Deconvolution in Networks

Estimating Distributions of Firm and Worker Effects*

Stéphane Bonhomme
University of Chicago

Adam Drozynski
University of Chicago

Thibaut Lamadon
University of Chicago

PRELIMINARY DRAFT
March 20, 2025

Abstract

We propose an estimator of the density of random coefficients in linear models with correlated heterogeneity. The setup covers network settings, a leading example being the estimation of the densities of worker and firm effects using matched employer-employee data. We show nonparametric identification of densities, and propose practical estimators based on a novel noise augmentation technique. Importantly, our approach does not restrict the dependence between the worker and firm effects and the network of worker-firm connections. We apply the method to Italian data on workers and firms from the Veneto region, and estimate marginal densities of worker and firm effects as well as joint densities.

JEL codes: C10. C50.
Keywords: Nonparametric identification, estimation, worker and firm effects, sorting.

*We thank Pat Kline, Lihua Lei, Evan Rose, and participants at various places for comments

1 Introduction

Linear regression models with fixed effects are increasingly used in economics. In many applications, researchers are interested in recovering the fixed effects. Leading examples are firm and worker effects in wage regressions (Abowd, Kramarz, and Margolis, 1999, AKM hereafter), and neighborhood effects for economic mobility (Chetty and Hendren, 2018).

Settings such as the AKM model have a network structure: effects are identified by exploiting movements across firms or neighborhoods (Bonhomme and Denis, 2024). The literature provides methods to recover the variance of effects, as well as covariances and regression coefficients (Andrews, Gill, Schank, and Upward, 2008, Kline, Saggio, and Sølvsten, 2020, Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler, 2023). The goal of this paper is to study the estimation of general distributional features of the effects in such contexts, including the densities of effects.

Our leading example is the study of how workers and firms impact wage determination in AKM. In the model, log wages are additive functions of worker and firm indicators, whose coefficients are the worker and firm effects. Key questions in the AKM literature include: How dispersed are the firm effects (often interpreted as the dispersion in firm-specific wage premia)? How correlated are worker and firm effects (often interpreted as reflecting sorting of heterogeneous workers into heterogeneous firms)? In this setting, recovering worker and firm effects requires exploiting movements between firms. The difference in log wages before and after a job move identifies the difference between the two firms' effects, and exploiting all job moves in the economy allows one to identify all firm effects in a connected set of firms (up to a normalization). Worker effects are identified in a similar way.

We argue that firm and worker effects are likely to be correlated, and to be correlated with the network of worker-firm connections. Such complex forms of dependence will arise if worker mobility and firm hiring decisions depend on worker and firm heterogeneity. As an example, this will happen in a job search model with firm heterogeneity (Mortensen, 2003). We use such a model as an illustration, and show that it implies that firm effects are increasing functions of firm productivity, firm productivities before and after a job search are correlated, and high-productivity firms are larger and have more movers than low-productivity ones. Hence, effects and network connections depend on each other in complex ways.

In this paper, our goal is to develop methods to recover marginal densities of worker and firm effects, and joint densities of effects, without restricting the conditional density of firm and worker effects given the network. Although the literature provides methods to estimate vari-

ances and covariances of effects, many important economic quantities depend on distributions. For example, an increase in the variance of firm effects over time has different implications depending on how the skewness changes, or how tail probabilities evolve. Documenting the joint density of worker and firm effects, or of a worker's effect with the effects of her coworkers, sheds light on sorting along the distribution. Moreover, comparing firm effects before and after a job move has implications for search models of the labor market.

However, recovering densities of effects is challenging for two main reasons. The first difficulty arises because of the dependence of the effects on the network. While the deconvolution literature provides nonparametric estimators of densities, the extensions of the classical model (Stefanski and Carroll, 1990, Fan, 1991) rely on restrictive assumptions about heteroskedasticity and dependence (e.g., Delaigle and Meister, 2008). Estimators of densities in panel data random coefficients models (Arellano and Bonhomme, 2012) similarly rule out dependence across individual units.

We overcome the dependence issue by focusing on the *marginal* densities of the effects, and showing that those can be recovered while leaving the *conditional* density of effects given the network unrestricted. Using this insight, we provide nonparametric estimators of characteristic functions, moments, and densities.

Our focus on marginal densities is motivated by their relevance in economic applications. This focus differs from the literature on Empirical Bayes (EB) shrinkage (Efron and Morris, 1973, Chen, 2023), since the conditional density of effects given the conditioning network would be required to compute EB predictors. A recent contribution by Cheng, Sheng, and Schorfheide (In Preparation) proposes a model of the conditional density and applies to it compute EB predictors.

The second difficulty is practical. In settings such as AKM regression models, the signal-to-noise ratio is often too low for nonparametric estimators to be reliable. To develop a method that performs well in those cases, we impose modeling assumptions on the marginal densities of worker and firm effects. Specifically, we rely on the log-spline approach of Efron (2016). However, due to the rich dependence structure that we wish to allow for, it is not possible to directly estimate the densities by maximum likelihood or other existing method. Indeed, the conditional density of effects given the network is a high-dimensional, nonparametric nuisance parameter in our model.

We address this second difficulty by introducing a *noise augmentation technique*. We show that, by adding artificial noise to the estimates of the effects, the model simplifies to one with independent homoskedastic noise for each individual unit. Based on the noise-augmented

estimates, we then estimate the model by following the penalized maximum likelihood approach proposed by Efron (2016).

We find that our density estimator performs well in simulations calibrated to real-world AKM settings. We report results for marginal densities of firm and worker effects, and for joint densities. We then apply our approach to worker-firm data from the Veneto region in Italy. Marginal densities of firm and worker effects are clearly non-normal, and different from kernel-smoothed densities based on AKM estimates. Joint densities of worker and firm effects indicate the presence of different sorting patterns along the distribution.

While our leading application is to worker and firm effects in wage determination, our methods can be used in other settings. For example, Chetty and Hendren (2018) estimate the effects of neighborhoods in the US (such as counties or commuting zones) on income at adulthood. Their exposure-time research design implies that the estimate of neighborhood j's effect depends on the outcome data on other neighborhoods j'. In this setting, our approach allows us to estimate the density of neighborhood effects without restricting how neighborhood effects and mobility patterns across neighborhoods depend on each other.

The outline of the rest of the paper is as folows. In Section 2 we describe the model. We study identification and nonparametric estimation in Section 3. We present our semi-parametric estimator based on noise augmentation in Section 4, and discuss its implementation in Section 5. We describe the application to worker-firm settings and report empirical results and simulations in Sections 6 and 7.

Finally, in Section 8 we describe a wage posting model to illustrate the role of dependence, and how the method can be combined with structural assumptions to recover policy-relevant quantities such as the density of firm-specific mark-downs.

2 Model of heterogeneous effects on a network

Consider the following linear random coefficient model

$$y_i = z_i'\boldsymbol{\eta} + x_i'\boldsymbol{\beta} + u_i, \quad i = 1, ..., n, \tag{1}$$

where $\boldsymbol{\eta} = (\eta_1, ..., \eta_m)'$ is an $m \times 1$ vector that contains the coefficients of interest, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients of control variables. This model has many applications in economics, see Bonhomme and Denis (2024) for a survey.

In vector form, model (1) reads

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}, \tag{2}$$

where $\boldsymbol{Y}$ and $\boldsymbol{U}$ are $n \times 1$ vectors, and $\boldsymbol{Z}$ and $\boldsymbol{X}$ are $n \times m$ and $n \times k$ matrices, respectively.

In model (1), $\eta_1, ...\eta_m$ are drawn from an unrestricted joint distribution, conditional on $\boldsymbol{Z}$. For reference, we state this as an assumption.

Assumption 1.

$$\boldsymbol{\eta} \,|\, \boldsymbol{Z}, \boldsymbol{X} \sim g,$$

where g is unrestricted. We denote the implied marginal density of η_j as f_j, for all $j = 1, ..., m$.

While we treat the effects η_j as random we do not impose any assumption on them, except for the existence of a density. Importantly, we do not restrict the density g. In particular, we allow for two features that are important in economic applications. First, η_j and $\eta_{j'}$ may be correlated given $\boldsymbol{Z}, \boldsymbol{X}$. Second, $\boldsymbol{\eta}$ may correlate with $\boldsymbol{Z}, \boldsymbol{X}$.

Let $\boldsymbol{D}$ be an $(n-k) \times n$ matrix such that $\boldsymbol{DX} = 0$. Then (2) implies

$$\boldsymbol{DY} = \boldsymbol{DZ\eta} + \boldsymbol{DU}. \tag{3}$$

We make the following assumption, which is required for the η_j's to be identified.

Assumption 2. *$\boldsymbol{DZ}$ has rank m.*

The OLS estimator of $\boldsymbol{\eta}$ based on (3) is then

$$\widehat{\boldsymbol{\eta}} = (\boldsymbol{Z'D'DZ})^{-1} \boldsymbol{Z'D'DY},$$

or, equivalently,

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \boldsymbol{V},$$

where

$$\boldsymbol{V} = (\boldsymbol{Z'D'DZ})^{-1} \boldsymbol{Z'D'DU} \tag{4}$$

is the estimation noise.

We make the following assumption about $\boldsymbol{V}$.

Assumption 3.

$$\boldsymbol{V} \,|\, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{Z}, \boldsymbol{X})$ does not depend on $\boldsymbol{\eta}$.

One can expect that the coefficient estimates $\widehat{\eta}_j$ are approximately normal under suitable conditions, which motivates Assumption 3. However, assuming exact normality is still restrictive. The assumption can be relaxed to some extent, and we will show how to proceed in a

future version of the paper. However, while the assumption that $\boldsymbol{V}$ is conditionally independent of $\boldsymbol{\eta}$ is not without loss of generality, it seems difficult to relax.

We will start by assuming that $\boldsymbol{\Sigma}$ is known. In practice, we estimate a parameterized version of it, in the spirit of Arellano and Bonhomme (2012), see Section 5. We denote the standard errors of $\widehat{\eta}_j$ as σ_j, for $j = 1, ..., m$; i.e., the diagonal elements of $\boldsymbol{\Sigma}^{\frac{1}{2}}$.

To motivate our treatment of η_j as draws from an unrestricted joint distribution, arbitrarily correlated with conditioning variables, consider our leading example. In the AKM model introduced by Abowd, Kramarz, and Margolis (1999), y_i is the log wage, z_i contains the firm indicators, and x_i contains the worker indicators. In this labor market setting, economic models of worker mobility and wage determination motivate thinking of g as arising from a network formation model with heterogeneity.

In Section 8 we describe such a model, following Burdett and Mortensen (1998) and Mortensen (2003), which implies the following three features. First, wages are increasing in firm productivity, which implies that firm effects are functions of firm productivities. Second, on-the-job search leads to correlation between the firm productivities before and after the move, which implies that η_j and $\eta_{j'}$ are correlated with each other conditional on $\boldsymbol{Z}, \boldsymbol{X}$. Lastly, high-productivity firms are larger and have more movers, which implies that η_j is correlated with $\boldsymbol{Z}, \boldsymbol{X}$.

More generally, in the AKM model and other related settings, economic decisions lead to complex forms of dependence involving the effects $\boldsymbol{\eta}$ and the conditioning variables $\boldsymbol{Z}, \boldsymbol{X}$. The goal of the paper is thus to propose a method for estimating densities of heterogeneity that accounts for general forms of dependence.

3 Nonparametric identification and estimation

In this section we show that the average density

$$\overline{f} = \frac{1}{m} \sum_{j=1}^{m} f_j \tag{5}$$

is identified in model (1), and we provide a nonparametric estimator of it. In the next section we will then describe a semi-parametric estimator that is better suited to deal with noisy empirical settings such as applications of AKM.

3.1 Identification

Let $\boldsymbol{i} = \sqrt{-1}$ be the imaginary number, and define the characteristic function of η_j, for all $j = 1, ..., m$, as

$$\psi_j(\tau) = \mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right].$$

We first show that the average characteristic function of the η_j's,

$$\psi(\tau) = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right],$$

is identified. All proofs can be found in Appendix A.

Proposition 1. *We have, for all τ,*

$$\psi(\tau) = \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right].$$

A by-product of Proposition 1 is that all moments of η_j, when they exist, are identified. Let

$$m_q = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[{\eta_j}^q\right]$$

be the average q-th moment of the η_j. We have

$$m_1 = \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\widehat{\eta}_j\right],$$
$$m_2 = \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\left(\widehat{\eta}_j^2 - \sigma_j^2\right)\right].$$

More generally, letting H_q be the q-th order ("probabilistic") Hermite polynomial, we have the following result.

Corollary 1.

$$m_q = \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\sigma_j^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right)\right].$$

Another implication of Proposition 1 is that the density $\overline{f}$ in (5) is identified.

Corollary 2.

$$\overline{f}(\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right]\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau.$$

As a result, the corresponding average distribution function is also identified, as (see Dattner, Goldenshluger, and Juditsky, 2011):

$$\overline{F}(\eta) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{\tau}\mathcal{I}\left[\psi(\tau)\exp\left(-\boldsymbol{i}\tau\eta\right)\right]d\tau,$$

where $\mathcal{I}(z)$ denotes the imaginary part of $z \in \mathbb{C}$. More generally, a nonlinear moment $w_H = \frac{1}{m}\sum_{j=1}^m \mathbb{E}[H(\eta_j)]$ is identified as

$$w_H = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(\tau)\mathcal{F}[H](-\tau)d\tau,$$

where $\mathcal{F}[H]$ is the Fourier transform of H.

3.2 Nonparametric estimation

Proposition 1 and Corollary 1 imply that the following estimators are unbiased for the characteristic function and moments of the η_j, respectively:

$$\widehat{\psi}(\tau) = \frac{1}{m}\sum_{j=1}^m \frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)},$$

and

$$\widehat{m}_q = \frac{1}{m}\sum_{j=1}^m \sigma_j^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right).$$

However, plugging in $\widehat{\psi}(\tau)$ in lieu of $\psi(\tau)$ in Corollary 2 is not feasible, since the resulting integral diverges. This is a common issue in nonparametric deconvolution (e.g., Stefanski and Carroll, 1990). A strategy to address it is to construct a regularized estimator. To do so, let $B_m > 0$ be some constant. A regularized estimator of the average density $\overline{f}(\eta)$ is

$$\widehat{f}(\eta) = \frac{1}{2\pi}\int_{-B_m}^{B_m} \frac{1}{m}\sum_{j=1}^m \frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau. \tag{6}$$

The presence of the regularization parameter B_m implies that $\widehat{f}(\eta)$ is not unbiased for $\overline{f}(\eta)$. However, by letting B_m tend to infinity as m increases, one can obtain a consistent estimator.

The following result characterizes the mean integrated squared error (MISE) of $\widehat{f}$ as a function of $\overline{f}$.

Proposition 2.

$$\mathbb{E}\left[\int\left(\widehat{f}(\eta) - \overline{f}(\eta)\right)^2 d\eta\right]$$
$$= \frac{1}{m^2}\sum_{j=1}^m \int_{-B_m}^{B_m} \mathbb{E}\left[\exp\left(\sigma_j^2\tau^2\right)\right]d\tau + 2\int_{B_m}^{+\infty}|\psi(\tau)|^2 d\tau + O\left(\frac{B_m}{m}\right) + \textit{Covariance Terms}, \tag{7}$$

where the expression for the covariance terms can be found in the proof.

Proposition 2 shows that the MISE consists of four terms, see (7). The last one reflects the dependence across observations, and it disappears when the η_j are mutually independent and the V_j are mutually independent (yet both may be related to $\boldsymbol{Z}$ and $\boldsymbol{X}$). The third term is typically negligible relative to the first two. The second term in (7) reflects the contribution of the bias of $\widehat{f}(\eta)$. It depends on the rate of decay of $\psi(\tau)$ as τ tends to infinity. In the "ordinary smooth" regime (Fan, 1991) where $|\psi(\tau)| \leqslant C|\tau|^{-\beta}$ for some $\beta > \frac{1}{2}$, this second term is proportinal to $B_m^{1-2\beta}$.

Lastly, the first term in (7) depends on quantities of the form $\int_{-B_m}^{B_m} \exp\left(\sigma_j^2 \tau^2\right) d\tau$, which grow exponentially fast as a function of B_m. This term reflects the variance of $\widehat{f}(\eta)$. Because of its presence, it is necessary to take B_m to be a logarithmic function of m in order for the MISE to tend to zero as m tends to infinity. However, since the second term in (7) is polynomial in B_m, this implies that the rate of the MISE is at best logarithmic in m. This is a standard feature in nonparametric deconvolution models with Gaussian errors. Moreover, notice that the error variances σ_j^2 greatly affect the first term in (7), since they enter the exponential. In network settings such as ours, some of the $\widehat{\eta}_j$ tend to be poorly estimated. This implies that some of the σ_j^2 are large, which adversely affects the convergence rate of the MISE.

In the settings we focus on in this paper, the convergence rate of the nonparametric density estimator $\widehat{f}(\eta)$ tends to be slow. One strategy to obtain an estimator with better convergence properties is to restrict the density g that relates the effects η_j to the network, i.e., to $\boldsymbol{Z}$ and $\boldsymbol{X}$. For example, Delaigle and Meister (2008) impose independence between η_j and σ_j^2, while Chen (2023) relaxes independence but relies on a location-scale model of the dependence. However, in our network setting with complex forms of dependence (as implied by common economic models), we wish to avoid such assumptions. We thus propose to take another route, as described in the next section.

4 Semi-parametric estimation

In this section we present our estimation strategy, where we model the marginal density of the η_j. We defer implementation issues, including how we estimate the error variances σ_j^2, until the next section.

4.1 A semi-parametric model

We model the average density $\overline{f}$ of $\eta_1, ..., \eta_m$ as a parametric function $f_{\boldsymbol{\theta}}(\eta)$ of a parameter vector $\boldsymbol{\theta} \in \Theta$.

Assumption 4. *There exists some* $\boldsymbol{\theta}_0 \in \Theta$ *such that* $\overline{f}(\eta) = f_{\boldsymbol{\theta}_0}(\eta)$.

An example of a parametric family, which we will use in our implementation, is the log-spline family used in Efron (2016):

$$f_{\boldsymbol{\theta}}(\eta) = \frac{\exp(P(\eta)'\boldsymbol{\theta})}{\int \exp(P(\widetilde{\eta})'\boldsymbol{\theta}) d\widetilde{\eta}}, \tag{8}$$

where $P(\eta)$ are polynomial splines. This flexible family of distributions has the advantage of leading to a log-concave likelihood function. However, since we do not model the conditional density of $\boldsymbol{\eta} \,|\, \boldsymbol{Z}$, but only the marginal density $\overline{f}$, the model is not fully parametric but only semi-parametric. As a result, it is not possible to directly use maximum likelihood estimation, as in Efron (2016) for example.

4.2 Noise augmentation

To introduce our estimation strategy, let $\boldsymbol{\Omega}$ be an $m \times m$ matrix satisfying the following two conditions.

Assumption 5. $\boldsymbol{\Omega}$ *is such that*
(i) $\boldsymbol{\Omega}$ *is positive semi-definite.*
(ii) All elements on the diagonal of $\boldsymbol{\Sigma} + \boldsymbol{\Omega}$ *are constant, equal to* $\overline{\sigma}^2$.

Given a matrix $\boldsymbol{\Omega}$ satisfying Assumption 5, let

$$\boldsymbol{\varepsilon} \,|\, \boldsymbol{\eta}, \boldsymbol{V}, \boldsymbol{Z}, \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{9}$$

and denote $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_m)'$. Our approach consists in adding the artificial noise ε_j to the estimates $\widehat{\eta}_j$. This leads to a simplification of the problem, which then becomes a classical deconvolution problem with independent normal homoskedastic errors (albeit with possibly dependent observations across j).

To see this, note that

$$\widehat{\boldsymbol{\eta}} + \boldsymbol{\varepsilon} = \boldsymbol{\eta} + (\boldsymbol{V} + \boldsymbol{\varepsilon}), \tag{10}$$

where, by Assumption 3 and Equation (9) we have

$$(\boldsymbol{V} + \boldsymbol{\varepsilon}) \,|\, \boldsymbol{\eta}, \boldsymbol{Z}, \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma} + \boldsymbol{\Omega}). \tag{11}$$

Since the diagonal elements of $\boldsymbol{\Sigma} + \boldsymbol{\Omega}$ are all equal to $\overline{\sigma}^2$, we thus have

$$(V_j + \varepsilon_j) \,|\, \boldsymbol{\eta}, \boldsymbol{Z}, \boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{0}, \overline{\sigma}^2\right), \quad j = 1, ..., m. \tag{12}$$

This suggests that a composite maximum likelihood approach (Lindsay, 1988), based on a model with independent normal homoskedastic errors, should be applicable.

Specifically, we propose to estimate $\boldsymbol{\theta}$ using the quasi-MLE estimator based on the composite log-likelihood

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \int \ln \left(\int f_{\boldsymbol{\theta}}(\eta) \phi_{\overline{\sigma}^2} \left(\widehat{\eta}_j + \varepsilon_j - \eta \right) d\eta \right) \phi_{\overline{\sigma}^2 - \sigma_j^2} \left(\varepsilon_j \right) d\varepsilon_j, \tag{13}$$

where ϕ_{σ^2} is the Gaussian density with mean zero and variance σ^2.

This method is quasi-MLE, as opposed to standard MLE, since we do not model the dependence between the elements $V_j + \varepsilon_j$ across j. Nevertheless, the following result shows that the expected log likelihood is maximized at the true parameter value $\boldsymbol{\theta}_0$.

Proposition 3. *For all $\boldsymbol{\theta} \in \Theta$ we have*

$$\mathbb{E}\left[L(\boldsymbol{\theta}) \right] \leqslant \mathbb{E}\left[L(\boldsymbol{\theta}_0) \right].$$

Example 1. *As an example, one can take $\overline{\sigma}^2 = \sigma^2_{\max}$ to be the maximal element on the diagonal of $\boldsymbol{\Sigma}$, that is, the maximal squared standard error of $\widehat{\eta}_j$ across $j = 1, ..., m$, and $\boldsymbol{\Omega}$ to be the diagonal matrix*

$$\boldsymbol{\Omega} = \sigma^2_{\max} \boldsymbol{I}_m - diag\left(\boldsymbol{\Sigma} \right), \tag{14}$$

where $\boldsymbol{I}_m$ is the identity matrix of size m. We follow this approach in our application.

Example 2. *As another example, one can take $\overline{\sigma}^2 = \lambda_{\max}$ to be the leading eigenvalue of $\boldsymbol{\Sigma}$. In that case, taking*

$$\boldsymbol{\Omega} = \lambda_{\max} \boldsymbol{I}_m - \boldsymbol{\Sigma}, \tag{15}$$

we obtain

$$\left(\boldsymbol{V} + \boldsymbol{\varepsilon} \right) \mid \boldsymbol{\eta}, \boldsymbol{Z}, \boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{0}, \lambda_{\max} \boldsymbol{I}_m \right). \tag{16}$$

To analyze the estimator when $\boldsymbol{\Omega}$ is chosen according to (15), suppose for simplicity that the η_j's are i.i.d. draws from $f_{\boldsymbol{\theta}_0}$. While this imposes a stronger restriction compared to Assumption 4, we do not further restrict the density of $\boldsymbol{\eta}$ given $\boldsymbol{Z}$. For example, the η_j could be i.i.d. firm productivity draws, and we do not restrict how workers sort across firms.

Suppose further than, as m tends to infinity, $\lambda_{\max}$ remains bounded. Note that this assumption may be restrictive, especially when the network is poorly connected (Jochmans and Weidner, 2019). Then, $\widehat{\boldsymbol{\theta}}$ is a usual maximum-likelihood estimator, which satisfies under standard regularity conditions

$$\sqrt{m} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\boldsymbol{0}} \right) \xrightarrow{d} \mathcal{N}\left(0, \left(\lim_{m \to \infty} \mathbb{E}\left[\frac{1}{m} \frac{\partial^2 L(\boldsymbol{\theta}_0)}{\partial \theta \partial \theta'} \right] \right)^{-1} \right).$$

5 Implementation

In this section we describe implementation issues for the density estimator that we have outlined in the previous section.

5.1 Penalized MLE

Following Efron (2016), we add a square penalty to the log-quasi-likelihood and minimize

$$-L(\boldsymbol{\theta}) + \xi\|\boldsymbol{\theta}\|^2$$

with respect to $\boldsymbol{\theta} \in \Theta$, where $\xi > 0$ is a penalty parameter and $\|\boldsymbol{\theta}\|$ denotes the Euclidean norm of the vector $\boldsymbol{\theta}$. Adding a penalty tends to improve the properties of the maximum likelihood estimator in finite samples, as we will document.

We set ξ by maximizing the marginal likelihood, as in Ruppert, Wand, and Carroll (2003), and aim to maximize

$$\int \exp\left(L(\boldsymbol{\theta})\right) \frac{1}{(2\pi)^{\frac{m}{2}}(2\xi)^{-\frac{m}{2}}} \exp\left(-\xi\|\boldsymbol{\theta}\|^2\right) d\boldsymbol{\theta}$$

with respect to $\xi \in \Xi$.

In practice, for computational reasons we use a Laplace approximation to the integral and compute

$$\widehat{\xi} = \underset{\xi\in\Xi}{\operatorname{argmax}} \left[L\left(\widehat{\boldsymbol{\theta}}(\xi)\right) - \frac{m}{2}\ln(2\pi) + \frac{m}{2}\ln(2\xi) - \xi\left\|\widehat{\boldsymbol{\theta}}(\xi)\right\|^2 + \frac{1}{2}\ln\det\left(\boldsymbol{H}(\xi)\right)\right], \tag{17}$$

where

$$\widehat{\boldsymbol{\theta}}(\xi) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left(-L(\boldsymbol{\theta}) + \xi\|\boldsymbol{\theta}\|^2\right)$$

is the penalized quasi-MLE of $\boldsymbol{\theta}$ for given ξ, and

$$\boldsymbol{H}(\xi) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\xi)} \left(L\left(\boldsymbol{\theta}\right) - \xi\left\|\boldsymbol{\theta}\right\|^2\right).$$

Our final estimator of $\overline{f}(\eta)$ is then

$$\widehat{f}(\eta) = f_{\widehat{\boldsymbol{\theta}}(\widehat{\xi})}(\eta).$$

Remark 1. *It is straightforward to modify our approach to estimate the weighted density of η_j, with weights $\omega_j \geqslant 0$ (for example, weighting firm effects by the number of person-year observations in the firm), simply by replacing (25) with*

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \omega_j \int \ln\left(\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(\widehat{\eta}_j + \varepsilon_j - \eta\right) d\eta\right) \phi_{\overline{\sigma}^2 - \sigma_j^2}\left(\varepsilon_j\right) d\varepsilon_j.$$

5.2 Estimation of error variances

Consider the linear model (2), under the following assumption.

Assumption 6. *We assume that*
$\mathbb{E}\left[\boldsymbol{U}\boldsymbol{U}' \mid \boldsymbol{Z}, \boldsymbol{X}\right] = \boldsymbol{\Gamma}_{\boldsymbol{\tau}}$, *where* $\boldsymbol{\Gamma}_{\boldsymbol{\tau}} = \boldsymbol{\Gamma}(\boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\tau})$ *depends on a parameter* $\boldsymbol{\tau}$.

An example satisfying Assumption 6 is the independent homoskedastic specification $\boldsymbol{\Gamma}_\tau = \tau^2 \boldsymbol{I}_n$, for a scalar parameter τ (as in Andrews, Gill, Schank, and Upward, 2008). More generally, Assumption 6 is compatible with any parametric model of the covariance matrix of errors (Arellano and Bonhomme, 2012).

Let $\boldsymbol{D}_2$ be a matrix such that $\boldsymbol{D}_2\boldsymbol{Z} = 0$ and $\boldsymbol{D}_2\boldsymbol{X} = 0$. By (2), we have $\boldsymbol{D}_2\boldsymbol{Y} = \boldsymbol{D}_2\boldsymbol{U}$. Hence, by Assumption 6, we obtain

$$\mathbb{E}\left[\boldsymbol{D}_2\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{D}_2' \mid \boldsymbol{Z}, \boldsymbol{X}\right] = \boldsymbol{D}_2 \mathbb{E}\left[\boldsymbol{U}\boldsymbol{U}' \mid \boldsymbol{Z}, \boldsymbol{X}\right] \boldsymbol{D}_2' = \boldsymbol{D}_2 \boldsymbol{\Gamma}_{\boldsymbol{\tau}} \boldsymbol{D}_2'.$$

As a result, given any $n \times n$ vector $\boldsymbol{c}$ (possibly a function of $\boldsymbol{Z}$ and $\boldsymbol{X}$), we obtain the moment restriction

$$\mathbb{E}\left[\boldsymbol{c}'\left((\boldsymbol{D}_2\boldsymbol{Y}) \otimes (\boldsymbol{D}_2\boldsymbol{Y}) - (\boldsymbol{D}_2 \otimes \boldsymbol{D}_2)\boldsymbol{\gamma}_{\boldsymbol{\tau}}\right)\right] = 0,$$

where $\boldsymbol{\gamma}_{\boldsymbol{\tau}} = \mathrm{vec}(\boldsymbol{\Gamma}_{\boldsymbol{\tau}})$. We estimate $\widehat{\boldsymbol{\tau}}$ using GMM, based on a set of such moment restrictions.

For example, in the independent homoskedastic case, taking $\boldsymbol{c} = \mathrm{vec}(\boldsymbol{I}_n)$ gives

$$\tau^2 = \frac{\mathbb{E}[\boldsymbol{Y}'\boldsymbol{D}_2'\boldsymbol{D}_2\boldsymbol{Y}]}{\mathbb{E}[\mathrm{Trace}(\boldsymbol{D}_2\boldsymbol{D}_2')]},$$

which is the population counterpart to the estimator in Andrews, Gill, Schank, and Upward (2008).

Next, note that (4) implies that

$$\boldsymbol{\Sigma} = \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{\Gamma}_{\boldsymbol{\tau}}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1}. \tag{18}$$

Hence, given an estimator $\widehat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$, we produce the following estimator of $\boldsymbol{\Sigma}$,

$$\widehat{\boldsymbol{\Sigma}} = \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{\Gamma}_{\widehat{\boldsymbol{\tau}}}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1}. \tag{19}$$

Remark 2. *In the presentation so far we have assumed that* $\mathrm{rank}(\boldsymbol{D}\boldsymbol{Z}) = m$. *However, in applications this rank condition may fail. For example, in the AKM model, securing identification requires restricting the analysis to a connected component of the worker-firm graph, and imposing a normalization.*

Suppose we are working within a connected component, and $\text{rank}(\boldsymbol{DZ}) = m - 1$. *In this case, following the literature we simply normalize one of the firm coefficients to zero. However, in our context this approach tends to create imbalances of the* σ_j^2 *along the network.*[1] *For this reason we focus on the de-meaned* η_j*'s, for which the covariance matrix estimate does not depend on the normalization. Specifically, removing a column in* $\boldsymbol{DZ}$ *and letting* $\boldsymbol{M} = \boldsymbol{I}_m - \boldsymbol{J}_m/m$, *for* $\boldsymbol{J}_m$ *the* $m \times m$ *matrix of ones, we construct*

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{M} \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{\Gamma}_{\widehat{\boldsymbol{\tau}}}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1} \boldsymbol{M}'. \tag{20}$$

Remark 3. *Kline, Saggio, and Sølvsten (2020) propose a leave-out estimator of quadratic forms of* $\boldsymbol{\eta}$ *that is unbiased under arbitrary forms of heteroskedasticity, under the assumption that errors are mutually independent. We use their estimator to assess our specification of the covariance matrix of* $\boldsymbol{U}$. *Specifically, for a symmetric* $m \times m$ *matrix* $\boldsymbol{Q}$ *we denote the leave-out estimator of* $V_{\boldsymbol{Q}} = \boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\eta}$ *as* $\widehat{V}_{\boldsymbol{Q}}^{\text{LO}}$, *and construct the difference*

$$\widehat{\boldsymbol{\eta}}'\boldsymbol{Q}\widehat{\boldsymbol{\eta}} - \text{Trace}\left(\boldsymbol{Q}\widehat{\boldsymbol{\Sigma}}\right) - \widehat{V}_{\boldsymbol{Q}}^{\text{LO}}. \tag{21}$$

Assuming that the u_i*'s are independent across* i, *a large absolute value of the difference in (21) relative to its sampling variation indicates misspecification of the parametric specification given by Assumption 6.*

5.3 Computation

In this subsection we describe our computational implementation, tailored to the problem of estimating the density of firm or worker effects in the AKM model.

First, we select a connected component of the worker-firm graph (for example, the largest one). We then focus on de-meaned effects $\boldsymbol{M}\boldsymbol{\eta}$, where $\boldsymbol{M} = \boldsymbol{I}_m - \boldsymbol{J}_m/m$, which with some abuse of notation here we simply refer to as $\boldsymbol{\eta}$, with associated estimate $\widehat{\boldsymbol{\eta}}$.

Assuming independent homoskedastic errors, we compute their variance as (Andrews, Gill, Schank, and Upward, 2008):

$$\widehat{\tau}^2 = \frac{\boldsymbol{Y}'\boldsymbol{D}_2'\boldsymbol{D}_2\boldsymbol{Y}}{n - m - k + 1},$$

and the implied variance of the fixed-effects estimates as

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\tau}^2 \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{Z}\right)^{-1},$$

[1] Indeed, the normalized firm, say j_0, has a variance $\sigma_{j_0}^2 = 0$ by construction, and firms well-connected to it tend to have small σ_j's as a result, while firms less well-connected to j_0 tend to have larger σ_j's.

from which we extract the diagonal elements $\widehat{\sigma}_j^2$, for $j = 1, ..., m$. Alternatively, we can rely on a different parametric model of errors and use (19). We also set $\overline{\sigma}^2$, for example, equal to the maximum $\widehat{\sigma}_j^2$'s. Lastly, we specify $f_{\boldsymbol{\theta}}$ as in Efron (2016), using Equation (8) where $P(\eta)$ is specified as a cubic spline on a set of knots.

Based on these inputs, we first compute the penalty parameter $\widehat{\xi}$ in (17). To do so, we use a grid search method. Computing the objective function requires maximizing the penalized quasi log-likelihood $L(\boldsymbol{\theta}) - \xi\|\boldsymbol{\theta}\|^2$. We use gradient ascent for this purpose, taking advantage that the objective function is concave in the case of Efron's log-spline specification. We now explain how we compute accurate approximations to $L(\boldsymbol{\theta})$, its gradient vector, and its Hessian matrix.

Letting

$$g_{\boldsymbol{\theta}}(u) = \ln\left(\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(u-\eta\right)d\eta\right),$$

we have, using the change in variables $v = \widehat{\eta}_j + \varepsilon_j$,

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{j=1}^{m}\int g_{\boldsymbol{\theta}}(\widehat{\eta}_j+\varepsilon_j)\phi_{\overline{\sigma}^2-\widehat{\sigma}_j^2}\left(\varepsilon_j\right)d\varepsilon_j = \sum_{j=1}^{m}\int g_{\boldsymbol{\theta}}(v)\phi_{\overline{\sigma}^2-\widehat{\sigma}_j^2}\left(v-\widehat{\eta}_j\right)dv \\ &= \int g_{\boldsymbol{\theta}}(v)\left(\sum_{j=1}^{m}\phi_{\overline{\sigma}^2-\widehat{\sigma}_j^2}\left(v-\widehat{\eta}_j\right)\right)dv = \int g_{\boldsymbol{\theta}}(v)\frac{\left(\sum_{j=1}^{m}\phi_{\overline{\sigma}^2-\widehat{\sigma}_j^2}\left(v-\widehat{\eta}_j\right)\right)}{\phi_{s^2}(v)}\phi_{s^2}(v)dv, \end{aligned}$$

where s^2 is the variance of the proposal distribution, set by the researcher. Therefore, $L(\boldsymbol{\theta})$ can be approximated with Gauss-Hermite quadrature with nodes v_k and weights w_k, as

$$\begin{aligned} L(\boldsymbol{\theta}) &\approx \sum_{k=1}^{K} g_\theta(v_k)\underbrace{w_k\frac{\sum_{j=1}^{m}\phi_{\overline{\sigma}^2-\widehat{\sigma}_j^2}\left(v_k-\widehat{\eta}_j\right)}{\phi_{\upsilon}(v_k)}}_{\widetilde{w}_k} \\ &= \sum_{k=1}^{K}\widetilde{w}_k\ln\left(\int f_\theta(\eta)\phi_{\overline{\sigma}^2}(v_k-\eta)d\eta\right). \end{aligned} \tag{22}$$

Equation (22) shows how the Gauss-Hermite-based approximation leads to a weighted version of the log-likelihood for parametric deconvolution (Efron, 2016), where the quadrature nodes v_k play the role of quasi-observations associated with the weights $\widetilde{w}_k$.

We initially choose a very large number of quadrature nodes, 100,000 in simulations. Then, before the quasi-likelihood maximization procedure, we remove those for which the weights are negligible (smaller than $1e-8$). Generating the nodes and computing $\widetilde{w}_k$ is very fast. We choose a proposal variance s^2 that is large relative to the variance of the distribution of $\widehat{\eta}_j + \varepsilon_j$, but not too large to avoid numerical issues. In simulations we use $s^2 = 10(\frac{1}{m}\sum_{j=1}^{m}(\widehat{\eta}_j - \overline{\widehat{\eta}})^2 +$

$\frac{1}{m}\sum_{j=1}^{m}(\overline{\sigma}^2 - \widehat{\sigma}_j^2)$). Across simulations, we found that these choices work well, and that the method is not very sensitive to them.

To approximate $g_{\boldsymbol{\theta}}(u)$, which is itself an integral, we use the discretization strategy proposed by Efron (2016). Lastly, we use a similar strategy to compute accurate approximations to the gradient and Hessian of $L(\boldsymbol{\eta})$.

6 Application to workers and firms

6.1 Quantities of interest in the AKM model

The AKM model proposed by Abowd, Kramarz, and Margolis (1999) specifies how the log wage y_{it} of worker i in period t, who is employed in firm $j(i,t)$, depends on a worker fixed effect α_i, a firm fixed effect $\psi_{j(i,t)}$, and an idiosyncratic error term u_{it},

$$y_{it} = \alpha_i + \psi_{j(i,t)} + u_{it}, \quad i = 1, ..., N,\ t = 1, ..., T. \tag{23}$$

To see how to recover densities of worker and firm effects in this model, suppose that the density of interest is the marginal density of the firm effects ψ_j, for firms $j = 1, ..., J$. One then simply defines z_i as the firm indicators, and x_i as the worker indicators. Likewise, suppose that the density of interest is the marginal density of the worker effects α_i, for workers $i = 1, ..., N$. In this case one defines z_i as the worker indicators, and x_i as the firm indicators.

One can also use our approach based on model (1) to estimate bivariate densities. Such densities are of interest in many settings. One example is the joint density of $(\alpha_i, \psi_{j(i,t)})$. A common interpretation in the literature of the covariance $\text{Cov}(\alpha_i, \psi_{j(i,t)})$ is that it reflects sorting between workers and firms. The joint density of $(\alpha_i, \psi_{j(i,t)})$ provides more information about the nature of sorting, for example, whether sorting happens mostly at the top or bottom of the distribution.

Another joint density of interest in applications is that of the effects α_i and $\alpha_{i'}$ or two workers i, i' in the same firm (i.e., such that $j(i,t) = j(i',t)$). This density is informative about the nature of worker's employment choices. It is informative about which workers tend to work together, or in different firms.

A third density of interest is the joint density of the firm effects before and after a job change, that is, the density of $(\psi_{j(i,t)}, \psi_{j(i,t+1)})$ when $j(i,t+1) \neq j(i,t)$. This density is informative about models of job mobility, and speaks to search models of the labor market where firm heterogeneity is a driver of dynamic job choices.

6.2 Estimating bivariate distributions

Our approach can be easily adapted to estimate bivariate distributions. We now outline how to proceed. For concreteness, consider the case where we wish to estimate the density of $(\alpha_i, \psi_{j(i,t)})$, although our approach is similar in the other examples we have just outlined. Denoting as $\widehat{\alpha}_i$ and $\widehat{\psi}_j$ the OLS estimates of α_i and ψ_j, we can write

$$\begin{pmatrix} \widehat{\alpha}_i \\ \widehat{\psi}_{j(i,t)} \end{pmatrix} = \begin{pmatrix} \alpha_i \\ \psi_{j(i,t)} \end{pmatrix} + \begin{pmatrix} v_{it}^{\alpha} \\ v_{it}^{\psi} \end{pmatrix}. \tag{24}$$

As in the univariate case, we obtain an estimate of the variance-covariance matrix $\mathbf{\Sigma}$ of the estimation errors $(v_{it}^{\alpha}, v_{it}^{\psi})'$ for all (i,t). Hence, $\mathbf{\Sigma}$ is a $(2NT) \times (2NT)$ matrix. We denote its 2×2 blocks on the diagonal as $\mathbf{\Sigma}_{it}$, i.e., the covariance matrix of $(v_{it}^{\alpha}, v_{it}^{\psi})$.

In the bivariate case, we replace Assumption 5 by the following.

Assumption 7. *$\mathbf{\Omega}$ is such that*
(i) $\mathbf{\Omega}$ is positive semi-definite.
(ii) All 2×2 blocks on the diagonal of $\mathbf{\Sigma} + \mathbf{\Omega}$ are constant, equal to $\overline{\mathbf{\Sigma}}$, where $\overline{\mathbf{\Sigma}}$ is a 2×2 matrix.

Given a $(2NT) \times (2NT)$ matrix $\mathbf{\Omega}$ satisfying Assumption 7, we then add artificial noise $\boldsymbol{\varepsilon}$ as in Equation (9) to the vector of estimates $\begin{pmatrix} \widehat{\alpha}_i \\ \widehat{\psi}_{j(i,t)} \end{pmatrix}$, and rely for estimation on the composite log-likelihood

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \iint \ln \left(\iint f_{\boldsymbol{\theta}}(\alpha, \psi) \phi_{\overline{\mathbf{\Sigma}}} \left(\widehat{\alpha}_i + \varepsilon_{it}^{\alpha} - \alpha, \widehat{\psi}_{j(i,t)} + \varepsilon_{it}^{\psi} - \psi \right) d\alpha d\psi \right) \phi_{\overline{\mathbf{\Sigma}} - \mathbf{\Sigma}_{it}} \left(\varepsilon_{it}^{\alpha}, \varepsilon_{it}^{\psi} \right) d\varepsilon_{it}^{\alpha} d\varepsilon_{it}^{\psi}, \tag{25}$$

where $\phi_{\widetilde{\mathbf{\Sigma}}}$ is the bivariate Gaussian density with mean zero and covariance matrix $\widetilde{\mathbf{\Sigma}}$.

For implementation, we rely on a bivariate extension of the log-spline specification proposed in Efron (2016), and we adapt the implementation described in Section 5 to the bivariate case.

7 Illustration using the Veneto data in Italy

7.1 Data and implementation

We use a sample from the Veneto region in Italy. The sample is the one used by Kline, Saggio, and Sølvsten (2020), and we refer to their paper for details about sample selection.[2] We provide descriptive statistics about the sample in Table 1. Note that there are fewer than four job movers per firm on average. Given this, we expect that the so-called "limited mobility bias" due to the noise in AKM estimates might be sizable.

Table 1: Summary Statistics

Number of Firms	42,492
Number of Movers	164,199
Number of Workers	660,019
Number of Observations	1,320,038
Mean log-wage	4.807
Variance of log-wages	0.184

We estimate densities on the largest connected component of the worker-firm graph. We report densities of de-meaned estimates. We implement the method as discussed in Section 5. All estimates are weighted by the number of person-year observations.

In this version of the paper we use a homoskedastic model for error variances in the log wage regression, as in Andrews, Gill, Schank, and Upward (2008). To estimate the common variance τ^2, we rely only on job movers since within-job observations are likely to be serially dependent. Given this approach, the bias-corrected variance components are close, though not identical, to the heteroskedastic bias-corrected components reported in Kline, Saggio, and Sølvsten (2020).

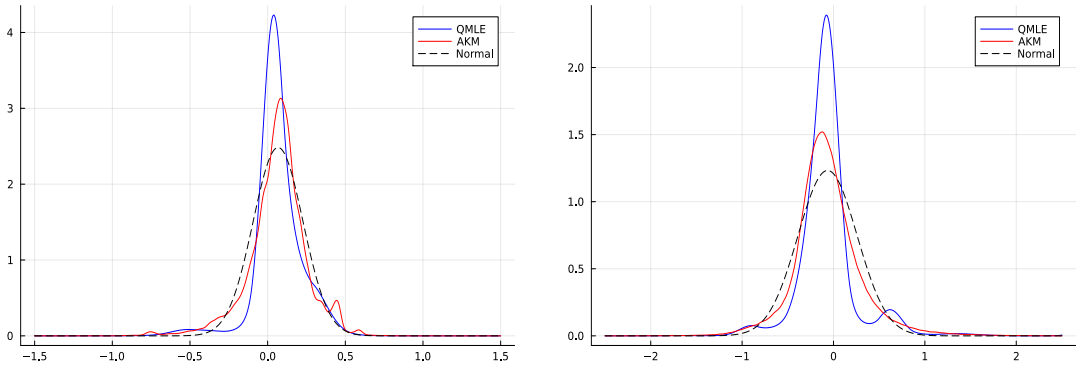
7.2 Univariate densities of worker and firm effects

We start by reporting estimates of densities of firm and worker effects. The left graph in Figure 1 shows that our estimate of the density of firm effects has less dispersion, and is more peaked, than the kernel-smoothed density of AKM estimates. We also notice that firm effects are not normally distributed.

[2] We focus on the largest leave-out connected component of the worker-firm graph. We were able to reproduce the sample in Kline, Saggio, and Sølvsten (2020) almost exactly, except for three firms.

Figure 1: Univariate densities of worker and firm effects.

(a) Firm effects (b) Worker effects



Notes: Estimates based on our composite quasi-likelihood method in blue, kernel-smoothed density estimator of AKM estimates in red, normal fit (to the density in blue) in dashed. Firm effects on the left, worker effects on the right.

The right graph in Figure 1 shows that our estimate of the density of worker effects is also less dispersed than the kernel-smoothed density of AKM estimates. We see that worker effects are not normally distributed either.

We report moment estimates for both worker and firm effects in Table 2. The bias-corrected moments based on the closed-form expression given in Corollary 1, and the ones implied by our density estimates, are close to each other. In contrast, uncorrected and bias-corrected moment estimates differ rather strongly from each other. The variance based on AKM estimates ("plug in") is 38% larger than the bias-corrected one for firm effects, 21% larger for worker effects. The other moments of the distribution are also affected by the bias correction. In particular, bias correction uncovers the presence of excess, non-Gaussian kurtosis for both worker and firm effects.

Lastly, we also report moment estimates under the assumption that $\boldsymbol{\eta}$ is independent of $\boldsymbol{\Sigma}$. We see that assuming independence gives different values, generally between the plug-in estimates and the estimates based on our approach that allows for dependence. This shows that assuming independence may be restrictive in applications.

Table 2: Estimated moments, person-year weighted.

	$Var[\psi]$	$Skew[\psi]$	$Kurt[\psi]$	$\mu_3[\psi]$	$\mu_4[\psi]$	$\mu_5[\psi]$
			Firm effects			
Plug-in	0.0358	−0.705	6.07	−0.00478	0.00778	−0.0035
Bias-corrected	0.0259	−0.829	8.16	−0.00345	0.00546	−0.00255
Density-implied	0.0258	−0.826	8.16	−0.00342	0.00544	−0.00251
Density-implied, ind.	0.0274	−0.733	8.28	−0.00333	0.00623	−0.00197
			Worker effects			
Plug-in	0.138	0.659	4.12	0.0338	0.136	0.112
Bias-corrected	0.105	1	10.2	0.0339	0.111	0.103
Density-implied	0.105	0.99	6.85	0.0336	0.108	0.0928
Density-implied, ind.	0.108	0.836	6.91	0.0297	0.115	0.0679

Notes: Plug-in moments are based on AKM estimates. Bias-corrected moments are based on the closed-form expression given in Corollary 1. Density-implied moments are based on our bias-corrected density estimates. We also report density-implied moments under the assumption that $\boldsymbol{\eta}$ is independent of $\boldsymbol{\Sigma}$ (under the label "density-implied, ind."). μ_n is the n-th central moment.

7.3 Bivariate densities

We then report estimates based on bivariate deconvolution. In this version of the paper we focus on the density of worker and firm effects, and on the density of firm effects after and before a job move.

In Figure 2, on the left panel, we report the conditional density of worker effects by quartiles of firm effects. The estimates suggest that the correlation between worker and firm effects mostly arises from the upper quartile of the firm effects distribution. In Figure 2, on the right panel, we report the conditional density of firm effects after a job move by quartile of firm effects before the move. The graph suggests some positive dependence.

To complement the density graphs, in Table 3 we report estimates of quartile conditional probabilities for the two cases (i.e., worker/firm and firm before/firm after). We also compare our quasi-MLE estimates to probability estimates based on AKM plug-in that uses $\widehat{\alpha}_i$'s and $\widehat{\psi}_j$'s as if they were the true α_i's and ψ_j's.

In the case of worker effects given firm effects, we see that the plug-in estimates suggest near-random sorting patterns, while the quasi-MLE estimates suggest stronger correlation. It is well-known since Andrews, Gill, Schank, and Upward (2008) that AKM estimates of the covariance between worker and firm effects tend to be downward-biased, and our findings for the entire joint distribution confirm this.

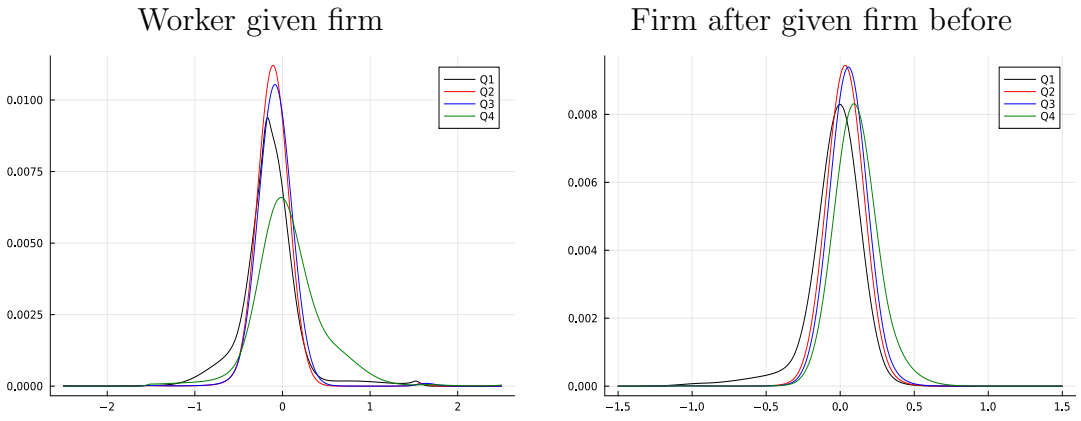
In the case of firm effects after a job move given firm effects before, Table 3 paints a different picture. Interestingly, in this case the plug-in estimate of the covariance is upward-biased, and we indeed see that the quasi-MLE estimates show less positive dependence than the plug-in ones. At the same time, the estimates still confirm the presence of positive dependence between the effects of the firm before and after a job move.[3]

7.4 Monte Carlo simulation

In this last part, we report the results of a small-scale Monte Carlo experiment. We take the worker-firm network connections from the Veneto sample, and generate the true values of worker and firm effects based on the semi-parametric deconvolution results. Specifically, we discretize the distribution implied by the deconvolution on a very fine grid (100,000 points), and solve a discrete optimal transport problem between this distribution and the discrete distribution of

[3]To provide intuition about the direction of the bias, let $\widehat{u}_{it}$ denote the AKM residuals. Since $y_{it} = \widehat{\alpha}_i + \widehat{\psi}_{j(i,t)} + \widehat{u}_{it}$, we expect high realizations of $\widehat{\alpha}_i$ (due to estimation noise) to be associated with low realizations of $\widehat{\psi}_{j(i,t)}$. In contrast, since $y_{i,t+1} - y_{it} = \widehat{\psi}_{j(i,t+1)} - \widehat{\psi}_{j(i,t)} + \widehat{u}_{i,t+1} - \widehat{u}_{it}$, we expect high realizations of $\widehat{\psi}_{j(i,t+1)}$ to be associated with high realizations of $\widehat{\psi}_{j(i,t)}$ upon a job move.

Figure 2: Conditional densities



Notes: Conditional densities of worker effects by quartile of firm effects (left), and conditional densities of firm effects after a job move by quartile of firm effects before the move (right).

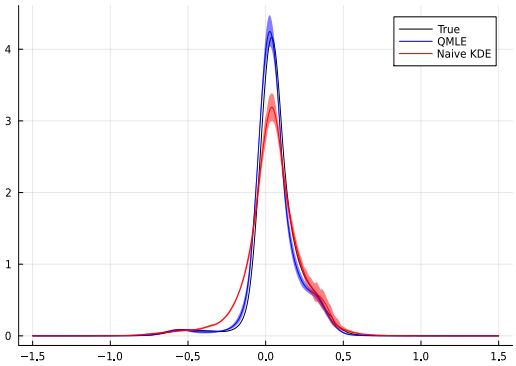
Table 3: Quartiles of worker effects (respectively, firm effects after a move) given quartiles of firm effects (resp., firm effects before the move)

	Worker effects given firm effects							
	QMLE				Plug-in			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Firm Q1	0.346	0.259	0.219	0.176	0.255	0.233	0.263	0.249
Firm Q2	0.239	0.302	0.293	0.166	0.225	0.272	0.268	0.236
Firm Q3	0.215	0.275	0.296	0.215	0.235	0.273	0.253	0.239
Firm Q4	0.179	0.159	0.201	0.461	0.26	0.229	0.225	0.286
	Firm effects after given firm effects before							
	QMLE				Plug-in			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Firm Before Q1	0.399	0.26	0.2	0.14	0.464	0.234	0.165	0.137
Firm Before Q2	0.252	0.283	0.259	0.207	0.231	0.342	0.253	0.174
Firm Before Q3	0.2	0.265	0.274	0.261	0.161	0.257	0.341	0.242
Firm Before Q4	0.138	0.206	0.253	0.403	0.136	0.175	0.235	0.454

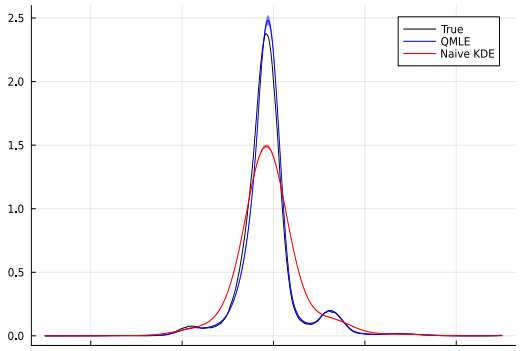
Notes: Quartile conditional probabilities of worker effects given firm effects (top panel), and quartile conditional probabilities of firm effects after a job move given firm effects before the move (bottom panel).

Figure 3: Monte-Carlo simulation, mean and 95% quantile bands

(a) Firm effects



(b) Worker effects



Notes: Monte Carlo simulations (100 simulations), with 95% confidence intervals. Estimates based on our composite quasi-likelihood method in blue, kernel-smoothed density estimator of AKM estimates in red, true distribution in black. Firm effects on the left, worker effects on the right.

AKM estimates, weighted at the person-year level. Then, based on the transport map we assign a fixed effect to each firm and each worker. This way, the true distribution of firm and worker effects (weighted by person-year) is identical to the result of the semi-parametric deconvolution. Then, for each simulation we draw the residuals u_{it} from a normal distribution with variance equal to the estimated variance $\hat{\tau}^2$, independently across i and t.

We report the findings of the simulation in Figure 3. We see that, for both worker and firm effects, the kernel-smoothed density estimator based on AKM estimates is heavily biased. In contrast, our penalized quasi-MLE estimator has low bias, and it recovers the shape of the true distribution well. Interestingly, there is only a small increase in variability between the two sets of estimates, suggesting that our approach is likely to give informative results in AKM settings.

8 An application to firm productivity and mark-downs (preliminary)

In this section (in progress) we describe a wage posting model with random search in the spirit of Burdett and Mortensen (1998) with ex-ante firm heterogeneity in productivity (following Mortensen, 2003). The model serves two goals. First, it motivates our treatment of η_j as

Table 4: Monte-Carlo simulation, moments for firm and worker effects

	True	Plug-in	Bias-corrected	Density-implied
		Firm Effects		
$Var[\psi]$	0.0258	0.0358 (0.0349, 0.0365)	0.0258 (0.0250, 0.0265)	0.0253 (0.0245, 0.0259)
$Skew[\psi]$	-0.826	-0.648 (-0.582, -0.733)	-0.828 (-0.957, -0.723)	-0.825 (-0.941, -0.733)
$Kurt[\psi]$	8.16	6.52 (6.29, 6.77)	8.15 (7.74, 8.61)	8.23 (7.79, 8.61)
		Worker Effects		
$Var[\psi]$	0.105	0.139 (0.138, 0.139)	0.105 (0.104, 0.106)	0.105 (0.104, 0.106)
$Skew[\psi]$	0.993	0.653 (0.640, 0.668)	0.992 (0.972, 1.01)	1.01 (0.989, 1.02)
$Kurt[\psi]$	9.83	6.98 (6.93, 7.03)	9.82 (9.73, 9.91)	9.79 (9.67, 9.91)

Notes: the moments of Monte Carlo simulations (100 simulations), with 95% confidence intervals. Firm effects in the top panel, worker effects in the bottom panel.

draws from an unrestricted joint distribution, arbitrarily correlated with conditioning variables. Second, it allows us to recover the densities of firm productivity and wage mark-downs, which are relevant quantities for policy.

We consider firms indexed by j each with productivity p_j. Workers have type k with productivity x_k. Firms have constant return to scale, and get $p_j x_k$ output per worker of type k. They post piece-rate wages $w_{jk} = \omega_j x_k$. Employment of worker type k in each firm j is given by h_{jk}. We consider a unit mass of worker, uniform across types.

Workers and firms are forward-looking. We have a search model, and, upon meeting, a preference shock is drawn for the worker. The worker can either accept or reject the poaching firm's offer.

Denote λ_k the probability that a worker meets a firm and μ_k the probability that a firm meets a worker. In this version of the paper, we will use a common λ and $\mu = 1$.

Given posted wages $(\omega_j)_{j=1..J}$, the worker present-value when getting some ω is given by:

$$(r + \lambda_k)V_k(\omega) = (1 + r)u(\omega \cdot x_k) + \lambda_k \sum_{j'} E[max\{V_k(\omega) + \epsilon_1, V_k(\omega_{j'}) + \epsilon_2\}],$$

which in the case of Type 1 extreme value shocks gives:

$$(r + \lambda_k)V_k(\omega) = (1 + r)u(\omega x_k) + \frac{\lambda}{n_j} \sum_{j'} \log(\exp V_k(\omega'_j) + \exp V_k(\omega)) + \gamma$$

Here we make an approximation as we include all firms in the poaching probability. Strictly speaking we should include all other firm $j' \neq j$ and define $V_{jk}(\omega)$, however the difference is negligible when there are enough firms.

We can then write the present-value for a firm j of having a worker k at wage ω:

$$(r + \lambda_k)\Pi_{jk}(\omega) = (1 + r)x_k(p_j - \omega) + \frac{\lambda_k}{n_j} \sum_{j'} G(V_k(\omega) - V_k(\omega_{j'}))\Pi_{jk}(\omega), \tag{26}$$

which, using $1 - G(x) = G(-x)$, gives us:

$$\Pi_{jk}(\omega) = \frac{(1 + r)x_k(p_j - \omega)}{r + \frac{\lambda_k}{n_j} \sum_{j'} G(V_k(\omega_{j'}) - V_k(\omega))}$$

Firms will post wages to maximize the probability of hiring times expected profit for each potential meeting. The probability of hire for worker k is given by:

$$\phi_k(\omega) = \frac{\mu_k}{n_j} \sum_j h_{jk} G(V_k(\omega) - V_k(\omega_j))$$

which represents the fact that we draw a worker at random from the cross-section. Such a worker compares the value offered by the new firm and compares it to the value of the poaching firm. The firm problem is then given by:

$$\max_{\omega} \sum_{k} (p_j - \omega) L_k(\omega),$$

where we introduce a present value of labor input defined as

$$L_k(\omega) = \frac{\frac{(1+r)x_k\mu_k}{n_j} \sum_{j'} h_{j'k} G(V_k(\omega) - V_k(\omega_{j'}))}{r + \frac{\lambda_k}{n_j} \sum_{j'} G(V_k(\omega_{j'}) - V_k(\omega))}.$$

We conclude by deriving a stationary condition for the equilibrium. For each firm j and each worker type k, the flow of workers coming into the firm must equal the flow coming out of the firm:

$$\frac{\mu_k}{n_j} \sum_{j'} h_{j'k} G(V_k(\omega_j) - V_k(\omega_{j'})) = h_{jk} \frac{\lambda_k}{n_j} \sum_{j'} G(V_k(\omega_{j'}) - V_k(\omega_j)),$$

which gives us a system for the h_{jk}.

An equilibrium is a set of wages ω_j, firm employments h_{jk} such that the stationary equilibrium is satisfied, $V_k(\omega)$ solves the worker problem and ω_j solves the firm maximization.

The model implies the following three features: (1) wages are increasing in firm productivity, (2) on-the-job search leads to correlation between the firm productivities before and after the move, and (3) high-productivity firms are larger and have more movers. In this model, economic decisions lead to complex forms of dependence.

In our application, we want to back out the parameters of the model. In practice, this means that we want the density of firm productivity, as well as the arrival rate and the variance of the preference shock. We proceed in the following steps:

1. Using the estimated density of wage premiums ω_j together with the allocation of workers to firms, we back out σ, λ, and $V_k(\omega)$.

2. Given V_k, we use the firm FOC to back out p_j from ω_j.

The first order condition of the firm implies that

$$p_j = \frac{\sum_k L_k(\omega_j) + \omega_j L_k'(\omega_j)}{\sum_k L_k'(\omega_j)},$$

where the only object needed is $L_k(\omega)$, which is only a function of the $V_k(\omega)$, λ, and μ.

Given these inputs, it is straightforward to recover the density of productivity p_j and the density of wage mark-downs $1 - \omega_j/p_j$.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67(2), 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?," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(3), 673–697.

Arellano, M., and S. Bonhomme (2012): "Identifying distributional characteristics in random coefficients panel data models," *The Review of Economic Studies*, 79(3), 987–1020.

Bonhomme, S., and A. Denis (2024): "Estimating Heterogeneous Effects: Applications to Labor Economics," *arXiv preprint arXiv:2404.01495*.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023): "How much should we trust estimates of firm effects and worker sorting?," *Journal of Labor Economics*, 41(2), 291–322.

Burdett, K., and D. T. Mortensen (1998): "Wage differentials, employer size, and unemployment," *International Economic Review*, pp. 257–273.

Chen, J. (2023): "Empirical Bayes When Estimation Precision Predicts Parameters," .

Cheng, X., C. H. Sheng, and F. Schorfheide (In Preparation): "Optimal Estimation of Two-Way Effects under Limited Mobility," *Working Paper*.

Chetty, R., and N. Hendren (2018): "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 133(3), 1163–1228.

Dattner, I., A. Goldenshluger, and A. Juditsky (2011): "On deconvolution of distribution functions," *The Annals of Statistics*, pp. 2477–2501.

Delaigle, A., and A. Meister (2008): "Density estimation with heteroscedastic error," *Bernoulli*, 14(2), 562–579.

Efron, B. (2016): "Empirical Bayes deconvolution estimates," *Biometrika*, 103(1), 1–20.

Efron, B., and C. Morris (1973): "Stein's estimation rule and its competitors—an empirical Bayes approach," *Journal of the American Statistical Association*, 68(341), 117–130.

Fan, J. (1991): "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, pp. 1257–1272.

Jochmans, K., and M. Weidner (2019): "Fixed-effect regressions on network data," *Econometrica*, 87(5), 1543–1560.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88(5), 1859–1898.

Lindsay, B. G. (1988): "Composite Likelihood Methods," in *Statistical Inference From Stochastic Processes*, ed. by N. U. Prabhu, pp. 221–239. American Mathematical Society, Providence, RI.

Mortensen, D. (2003): *Wage dispersion: why are similar workers paid differently?* MIT press.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003): *Semiparametric regression*, no. 12. Cambridge university press.

Stefanski, L. A., and R. J. Carroll (1990): "Deconvolving kernel density estimators," *Statistics*, 21(2), 169–184.

APPENDIX

A Proofs

A.1 Proof of Proposition 1

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right] = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\frac{\exp\left(\boldsymbol{i}\tau(\eta_j+V_j)\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\frac{\mathbb{E}\left(\exp\left(\boldsymbol{i}\tau V_j \,|\, \boldsymbol{\eta}, \boldsymbol{Z}, \boldsymbol{X}\right)\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right] \text{ (law of iterated expectations)} \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\frac{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right] \text{ (Assumption 3)} \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right] = \psi(\tau).
\end{aligned}
$$

A.2 Proof of Corollary 1

By Proposition 1 we have

$$
\begin{aligned}
\psi(\tau) =& \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\frac{\exp\left(\boldsymbol{i}\widehat{\eta}_j\tau\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right] \\
=& \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\widehat{\eta}_j\tau + \frac{1}{2}\sigma_j^2\tau^2\right)\right].
\end{aligned}
$$

Moreover, we have

$$
m_q = \frac{1}{\boldsymbol{i}^q}\frac{d^q}{d\tau^q}\bigg|_{\tau=0}\psi(\tau).
$$

Now, by Faà di Bruno's formula, we have

$$
\frac{d^q}{d\tau^q}\bigg|_{\tau=0}\exp\left(\boldsymbol{i}\widehat{\eta}_j\tau + \frac{1}{2}\sigma_j^2\tau^2\right) = (\boldsymbol{i}\sigma_j)^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right).
$$

It follows that

$$
\begin{aligned}
m_q &= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\frac{1}{\boldsymbol{i}^q}(\boldsymbol{i}\sigma_j)^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right)\right] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\sigma_j^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right)\right].
\end{aligned}
$$

A.3 Proof of Corollary 2

By inverse Fourier transformation we have

$$\overline{f}(\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(\tau)\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau,$$

where $\psi(\tau)$ is identified by Proposition 1.
Hence,

$$\overline{f}(\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right]\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau.$$

A.4 Proof of Proposition 2

We have

$$\begin{aligned}
&\widehat{f}(\eta) - \overline{f}(\eta) \\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{2\pi}\int_{-B_m}^{B_m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau - \frac{1}{2\pi}\int_{-\infty}^{+\infty}\psi(\tau)\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\{\tau\in[-B_m,B_m]\}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)} - \psi(\tau)\right)\exp\left(-\boldsymbol{i}\tau\eta\right)d\tau.
\end{aligned}$$

By Parseval's identity we obtain

$$\begin{aligned}
&\int\left(\widehat{f}(\eta) - \overline{f}(\eta)\right)^2 d\eta \\
&= \int\left|\frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\{\tau\in[-B_m,B_m]\}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)} - \psi(\tau)\right|^2 d\tau,
\end{aligned}$$

so

$$\begin{aligned}
&\mathbb{E}\left[\int\left(\widehat{f}(\eta) - \overline{f}(\eta)\right)^2 d\eta\right] \\
&= \int_{-B_m}^{B_m}\mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)} - \psi(\tau)\right|^2\right]d\tau + \int_{-\infty}^{-B_m}|\psi(\tau)|^2 d\tau + \int_{B_m}^{+\infty}|\psi(\tau)|^2 d\tau.
\end{aligned}$$

Now, by Proposition 1,

$$\int_{-B_m}^{B_m}\mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)} - \psi(\tau)\right|^2\right]d\tau = \int_{-B_m}^{B_m}\mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right)d\tau.$$

Moreover, we have, denoting as $\rho_{jj'}$ the conditional correlation of V_j and $V_{j'}$,

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right)\\
&=\operatorname{Var}\left(\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\,|\,\boldsymbol{\eta},\boldsymbol{Z},\boldsymbol{X}\right]\right)+\mathbb{E}\left[\operatorname{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\,|\,\boldsymbol{\eta},\boldsymbol{Z},\boldsymbol{X}\right)\right]\\
&=\operatorname{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)\\
&+\mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'=1}^{m}\exp\left(\boldsymbol{i}\tau(\eta_j-\eta_{j'})\right)\exp\left(\rho_{jj'}\sigma_j\sigma_{j'}\tau^2\right)\right]-\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(-\boldsymbol{i}\tau\eta_j\right)\right)\right]\\
&=\operatorname{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)+\mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'=1}^{m}\exp\left(\boldsymbol{i}\tau(\eta_j-\eta_{j'})\right)\left(\exp\left(\rho_{jj'}\sigma_j\sigma_{j'}\tau^2\right)-1\right)\right].
\end{aligned}
$$

Combining results gives

$$
\begin{aligned}
\mathbb{E}\left[\int\left(\widehat{f}(\eta)-\overline{f}(\eta)\right)^2 d\eta\right]&=\int_{-B_m}^{B_m}\mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^{m}\left(\exp\left(\sigma_j^2\tau^2\right)-1\right)\right]d\tau+\int_{-\infty}^{-B_m}|\psi(\tau)|^2d\tau+\int_{B_m}^{+\infty}|\psi(\tau)|^2d\tau\\
&+\int_{-B_m}^{B_m}\operatorname{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)d\tau+\int_{-B_m}^{B_m}\mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'\neq j}\exp\left(\boldsymbol{i}\tau(\eta_j-\eta_{j'})\right)\left(\exp\left(\rho_{jj'}\sigma_j\sigma_{j'}\tau^2\right)-1\right)\right]d\tau.
\end{aligned}
$$

That is, noting that $|\psi(\tau)|=|\psi(-\tau)|$,

$$
\begin{aligned}
&\mathbb{E}\left[\int\left(\widehat{f}(\eta)-\overline{f}(\eta)\right)^2 d\eta\right]\\
&=\frac{1}{m^2}\sum_{j=1}^{m}\int_{-B_m}^{B_m}\left(\mathbb{E}\left[\exp\left(\sigma_j^2\tau^2\right)\right]+\operatorname{Var}\left(\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)-1\right)d\tau+2\int_{B_m}^{+\infty}|\psi(\tau)|^2d\tau\\
&\underbrace{+\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'\neq j}\int_{-B_m}^{B_m}\operatorname{Cov}\left(\exp\left(\boldsymbol{i}\tau\eta_j\right),\exp\left(\boldsymbol{i}\tau\eta_{j'}\right)\right)d\tau}_{\text{dependence among the }\eta_j}\\
&\underbrace{+\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'\neq j}\int_{-B_m}^{B_m}\mathbb{E}\left[\exp\left(\boldsymbol{i}\tau(\eta_j-\eta_{j'})\right)\left(\exp\left(\rho_{jj'}\sigma_j\sigma_{j'}\tau^2\right)-1\right)\right]d\tau.}_{\text{dependence among the }V_j}
\end{aligned}
$$

Lastly, note that $0\leqslant\operatorname{Var}\left(\exp\left(\boldsymbol{i}\tau\eta_j\right)\right)\leqslant 1$. Proposition 2 then follows.

A.5 Proof of Proposition 3

We have

$$\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta})\right]-\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta}_0)\right]=\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\int\ln\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(\widehat{\eta}_j+\varepsilon_j-\eta\right)d\eta}{\int f_{\boldsymbol{\theta}_0}(\eta)\phi_{\overline{\sigma}^2}\left(\widehat{\eta}_j+\varepsilon_j-\eta\right)d\eta}\right)\phi_{\overline{\sigma}^2-\sigma_j^2}\left(\varepsilon_j\right)d\varepsilon_j\right],$$

where the expectations are with respect to the joint density of $\widehat{\eta}_j$ and σ_j^2, for $j\in\{1,...,m\}$. Hence, using (9), we have

$$\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta})\right]-\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta}_0)\right]=\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\ln\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(\widehat{\eta}_j+\varepsilon_j-\eta\right)d\eta}{\int f_{\boldsymbol{\theta}_0}(\eta)\phi_{\overline{\sigma}^2}\left(\widehat{\eta}_j+\varepsilon_j-\eta\right)d\eta}\right)\right],$$

where now the expectations are with respect to the marginal density of $\widehat{\eta}_j+\varepsilon_j$, for $j\in\{1,...,m\}$. By (12), the marginal density of $w_j=\widehat{\eta}_j+\varepsilon_j$ is

$$w\mapsto\int f_j(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta,$$

from which it follows that

$$\begin{aligned}\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta})\right]-\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta}_0)\right]&=\frac{1}{m}\sum_{j=1}^{m}\int\ln\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}{\int f_{\boldsymbol{\theta}_0}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}\right)\left(\int f_j(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta\right)dw\\&=\int\ln\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}{\int f_{\boldsymbol{\theta}_0}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}\right)\left(\int\left(\frac{1}{m}\sum_{j=1}^{m}f_j(\eta)\right)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta\right)dw.\end{aligned}$$

Now, since

$$f_{\boldsymbol{\theta}_0}(\eta)=\overline{f}(\eta)=\frac{1}{m}\sum_{j=1}^{m}f_j(\eta),$$

we obtain, using Jensen's inequality, that

$$\begin{aligned}\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta})\right]-\frac{1}{m}\mathbb{E}\left[L(\boldsymbol{\theta}_0)\right]&=\int\ln\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}{\int\overline{f}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}\right)\left(\int\overline{f}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta\right)dw\\&\leqslant\ln\left(\int\left(\frac{\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}{\int\overline{f}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta}\right)\left(\int\overline{f}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta\right)dw\right)\\&=\ln\left(\int\left(\int f_{\boldsymbol{\theta}}(\eta)\phi_{\overline{\sigma}^2}\left(w-\eta\right)d\eta\right)dw\right)=0.\end{aligned}$$