-fold cross-validation on $\mathcal{L}$, minimizing held-out prediction loss. We further assess temporal stability via out-of-time validation: the model is trained on an earlier sub-period of $\mathcal{L}$ and evaluated on a later one, directly testing whether the activity-to-payroll relationship is sufficiently stable to support real-time estimation (Ginsberg et al., 2009).

6.3 Sample Churn in the Unlabeled Activity Data

A complication arises for unlabeled establishments that enter or exit the activity data sample frame between $t-1$ and t. For such establishments, one of $x_{\bar{\ell},t}$ or $x_{\bar{\ell},t-1}$ is unobserved, and it is not clear whether these entries or exits reflect a genuine business opening or closing, or merely a change in coverage (e.g., because a payments processor acquires new clients or loses existing ones). This is precisely the sample churn problem that Section 3.1 addresses for the payroll sample, where self-reported creation dates and exit reasons provide intrinsic signals of true business status at the moment of entry or exit. No equivalent internal signals exist for the unlabeled activity data, and treating all entries and exits as openings and closings potentially gives rise to exactly the measurement problems that motivates the methodology of Section 3.1.

That said, the sample churn issue is likely to be a quantitatively smaller issue for the typically much larger business activity data frames, simply because these samples already cover a large portion of the population. Indeed, for continuously active unlabeled establishments, both $x_{\bar{\ell},t}$ and $x_{\bar{\ell},t-1}$ are observed, and the composite outcome estimates $\hat{f}_n$ and $\hat{f}_d$ can be applied directly. For entrants and exits, in turn, we can at least partially address the issue through two complementary approaches.

First, we can exploit the information content of activity signals themselves. If more than one independent activity signal is available (e.g., if we have access to both visits and transactions data), then we can compare the signals: if an establishment enters or exits the datasets simultaneously, respectively if the different signals comove strongly, then it is far more likely that they reflect a genuine opening or closing than churn. If instead only one activity signal is available, then we can leverage temporal patterns in activity levels around the moment of entry or exit. Suppose, for instance, that true closings in the labeled sample $\mathcal{L}$

are systematically preceded by declining visit counts or transaction volumes, while platform exits due to provider churn tend to show no such deterioration; and that conversely, true openings in the labeled sample exhibit a systematic ramp-up in activity following entry, while new client acquisitions by the data provider do not. Then we can use this information to train a status classifier on $\mathcal{L}$ and apply it entrants and exits in the unlabeled sample $\bar{\mathcal{L}}$ to estimate the probability that each represents a genuine business event. This approach directly parallels the use of creation dates and exit reasons in Section 3.1, replacing intrinsic payroll signals with learned activity-based signals. Its limitation is that the classifier is trained on Homebase businesses, whose exit dynamics may differ from churn dynamics in transaction or visits data, where provider client acquisition and attrition follow different patterns. The cross-dataset corroboration strategy is more robust but applies only to the subset of $\bar{\mathcal{L}}$ covered by multiple independent providers.

The second approach bypasses establishment-level classification entirely and instead calibrates the aggregate extensive margin contribution of the unlabeled sample to population-level benchmarks. Specifically, we can extend the cell-level adjustment factors $\theta^O_{i,t}$ and $\theta^C_{i,t}$ from Section 3.1 to the unlabeled sample, constraining the implied birth and death rates to match their administrative counterparts from the BED and BDS. This is conceptually similar to the CES birth-death model, though applied at a considerably more disaggregated level—by cell i rather than broad industry division—and using contemporaneous activity signals to distribute the calibrated totals across establishments within each cell. The approach has an important limitation, however: the adjustment factors are calibrated to administrative data that become available only with a lag, and the implicit assumption that the relationship between sample entry/exit and true business dynamics is stable over time may break down precisely when it matters most—at sharp cyclical turning points, where the composition of exits between true closings and platform churn may shift rapidly. This is analogous to the well-documented failure of the CES birth-death model around recessions (Decker, 2026), albeit at a much more detailed level of aggregation.

6.4 The Key Identifying Assumption

A key concern in applying $\hat{f}$ out-of-sample is that the labeled sample (Homebase establishments) and the unlabeled target population (all small business POIs) may be drawn from different distributions. In the machine learning literature, this is known as *covariate shift*: the marginal distribution of features differs between the training and target populations, $P_{\text{train}}(\mathbf{x}) \neq P_{\text{target}}(\mathbf{x})$, even if the conditional relationship $P(y \mid \mathbf{x})$ is the same across the two populations (Shimodaira, 2000). The extrapolated estimator (9) is consistent provided

that the conditional expectations $\mathbb{E}[n_{\ell,t} \mid x_{\ell,t}]$ and $\mathbb{E}[d_{\ell,t} \mid x_{\ell,t}]$ are the same in the labeled and unlabeled populations—that is, that the activity-to-payroll mapping $P(y \mid x)$ is stable across establishments regardless of whether they belong to the payroll sample. Differences in the marginal distribution of features $P(x)$ between $\mathcal{L}$ and $\bar{\mathcal{L}}$ do not themselves bias the predictions; as long as $P(y \mid x)$ is homogeneous, $\hat{f}$ trained on the labeled sample remains a valid predictor for the unlabeled population (Shimodaira, 2000; Yang et al., 2021). The stability of $P(y \mid x)$ is economically plausible: visit counts, dwell times, and transaction volumes are physical signals of business operation that should map to employment similarly across establishments with comparable activity profiles, regardless of which scheduling platform they use. Nonetheless, this is an assumption that is not directly testable and constitues the the primary threat to the validity of the proposed extended estimator.

We assess its plausibility in two ways. First, we test whether the activity-to-payroll relationship is stable across observable subgroups of $\mathcal{L}$—by industry, size, geography, and tenure in the payroll sample—which would be violated if selection into the payroll sample were correlated with $P(y \mid x)$. Second, we benchmark the aggregated extrapolated estimates against QCEW and BED population totals; systematic discrepancies would indicate either violations of $P(y \mid x)$ homogeneity or errors in the activity data. Where the marginal distribution of x differs substantially between $\mathcal{L}$ and $\bar{\mathcal{L}}$, we additionally apply importance reweighting (Shimodaira, 2000) as a variance-reduction device, improving prediction precision in regions of feature space where the labeled sample is sparse relative to the target population.

6.5 Implementation, Validation, and Results

Work in progress...

7 Conclusion

Work in progress...

References

Athey, Susan and Guido W. Imbens, "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 2019, *11*, 685–725.

Bartik, Alexander W., Marianne Bertrand, Feng Lin, Jesse Rothstein, and Matt Unrath, "Measuring the Labor Market at the Onset of the COVID-19 Crisis: Evidence from Traditional and Non-Traditional Data," *Brookings Papers on Economic Activity, Special Edition: COVID-19 and the Economy*, 2020, pp. 239–268.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, Adrian Hamins-Puertolas, and Christopher J. Kurz, "Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data," in "Big Data for Twenty-First-Century Economic Statistics," Vol. 79 of *NBER Studies in Income and Wealth*, University of Chicago Press, 2022.

_ , _ , _ , John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz, "The U.S. Labor Market during the Beginning of the Pandemic Recession," *Brookings Papers on Economic Activity, Special Edition: COVID-19 and the Economy*, 2020, pp. 3–33.

Chetty, Raj, John N. Friedman, and Michael Stepner, "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data," *Quarterly Journal of Economics*, 2024, *139* (2), 829–889.

Crane, Leland D., Ryan A. Decker, Aaron Flaaen, Adrian Hamins-Puertolas, and Christopher Kurz, "Business Exit During the COVID-19 Pandemic: Non-Traditional Measures in Historical Context," *Journal of Macroeconomics*, 2022, *72*, 103419.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, Cambridge, MA: MIT Press, 1996.

Decker, Ryan A., "Measuring Payroll Employment: A Note on the Current Employment Statistics," Working Paper 34924, National Bureau of Economic Research 2026.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, 2009, *457*, 1012–1014.

Haltom, Nadezhda, "Payroll Employment Data: Measuring the Effects of Annual Benchmark Revisions," *Federal Reserve Bank of Richmond Economic Quarterly*, 2005, *91* (1).

Kim, Jae Kwang and Jon N. K. Rao, "Combining Data from Two Independent Surveys: A Model-Assisted Approach," *Biometrika*, 2012, *99* (1), 85–100.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction Policy Problems," *American Economic Review*, 2015, *105* (5), 491–495.

Kurmann, André, Etienne Lalé, and Lien Ta, "Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data," *Journal of Public Economics*, 2025, *250*, 105477.

Mueller, Kirk, "Impact of Business Births and Deaths in the Payroll Survey," *Monthly Labor Review*, 2006, *129* (5), 28–34.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

Phillips, Keith R. and Jamie Nordlund, "The Efficiency of the Benchmark Revisions to the Current Employment Statistics (CES) Data," *Economics Letters*, 2012, *115* (3), 431–434.

Quinlan, Ryan and Roberto Pinheiro, "BLS Payroll Revisions: Forecasting Recessions," *Economics Letters*, 2026, *261*, 112859.

Rao, Jon N. K. and Isabel Molina, *Small Area Estimation*, 2nd ed., Hoboken, NJ: Wiley, 2015.

Shimodaira, Hidetoshi, "Improving Predictive Inference Under Covariate Shift by Weighting the Log-Likelihood Function," *Journal of Statistical Planning and Inference*, 2000, *90* (2), 227–244.

Yang, Shu, Jae Kwang Kim, and Youngdeok Hwang, "Integration of Data from Probability Surveys and Big Found Data for Finite Population Inference Using Mass Imputation," *Survey Methodology*, 2021, *47* (1), 29–58.