

Branching fixed effects: A proposal for communicating uncertainty

Patrick Kline*

UC Berkeley

August 19, 2025

Abstract

Economists often rely on estimates of linear fixed effects models developed by other teams of researchers. Assessing the uncertainty in these estimates can be challenging. I propose a form of sample splitting for network data that breaks two-way fixed effects estimates into statistically independent “branches,” each of which provides an unbiased estimate of the parameters of interest. These branches facilitate uncertainty quantification, moment estimation, and shrinkage. Algorithms are developed for efficiently extracting branches from large datasets. I illustrate these techniques using a benchmark dataset from Veneto, Italy that has been widely used to study firm wage effects.

*This paper was prepared for the Econometric Society World Congress 2025. I thank Kevin Chen, Raffaele Saggio, Andres Santos, and Chris Walters for helpful comments. This paper makes use of the Veneto Work Histories dataset developed by the Economics Department in Università Ca’ Foscari Venezia under the supervision of Giuseppe Tattara.

1 Introduction

Fixed effects methods have emerged as an important tool for scientific communication in empirical economics, allowing researchers to summarize complex empirical patterns found in vast administrative databases with minimal loss of fidelity. A single research team can rarely envision, much less exploit, all of the potential uses of such granular summaries. It has therefore become routine for estimates to be publicly posted online or hosted by statistical agencies for use by other researchers.

For instance, the Opportunity Insights (OI) group provides public estimates of intergenerational mobility at the census tract level derived from fitting complex fixed effects models to US tax records (Chetty and Hendren, 2018). Likewise, estimates of firm wage fixed effects derived from German social security records have been developed for secondary use by research teams in partnership with the Institute for Employment Research (Card, Heining, and Kline, 2015; Bellmann et al., 2020). Estimates enabling secondary analysis of teacher and school value added have also been shared between research teams (e.g., Chamberlain, 2013).

In each of these literatures, the workhorse models are linear and involve two (or more) high-dimensional sets of fixed effects. Least squares estimates of these parameters are unbiased but will inevitably be noisy. Unfortunately, the properties of this noise are often poorly documented. It is typically infeasible to release full covariance matrices for more than a few thousand estimated effects, as the number of distinct entries in such matrices grows quadratically. Moreover, when the number of parameters being estimated is proportional to the underlying sample size, the noise in the estimates may not be normally distributed, in which case variance matrices provide incomplete guides to uncertainty.

This ambiguity regarding the stochastic properties of published fixed effects estimates presents difficulties for secondary analysis. For example, Mogstad et al. (2024) note that lack of information on the covariance of the quasi-experimental place fixed effects developed by Chetty and Hendren (2018) is one factor that leads them to focus their analysis on simpler cross-sectional estimates of intergenerational mobility. In contrast to the OI place effects, estimates of firm wage fixed effects are almost always provided without standard errors. As discussed in Kline (2025), researchers often treat these estimated firm effects as outcomes in second-step regressions, reporting (downstream) standard errors that may be severely biased by the implicit assumption that the estimated effects are mutually independent.

This paper proposes breaking two-way fixed effects estimates into statistically independent “branches” as a means of transparently communicating uncertainty. Each branch corresponds to a distinct subsample of the micro-data within which unbiased estimates of all target parameters can be constructed. Building on a graph-theoretic interpretation of two-way fixed effects models developed in Kline (2025), the branches are constructed from edge-disjoint spanning trees of the “mobility network” – a graph representing the evolution

of group memberships in the data. Though I will focus on a setting where this network captures the structure of worker moves between firms, the spanning tree construction applies broadly to contexts involving two-dimensional heterogeneity, including student-teacher, patient-doctor, and judge-district pairings. Edges not included in any spanning tree are appended to the final branch, which ensures the full-sample fixed effects estimator decomposes linearly into branch-specific estimates.

Publicly releasing branches allows outside researchers to transparently assess uncertainty in published fixed effect estimates and in projections of those estimates onto observable covariates. Branches also simplify more advanced tasks such as estimating moments of the fixed effects and shrinkage, tasks that have proved challenging to extend to high-dimensional fixed effects settings with heteroscedasticity (Kwon, 2023; Cheng, Ho, and Schorfheide, 2025; He and Robin, 2025). I illustrate these ideas using as an example the estimation of firm fixed effects in a benchmark dataset from the Italian province of Veneto.

The idea of releasing branches of fixed effects has several precedents in the literature. First, it mirrors the common practice among statistical agencies of releasing “replicates” to assess uncertainty in published estimates derived from surveys. For example, the US Bureau of Labor Statistics uses two independent replicates of inflation measures in each geographic area to evaluate the sampling variance of inflation measures reported in the Consumer Price Index (U.S. Bureau of Labor Statistics, 2025). Likewise, the American Community Survey uses 80 replicates to assess margins of error in published estimates (U.S. Census Bureau, 2025). While replicates are designed to be independent and identically distributed, the branched estimates considered here will generally not be identically distributed, as different subsamples will tend to exhibit different noise levels.

A second precedent comes from recent empirical work with two-way fixed effects models exploiting random splits of the micro-data. Sorkin (2018), Goldschmidt and Schmieder (2017), Drenik et al. (2023), and Jäger et al. (2024) all randomly split worker-firm datasets into half-samples in order to obtain independent estimates of the subset of firm fixed effects identified in both samples. Likewise, Silver (2021) uses a split-sample approach to estimate the variance of physician value added, while Card, Rothstein, and Yi (2024) use a similar approach to estimate the variance of industry wage effects. The popularity of random splitting in these contexts derives from the simplicity of the covariance-based methods that can be employed to account for estimation error. However, an important limitation of random splitting approaches is that cross-split covariances can only be computed among the random subset of parameters for which multiple estimates turn out to be available.

Rather than consider a random estimand, the branching approach deterministically partitions the data in a way that preserves the estimability of all target parameters in each split. The partitioning strategy builds on results from the graph theory literature on so-called “tree packing” problems (Nash-Williams, 1961;

Tutte, 1961; Roskind and Tarjan, 1985). First, the micro-data are pruned to the largest k -edge-connected-component (k -ECC): a subgraph that remains connected if any k edges are removed. In the worker-firm setting, restricting to a k -ECC with $k > 1$ is a natural refinement of the ubiquitous practice of pruning to the largest connected component (Abowd, Creedy, and Kramarz, 2002). After pruning, the k -ECC is “packed” with edge-disjoint spanning trees. These trees are then used to build branches that fully partition the micro-data.

In addition to yielding a deterministic estimand, this “prune and pack” approach ensures that each fixed effect corresponding to a node in the k -ECC has as many independent estimates as there are branches of the graph. I show that the pooled two-way fixed effects estimator can be written as a linear combination of branch-specific estimates. These linear combinations can be released alongside the branch-specific estimates to simplify the quantification of uncertainty in the full-sample estimates.

The Veneto data provide a challenging test case for the branching approach, as many firms are connected by a single worker move. In this low connectivity environment, constructing even two branches requires limiting attention to 39% of the firms in the largest connected set. However, the estimates turn out to be remarkably insensitive to further pruning of the sample despite the fact that larger firms tend to be better connected. Among other interesting findings, the analysis reveals that the distribution of firm wage effects in these pruned samples are skew left and exhibit heavy tails. In other empirical literatures where the relevant graphs tend to exhibit higher edge-connectivity – e.g., work on teacher value added or models of place effects – it may be feasible to extract dozens of branches without meaningfully narrowing the scope of investigation.

The rest of the paper is structured as follows. The next section previews the basic properties of branches and describes their potential uses in greater detail. Section 3 reviews the algebra of two-way fixed effects estimators. Section 4 introduces a graph-theoretic interpretation of the model. Section 5 defines branches algebraically and derives a representation of the fixed effects estimator as a linear combination of branch-specific estimators. Section 6 reviews the theory underlying tree packing problems and outlines the prune and pack algorithm. Section 7 illustrates the use of branches for quantifying uncertainty, moment estimation, and shrinkage via exercises involving firm wage fixed effects derived from administrative data in Veneto, Italy.

2 The blessings of branches

Before getting into the weeds of how to build branches, it is useful to preview how they can simplify many modern estimation tasks. Suppose that we are interested in some $J \times 1$ parameter vector ψ and have constructed a least squares estimator $\hat{\psi} \in \mathbb{R}^J$ of ψ that is unbiased.

Branches are formed by partitioning the micro-data in a way that allows the construction of M mutually

independent estimates $\{\hat{\psi}_b\}_{b=1}^M$, each of which obeys $\mathbb{E}[\hat{\psi}_b] = \psi$. While each branch has the same mean, their sampling distributions may differ. In particular, we do not assume that the $J \times J$ variance matrix $\mathbb{V}[\hat{\psi}_b]$ is diagonal or identical across branches.

Another important property of branches is that the full-sample estimator can be decomposed as the sum

$$\hat{\psi} = \sum_{b=1}^M \mathbf{C}_b \hat{\psi}_b \equiv \sum_{b=1}^M \varphi_b, \quad (1)$$

where $\mathbf{C}_b$ is a known $J \times J$ matrix that obeys $\sum_{b=1}^M \mathbf{C}_b = \mathbf{I}$. Note that each element of $\varphi_b \in \mathbb{R}^J$ is a linear combination of the unbiased estimates in $\hat{\psi}_b \in \mathbb{R}^J$. Thus, the full-sample fixed effects estimator effectively utilizes each branch to estimate a different linear combination of ψ .

Finally, note that both $\hat{\psi}_b$ and φ_b are $J \times 1$ vectors that can be stored as columns in a spreadsheet alongside the full-sample estimate $\hat{\psi}$. As detailed in Section 5, after the data have been partitioned into branches, computing each φ_b is no more difficult than computing $\hat{\psi}$. By the decomposition property in (1), releasing the full-sample estimate is redundant once columns corresponding to the $\{\varphi_b\}_{b=1}^M$ vectors have been provided.

Having outlined what branches are, I now preview how they can facilitate three common empirical tasks:

1. Quantifying uncertainty

Since the branches are independent, the $J \times J$ covariance matrix of $\hat{\psi}$ can be written

$$\mathbb{V}[\hat{\psi}] = \sum_{b=1}^M \mathbb{V}[\varphi_b] = \sum_{b=1}^M \mathbf{C}_b \mathbb{V}[\hat{\psi}_b] \mathbf{C}_b' \equiv \Sigma.$$

Independence also guarantees that for any two branches b and ℓ ,

$$\mathbb{E}[(\varphi_b - \varphi_\ell)(\varphi_b - \varphi_\ell)'] = \mathbf{C}_b \mathbb{V}[\hat{\psi}_b] \mathbf{C}_b' + \mathbf{C}_\ell \mathbb{V}[\hat{\psi}_\ell] \mathbf{C}_\ell'.$$

Summing across all $\binom{M}{2} = M(M-1)/2$ pairs of branches, and rescaling to account for the fact that each branch appears exactly $M-1$ times, yields the variance estimator

$$\hat{\Sigma} = \frac{1}{M-1} \sum_{b=1}^M \sum_{\ell < b} (\varphi_b - \varphi_\ell)(\varphi_b - \varphi_\ell)'$$

While this estimator is unbiased ($\mathbb{E}[\hat{\Sigma}] = \Sigma$), the individual entries in $\hat{\Sigma}$ will tend to be imprecise when M is small. Fortunately, interest often centers on low-dimensional quadratic functions of Σ , which are easier to estimate than any particular entry in this matrix.

Suppose, for example, that one wishes to estimate the coefficient vector γ from a projection of ψ onto the column space of a matrix of covariates $\mathbf{X}$. Treating $\mathbf{X}$ as fixed and assuming that $\mathbf{S}_{xx} = \mathbf{X}'\mathbf{X}$ is full rank, we can write the estimand $\gamma = \mathbf{S}_{xx}^{-1}\mathbf{X}'\psi$ and the corresponding estimator $\hat{\gamma} = \mathbf{S}_{xx}^{-1}\mathbf{X}'\hat{\psi}$. It follows that $\mathbb{E}[\hat{\gamma}] = \gamma$ and $\mathbb{V}[\hat{\gamma}] = \mathbf{S}_{xx}^{-1}\mathbf{X}'\Sigma\mathbf{X}\mathbf{S}_{xx}^{-1}$. Hence, an unbiased estimate of the variance of a second-step linear projection is $\hat{\mathbb{V}}[\hat{\gamma}] = \mathbf{S}_{xx}^{-1}\mathbf{X}'\hat{\Sigma}\mathbf{X}\mathbf{S}_{xx}^{-1}$. The entries of $\hat{\mathbb{V}}[\hat{\gamma}]$ will tend to be significantly more precise than the entries of $\hat{\Sigma}$ when J is large relative to $\dim(\mathbf{S}_{xx})$.¹

2. Moment estimation

Let $\odot$ denote the element-wise product operator. For any two branches b and ℓ , independence implies $\mathbb{E}[\hat{\psi}_b \odot \hat{\psi}_\ell] = \psi_b \odot \psi_\ell$. Hence, an unbiased estimate of the average squared entry in ψ (i.e., the second uncentered moment of ψ) is $\frac{1}{J}\mathbf{1}'\left(\hat{\psi}_b \odot \hat{\psi}_\ell\right)$, where $\mathbf{1}$ is a $J \times 1$ vector of ones. A more precise estimator can be had by averaging across all $\binom{M}{2}$ distinct pairs of branches:

$$\frac{1}{J}\mathbf{1}'\left[\frac{2}{M(M-1)}\sum_{b=1}^M\sum_{\ell < b}\hat{\psi}_b \odot \hat{\psi}_\ell\right].$$

Likewise, third moments can be estimated by averaging over all $\binom{M}{3}$ distinct triples of branches:

$$\frac{1}{J}\mathbf{1}'\left(\binom{M}{3}^{-1}\sum_{b_1=1}^M\sum_{b_2 < b_1}\sum_{b_3 < b_2}\hat{\psi}_{b_1} \odot \hat{\psi}_{b_2} \odot \hat{\psi}_{b_3}\right).$$

By induction, moments up to order M can be estimated via leveraging moment conditions of the form:

$$\mathbb{E}\left[\hat{\psi}_1 \odot \hat{\psi}_2 \odot \cdots \odot \hat{\psi}_M\right] = \psi_1 \odot \psi_2 \odot \cdots \odot \psi_M.$$

3. Shrinkage

Standard empirical bayes shrinkage arguments are predicated on the assumption that the distribution of noise is known ex-ante (Walters, 2024). A recent paper by Ignatiadis et al. (2023) proposes a best predictor based on independent and identically distributed replicates that adapts to the unknown noise distribution. I show that important insights from this paper carry over to the problem of predicting the elements of ψ given independent branch estimates $\left\{\hat{\psi}_b\right\}_{b=1}^M$ that are not identically distributed. The proposed procedure consists of running a series of regressions of each entry of $\hat{\psi}_b$ against the values of that entry in the other branches. This process is repeated for each choice of b . Finally, the predicted values, which generally shrink the noisy branch-specific estimates, are averaged.

¹Assume that $\lambda_{\min}(\mathbf{S}_{xx}) > \kappa > 0$, where $\lambda_{\min}(\mathbf{S}_{xx})$ gives the smallest eigenvalue of $\mathbf{S}_{xx}$. Then, as $J \rightarrow \infty$ (with $\dim(\mathbf{S}_{xx})$ fixed), the noise in $\hat{\mathbb{V}}[\hat{\gamma}]$ will become negligible relative to the noise in $\hat{\Sigma}$.

3 The algebra of two-way fixed effects

Constructing branches involves carefully splitting the sample so as to preserve estimability of the model. It is useful to review the basic algebra of two-way fixed effects estimators using as our running example the worker-firm setting of Abowd, Kramarz, and Margolis (1999), henceforth “the AKM model.” The treatment here will depart minimally from the setup in Kline (2025). Suppose we have N workers and J firms. For simplicity, we assume there are only 2 time periods, that all workers switch employers between these periods, and that there are no time varying covariates.² The AKM model can be written:

$$y_{it} = \alpha_i + \psi_{\mathbf{j}(i,t)} + \varepsilon_{it}$$

where y_{it} is the log wage of worker $i \in \{1, 2, \dots, N\} \equiv [N]$ in period $t \in \{1, 2\}$ and the function $\mathbf{j} : [N] \times \{1, 2\} \rightarrow \{1, 2, \dots, J\} \equiv [J]$ gives the identity of the firm that worker i is paired with in period t . The $\{\alpha_i\}_{i=1}^N$ are *person effects* that can be ported from one employer to another, whereas the $\{\psi_j\}_{j=1}^J$ capture *firm effects* that must be forfeited when leaving employer j . Both the person and firm effects are parameters – i.e., they are “fixed effects.” In contrast, the time-varying errors $\{\varepsilon_{it}\}_{i=1, t=1}^{N, 2}$ are stochastic. Following the convention in the literature, the errors are assumed to be mutually independent and to each have mean zero, which amounts to a strict exogeneity requirement.

Interest often centers on the firm effects, which, under assumptions described in Kline (2025), can be thought of as capturing average treatment effects on wages of moving between particular firm pairs. Adopting this perspective, we can eliminate the person effects with a first differencing transformation

$$y_{i2} - y_{i1} = \psi_{\mathbf{j}(i,2)} - \psi_{\mathbf{j}(i,1)} + \varepsilon_{i2} - \varepsilon_{i1}. \quad (2)$$

Thus, each worker’s wage change offers a noisy estimate of the difference in firm effects between an origin firm $\mathbf{j}(i, 1)$ and a destination firm $\mathbf{j}(i, 2)$. Denote the set of origin-destination pairs traversed by workers between the two periods as

$$\mathcal{P} = \left\{ (o, d) \in [J]^2 : \mathbf{j}(i, 1) = o, \mathbf{j}(i, 2) = d \text{ for some } i \in [N] \text{ and } o \neq d \right\}.$$

The number of ordered pairs in this set is $|\mathcal{P}|$ and we denote the p ’th pair in the set by (o_p, d_p) .

²Adjustments for covariates can be conducted in a first step, in which case the relevant outcomes, y_{it} , becomes an estimated residual. The AKM model has no implications for wage dynamics within a worker-firm match. Therefore, when additional time periods are available, nothing is lost by collapsing y_{it} down to worker-firm match means.

3.1 The least squares estimator

We can now rewrite (2) in matrix notation as

$$\begin{aligned} \mathbf{y}_2 - \mathbf{y}_1 &= (\mathbf{F}_2 - \mathbf{F}_1) \psi + \boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1 \\ &\equiv \mathbf{P}\mathbf{B}'\psi + \mathbf{u}, \end{aligned} \tag{3}$$

where $\mathbf{F}_t$ is an $N \times J$ matrix of firm indicators with i 'th row and j 'th column entry $1\{\mathbf{j}(i, t) = j\}$, ψ is a $J \times 1$ vector of firm effects, and $\boldsymbol{\varepsilon}_2$, $\boldsymbol{\varepsilon}_1$, and $\mathbf{u} = \boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1$ are each $N \times 1$ vectors of errors.

The $J \times |\mathcal{P}|$ matrix $\mathbf{B}$ is known in the graph theory literature as the (signed) *incidence matrix*. Each row of $\mathbf{B}$ corresponds to a particular firm, while each column corresponds to an ordered pair p . The entry in the j 'th row and p 'th column of $\mathbf{B}$ can be written $1\{d_p = j\} - 1\{o_p = j\}$. Note that each column of $\mathbf{B}$ has exactly two non-zero entries, one of which takes the value -1 , representing departure from some origin firm, and one of which takes the value $+1$, representing arrival at a destination firm.

The $N \times |\mathcal{P}|$ matrix $\mathbf{P}$ is comprised of origin-destination pair indicators. The entry in the i 'th row and p 'th column of $\mathbf{P}$ can be written $1\{\mathbf{j}(i, 1) = o_p\} \cdot 1\{\mathbf{j}(i, 2) = d_p\}$. Thus,

$$\hat{\Delta} = (\mathbf{P}'\mathbf{P})^{-1} \mathbf{P}'(\mathbf{y}_2 - \mathbf{y}_1)$$

gives the $|\mathcal{P}| \times 1$ vector of mean wage changes between each origin-destination pair. Note that $\mathbf{P}'\mathbf{P} \equiv \mathbf{W}$ is a diagonal $|\mathcal{P}| \times |\mathcal{P}|$ matrix giving the number of workers who move from each origin to each destination firm. Hence, $\mathbf{P}'(\mathbf{y}_2 - \mathbf{y}_1) = \mathbf{W}\hat{\Delta}$.

The identity $\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{P}\mathbf{B}'$ provides a useful way to separate the sorts of moves present in the data from how many workers move between each firm pair. Premultiplying the system in (3) by $\mathbf{B}\mathbf{P}'$ yields

$$\mathbf{B}\mathbf{W}\hat{\Delta} = \mathbf{B}\mathbf{W}\mathbf{B}'\psi + \mathbf{B}\mathbf{P}'\mathbf{u}.$$

Strict exogeneity of the errors implies $\mathbb{E}[\mathbf{u} \mid \mathbf{B}, \mathbf{P}] = 0$. Hence, we have the moment condition

$$\mathbb{E}[\mathbf{B}\mathbf{W}\hat{\Delta}] = \mathbf{B}\mathbf{W}\mathbf{B}'\psi \equiv \mathbf{L}\psi.$$

The $J \times J$ matrix $\mathbf{L} = \mathbf{B}\mathbf{W}\mathbf{B}'$ is known in graph theory as the (weighted) *Laplacian matrix*. When the mobility network is *connected* – a concept we will review in more detail below – the Laplacian will have rank $J - 1$, implying ψ is identified up to a constant. It is common to resolve this indeterminacy by normalizing

one of the firm effects to zero. We will do so by setting the first entry of ψ to zero, which yields the reduced system

$$\mathbb{E} \left[\mathbf{B}_{(1)} \mathbf{W} \hat{\Delta} \right] = \mathbf{L}_{(1)} \psi_{(1)},$$

where $\mathbf{B}_{(1)}$ is the submatrix generated by removing the first row from $\mathbf{B}$, $\mathbf{L}_{(1)} = \mathbf{B}_{(1)} \mathbf{W} \mathbf{B}_{(1)}'$, and $\psi_{(1)}$ is a vector of length $J - 1$ formed by removing the first entry from ψ . Solving this reduced system for $\psi_{(1)}$ yields the least squares estimator

$$\hat{\psi}_{(1)} = \mathbf{L}_{(1)}^{-1} \mathbf{B}_{(1)} \mathbf{W} \hat{\Delta}. \quad (4)$$

Thus, the estimated firm effects are a Laplacian normalized linear combination of the average wage changes of movers along all origin-destination pairs.

3.2 Pooling data on firm pairs

When worker moves are present in both directions between a pair of firms, some columns of the incidence matrix $\mathbf{B}$ will be mirror images of each other. In such cases, it is possible to further simplify (4) by expressing the estimated firm effects as a linear combination of oriented average wage changes $\vec{\Delta}$ that difference the entries of $\hat{\Delta}$ along opposite directions of worker flow. In addition to simplifying computation of $\hat{\psi}$, this pooled representation will provide a foundation for the next section, which develops an interpretation of the AKM model as an undirected graph. As detailed in Section 5, each branch-specific estimate $\hat{\psi}_b$ will ultimately correspond to a linear combination of a mutually exclusive subset of the entries of $\vec{\Delta}$.

To illustrate the basic idea, suppose that our data only measure moves between two firms. In such a case, we can write $\hat{\Delta} = (\hat{\Delta}_+, \hat{\Delta}_-)'$, where $\hat{\Delta}_+$ is the average wage change of workers moving from firm 2 to firm 1, and $\hat{\Delta}_-$ is the average wage change of workers moving from firm 1 to firm 2. Since $\mathbb{E} [\hat{\Delta}_+] = \psi_1 - \psi_2$ and $\mathbb{E} [\hat{\Delta}_-] = \psi_2 - \psi_1$, it is natural to pool this information. Letting n_+ be the number of movers from firm 2 to firm 1 and n_- the number of movers from firm 1 to firm 2, we can define $\vec{\Delta} = (n_+ \hat{\Delta}_+ - n_- \hat{\Delta}_-) / (n_+ + n_-)$, which provides an unbiased estimate of $\psi_1 - \psi_2$. This mover-weighted average provides an efficient pooling of the information in $\hat{\Delta}_+$ and $\hat{\Delta}_-$ under the auxiliary assumption that the worker level errors $\mathbf{u}$ are homoscedastic.

When $Q \leq |\mathcal{P}|$ firm pairs have worker flows in both directions, the incidence matrix can be partitioned as

$$\underbrace{\mathbf{B}}_{J \times |\mathcal{P}|} = \left[\underbrace{\mathbf{D}}_{J \times Q}, \underbrace{-\mathbf{D}}_{J \times Q}, \underbrace{\mathbf{R}}_{J \times (|\mathcal{P}| - 2Q)} \right],$$

where the matrices $(\mathbf{D}, -\mathbf{D})$ capture moves in opposite directions between firm pairs and the “residual” incidence matrix $\mathbf{R}$ represents moves between firm pairs that only experience flows in one direction. Note

that this representation is not unique: there are 2^Q possible choices of $\mathbf{D}$, each of which amounts to treating one member of the firm pair as the origin and the other as the destination.

Given any orientation $\mathbf{D}$, the vector of wage changes between origin-destination pairs can be partitioned

$$\underbrace{\hat{\Delta}}_{|\mathcal{P}| \times 1} = \begin{bmatrix} \underbrace{\hat{\Delta}'_+}_{1 \times Q}, & \underbrace{\hat{\Delta}'_-}_{1 \times Q}, & \underbrace{\hat{\Delta}'_R}_{1 \times (|\mathcal{P}| - 2Q)} \end{bmatrix}',$$

where $\hat{\Delta}_+$ gives the average wage changes of workers moving from origins to destinations dictated by $\mathbf{D}$, $\hat{\Delta}_-$ gives the average wage changes workers moving between those same firms in the opposite direction, and $\hat{\Delta}'_R$ gives the average wage changes associated with $\mathbf{R}$. The corresponding weight matrix can be partitioned

$$\underbrace{\mathbf{W}}_{|\mathcal{P}| \times |\mathcal{P}|} = \begin{bmatrix} \underbrace{\mathbf{W}_+}_{Q \times Q} & 0 & 0 \\ 0 & \underbrace{\mathbf{W}_-}_{Q \times Q} & 0 \\ 0 & 0 & \underbrace{\mathbf{W}_R}_{(|\mathcal{P}| - 2Q) \times (|\mathcal{P}| - 2Q)} \end{bmatrix},$$

where $\mathbf{W}_+$ captures the number of workers moving in one direction between a pair and $\mathbf{W}_-$ the number moving in the opposite direction. Hence, $\mathbf{B}_{(1)} = [\mathbf{D}_{(1)}, -\mathbf{D}_{(1)}, \mathbf{R}_{(1)}]$ and $\mathbf{B}_{(1)} \mathbf{W} \hat{\Delta} = \mathbf{D}_{(1)} (\mathbf{W}_+ \hat{\Delta}_+ - \mathbf{W}_- \hat{\Delta}_-) + \mathbf{R}_{(1)} \mathbf{W}_R \hat{\Delta}_R$.

Plugging the latter expression into (4) reveals that the OLS estimator can be written

$$\begin{aligned} \hat{\psi}_{(1)} &= \mathbf{L}_{(1)}^{-1} \left[\mathbf{D}_{(1)} (\mathbf{W}_+ \hat{\Delta}_+ - \mathbf{W}_- \hat{\Delta}_-) + \mathbf{R}_{(1)} \mathbf{W}_R \hat{\Delta}_R \right] \\ &= \mathbf{L}_{(1)}^{-1} \begin{bmatrix} \mathbf{D}_{(1)} & \mathbf{R}_{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{W}_D & 0 \\ 0 & \mathbf{W}_R \end{bmatrix} \begin{pmatrix} \hat{\Delta}_D \\ \hat{\Delta}_R \end{pmatrix} \\ &\equiv \mathbf{L}_{(1)}^{-1} \underbrace{\bar{\mathbf{B}}_{(1)}}_{(J-1) \times (|\mathcal{P}| - Q)} \underbrace{\bar{\mathbf{W}}}_{(|\mathcal{P}| - Q) \times (|\mathcal{P}| - Q)} \underbrace{\vec{\Delta}}_{(|\mathcal{P}| - Q) \times 1}, \end{aligned} \tag{5}$$

where $\mathbf{W}_D = \mathbf{W}_- + \mathbf{W}_+$ records the total number of workers moving between each of the Q firm pairs with flows in both directions and $\hat{\Delta}_D = \mathbf{W}_D^{-1} (\mathbf{W}_+ \hat{\Delta}_+ - \mathbf{W}_- \hat{\Delta}_-)$ gives the average wage change between these firm pairs across an orientation $\mathbf{D}$.

Equation (5) reveals that no information is lost by collapsing the vector $\hat{\Delta}$ of average wage changes between origin-destination pairs down to a shorter vector $\vec{\Delta}$ of oriented average wage changes between firm pairs. Importantly, this representation holds for any choice of $\mathbf{D}$. The Laplacian is likewise invariant to the

choice of orientation $\mathbf{D}$ as:

$$\mathbf{L}_{(1)} = \mathbf{B}_{(1)} \mathbf{W} \mathbf{B}'_{(1)} = \mathbf{D}_{(1)} (\mathbf{W}_+ + \mathbf{W}_-) \mathbf{D}'_{(1)} + \mathbf{R}_{(1)} \mathbf{W}_R \mathbf{R}'_{(1)} = \bar{\mathbf{B}}_{(1)} \bar{\mathbf{W}} \bar{\mathbf{B}}'_{(1)}.$$

The next section details how the Laplacian can be used to construct an undirected graphical representation of worker mobility patterns.

4 A graph-theoretic interpretation

I will now introduce a graph-theoretic interpretation of the AKM model and the Laplacian $\mathbf{L}$, which we have already seen plays a key role in determining estimability of the firm effects. By clarifying when the least squares estimator $\hat{\psi}_{(1)}$ can be computed, this discussion will provide a principled approach to partitioning the sample into branches.

Define the *mobility graph* $G = (V, E)$ of a two-way fixed effects model as a finite set of vertices V and a collection of edges $E = \{(s, v) \in V \times V : s \neq v\}$ that connect pairs of vertices. I will refer to individual edges by e_{sv} . In contrast to the treatment in Kline (2025), the edges will be viewed here as undirected, which implies that $e_{sv} = e_{vs}$.

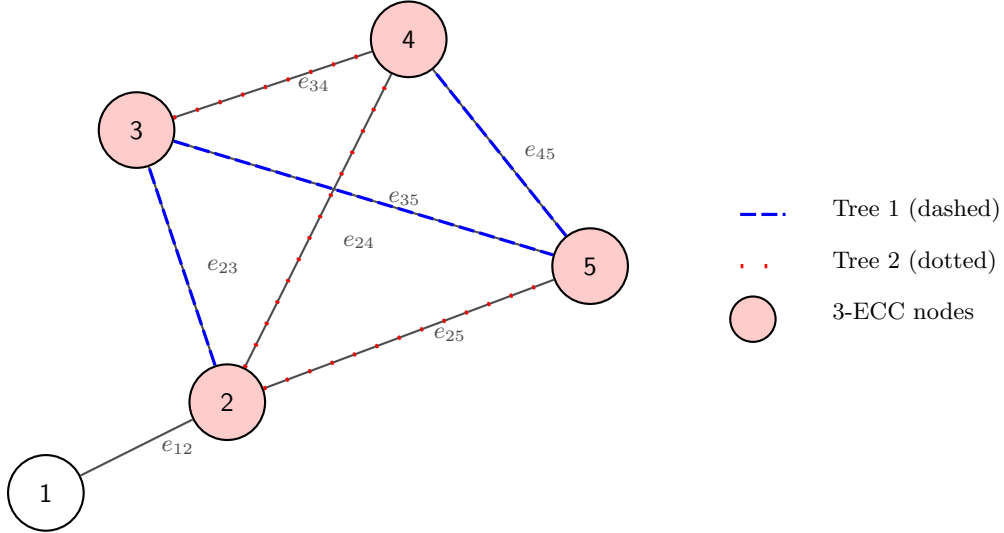


Figure 1: A connected mobility graph with $J = 5$ and $|E| = 7$

In the AKM model, the vertices are firms. Hence, $V = [J]$ and $|V| = J$. An edge joins a pair of firms s and v whenever at least one worker moves between the pair in either direction. That is, in our earlier panel

notation,

$$E = \{(s, v) \in [J]^2 : \mathbf{j}(i, 1) = s, \mathbf{j}(i, 2) = v \text{ or } \mathbf{j}(i, 1) = v, \mathbf{j}(i, 2) = s \text{ for some } i \in [N] \text{ and } s \neq v\}.$$

Thus, the mobility graph has $|E| = |\mathcal{P}| - Q$ edges, each of which corresponds to a column of $\bar{\mathbf{B}} = [\mathbf{D}, \mathbf{R}]$ and can be associated with an entry in $\vec{\Delta}$.

The graph is represented algebraically by the unweighted Laplacian matrix $\mathbf{L}^u = \bar{\mathbf{B}}\bar{\mathbf{B}}'$. The j 'th diagonal entry of $\mathbf{L}^u$ gives firm j 's *degree* in the network: the total number of edges incident to firm j . If firm j and k have an edge between them, then the off-diagonal entry $\mathbf{L}_{jk}^u$ will equal -1. Otherwise it will equal zero.

Figure 1 depicts a graph with 5 firms and 7 edges. The unweighted Laplacian of this graph takes the form:

$$\mathbf{L}^u = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & -1 & 3 \end{bmatrix}.$$

In the figure, firms are depicted with circles and edges by lines. The edge names are displayed to the right of each edge.

I will now introduce some terms of art that are useful for describing networks. A *walk* is a sequence of edges that join a set of firms, a *trail* is a walk with no repeated edges, and a *path* is a trail with no repeated firms. Hence, the sequence $\{e_{23}, e_{35}, e_{25}\}$ is a trail that is not a path.

A graph is *connected* if there is a path from any firm to any other firm. The *edge connectivity* $\lambda(G)$ of a graph is the number of edges that need to be removed from the graph for it to become disconnected. Formally,

$$\lambda(G) = \min\left\{|S| : S \subseteq E, \text{ and } (V, E \setminus S) \text{ is disconnected}\right\}.$$

The graph depicted in Figure 1 has $\lambda(G) = 1$ because removing edge e_{12} severs firm 1 from the network.

A graph is k -edge-connected if $\lambda(G) \geq k$. Equivalently, a k -edge-connected graph has at least k paths between any two vertices (Menger, 1927). A k -edge-connected component (k -ECC) is a maximal subgraph – the largest collection of vertices and edges in the original graph – that is k -edge-connected. Figure 1 has a single 3-ECC, the vertices of which are shaded red. Removing any two edges from the 3-ECC fails to disconnect the graph, while removing the edges $\{e_{23}, e_{24}, e_{25}\}$ partitions the network into two separate

connected components. Hence, this 3-ECC is not also a 4-ECC.

A *tree* is a connected graph for which there is a unique path between any pair of firms. A *spanning tree* is any subset of a connected graph that contains all firms and is a tree. Kirchhoff’s matrix tree theorem states that any cofactor of the unweighted Laplacian equals the number of spanning trees in a graph (Kirchhoff, 1847; Spielman, 2019). An implication of this result is that $\mathbf{L}_{(1)}$ is guaranteed to be invertible – and hence, $\hat{\psi}_{(1)}$ computable – whenever the graph is connected.

The graph depicted in Figure 1 contains 16 spanning trees. One can use the matrix tree theorem to verify this claim: computing the determinant of the matrix formed by deleting any row and column from the unweighted Laplacian $\mathbf{L}^u$ will yield 16. However, all 16 of these trees contain the edge e_{12} . Thus, we can only *pack* a single spanning tree into this graph without having to share an edge.

The 3-ECC in Figure 1 also has 16 spanning trees but we can pack more than one edge-disjoint spanning tree into this component. Two edge-disjoint trees that span the 3-ECC are depicted by the blue and red lines. Clearly, we cannot pack a third spanning tree into this component as the two depicted trees consume all of the available edges. Evidently, it is not always possible to pack k edge-disjoint spanning trees into an k -ECC. We will return to this insight in Section 6, where tree packing is discussed in greater detail.

5 Trees and branches

Returning to the algebraic representation of the fixed effects estimator in (5), an important simplification arises when $\bar{\mathbf{B}}$ has dimension $J \times (J - 1)$, which implies there are only $J - 1$ edges connecting the J firms. An incidence matrix of this form represents a spanning tree. Hence, $\bar{\mathbf{B}}_{(1)}$ is a square invertible matrix and $\mathbf{L}_{(1)}^{-1} = \left(\bar{\mathbf{B}}'_{(1)}\right)^{-1} \bar{\mathbf{W}}^{-1} \bar{\mathbf{B}}_{(1)}^{-1}$. Plugging this expression into (5) and simplifying reveals that the firm effects estimator reduces in this case to

$$\hat{\psi}_{(1)} = \left(\bar{\mathbf{B}}'_{(1)}\right)^{-1} \bar{\Delta}.$$

Note that, relative to (5), the weighting matrix $\bar{\mathbf{W}}$ has disappeared, which reflects that the firm effects are just-identified in this scenario. This interpretation is pursued at greater length in Kline (2025), where it is shown that the fundamental restriction of the AKM model is that cycle sums of the average wage changes $\bar{\Delta} \in \mathbb{R}^{J-1}$ must equal zero.

In general, when the mobility network is connected, $\bar{\mathbf{B}}$ can be partitioned into submatrices capturing spanning trees and a set of “leftover” edges that fail to connect some of the firms. In particular, if M spanning trees are capable of being packed into the mobility graph, with corresponding incidence matrices $\mathbf{T}_1, \dots, \mathbf{T}_M$, then we can write $\bar{\mathbf{B}} = [\mathbf{T}_1, \dots, \mathbf{T}_{M-1}, \mathbf{T}_M^+]$, where $\mathbf{T}_M^+$ appends columns to $\mathbf{T}_M$ that capture

any leftover edges. This set of incidence matrices $\{\mathbf{T}_1, \dots, \mathbf{T}_{M-1}, \mathbf{T}_M^+\}$ define our *branches*: each branch is a subgraph of the data that connects all firms in the mobility network.

Removing the first row of $\bar{\mathbf{B}}$ yields $\bar{\mathbf{B}}_{(1)} = [\mathbf{T}_{(1),1}, \dots, \mathbf{T}_{(1),M-1}, \mathbf{T}_{(1),M}^+]$. It is useful to also partition $\vec{\Delta} = (\vec{\Delta}'_1, \dots, \vec{\Delta}'_M)'$ and $\bar{\mathbf{W}} = \text{diag}(\bar{\mathbf{W}}_1, \dots, \bar{\mathbf{W}}_M)$. We can now define branch-specific estimates

$$\hat{\psi}_{(1),b} = \begin{cases} \left(\mathbf{T}_{(1),b} \mathbf{T}'_{(1),b} \right)^{-1} \mathbf{T}_{(1),b} \vec{\Delta}_b & \text{if } b < M \\ \left(\mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \left(\mathbf{T}_{(1),M}^+ \right)' \right)^{-1} \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \vec{\Delta}_M & \text{if } b = M \end{cases}.$$

This representation reflects the earlier finding that the weights are irrelevant for trees. However, the weights may matter for the last branch ($b = M$) because the graph may contain edges that are not a part of any spanning tree. Including these leftover edges forms cycles – disjoint paths between vertices – on the subgraph corresponding to the last branch. The inclusion of these cycles will tend to improve the precision of the last branch relative to the others.

With all the edges assigned to a branch, no information in the microdata data is wasted. The following proposition formally establishes that the full-sample OLS estimator $\hat{\psi}_{(1)}$ can be written as a linear combination of the branch-specific estimates $\left\{ \hat{\psi}_{(1),b} \right\}_{b=1}^M$.

Proposition 1. *Suppose the mobility network is connected and contains M edge-disjoint spanning trees.*

Then $\hat{\psi}_{(1)} = \sum_{b=1}^M \mathbf{C}_{(1),b} \hat{\psi}_{(1),b}$, where each $\mathbf{C}_{(1),b}$ is a $(J-1) \times (J-1)$ matrix and $\sum_{b=1}^M \mathbf{C}_{(1),b} = \mathbf{I}$.

Proof. We can write

$$\mathbf{L}_{(1)} = \sum_{b=1}^{M-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \mathbf{T}'_{(1),b} + \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \left(\mathbf{T}_{(1),M}^+ \right)',$$

$$\bar{\mathbf{B}}_{(1)} \bar{\mathbf{W}} \vec{\Delta} = \sum_{b=1}^{M-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \vec{\Delta}_b + \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \vec{\Delta}_M.$$

Now define

$$\mathbf{C}_{(1),b} = \begin{cases} \mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \mathbf{T}'_{(1),b} & \text{if } b < M \\ \mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \left(\mathbf{T}_{(1),M}^+ \right)' & \text{if } b = M. \end{cases}$$

Hence, $\sum_{b=1}^M \mathbf{C}_{(1),b} = \mathbf{L}_{(1)}^{-1} \left(\sum_{b=1}^{M-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \mathbf{T}'_{(1),b} + \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \left(\mathbf{T}_{(1),M}^+ \right)' \right) = \mathbf{L}_{(1)}^{-1} \mathbf{L}_{(1)} = \mathbf{I}$.

For each $b < M$,

$$\mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \vec{\Delta}_b = \left[\mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \mathbf{T}'_{(1),b} \right] \hat{\psi}_{(1),b} = \mathbf{C}_{(1),b} \hat{\psi}_{(1),b},$$

while, for $b = M$,

$$\mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \vec{\Delta}_M = \left[\mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \left(\mathbf{T}_{(1),M}^+ \right)' \right] \hat{\psi}_{(1),M} = \mathbf{C}_{(1),M} \hat{\psi}_{(1),M}.$$

Hence,

$$\hat{\psi}_{(1)} = \mathbf{L}_{(1)}^{-1} \bar{\mathbf{B}}_{(1)} \bar{\mathbf{W}} \vec{\Delta} = \sum_{b=1}^M \mathbf{C}_{(1),b} \hat{\psi}_{(1),b}.$$

□

The matrix $\mathbf{C}_{(1),b}$ can be shown to capture the relative precision of the branch-specific estimates relative to the full-sample estimate when the wage errors are homoscedastic (Jochmans and Weidner, 2019). Thus, $\hat{\psi}_{(1)}$ can be thought of as a precision-weighted average of the branch estimates, albeit with the potential for negative weights, as $\mathbf{C}_{(1),b}$ is positive semi-definite but not diagonal.

Now letting

$$\hat{\psi} = \left(0, \hat{\psi}_{(1)}' \right)', \quad \hat{\psi}_b = \left(0, \hat{\psi}_{(1)b}' \right)', \quad \mathbf{C}_b = \text{diag} \left(0, \mathbf{C}_{(1),b} \right), \quad \text{and} \quad \varphi_b = \mathbf{C}_b \hat{\psi}_b,$$

Proposition 1 yields the additive decomposition introduced in (1):

$$\hat{\psi} = \sum_{b=1}^M \varphi_b.$$

This representation reveals that each φ_b measures the influence of a branch on the full-sample estimator. In fact, one can show that each φ_b gives the branch sum of observation-level contributions to the recentered influence function of the full-sample estimator. Since the $\{\varphi_b\}_{b=1}^M$ are constructed from disjoint sets of microdata, they are statistically independent, providing a transparent foundation for assessing uncertainty in $\hat{\psi}$.

Computation of $\varphi_b = \left(0, \varphi_{(1),b}' \right)'$ is greatly aided by the observation that

$$\varphi_{(1),b} = \begin{cases} \mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),b} \bar{\mathbf{W}}_b \vec{\Delta}_b & \text{if } b < M \\ \mathbf{L}_{(1)}^{-1} \mathbf{T}_{(1),M}^+ \bar{\mathbf{W}}_M \vec{\Delta}_M & \text{if } b = M. \end{cases}$$

Each of these expressions is simply a least squares regression that is no more computationally expensive to solve than the weighted least squares problem in (5).

6 Building branches

The key challenge in constructing branch-specific estimates is building the incidence matrices $T_1, \dots, T_M$ corresponding to the graph’s spanning trees. Doing so requires first determining how many spanning trees can be packed into the graph. The following result due to Nash-Williams (1961) and Tutte (1961) provides a useful foundation for answering this question.

Theorem (Nash-Williams-Tutte, 1961). *A graph $G = (V, E)$ can pack M spanning trees if and only if, for every partition $\{V_1, \dots, V_r\}$ of V into $r \geq 2$ blocks, the number of edges connecting vertices in different blocks is at least $M(r - 1)$.*

Intuitively, a spanning tree must cross any partition of the vertices to span the graph. The Nash-Williams-Tutte theorem establishes that if there is a partition of the vertices that yields fewer than $M(r - 1)$ crossings, then there must be fewer than M spanning trees capable of being packed into the graph. Conversely, if M spanning trees can be packed into the graph, then the number of edges crossing any partition must be at least $M(r - 1)$.

The following corollary of the Nash-Williams-Tutte theorem provides us with bounds on the number $\tau(G)$ of spanning trees capable of being packed into a graph in terms of its edge connectivity $\lambda(G)$.

Corollary. *The number $\tau(G) \in \mathbb{N}$ of spanning trees that can be packed into a connected graph $G = (V, E)$ obeys $\lfloor \lambda(G)/2 \rfloor \leq \tau(G) \leq \lambda(G)$, where $\lfloor \cdot \rfloor$ denotes the floor operator.*

A proof of this corollary can be found in Kundu (1974). The upper bound $\tau(G) \leq \lambda(G)$ follows immediately from the theorem by setting the number of blocks $r = |V|$. The lower bound $\tau(G) \geq \lfloor \lambda(G)/2 \rfloor$ comes from a double-counting of edges across any r -way partition. Since each of the r blocks has at least $\lambda(G)$ edges leaving it, summing over all blocks counts each crossing edge twice. Hence, the total weight associated with all partition crossings must be at least $r\lambda(G)/2$. In the case of Figure 1, this corollary tells us that $\tau(G) \in [1, 3]$ in the graph’s 3-ECC. As the Figure illustrates, there are exactly two spanning trees in that component.

In the prototypical worker-firm dataset, the edge connectivity is zero: there are usually multiple connected components of firms. Since the work of Abowd, Creedy, and Kramarz (2002), the convention in the literature has been to focus on the largest connected component of the mobility graph (i.e., the largest 1-ECC). Since many of firms in the largest 1-ECC are connected by only a single edge, no more than one edge-disjoint spanning tree can be packed into such graphs. The Nash-Williams-Tutte theorem suggests a reasonable way to find multiple disjoint spanning trees is to narrow the scope of investigation from the largest 1-ECC to the largest k -ECC for $k > 1$. Doing so requires first finding the largest k -ECC and then packing this subgraph.

6.1 Finding the largest k -ECC

A key tool that facilitates finding a k -ECC is the *Gomory-Hu tree* (Gomory and Hu, 1961), which is a weighted tree on V . The edge weights of this tree provide the number of edges removals required to disconnect a pair of vertices. Figure 2 provides the Gomory-Hu tree associated with the graph in Figure 1. The edge connecting Firm 1 to Firm 2 has a weight of 1, reflecting that Firm 1 can be disconnected from the network by removing e_{12} . In contrast, disconnecting Firm 2 from Firm 3 would require removing three edges $\{e_{23}, e_{24}, e_{25}\}$, which is why the Gomory-Hu edge between these vertices has a weight of 3.

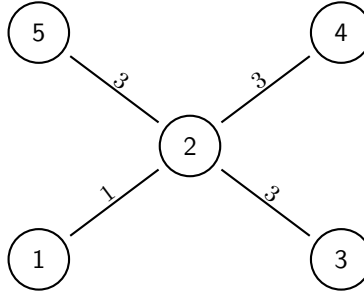


Figure 2: Gomory-Hu Tree of graph in Figure 1

A useful feature of Gomory-Hu trees is that the minimal edge weight on the path between any pair of vertices $(s, v) \in V \times V$ gives the number of edge removals required to disconnect those vertices. As a result, dropping all edges with weight less than k reveals the set of all k -ECCs.

For example, if we threshold the tree in Figure 2 at $m = 2$, we are left with a single 2-ECC formed by the vertex set $\{2, 3, 4, 5\}$. Several algorithms exist for rapidly building the Gomory-Hu tree associated with a graph. I will use an algorithm due to Gusfield (1990), implemented in the `igraph` package (Csárdi and Nepusz, 2023), which has worst-case computational complexity of $O(J^4)$.³ Once we have pruned the Gomory-Hu tree, we can select the k -ECC with the most vertices.

6.2 Packing a k -ECC

Once the largest k -ECC has been found, we would like to extract as many edge-disjoint spanning trees from the graph as possible. This is known as the *tree-packing problem*. Formally, the tree packing problem is an integer linear programming (ILP) problem that can be expressed as follows:

³For very large graphs, building the full Gomory-Hu tree will be excessively costly. In those settings, one can work with an approach due to Nagamochi and Ibaraki (1992a) and Nagamochi and Ibaraki (1992b) that avoids computing the entire tree and directly finds the k -ECC.

$$\begin{aligned}
& \max_x \sum_{T \in \mathcal{T}} x_T \\
& \text{s.t.} \quad \sum_{T \ni e} x_T \leq 1 \quad \forall e \in E, \\
& \quad x_T \in \{0, 1\} \quad \forall T \in \mathcal{T},
\end{aligned}$$

where $\mathcal{T}$ is the collection of all trees in the graph and x_T is an indicator for whether tree $T \in \mathcal{T}$ is chosen. The matrix tree theorem tells us that $|\mathcal{T}|$ can grow exponentially with the scale of the graph, making direct enumeration infeasible.

Rather than tackle the ILP directly, I rely on an iterative approach proposed by Roskind and Tarjan (1985). Their algorithm requires pre-specifying the number $M > 1$ of edge-disjoint spanning trees that one is seeking to extract from a k -ECC. When $M \leq \tau(G)$, the procedure is guaranteed to exactly M trees. In contrast, greedily extracting trees sequentially (e.g., via Kruskal’s algorithm) may fail to find $\tau(G)$ trees. An example is provided in Appendix A:

The Roskind-Tarjan algorithm, which is implemented in SageMath, is well suited to large graphs, exhibiting computational complexity $O(J^2 M^2)$. If we apply the algorithm to a k -ECC, then the Corollary assures us that we will be able to find at least $\lfloor k/2 \rfloor$ trees.

6.3 A “prune and pack” algorithm

Algorithm 1 provides a sketch of the composite procedure for pruning to the k -ECC and extracting the disjoint spanning trees.

Algorithm 1: Pack k -ECC

- 1 Fix a $k \in \mathbb{N}$
 - 2 Drop all vertices with degree $< k$. Drop all but largest connected component. Repeat until no more vertices dropped.
 - 3 Build a Gomory-Hu tree. Remove all tree edges with weight $< k$, then take the largest connected component.
 - 4 Initialize $M = \lfloor k/2 \rfloor$. While Roskind-Tarjan returns M disjoint spanning trees, increment M , stopping when no more spanning trees can be found.
-

Step 2 of this algorithm is a screen that exploits the fact that a vertex with degree $< k$ cannot be a member of the k -ECC. The graph’s connected subgraph of a network with degree $\geq k$ is known as the “ k -core.” Step 3 searches within the k -core for the k -ECC. In step 4, we first try $M = \lfloor k/2 \rfloor$, and then explore higher values of M until no more trees can be found or we reach the upper bound of k .

7 Empirical application

I now turn to analyzing an extract of the benchmark Veneto dataset that was also studied in Kline, Saggio, and Sølvesten (2020) and Kline (2025). The extract is comprised of 1,859,459 person-year observations from the years 1999 and 2001. The largest connected component contains 73,933 firms and 747,205 workers, 197,572 of whom switch employers between the two years. The worker moves between these two years involve 150,417 ordered origin-destination pairs $|\mathcal{P}|$. Exactly 1,500 pairs of firms have flows in both directions. Hence, the mobility network formed by these pairs is compromised of $|E| = 150,417 - 1,500 = 148,917$ undirected edges.

Note that the average number of movers per edge is quite low as $197,572 / 148,917 \approx 1.3$. Thus, branching firm effect estimates based upon undirected edges is likely to provide a reasonable approximation to what is possible when treating each individual worker move as an edge. This is fortunate, as packing spanning trees into multi-edge graphs turns out to be significantly more complex than packing them into simple graphs (Barahona, 1995).

Table 1 reports the results of pruning the mobility graph to the largest k -ECC for different choices of k as in Algorithm 1.

k	1	2	3	4	5	6
Firms in k -core	73,933	41,093	21,570	11,145	5,682	3,128
Firms in k -ECC (J)	73,933	41,054	21,565	11,145	5,682	3,128
Edges in k -ECC ($ E $)	148,917	116,026	80,561	51,824	31,677	19,796
Movers in k -ECC (N)	197,572	158,149	114,717	78,908	53,131	36,903
Spanning Trees (M)	1	1	2	3	3	4

Table 1: Network properties of k -ECCs

The size of the k -core contracts rapidly with k , reflecting that most firms have low degree. To some extent this phenomenon is an artifact of studying mobility over only a 2-year horizon. However, adding further years of mobility data may introduce new firms that only contain a single edge, which could actually lower the average degree of the network. An interesting approach for future work would be to fix a set of firms of interest in a base year and “infill” the edges between them with additional years of mobility data.

The sparse nature of the Veneto network makes the k -core a very close approximation to the largest k -ECC. While this finding need not generalize to larger, or more densely connected, networks, these patterns suggest the k -core is likely to provide a useful starting guess for the k -ECC in larger graphs where full computation of a Gomory-Hu tree would be impractical. In truly massive graphs, it may be advisable to simply pack spanning trees into the k -core via the Roskind-Tarjan algorithm rather than refining to the largest k -ECC.

In the Veneto data, pruning the sample to the 2-ECC leads to the loss of about 44% of the firms but

fails to yield a second spanning tree. The Nash-Williams-Tutte theorem suggests we should not be surprised by this finding as a 2-ECC can have at most 2 spanning trees. The 3-ECC, which retains about 29% of the firms but 58% of the movers in the original graph, does contain a second spanning tree. A third tree is found in the 4-ECC, which has 15% of the firms and 39% of the movers in the original graph. A fourth tree emerges with the 6-ECC, which contains only 4% of firms but 17% of the movers in the original graph.

It is interesting to compare these sample dimensions to what one would obtain by randomly splitting the sample at the mover level. To build this benchmark, I randomly assign each worker to one of M splits with equal probability, varying M from one to four. In each split, I calculate the largest connected component and then intersect the vertex set of those components across splits to arrive at the number of firms for which M independent unbiased estimates can be computed. Table 2 reports the results of repeating this process 500 times.

Splits (M)	1	2	3	4
Number of firms				
25th Percentile	73,933	28,680	13,034	6,537
Median	73,933	28,737	13,077	6,576
75th Percentile	73,933	28,797	13,127	6,607
Overlap across simulations				
25th Percentile	73,933	22,745	9,328	4,482
Median	73,933	22,804	9,366	4,509
75th Percentile	73,933	22,861	9,406	4,536

Table 2: Firm effects estimable in each of M random splits (500 simulations)

Splitting the sample in half yields a pair of independent estimates for roughly one third of the firms. A three-way split yields three estimates for roughly 18% of firms. A four-way split yields four estimates for about 9% of firms. While little variability emerges across the 500 simulation draws in the number of firms for which M estimates can be computed, the composition of these sets of firms varies considerably. The bottom panel of the table reports quantiles of the degree of overlap in the sets of firms for which M estimates can be computed across all $\binom{500}{2}$ pairs of simulations. The table reveals that reshuffling a split into $M = 2$ groups would yield less than 22,804 firms in common with the previous split half of the time. Likewise, re-randomizing a three-way split would yield less than 9,366 of the firms present in a previous random split half of the time.

Comparing these findings to Table 1 reveals that randomly splitting workers into M groups yields more firms with M estimates than does pruning to the largest k -ECC with M trees. For example, the largest 3-ECC has 21,565 firms, while randomly assigning workers to one of two splits yields a median of 28,737 firms with two independent estimates. This discrepancy suggests that there exist subgraphs of the largest 2-ECC that strictly nest the largest 3-ECC yet contain two edge-disjoint spanning trees. How to systematically explore

the space of such intermediate subgraphs is an interesting question. For example, one could potentially append pruned edges to the branches of the 3-ECC, growing the size of the network in way that ensures each branch connects the same set of vertices. I leave such investigations to future research.

While random splitting delivers independent estimates for more firms, a major advantage of the pruning and packing strategy is that the target population is deterministic. Another advantage is that the full sample estimator decomposes linearly into M branch-specific estimates, which facilitates uncertainty quantification. In contrast, the random splitting approach yields a full-sample estimator that includes a fixed effect for every firm in the *union* of the M connected sets. While some of these parameters have M independent estimates, others have may have only a single estimate, which hinders uncertainty quantification even when ignoring the randomness in the split itself.

The next subsections illustrate how branches constructed from packing k -ECC's can be put to work. In what follows, I compute the full-sample estimator $\hat{\psi}$ for each k -ECC using the adjusted wage changes $\hat{\Delta}$ between origin-destination pairs described in Kline (2025), which only adjust for a year fixed effect. Note that each of the 197,572 worker moves contribute to at most one of these wage changes, implying the entries of $\hat{\Delta}$ are mutually independent. To facilitate computation, the vector $\hat{\Delta}$ is collapsed to a shorter vector of oriented wage changes $\vec{\Delta}$, which is then used to construct the branch-specific estimates $\hat{\psi}_b$ and φ_b . While the methods described in the next subsection only apply to branches, the approaches in Sections 7.2 and 7.3 can just as easily be applied to estimates constructed from random splits.

7.1 Quantifying uncertainty

The arguments of Section (2) suggest estimating the variance matrix Σ of the vector $\hat{\psi}$ of full-sample firm effect estimates with

$$\hat{\Sigma} = \frac{1}{M-1} \sum_{b=1}^M \sum_{\ell < b} (\varphi_b - \varphi_\ell) (\varphi_b - \varphi_\ell)'.$$

To illustrate the potential usefulness of this matrix, I consider the projection of ψ onto the matrix $\mathbf{X} = [\mathbf{1}, \ln f]$, which consists of an intercept and the logarithm of average firm size across the two time periods. The second entry of $\gamma = \mathbf{S}_{xx}^{-1} \mathbf{X}' \psi$, which I will denote $\gamma_{(1)}$, approximates the elasticity of firm wage effects with respect to firm size.

As documented in Bloom et al. (2018) and Kline (2025) firm wage effects do not seem to be linear in log firm size, instead exhibiting a concave relationship in which the largest firms exhibit wages effects roughly equivalent to or lower than their slightly smaller peers. Consequently, this linear projection will not coincide exactly with the conditional expectation function. Nonetheless, the plug-in estimator $\hat{\gamma} = \mathbf{S}_{xx}^{-1} \mathbf{X}' \hat{\psi}$ remains unbiased for the projection coefficient γ conditional on the covariates $\mathbf{X}$, which we treat as fixed.

The branch-based estimator of the variance of the projection coefficients is:

$$\hat{\mathbf{V}}_{branch} [\hat{\gamma}] = \mathbf{S}_{xx}^{-1} \mathbf{X}' \hat{\Sigma} \mathbf{X} \mathbf{S}_{xx}^{-1}.$$

It is instructive to compare this unbiased variance estimator to the naive standard errors that come from treating the entries of $\hat{\psi}$ as mutually independent in a second-step regression:

$$\hat{\mathbf{V}}_{HC0} [\hat{\gamma}] = \mathbf{S}_{xx}^{-1} \left(\mathbf{X} \odot (\hat{\psi} - \mathbf{X} \hat{\gamma}) \right)' \left(\mathbf{X} \odot (\hat{\psi} - \mathbf{X} \hat{\gamma}) \right) \mathbf{S}_{xx}^{-1}.$$

While $\hat{\mathbf{V}}_{HC0} [\hat{\gamma}]$ is robust to misspecification of the conditional expectation function, it neglects dependence between the estimated firm effects, which can lead to large biases in either direction.

Table (3) reports the estimated projection coefficients, the two sets of standard errors, and the mean firm size in each k -ECC. As k grows larger, the sample is restricted to larger firms, which tend to have greater degree in the mobility network. The estimated projection coefficient declines substantially with firm size, reflecting that the relationship between firm wage effects and log firm size is concave.

k	1	2	3	4	5	6
Full-sample estimate ($\hat{\gamma}_{(1)}$)	0.0446	0.0329	0.0235	0.0199	0.0191	0.0209
Standard error ($\hat{\mathbf{V}}_{branch} [\hat{\gamma}_{(1)}]^{1/2}$)	-	-	0.0080	0.0045	0.0084	0.0091
Naive std err ($\hat{\mathbf{V}}_{HC0} [\hat{\gamma}_{(1)}]^{1/2}$)	0.0009	0.0010	0.0013	0.0018	0.0023	0.0032
Mean firm size	13.6	21.6	34.6	55.1	88.0	131.6

Table 3: Elasticity of firm wage effects with respect to firm size

A researcher relying on naive two-step standard errors might feel quite confident about the likely value of the projection slope in each sample, with t-ratios ranging from roughly 7 to 50. However, the standard errors of the projection based on branches are several times larger, yielding t-ratios ranging from 2.3 to 2.9.

Some of the increase in the branch-based standard errors over their naive counterparts may reflect misspecification of the AKM model. If each branch estimator $\hat{\psi}_b$ is unbiased for a slightly different target parameter $\psi + \xi_b$, then one can show that $\hat{\Sigma}$ measures the composite variance in $\hat{\psi}$ one would face if the branches were themselves randomly sampled with equal probability. Though branches have not been randomly sampled here, this heuristic may nonetheless be helpful for assessing which results are likely to generalize to other environments. I leave the study of branch-based estimators under misspecification to future work.

7.2 Moment estimation

Moments of firm effects are a common summary statistic that have long played an important role in the literature. Since the firm effects ψ are only identified up to a location shift, it is traditional to report centered moments as in Abowd, Kramarz, and Margolis (1999). Researchers typically weight these central moments by activity measures such as firm size.

Let $\omega \in \mathbb{R}_{\geq 0}^J$ be a vector of firm weights that sums to one ($\mathbf{1}'\omega = 1$). For any $\ell \geq 2$, we can write the plug-in estimator of the ℓ 'th central weighted moment of firm effects as

$$\hat{\mu}_{\ell,PI} = \omega' \left(\hat{\psi} - \omega' \hat{\psi} \mathbf{1} \right)^{\circ \ell},$$

where the $\circ \ell$ superscript denotes raising the entries in a vector to the ℓ 'th power element-wise. Generalizing the approach suggested in Section 2, I use branches to construct the corresponding unbiased estimator:

$$\hat{\mu}_{\ell} = \binom{M}{\ell}^{-1} \sum_{(b_1, \dots, b_{\ell}) \in \mathcal{B}} \omega' \left(\hat{\psi}_{b_1} - \omega' \hat{\psi}_{b_1} \mathbf{1} \right) \odot \dots \odot \left(\hat{\psi}_{b_{\ell}} - \omega' \hat{\psi}_{b_{\ell}} \mathbf{1} \right),$$

where $\mathcal{B} = \{(b_1, \dots, b_{\ell}) : 1 \leq b_1 < \dots < b_{\ell} \leq M\}$ is the set of all combinations of ℓ distinct branches. Note that, when $M = 2$, $\hat{\mu}_2$ is the covariance estimator used in split sample approaches. When $M > 2$, $\hat{\mu}_2$ averages across covariance matrices formed by all possible sample splits. When $\ell > 2$, higher-order products are taken.

k	1	2	3	4	5	6
Std dev ($\sigma = \sqrt{\mu_2}$)						
Plug-in	0.2215	0.1951	0.1841	0.1794	0.1751	0.1756
Branches	-	-	0.1591	0.1669	0.1917	0.1555
Skew (μ_3/σ^3)						
Plug-in	-0.8332	-0.7403	-0.8223	-0.8507	-0.7198	-0.6998
Branches	-	-	-	-1.217	-0.7911	-0.7368
Kurtosis (μ_4/σ^4)						
Plug-in	7.309	5.824	6.137	6.191	5.776	5.416
Branches	-	-	-	-	-	6.617

Table 4: Moments of firm effects (size-weighted)

Table 4 reports plug-in and branch-based estimates of moments of the firm effects weighting by total firm size across the two periods. The plug-in estimates of the standard deviation are upward biased due to sampling error (Andrews et al., 2008). The branch-based estimates suggest this bias is fairly mild. For example, in the 3-ECC, we find $\sqrt{\hat{\mu}_{2,PI}} - \sqrt{\hat{\mu}_2} = .025$. This finding is broadly in line with the results of Kline, Saggio, and Sølvesten (2020) who report a plug-in standard deviation of 0.189 and a bias-corrected

standard deviation of 0.155 in a closely related sample.⁴

Note that for $k = 5$ we find $\hat{\mu}_2 > \hat{\mu}_{2,PI}$, which may reflect noise in both the plug-in and covariance based estimates. In general, the unbiased branch-based estimators need not respect logical constraints on the parameter space. For instance, $\hat{\mu}_2$ can take on negative values. To rule such behavior out, one can impose constraints in a second step, for example, by considering biased estimators of the form $\min \{\max \{0, \hat{\mu}_2\}, \hat{\mu}_{2,PI}\}$.

As described in Appendix B, when four or more branches are present, it becomes possible to estimate the variance of the second moment estimator $\hat{\mu}_2$ by exploiting the covariance between disjoint pairs of differences in the branch-specific estimates. Applying this variance estimator to the 6-ECC yields a standard error on $\hat{\mu}_2$ of 0.0014. Thus, the ratio of $\hat{\mu}_2$ to its standard error yields a z -score of $(.1555)^2/.0014 \approx 17$, suggesting the standard deviation σ is estimated quite precisely. A similar finding was reported by Kline, Saggio, and Sølvesten (2020, Table IV), who obtained a standard error of 0.0006 on a closely related estimator of μ_2 , yielding a z -score of roughly 40.

The precision with which $\hat{\mu}_2$ is estimated suggests that it can safely be used to scale the estimates of higher-order moments without inducing weak identification problems. The scaled higher moment estimates in Table 4 suggest that the size-weighted distribution of firm effects exhibits a mild left skew. The four branches of the 6-ECC allow us to estimate the sized-weighted fourth moment. While a normal distribution would exhibit a kurtosis of 3, the branch-based estimate of kurtosis is 6.2, indicating heavy tails. Evidently, there are more very low and very high paying firms than one would surmise based upon a Gaussian benchmark.

Table 5 reports moment estimates that weight all firms equally (i.e., $\omega = 1/J$). While the results are quite similar, the discrepancies between the plug-in and branch-based estimates are somewhat larger, reflecting that greater weight is being put on small firms that are less connected in the mobility network.

k	1	2	3	4	5	6
Std dev ($\sigma = \sqrt{\mu_2}$)						
Plug-in	0.3088	0.2223	0.188	0.1774	0.175	0.1792
Branches	-	-	0.1315	0.1325	0.1696	0.1642
Skew (μ_3/σ^3)						
Plug-in	-0.573	-0.5504	-0.6029	-0.8343	-0.9726	-1.077
Branches	-	-	-	-1.873	-1.255	-1.856
Kurtosis (μ_4/σ^4)						
Plug-in	6.156	5.171	5.43	5.771	5.899	5.663
Branches	-	-	-	-	-	5.164

Table 5: Moments of firm effects (equally weighted)

In the 6-ECC, the discrepancy $\sqrt{\hat{\mu}_{2,PI}} - \sqrt{\hat{\mu}_2}$ remains small because this set of large firms exhibits

⁴The Kline, Saggio, and Sølvesten (2020) estimates pertain to a sample under which the mobility graph remains connected whenever any individual *mover* is removed. This sample should be thought of as lying between the 1-ECC and the 2-ECC as each undirected edge may involve one or more movers.

degree of at least six, leading to precise estimates. Confirming this impression, the variance estimator described in Appendix B: yields a standard error for $\hat{\mu}_2$ under this weighting scheme of 0.0015, which differs negligibly from the standard error found for the earlier estimator utilizing firm-size weights. The z -score of $(.1642)^2 / .0015 \approx 18$ again suggests the standard deviation σ is estimated with great precision. Relative to the firm size weighted results, the left skew in the equally weighted distribution of firm effects is stronger but the kurtosis is less pronounced. These discrepancies may indicate that the right tail of the firm effect distribution is populated by very large firms.

7.3 Shrinkage

It is often of interest, either for forecasting or policy targeting purposes, to construct a best predictor of the firm effects ψ based upon the full-sample estimates $\hat{\psi}$. Standard empirical bayes shrinkage arguments are predicated on the assumption that the distribution of noise is known ex-ante (Walters, 2024). Given that the number of movers per edge is only 1.3, we cannot appeal to a central limit theorem to defend a Gaussian approximation to the noise in the branch-specific estimates $\{\hat{\psi}_b\}_{b=1}^M$. It turns out, however, that independence across branches facilitates the construction of optimal predictors via standard regression methods that remain valid regardless of the noise distribution.

To understand the basic logic of this argument, consider the case where $M = 2$. Let $\hat{\psi}_{bj}$ denote the j 'th entry of $\hat{\psi}_b$ and ψ_j the j 'th entry of ψ . Now consider a random effects model wherein $\psi_j \stackrel{i.i.d}{\sim} G$ and $\hat{\psi}_b \mid \psi \sim F_b$, with F_b obeying $\mathbb{E}_{F_b} [\hat{\psi}_b] = \psi$. By iterated expectations,

$$\mathbb{E} [\hat{\psi}_{2j} \mid \hat{\psi}_{1j}] = \mathbb{E} \left[\mathbb{E} [\hat{\psi}_{2j} \mid \psi_j, \hat{\psi}_{1j}] \mid \hat{\psi}_{1j} \right] = \mathbb{E} \left[\mathbb{E} [\hat{\psi}_{2j} \mid \psi_j] \mid \hat{\psi}_{1j} \right] = \mathbb{E} [\psi_j \mid \hat{\psi}_{1j}],$$

where the second equality follows by independence of the branch measurement errors. Thus, a regression of the estimates in one branch on those from another yields a minimum mean squared error optimal predictor of the latent firm effect regardless of the noise distributions $\{F_1, F_2\}$. Importantly, this result holds under arbitrary patterns of heteroscedasticity, both across and within branches.

Figure 3 illustrates this idea using the two branches in the 3-ECC. Each panel depicts estimates of $\mathbb{E} [\hat{\psi}_{bj} \mid \hat{\psi}_{\ell j} = x] \equiv m_b(x)$ via the binscatter methods described in Cattaneo et al. (2024). In both panels, a second-order global polynomial approximation $\hat{m}_b(x)$ to $m_b(x)$ appears to fit well. Note that the scale of the x-axis differs across the two sub-panels, reflecting that the entries of $\hat{\psi}_2$ are generally less variable than the corresponding entries of $\hat{\psi}_1$. This is an artifact of the second branch utilizing leftover edges that add additional information about the firm effects.

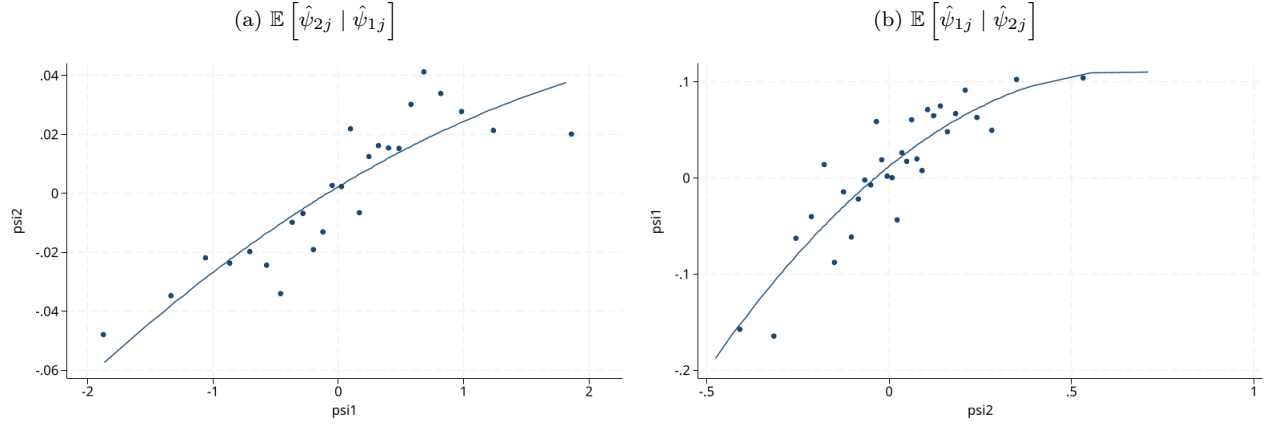


Figure 3: Binscatters of polynomial fit to opposite branch

It is clear from the figure that the slope $\frac{d}{dx} \hat{m}_b(x)$ is less than one for both branches $b \in \{1, 2\}$, which is an example of the usual shrinkage phenomenon in optimal prediction with noisy measurements. Rather than rely on one predictor or the other, it makes sense to average them. For any j , an improved predictor of ψ_j is the average $\bar{m}(\hat{\psi}_{1j}, \hat{\psi}_{2j}) = \frac{1}{2} \hat{m}_1(\hat{\psi}_{2j}) + \frac{1}{2} \hat{m}_2(\hat{\psi}_{1j})$.

This idea generalizes to the case with more than two branches by regressing the elements of each branch estimate on the corresponding entries of all remaining branches. For example, with $M = 3$, one would estimate the functions $\mathbb{E} [\hat{\psi}_{b_1j} | \hat{\psi}_{b_2j} = x_2, \hat{\psi}_{b_3j} = x_3] \equiv m_{b_1}(x_2, x_3)$ across all three distinct triples of branches $(b_1, b_2, b_3) \in \{1, 2, 3\} : b_1 \neq b_2 \neq b_3$. The resulting average prediction across branches can be written

$$\bar{m}(\hat{\psi}_{1j}, \hat{\psi}_{2j}, \hat{\psi}_{3j}) = \frac{1}{3} \hat{m}_1(\hat{\psi}_{2j}, \hat{\psi}_{3j}) + \frac{1}{3} \hat{m}_2(\hat{\psi}_{1j}, \hat{\psi}_{3j}) + \frac{1}{3} \hat{m}_3(\hat{\psi}_{1j}, \hat{\psi}_{2j}).$$

Note that if the noise in the branches were identically distributed it would necessarily be the case that $m_1(x_2, x_3) = m_2(x_2, x_3) = m_3(x_2, x_3)$, in which case one might as well collapse the left out branch estimates down to their mean, and estimate the pooled leave-out conditional expectation function $\mathbb{E} [\hat{\psi}_{b_1j} | \hat{\psi}_{b_2j} + \hat{\psi}_{b_3j} = x]$.

A recent paper by Ignatiadis et al. (2023) establishes that it is possible to improve on such collapsed estimators when the noise in each branch is identically distributed and non-Gaussian. Their proposal involves regressing the estimates in each branch b on *sorted values* of the estimates in all other branches. In the case of $M = 3$ branches, one would first regress $\hat{\psi}_{1j}$ on $\min\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\}$ and $\max\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\}$ to estimate the function $\mathbb{E} [\psi_j | \min\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\} = \underline{x}, \max\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\} = \bar{x}] \equiv h_1(\underline{x}, \bar{x})$. Next, $\hat{\psi}_{2j}$ is regressed on $\min\{\hat{\psi}_{1j}, \hat{\psi}_{3j}\}$ and $\max\{\hat{\psi}_{1j}, \hat{\psi}_{3j}\}$ to estimate $h_2(\underline{x}, \bar{x})$. Finally, $\hat{\psi}_{3j}$ is regressed on $\min\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\}$ and $\max\{\hat{\psi}_{2j}, \hat{\psi}_{3j}\}$ to estimate $h_3(\underline{x}, \bar{x})$. Averaging these three cross-branch fits $(\hat{h}_1, \hat{h}_2, \hat{h}_3)$ yields the AURORA estimator $\bar{h}$, which Ignatiadis et al. (2023) show performs nearly as well as an oracle that knows the noise distribution.

The rationale for using sorted values as regressors stems from the observation that order statistics are sufficient for iid samples. In the present setting, the noise distribution likely differs across branches, which may undermine the advantages of relying on order statistics.

Table 6 summarizes the results of a forecasting exercise in which the first three branch estimates $(\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3)$ of the 6-ECC are used to form best predictors of ψ . I compute both the AURORA estimator and the simpler regression based estimator $\bar{m}(\hat{\psi}_{1j}, \hat{\psi}_{2j}, \hat{\psi}_{3j})$ that ignores the order of the estimates. The conditional expectation functions $\{\hat{m}_b\}_{b=1}^3$ and $\{\hat{h}_b\}_{b=1}^3$ underlying these approaches are computed via non-parametric series regressions on B-spline bases tuned by cross-validation using Stata’s `npregress` function. For comparison, I also report the naive predictor $\bar{\psi} = (\hat{\psi}_1 + \hat{\psi}_2 + \hat{\psi}_3)/3$, which serves as an unshrunk benchmark. To assess the quality of these forecasts, the predictions are compared to $\hat{\psi}_4$, which should be unbiased for ψ . For each predictor $\hat{\psi} \in \{\bar{\psi}, \bar{m}, \bar{h}\}$, the mean squared error of the prediction is computed as $MSE(\hat{\psi}_4, \hat{m}) = (\hat{\psi}_4 - \hat{\psi})'(\hat{\psi}_4 - \hat{\psi})/J$.

	Naive (ψ)	Ignore order ($\bar{m}$)	AURORA ($\bar{h}$)
MSE	0.129	0.041	0.042

Table 6: Predicting $\hat{\psi}_4$ using $(\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3)$

As expected, the naive predictor incurs a large MSE because of the noise in the branches. Relative to this benchmark, both the AURORA estimator and its alternative that ignores order yield dramatic improvements due to shrinkage. However, the order statistic approach does not appear to convey any advantage over the simpler approach based on levels. Whether the disappointing performance of AURORA is primarily attributable to heteroscedasticity across branches or other factors (e.g., nearly Gaussian noise) is an interesting question for future research.

8 Conclusion

As these examples illustrate, branches can dramatically simplify the work of parsing signal from noise in fixed effects estimates. By quantifying uncertainty, branches also enable advanced downstream tasks such as moment estimation and shrinkage that form key aspects of the empirical bayes toolkit (Walters, 2024).

More generally, breaking over-identified estimates down into simpler branches is appealing on transparency grounds. In this analysis, the branches corresponding to trees of the mobility graph are simple linear transformations of the cell means. In addition to demystifying the origin of the full-sample estimates being reported, branches may be useful for assessing the degree to which over-identifying restrictions are violated in practice, a topic that I leave for future work.

As our discussion of the Nash-Williams-Tutte theorem revealed, the number of branches that can be extracted from a dataset depends crucially on the edge connectivity of the mobility network. The findings reported here suggest that a useful approximation to a k -ECC can be had from a networks' k -core. In many settings (e.g., studies of migration, bilateral trade, or friendship networks) the average degree of graph vertices is very high, which should enable the extraction of dozens of branches without substantial pruning of the network. The development of algorithms capable of rapidly extracting all spanning trees from enormous datasets with high degree is an interesting area for future research.

References

- Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). *Computing person and firm effects using linked longitudinal employer-employee data*. Tech. rep. Center for Economic Studies, US Census Bureau.
- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). “High wage workers and high wage firms”. In: *Econometrica* 67.2, pp. 251–333.
- Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). “High wage workers and low wage firms: negative assortative matching or limited mobility bias?” In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.3, pp. 673–697.
- Barahona, F. (1995). “Packing spanning trees”. In: *Mathematics of Operations Research* 20.1, pp. 104–115.
- Bellmann, L., B. Lochner, S. Seth, and S. Wolter (2020). *AKM effects for German labour market data*. Tech. rep. Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for ...
- Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018). “The disappearing large-firm wage premium”. In: *AEA Papers and Proceedings*. Vol. 108, pp. 317–22.
- Card, D., J. Heining, and P. Kline (2015). “CHK effects”. In: *FDZ-Methodenreport* 6, p. 2015.
- Card, D., J. Rothstein, and M. Yi (2024). “Industry wage differentials: A firm-based approach”. In: *Journal of Labor Economics* 42.S1, S11–S59.
- Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024). “On binscatter”. In: *American Economic Review* 114.5, pp. 1488–1514.
- Chamberlain, G. E. (2013). “Predictive effects of teachers and schools on test scores, college attendance, and earnings”. In: *Proceedings of the National Academy of Sciences* 110.43, pp. 17176–17182.
- Cheng, X., S. C. Ho, and F. Schorfheide (2025). “Optimal estimation of two-way effects under limited mobility”. In: *arXiv preprint arXiv:2506.21987*.
- Chetty, R. and N. Hendren (2018). “The impacts of neighborhoods on intergenerational mobility II: County-level estimates”. In: *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

- Csárdi, G. and T. Nepusz (2023). *igraph: Network Analysis and Visualization*. R package version 1.4.1. URL: <https://igraph.org>.
- Drenik, A., S. Jäger, P. Plotkin, and B. Schoefer (2023). “Paying outsourced labor: Direct evidence from linked temp agency-worker-client data”. In: *Review of Economics and Statistics* 105.1, pp. 206–216.
- Goldschmidt, D. and J. F. Schmieder (2017). “The rise of domestic outsourcing and the evolution of the German wage structure”. In: *The Quarterly Journal of Economics*, qjx008.
- Gomory, R. E. and T. C. Hu (Dec. 1961). “Multi-Terminal Network Flows”. In: *Journal of the Society for Industrial and Applied Mathematics* 9.4, pp. 551–570. DOI: 10.1137/0109047.
- Gusfield, D. (1990). “Very simple methods for all pairs network flow analysis”. In: *SIAM Journal on Computing* 19.1, pp. 143–155. DOI: 10.1137/0219009.
- He, J. and J.-M. Robin (2025). “Ridge Estimation of High Dimensional Two-way Fixed Effect Regression”. In: *working paper*.
- Ignatiadis, N., S. Saha, D. L. Sun, and O. Muralidharan (2023). “Empirical Bayes mean estimation with nonparametric errors via order statistic regression on replicated data”. In: *Journal of the American Statistical Association* 118.542, pp. 987–999.
- Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2024). “Worker beliefs about outside options”. In: *The Quarterly Journal of Economics*, qjae001.
- Jochmans, K. and M. Weidner (2019). “Fixed-effect regressions on network data”. In: *Econometrica* 87.5, pp. 1543–1560.
- Kirchhoff, G. R. (1847). “Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird”. In: *Annalen der Physik und Chemie* 72.12, pp. 497–508.
- Kline, P. (2025). “Firm Wage Effects”. In: *Handbook of Labor Economics*.
- Kline, P., R. Saggio, and M. Sølvesten (2020). “Leave-out estimation of variance components”. In: *Econometrica* 88.5, pp. 1859–1898.
- Kundu, S. (1974). “Bounds on the number of disjoint spanning trees”. In: *Journal of Combinatorial Theory, Series B* 17.2, pp. 199–203. DOI: 10.1016/S0095-8956(74)90087-2.
- Kwon, S. (2023). “Optimal shrinkage estimation of fixed effects in linear panel data models”. In: *arXiv preprint arXiv:2308.12485*.
- Menger, K. (1927). “Zur allgemeinen kurventheorie”. In: *Fundamenta Mathematicae* 10.1, pp. 96–115.
- Mogstad, M., J. P. Romano, A. M. Shaikh, and D. Wilhelm (2024). “Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries”. In: *Review of Economic Studies* 91.1, pp. 476–518.

- Nagamochi, H. and T. Ibaraki (1992a). “A linear-time algorithm for finding a sparse k -connected spanning subgraph of ak -connected graph”. In: *Algorithmica* 7.1, pp. 583–596.
- (1992b). “Computing edge-connectivity in multigraphs and capacitated graphs”. In: *SIAM Journal on Discrete Mathematics* 5.1, pp. 54–66.
- Nash-Williams, C. S. J. A. (1961). “Edge-Disjoint Spanning Trees of Finite Graphs”. In: *Journal of the London Mathematical Society*. Series 1 36.1, pp. 445–450.
- Roskind, J. and R. E. Tarjan (1985). “A Note on Finding Minimum-Cost Edge-Disjoint Spanning Trees”. In: *Mathematics of Operations Research* 10.4, pp. 701–708. DOI: 10.1287/moor.10.4.701.
- Silver, D. (2021). “Haste or waste? Peer pressure and productivity in the emergency department”. In: *The Review of Economic Studies* 88.3, pp. 1385–1417.
- Sorkin, I. (2018). “Ranking firms using revealed preference”. In: *The quarterly journal of economics* 133.3, pp. 1331–1393.
- Spielman, D. (2019). *Spectral and algebraic graph theory*. URL: <http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf>.
- Tutte, W. T. (1961). “On the Problem of Decomposing a Graph into n Connected Factors”. In: *Journal of the London Mathematical Society*. Series 1 36.1, pp. 221–230.
- U.S. Bureau of Labor Statistics (Jan. 2025). *Consumer Price Index: Calculation*. <https://www.bls.gov/opub/hom/cpi/calculation.htm>. Last modified January 30, 2025; accessed June 13, 2025.
- U.S. Census Bureau (Jan. 2025). *American Community Survey Variance Replicate Estimate Tables*. <https://www.census.gov/programs-surveys/acs/data/variance-tables.html>. Last revised January 6, 2025; accessed June 13, 2025.
- Walters, C. R. (2024). “Empirical Bayes Methods in Labor Economics”. In: *Handbook of Labor Economics*.

Appendix A: An example where greedy extraction fails

This appendix gives an example of how the order in which spanning trees are extracted from a graph can influence the total number of trees packed. The graph depicted in Figure A.1 has 6 vertices and 30 edges. Note that every vertex has degree 5. Hence, this is a 5-ECC. By the Nash-Williams-Tutte theorem, the graph must contain at least $\lfloor 5/2 \rfloor = 2$ trees. As we will see, however, this graph's actual tree packing number, $\tau(G)$, is 3.



Figure A.1: Two attempts to pack the same graph

The perils of greedy sequential extraction are depicted in panel (a), which shows a possible choice of two initial spanning trees that have been colored blue and red. The remaining edges fail to connect vertex 1, implying no remaining trees can be extracted after these two have been removed. Panel (b) demonstrates that, in fact, three edge-disjoint spanning trees that can be packed into the graph. While the three shaded trees depicted are not unique, applying the Roskind-Tarjan algorithm to this 5-ECC with $M = 3$ would ensure that three spanning trees are found.

Appendix B: Estimating the variance of $\hat{\mu}_2$

The unbiased second central moment estimator described in Section 7.2 can be written

$$\hat{\mu}_2 = \binom{M}{2}^{-1} \omega' \left(\sum_{b=1}^M \sum_{\ell < b} \hat{\mu}_{2,b,\ell} \right),$$

$$\hat{\mu}_{2,b,\ell} = \left(\hat{\psi}_b - \omega' \hat{\psi}_b \mathbf{1} \right) \odot \left(\hat{\psi}_\ell - \omega' \hat{\psi}_\ell \mathbf{1} \right).$$

From independence across branches we have

$$\begin{aligned} \mathbb{V}[\hat{\mu}_2] &= \binom{M}{2}^{-2} \omega' \mathbb{V} \left[\sum_{b=1}^M \sum_{\ell < b} \hat{\mu}_{2,b,\ell} \right] \omega \\ &= \binom{M}{2}^{-2} \omega' \left\{ \sum_{b=1}^M \sum_{\ell < b} \mathbb{V}[\hat{\mu}_{2,b,\ell}] + 2 \sum_{b_1=1}^M \sum_{b_2 < b_1} \sum_{b_3 < b_2} \left(\mathbb{E}[\hat{\mu}_{2,b_1,b_2} \hat{\mu}'_{2,b_1,b_3}] + \mathbb{E}[\hat{\mu}_{2,b_1,b_2} \hat{\mu}'_{2,b_2,b_3}] \right) \right\} \omega. \end{aligned}$$

Because each of the $\binom{M}{2}$ terms $\hat{\mu}_{2,b,\ell}$ is independent except when they share a branch, this variance expression separates into two parts: the variance of any single $\hat{\mu}_{2,b,\ell}$, and the covariance of two such terms when they overlap in exactly one branch. The first sort of term can be estimated by examining the covariance of disjoint pairs (b_1, b_2) and (b_3, b_4) as

$$\mathbb{E}[(\hat{\mu}_{2,b_1,b_2} - \hat{\mu}_{2,b_3,b_4})(\hat{\mu}_{2,b_1,b_2} - \hat{\mu}_{2,b_3,b_4})'] = \mathbb{V}[\hat{\mu}_{2,b_1,b_2}] + \mathbb{V}[\hat{\mu}_{2,b_3,b_4}].$$

The covariance terms can be estimated directly by their sample counterparts.

Averaging these estimated terms together yields an unbiased estimator for the variance of $\hat{\mu}_2$:

$$\begin{aligned} \hat{\mathbb{V}}[\hat{\mu}_2] &= \binom{M}{2}^{-2} \omega' \hat{V} \omega, \\ \hat{V} &= \frac{1}{3 \binom{M}{4}} \sum_{\substack{b_1 > b_2, b_3 > b_4 \\ (b_1, b_2) \cap (b_3, b_4) = \emptyset}} (\hat{\mu}_{2,b_1,b_2} - \hat{\mu}_{2,b_3,b_4})(\hat{\mu}_{2,b_1,b_2} - \hat{\mu}_{2,b_3,b_4})' \\ &\quad + \frac{2}{J} \sum_{1 \leq b_3 < b_2 < b_1 \leq M} (\hat{\mu}_{2,b_1,b_2} \hat{\mu}'_{2,b_1,b_3} + \hat{\mu}_{2,b_1,b_2} \hat{\mu}'_{2,b_2,b_3}). \end{aligned}$$

As the $\binom{M}{4}$ term in this formula reveals, constructing $\hat{\mathbb{V}}[\hat{\mu}_2]$ requires $M \geq 4$.