

Estimating Heterogeneous Effects: Applications to Labor Economics*

Stéphane Bonhomme
University of Chicago

Angela Denis
Banco de España

April 2, 2024

Abstract

A growing number of applications involve settings where, in order to infer heterogeneous effects, a researcher compares various units. Examples of research designs include children moving between different neighborhoods, workers moving between firms, patients migrating from one city to another, and banks offering loans to different firms. We present a unified framework for these settings, based on a linear model with normal random coefficients and normal errors. Using the model, we discuss how to recover the mean and dispersion of effects, other features of their distribution, and to construct predictors of the effects. We provide moment conditions on the model's parameters, and outline various estimation strategies. A main objective of the paper is to clarify some of the underlying assumptions by highlighting their economic content, and to discuss and inform some of the key practical choices.

JEL codes: C10. C50.

Keywords: Heterogeneity, neighborhoods, firms, workers, variance components, shrinkage.

*We thank Evan Rose for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Banco de España or the Eurosystem.

1 Introduction

Documenting heterogeneity across individuals, firms, or space, has become a central theme in applied economics. In earlier research, heterogeneous parameters used to be treated as a nuisance, and “differenced out” by means of fixed-effects regressions. Increasingly, however, estimating and studying heterogeneous effects has become the main goal of the analysis.

This focus on heterogeneity is enabled by the availability of richer data sets. A leading example is given by administrative data sets that feature many units (such as firms, workers, or neighborhoods), and at the same time provide information about each unit (such as multiple workers in a firm, multiple time periods on a worker, or multiple individuals in a neighborhood).

Settings with heterogeneous parameters are often studied using panel data techniques. However, traditional panel data methods treat units, such as firms or neighborhoods, independently of each other. A growing number of applications instead involve settings where, in order to infer heterogeneous effects, the researcher specifies a research design that compares various units. For example, to estimate firm or neighborhood effects, researchers exploit workers who move between firms or children who move to a new neighborhood, respectively.

We will refer to three leading examples as illustrations. In the first one, [Kline, Rose, and Walters \(2022\)](#) send multiple job applications to several large firms, and estimate firm-specific call-back rates as a function of applicants’ characteristics. Documenting differences across firms in call-back rates allows the researchers to study hiring discrimination at the firm level, hence complementing the literature using résumé correspondence experiments (e.g., [Bertrand and Duflo, 2017](#)) by providing firm-level estimates. In this application, discrimination parameters can be estimated independently, firm by firm. Hence, this setting is akin to a traditional panel data or grouped data setting.

In the second example that we analyze, [Chetty and Hendren \(2018b\)](#) study how the income at adulthood of children depends on the place where they grew up. To estimate the effects of neighborhoods on income, the researchers exploit mobility of families *across* neighborhoods. Hence, the effect of neighborhood j on income at adulthood is constructed by comparing the incomes of individuals who grew up in various neighborhoods j' . This setting has been studied by a large subsequent literature, including [Chetty and Hendren \(2018a\)](#), [Laliberté \(2021\)](#), [Bergman, Chetty, DeLuca, Hendren, Katz, and Palmer \(2019\)](#), and [Aloni and Avivi \(2023\)](#).

Our third leading example is the so-called AKM regression framework for matched employer-employee data introduced by [Abowd, Kramarz, and Margolis \(1999\)](#). The researchers’ goal is to estimate how worker and firm effects contribute to wage dispersion. This setting similarly

features comparisons between multiple units, since the differences in wage premia offered by two firms j and j' is informed by workers moving between j and j' . The AKM approach has become central to the study of workers and firms, see among others [Card, Heining, and Kline \(2013\)](#), [Card, Cardoso, Heining, and Kline \(2018\)](#), and [Song, Price, Guvenen, Bloom, and Von Wachter \(2019\)](#). Moreover, the methodology in this example has been used to study other questions, such as differences in health care utilization across space inferred from patients moving between cities ([Finkelstein, Gentzkow, and Williams, 2016](#)), or the impacts of banks and firms on credit growth inferred from banks' loans to multiple firms ([Amiti and Weinstein, 2018](#)).

In these applications, the data sets are complex and the models are high-dimensional. Practitioners need to make a large number of choices for modeling, practical specification, and estimation. Since we have not seen any survey on the econometric analysis of these settings as of yet,¹ we have decided to write a paper on the topic. Our main goal is to lay out a framework to analyze these settings, and to clarify some of the underlying assumptions and key practical choices. While doing so, we will highlight the economic content behind the main assumptions.

The model that underlies most applications to these settings is a linear normal random coefficients (RC) model. The “random coefficients” refer to the heterogeneous effects that are the focus of the analysis, such as the effects of neighborhoods, or the worker and firm effects. Those coefficients are associated with specific covariates: in [Chetty and Hendren \(2018b\)](#) the effect of a neighborhood is simply the coefficient of the exposure to that neighborhood (i.e., of how long the family stayed in that neighborhood), whereas in [Abowd, Kramarz, and Margolis \(1999\)](#) the effect of firm j is the coefficient of the j -th firm indicator. In both cases, the model involves a very large number of such covariates (e.g., many thousands of firm and worker indicators or neighborhood exposures).

The primitive parameters of the RC model are the means and variances of the coefficients (e.g., the neighborhood effects, or the worker and firm effects), as well as the variance of the errors. All of these parameters are potentially functions of all the covariates. They satisfy first and second moment conditions that we present. These moment conditions, which remain valid absent normality, build on moment conditions previously derived for panel data settings (e.g., [Chamberlain, 1992](#), [Arellano and Bonhomme, 2012](#)).

While means and variances are useful to answer substantive questions, such as how much dispersion in outcomes is explained by neighborhood, worker, or firm heterogeneity, other important questions require additional information. As we describe, higher-order moments (such

¹[Abowd, Kramarz, and Woodcock \(2008\)](#) and [Bonhomme \(2020\)](#) survey methods for bipartite networks and matched employer-employee data.

as skewness or kurtosis), nonlinear moments, as well as marginal and multivariate distributions, can all be inferred under the normality assumption of the RC model. For example, researchers may be interested in the distribution of neighborhood effects across space, or in the densities of worker and firm effects. In addition, the normal RC model can be used to construct optimal predictors of the effects. Such predictors or “forecasts” are also of considerable interest in various literatures outside our three leading examples, including in the work on teacher quality (e.g., [Kane and Staiger, 2008](#), [Chetty, Friedman, and Rockoff, 2014](#)).

Taking the RC model to data, however, is challenging in the settings that we study, given the large number of covariates (such as neighborhood exposures, or worker and firm effects) and the complex forms of dependence implied by the model. We show that commonly used specifications effectively impose conditional independence assumptions that may be economically restrictive. For example, common strategies in the analysis of neighborhood effects require location choice to be independent of neighborhood heterogeneity conditional on a set of neighborhood-specific covariates. Estimating flexible models of means, covariances, and distributions in these settings is an important yet still relatively unexplored research area.

Lastly, while the normal RC model provides a unified, self-contained framework to estimate heterogeneous parameters, normality and linearity may both be restrictive. In the last part of the paper, we briefly explore how these assumptions could be relaxed.

Relation to the literature. The framework we present builds on a vast methodological literature in statistics and econometrics. This includes the statistical literature on mixed models (e.g., [Jiang and Nguyen, 2007](#), [McCulloch and Searle, 2004](#)), the literature on random coefficients models and correlated random-effects approaches in panel data (e.g., [Chamberlain, 1992](#), [Arellano and Bonhomme, 2012](#)) and related panel data work based on decision-theoretic approaches ([Chamberlain and Moreira, 2009](#), [Chamberlain, 2016](#)), as well as the empirical Bayes literature (e.g., [Efron, 2012](#)) and the Bayesian and frequentist interpretations of best linear unbiased predictors (BLUP, e.g., [Robinson, 1991](#)).

2 Regressions with fixed effects

It is common in empirical work to focus on the relationship between some (typically scalar) outcomes y_i and some covariates x_i and z_i , and to specify a linear regression model of the form

$$y_i = x_i' \beta + z_i' \eta + u_i, \quad i = 1, \dots, n. \quad (1)$$

We assume that the researcher has access to multiple observations i about some economic units j . The units may be firms, workers, or neighborhoods, depending on the application. Our focus is on settings where z_i contains unit-specific variables, as well as interactions of unit-specific variables with other covariates. Let p denote the number of covariates in z_i . Hence, the coefficients η_j , $j = 1, \dots, p$, are unit-specific “fixed effects”.² Let q denote the number of covariates in x_i . We will focus on settings where p is large, and, depending on the application of interest, q may be small or large.

Estimating high-dimensional regressions as in (1) is made possible by the availability of increasingly large and detailed data sets, and by the development of powerful computational methods and software. This enables researchers to “zoom in” on the effects of particular units in the sample.

Under the assumption that u_i are uncorrelated with x_i and z_i in (1), researchers typically estimate β and η using OLS. This regression delivers parameter estimates $\hat{\beta}$ and $\hat{\eta}$. The researcher’s goal is then to use the estimates $\hat{\eta}_1, \dots, \hat{\eta}_p$ to learn about the true effects η_1, \dots, η_p .

We now describe three examples where this setup arises.

Example 1. (*Firm-specific discrimination*) *Kline, Rose, and Walters (2022) construct firm-specific measures of racial discrimination in hiring. They send various fictitious job applications with randomized characteristics to some of the largest employers in the US. They then measure call-back rates by race, for every firm in the sample. To map this setting to the current setup, denote job applications as i and firms as j . The outcome y_i is whether the firm calls back the applicant, and z_i contains the firm’s indicator and the interaction of the firm’s indicator with the race of applicant (white or black). There are no additional covariates x_i . The researchers are interested in the coefficient in η_j of the interaction between firm j ’s indicator with the race of the applicant, which they interpret as reflecting firm j ’s discrimination in hiring.*

In this particular setting, the regression coefficient $\hat{\eta}_j$ is constructed using the observations from firm j only, not other firms’ observations. The data structure is analogous to panel (or grouped) data, and indeed, in this application, model (1) can be interpreted as a grouped data model with group-specific intercepts and slopes, where the slope coefficients are associated with the applicant’s race.

Increasingly, researchers estimate regressions where learning about η_j requires comparing various units. In that case, the estimate $\hat{\eta}_j$ for unit j is constructed using the observations from

²We will refer to the η_j ’s as “fixed effects”, in line with the usual terminology in applied economics. In contrast, in the statistical literature on mixed models β are often referred to as “fixed effects”, and η as “random effects”. See, e.g., the monograph by [Jiang and Nguyen \(2007\)](#).

several (and possibly a large number of) other units j' . Our next two examples are in this vein.

Example 2. (Neighborhood effects and intergenerational mobility) *Chetty and Hendren (2018b)* estimate the effects of neighborhoods in the US (such as counties or commuting zones) on income at adulthood. In the setting they propose, i are children, the outcome y_i is the income of a child at age 26 (specifically, her rank in the overall income distribution at age 26), z_i contains the time spent by the child in every neighborhood when young as well as its interaction with parental income, and x_i contains origin times destination indicators, also interacted with parental income, as well as interactions between children's cohort indicators and parental income. The times spent by the child in every neighborhood, which the authors refer to as the "exposures" to the neighborhoods, are key covariates in the model. Note that, in this case, both z_i and x_i are high-dimensional.

The researchers are interested in neighborhood j 's effect on adult outcomes at some particular level of parental income \bar{p} . This neighborhood-specific parameter is a linear combination of the η_j parameters that pertain to neighborhood j . The exposure-time research design, which builds on *Chetty and Hendren (2018a)*, implies that the estimate of neighborhood j 's effect depends on the outcome data on other neighborhoods j' . The authors rely on this design to estimate the causal effect of neighborhoods, under the assumption that the age at which children move across neighborhoods does not directly affect adult outcomes.

Example 3. (Firm and worker effects in wage determination) *Abowd, Kramarz, and Margolis (1999)* study how worker and firm heterogeneity contribute to wage dispersion. In the setup they propose, y_i are log wages in a given period, x_i includes age and time indicators as well as other demographics, and z_i includes worker indicators and firm indicators. The worker and firm indicators are key covariates in the model. The researchers are interested in the coefficients η_j associated with workers and firms. A first question of interest is how dispersed are those worker and firm effects? By answering this question, the AKM model sheds light into how much of wage dispersion can be attributed to workers earning different wages irrespective of where they work, versus firms paying similar workers differently. A second question is how correlated are worker and firm effects? By answering this second question, the model sheds light on the nature of sorting patterns and how sorting contributes to wage dispersion.

Recovering worker and firm effects requires exploiting movements between firms. Intuitively, if workers remain in the same firms over time it is not possible to tell whether a high wage reflects a high worker effect or a high firm effect. Hence, the estimates of the effects depend on the network of employment relationships between workers and firms. Indeed, identification of the effects requires the network to be connected (*Abowd, Creecy, and Kramarz, 2002*).

3 Quantities of interest and noisy estimates

In the three applications that we have outlined, the researcher’s goal is to use the estimates $\hat{\eta}_1, \dots, \hat{\eta}_p$ to learn about the true effects η_1, \dots, η_p . We now provide examples of quantities of interest taken from the literature.

Throughout the paper we will treat η_1, \dots, η_p as random. Our goal will be to estimate features of the joint distribution of η_1, \dots, η_p , and construct predictors of those effects, without imposing *a priori* that the η_j ’s are independent of each other or independent of the x_i ’s and z_i ’s. When the conditional distribution of η_1, \dots, η_p given $x_1, \dots, x_n, z_1, \dots, z_n$ is left unrestricted, proceeding in this way does not materially differ from a setup where the η_j ’s are treated as fixed parameters (i.e., as “fixed effects”). The model we will present in Section 4 will impose restrictions on this conditional distribution, however.

3.1 Quantities of interest

Researchers may have various objectives. A first goal may be to estimate some moments of η_j . For simplicity we will focus on the case where η_j is scalar, although the expressions below are easily adapted to the case of vector-valued η_j ’s.

Moments that are expectations of linear combinations of the η_j ’s can be written as

$$m_c = \mathbb{E}[c'\eta], \tag{2}$$

where c is a $p \times 1$ vector. An example is the mean of the η_j ’s,

$$\mathbb{E}\left[\frac{1}{p} \sum_{j=1}^p \eta_j\right].$$

Consider the coefficient in the linear regression of η_j on some covariates W_j . For example, one may be interested in regressing firm effects on firm size or industry indicators in Example 3, or in regressing neighborhood effects on average income in the neighborhood in Example 2. The regression coefficient can be written as $(\mathbb{E}[W_j W_j'])^{-1} \mathbb{E}[W_j \eta_j]$, which takes the form (2) for $c = (\mathbb{E}[W_j W_j'])^{-1} W_j$.

Moments that are quadratic in η can be written as

$$v_Q = \mathbb{E}[\eta'Q\eta], \tag{3}$$

where Q is a $p \times p$ matrix. For example, the variance of the η_j ’s,

$$\text{Var}(\eta_j) = \mathbb{E}\left[\frac{1}{p} \sum_{j=1}^p \left(\eta_j - \frac{1}{p} \sum_{j'=1}^p \eta_{j'}\right)^2\right],$$

can be written in this form for a suitable matrix Q . In Example 3, interest often centers on the variances of worker and firm effects and on the covariance between worker and firm effects, all of which can be written as (3).

Alternatively, one may be interested in the coefficient in the linear regression of some scalar variable W_j on η_j . For example, in Example 3 one may be interested in regressing promotion opportunities in a firm on the firm effects. The regression coefficient is $(\mathbb{E}[\eta_j^2])^{-1} \mathbb{E}[\eta_j W_j]$, which can be written as the ratio between some m_c in (2) and some v_Q in (3), for suitable c vector and Q matrix.

More general, nonlinear moments can be written as

$$w_H = \mathbb{E}[H(\eta)], \quad (4)$$

for some function $H : \mathbb{R}^p \rightarrow \mathbb{R}$. For example, one may be interested in the skewness or kurtosis of the η_j 's, which can be written as ratios of quantities of the form (4) for suitable H functions.

Learning about distributions of effects, beyond their means and variances, is important in all the examples we have mentioned. In Example 1, Kline, Rose, and Walters (2022) report estimates of the distribution of racial discrimination in hiring across firms. In Example 2, researchers are often interested in documenting the distribution of neighborhood effects. In Example 3, an increase in the variance of firm effects, say, has different implications for inequality whether it comes from a deepening of the left tail, an expansion of the right tail, or a symmetric increase in spread.

The weighted cumulative distribution function, for some weights ω_j that sum up to one, can be written as the following nonlinear moment

$$F_\omega(a) = \mathbb{E} \left[\sum_{j=1}^p \omega_j \mathbf{1} \{ \eta_j \leq a \} \right]. \quad (5)$$

One may also be interested in the (weighted) density of the η_j 's, $f_\omega(a) = \frac{\partial F_\omega(a)}{\partial a}$. Bivariate counterparts to $F_\omega(a)$ and $f_\omega(a)$, which reflect the bivariate distribution of workers and firms, are of interest in Example 3 in order to document sorting patterns along the worker and firm distributions.

Another common goal in applications is to construct predictors of the η_j 's. The optimal predictor that minimizes the expected sum of squared errors is a set of functions $\varphi_1, \dots, \varphi_p$ that solves

$$\min_{\varphi_1, \dots, \varphi_p} \mathbb{E} \left[\sum_{j=1}^p (\eta_j - \varphi_j(y_1, \dots, y_n, x_1, \dots, x_n, z_1, \dots, z_n))^2 \right], \quad (6)$$

the solution of which is the set of conditional means

$$\varphi_j(y_1, \dots, y_n, x_1, \dots, x_n, z_1, \dots, z_n) = \mathbb{E}[\eta_j | y_1, \dots, y_n, x_1, \dots, x_n, z_1, \dots, z_n], \quad j = 1, \dots, p. \quad (7)$$

In Example 2, [Chetty and Hendren \(2018a\)](#) and [Chetty and Hendren \(2018b\)](#) construct predictors of neighborhood-specific income effects. [Bergman, Chetty, DeLuca, Hendren, Katz, and Palmer \(2019\)](#) use effect predictors to select the top census tracts for income mobility.³ A key input to answering these questions is the set of conditional means given by (7).

3.2 Noisy estimates

The data is not directly informative about the η_j 's, but delivers estimates

$$\hat{\eta}_j = \eta_j + v_j, \quad j = 1, \dots, p, \quad (8)$$

where $v_j = \hat{\eta}_j - \eta_j$ reflects estimation noise. In many applications involving large data sets (large n) and many unit-specific parameters (large p), the noise v_j is substantial enough to be of practical concern. Inferring η_j from $\hat{\eta}_j$ then requires solving a filtering (or “de-noising”) problem.

It is important to note that directly using the estimates $\hat{\eta}_j$ in place of η_j may be misleading. The noise v_j in (8) reflects the presence of a form of measurement error. When the quantity of interest is nonlinear in η_j , such as a variance, a higher-order moment, a cumulative distribution function, or a density, the presence of measurement error often leads to unreliable, biased estimates of the quantities of interest.

As an example, suppose the researcher is interested in the variance of the η_j 's, and that she reports the following “plug in” estimate based on the $\hat{\eta}_j$'s,

$$\widehat{\text{Var}}(\hat{\eta}_j) = \frac{1}{p} \sum_{j=1}^p \left(\hat{\eta}_j - \frac{1}{p} \sum_{j'=1}^p \hat{\eta}_{j'} \right)^2. \quad (9)$$

Does a large estimate $\widehat{\text{Var}}(\hat{\eta}_j)$ indicate that the variance of the true effects η_j , $\text{Var}(\eta_j)$, is large? Or does the presence of measurement error v_j artificially inflate the dispersion in the estimates $\hat{\eta}_j$?

As another example, suppose the researcher uses $\hat{\eta}_j$ as a predictor of η_j . Since the p parameters η_1, \dots, η_p are estimated in the available sample and p is large (or, alternatively, the noise

³See [Gu and Koenker \(2023\)](#) for a broad account of ranking problems and selection of groups of units based on noisy estimates.

in (8) is substantial), it is likely that $\hat{\eta}_j$ “overfits”, in the sense that it reflects too much of the noise v_j and too little of the true effect η_j . In such cases, one may wish to construct predictors that lead to a lower expected sum of squared errors in (6).

To illustrate these points, consider a simple setting inspired by Example 1, where η_j and v_j in (8) are independent, i.i.d. across j , and $\mathcal{N}(\mu_\eta, \sigma_\eta^2)$ and $\mathcal{N}(0, \sigma_v^2)$, respectively. In this case, we have

$$\mathbb{E} \left[\widehat{\text{Var}}(\hat{\eta}_j) \right] = \text{Var}(\eta_j) + \frac{p-1}{p} \sigma_v^2, \quad (10)$$

which shows that the variance of the estimates $\hat{\eta}_j$ is upward-biased for the variance of the true η_j 's. Moreover, while the expected sum of squared errors of $\hat{\eta}_j$ is equal to $p\sigma_v^2$, the expected sum of squared errors of the conditional means

$$\mathbb{E}(\eta_j | \hat{\eta}_j) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2} \hat{\eta}_j + \frac{\sigma_v^2}{\sigma_\eta^2 + \sigma_v^2} \mu_\eta \quad (11)$$

is smaller, equal to $\frac{p\sigma_v^2\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2}$. This shows that the “shrunk” quantities $\mathbb{E}(\eta_j | \hat{\eta}_j)$ in (11) have a lower expected sum of squared errors than the original estimates $\hat{\eta}_j$. The difference between the two is greater when the noise variance σ_v^2 is large relative to the variance σ_η^2 of the true effects. Bias-correction methods for variance components based on equations in the spirit of (10), and linear shrinkage predictors akin to (11), are now widespread in applied economics.

This simple example is too stylized to accurately describe the situations in Examples 2 and 3, however. In such settings, estimates $\hat{\eta}_j$ are constructed using observations from other units $j' \neq j$. Hence, the $\hat{\eta}_j$'s are not independent. This gives rise to more complex forms for the estimation noise v_j in (8), and complicates the way the noise affects the quantities of interest. Jochmans and Weidner (2019) study how, in settings where the matrix of covariates z_i has a network structure (such as Example 3, where z_i represent worker and firm employment relationships), the properties of the network, such as how connected it is, affect the precision of the estimates $\hat{\eta}_j$. Moreover, in the settings of Examples 2 and 3, unit-specific parameters η_j may not be independent. We next present a framework that applies to an arbitrary matrix of covariates z_i and allows for dependence between units.

4 The normal random coefficients model

In this section we describe a normal Random Coefficients (RC) model, which allows researchers to answer the questions introduced in the previous section.

4.1 Model and assumptions

For the presentation we will remove the term $x'_i\beta$ from equation (1). Depending on the setting, β can be estimated using OLS or differenced out, see Remark 1 below. We then write (1) in vector form, removing $x'_i\beta$ from the equation, as follows,

$$Y = Z\eta + U, \quad (12)$$

where Y is an $n \times 1$ vector with generic element y_i , Z is an $n \times p$ matrix with generic row z'_i , and U is an $n \times 1$ vector with generic element u_i .

The form of the design matrix Z differs across applications. In Example 1, Z takes the form

$$Z = \begin{pmatrix} Z_1 & 0 & 0 & \dots & 0 \\ 0 & Z_2 & 0 & \dots & 0 \\ 0 & 0 & Z_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & Z_p \end{pmatrix},$$

where Z_j contains two columns, the first one being a column of 1's, and the second one being a column of 0's and 1's, depending on the race of the applicant. This block-diagonal structure characterizes panel data and grouped data settings.

In Examples 2 and 3, Z takes more complex forms. To present Example 2, let us abstract from the dependence on parental income for simplicity. Then, z_{ij} in (1) is the exposure of i to neighborhood j . In every row of the $n \times p$ matrix with elements z_{ij} , all but a handful of elements are equal to zero.⁴ However, the matrix does not have a block-diagonal form. Then, differencing out the origin-and-destination indicators x_i (as explained in Remark 1 below) leads to the matrix

$$Z = \begin{pmatrix} \tilde{z}_{11} & \tilde{z}_{12} & \tilde{z}_{13} & \dots & \tilde{z}_{1J} \\ \tilde{z}_{21} & \tilde{z}_{22} & \tilde{z}_{23} & \dots & \tilde{z}_{2J} \\ \tilde{z}_{31} & \tilde{z}_{32} & \tilde{z}_{33} & \dots & \tilde{z}_{3J} \\ \dots & \dots & \dots & \dots & \dots \\ \tilde{z}_{n1} & \tilde{z}_{n2} & \tilde{z}_{n3} & \dots & \tilde{z}_{nJ} \end{pmatrix}, \quad (13)$$

where \tilde{z}_{ij} is equal to z_{ij} minus the mean of $z_{i'j}$ for all individuals i' who experience the same neighborhood moves in childhood as individual i (though the time i' and i stay in each neighborhood may differ). Z is sparse, and not block-diagonal.

⁴For example, [Chetty and Hendren \(2018b\)](#) focus on families that move exactly once, so every row in the matrix has exactly two non-zero elements.

In Example 3, Z stacks worker and firm indicators together. For example, with two periods, K workers, and J firms, Z reads

$$Z = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & f_{11}^1 & \dots & f_{11}^J \\ 1 & 0 & 0 & \dots & 0 & f_{12}^1 & \dots & f_{12}^J \\ 0 & 1 & 0 & \dots & 0 & f_{21}^1 & \dots & f_{21}^J \\ 0 & 1 & 0 & \dots & 0 & f_{22}^1 & \dots & f_{22}^J \\ 0 & 0 & 1 & \dots & 0 & f_{31}^1 & \dots & f_{31}^J \\ 0 & 0 & 1 & \dots & 0 & f_{32}^1 & \dots & f_{32}^J \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & f_{K1}^1 & \dots & f_{K1}^J \\ 0 & 0 & 0 & \dots & 1 & f_{K2}^1 & \dots & f_{K2}^J \end{pmatrix}, \quad (14)$$

where $f_{kt}^j = 1$ if worker k is employed in firm j in period t , and $f_{kt}^j = 0$ otherwise. Note that Z is not block-diagonal in this case. However, it is typically a sparse matrix since each row has exactly two non-zero elements. The form of Z reflects the network of workers' and firms' employment relationships (Andrews, Gill, Schank, and Upward, 2008, Jochmans and Weidner, 2019).

Throughout, we assume that Z has full column rank, so $(Z'Z)$ is non-singular. When p is large, this assumption may be restrictive. In Example 3, ensuring non-singularity requires imposing a normalization on the parameters η_j (e.g., that the firm-specific effects sum up to zero), and focusing on a connected component of the firm-worker network (Abowd, Creedy, and Kramarz, 2002). Similarly, in Example 2, a normalization is needed since neighborhood effects are identified relative to the national average (Chetty and Hendren, 2018b).

The defining assumptions for the normal random coefficients model are as follows.

Assumption 1. (*normal RC model*)

- (i) $U | Z, \eta \sim \mathcal{N}(0, \Omega(Z))$.
- (ii) $\eta | Z \sim \mathcal{N}(\mu(Z), \Sigma(Z))$.

In part (i) of Assumption 1 we assume that the error terms u_i in (1) are normally distributed, with zero mean and some $n \times n$ covariance matrix $\Omega(Z)$, independent of η . In part (ii) we specify a normal model for η given Z , with mean $\mu(Z)$ (a $p \times 1$ vector) and variance $\Sigma(Z)$ (a $p \times p$ matrix). Hence, we treat the parameters η_j as random coefficients, and we specify their conditional distribution given Z . This modeling device is often used in panel data and complex data settings. Note that Assumption 1 implies that the conditional distribution of Y given Z

is fully specified, and normal, given the parameters $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$,

$$Y | Z \sim \mathcal{N}(Z\mu(Z), Z\Sigma(Z)Z' + \Omega(Z)).$$

Assumption 1 requires the following *strict exogeneity* assumption: $\mathbb{E}[U | Z, \eta] = 0$. This assumption imposes substantive restrictions on the economic environment. In Example 2, it requires that the times families spend in every neighborhoods are unrelated to the unobserved determinants of adult outcomes. In Example 3, it imposes an assumption of so-called “exogenous mobility”, through which workers’ decisions to change jobs may be driven by worker and firm effects η_j but not by idiosyncratic time-varying shocks u_i . Although some authors have attempted to relax strict exogeneity in the setting of Example 3 (e.g., [Abowd, McKinney, and Schmutte, 2019](#), [Bonhomme, Lamadon, and Manresa, 2019](#)), most research to date relying on model (1) makes this assumption.

The concrete specification of $\mu(Z)$, $\Sigma(Z)$, and $\Omega(Z)$ will depend on the application. In panel or grouped data settings, such as in Example 1, a common assumption is that (η_j, Z_j) are independent across j , where Z_j denotes the subset of observations z_i that pertain to unit j . In that case, $\mu_j(Z)$ depends on Z only through Z_j , $\Sigma(Z)$ is diagonal, and $\Sigma_{j,j}(Z)$ only depends on Z_j as well.

However, in Examples 2 and 3, it may be more plausible to allow for a rich dependence of $\mu_j(Z)$ and $\Sigma_{j,j'}(Z)$ on the elements of Z , and to allow $\Sigma(Z)$ to be a general, non-diagonal symmetric matrix. Indeed, in Example 2, the de-meaned exposures $\tilde{z}_{ij'}$ to other neighborhoods $j' \neq j$ in (13) are unlikely to be independent of neighborhood effects η_j , unless mobility across neighborhoods is unrelated to neighborhood heterogeneity. Likewise, in Example 3, indicators of firms $j' \neq j$ in (14) will generally correlate with the effect of firm j unless workers’ sorting patterns are independent of firm heterogeneity. We will return to specification issues in the next section.

Given Assumption 1, the OLS estimator of η in (12) satisfies

$$\hat{\eta} = \eta + V, \tag{15}$$

where, denoting the variance-covariance matrix of the OLS estimates $\hat{\eta}$ as

$$S(Z) = (Z'Z)^{-1}Z'\Omega(Z)Z(Z'Z)^{-1}, \tag{16}$$

we have

$$V = (Z'Z)^{-1}Z'U | Z, \eta \sim \mathcal{N}(0, S(Z)).$$

Model (12) under Assumption 1 is a normal linear mixed model (see, e.g., Jiang and Nguyen, 2007 and McCulloch and Searle, 2004). As we will review below, normality is not needed to obtain informative moment conditions on $\mu(Z)$, $\Sigma(Z)$, and $\Omega(Z)$, and the following restrictions on first and second moments suffice.

Assumption 2. (*RC model*)

- (i) $\mathbb{E}[U | Z, \eta] = 0$, $\text{Var}[U | Z] = \Omega(Z)$.
- (ii) $\mathbb{E}[\eta | Z] = \mu(Z)$, $\text{Var}[\eta | Z] = \Sigma(Z)$.

Note that, while Assumption 2 relaxes normality, it does maintain the strict exogeneity condition $\mathbb{E}[U | Z, \eta] = 0$. If the researcher is only interested in means, variances or covariances of the η_j 's, or alternatively in coefficients of regressions where η_j appears on the left- or right-hand side, then Assumption 1 can be replaced by the weaker Assumption 2 that only restricts first and second moments. However, normality is needed to answer questions related to the higher-order and nonlinear moments of the η_j 's, their distributions, and to construct optimal predictors of the η_j 's.

Remark 1. In (12) we have removed the term $x_i'\beta$. In applications with covariates x_i , there are two common ways of handling the presence of the unknown parameter β . When the dimension q of β is low, one can often reliably estimate $\hat{\beta}$ jointly with $\hat{\eta}$ using OLS in (1), and then replace $Y = (y_i)$ with $Y = (y_i - x_i'\hat{\beta})$ in (12). The analysis below is then essentially unchanged relative to the case without x_i 's. This approach is commonly used to handle the presence of age, time, and other demographics in Example 3.

In applications where q is large, the parameter β can be differenced out as follows. Write model (1) in vector form, as

$$Y = X\beta + Z\eta + U, \tag{17}$$

where X is an $n \times q$ matrix with generic row x_i' . Let M be an $n \times n$ matrix such that $MX = 0$ (e.g., a projection matrix). Left-multiplying (17) by M then gives

$$MY = MZ\eta + MU, \tag{18}$$

which takes the same form as (12). In Example 2, taking the within-group transformation as matrix M , this differencing approach provides a way to handle the presence of the high-dimensional origin-and-destination indicators in x_i .⁵

⁵Also, in Example 3, if one includes worker indicators in X , and Z contain only firm indicators, then first-differencing yields an equation of the form (18).

4.2 Quantities of interest in the normal RC model

Suppose that the normal RC model (12) holds and Assumption 1 is satisfied, and suppose the covariance and mean functions $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$ are known. Then the model implies closed-form expressions for all the quantities that we mentioned in Section 3.

For example, the first and second moments m_c in (2) and v_Q in (3) are given, respectively, by

$$m_c = \mathbb{E} [c' \mu(Z)] \quad (19)$$

and

$$v_Q = \mathbb{E} [\mu(Z)' Q \mu(Z) + \text{Trace} (Q \Sigma(Z))]. \quad (20)$$

The expressions (19) and (20) do not rely on normality, and hold under the weaker Assumption 2.

Under Assumption 1, we further can write the nonlinear moment w_H in (4), and the cumulative distribution function $F_\omega(a)$ in (5), in closed form as, respectively,

$$w_H = \mathbb{E} \left[\int_{\mathbb{R}^p} H(\eta) \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma(Z)|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\eta - \mu(Z))' [\Sigma(Z)]^{-1} (\eta - \mu(Z)) \right) d\eta \right], \quad (21)$$

and

$$F_\omega(a) = \mathbb{E} \left[\sum_{j=1}^p \omega_j \Phi \left(\frac{a - \mu_j(Z)}{\sqrt{\Sigma_{j,j}(Z)}} \right) \right], \quad (22)$$

where $\mu_j(Z)$ is the j -th element of $\mu(Z)$, $\Sigma_{j,j}(Z)$ is the j -th diagonal element of $\Sigma(Z)$, and Φ denotes the standard normal cumulative distribution function. Note that, while the conditional distribution of $\eta | Z$ is normal under Assumption 1, the unconditional distribution of η is not normal.

Lastly, under Assumption 1, one can also derive a closed-form expression for the conditional mean of the vector η given the data (Y, Z) , as

$$\mathbb{E} [\eta | Y, Z] = G(Z) (S(Z)^{-1} \hat{\eta} + \Sigma(Z)^{-1} \mu(Z)), \quad (23)$$

where $S(Z)$ is given by (16), and

$$G(Z) = (S(Z)^{-1} + \Sigma(Z)^{-1})^{-1}$$

is the conditional variance of η given (Y, Z) . The conditional density of $\eta | Y, Z$ is then

$$\eta | Y, Z \sim \mathcal{N} (\mathbb{E} [\eta | Y, Z], G(Z)). \quad (24)$$

5 Specification choices

In this section, we discuss several possibilities to specify $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$.

5.1 Specification of $\Omega(Z)$

The simplest specification for $\Omega(Z)$ is independent homoskedastic, that is,

$$\Omega(Z) = \sigma^2 I_n,$$

for a constant variance parameter σ^2 . In Example 3, Andrews, Gill, Schank, and Upward (2008) rely on this assumption and construct an unbiased estimator of σ^2 for applications to matched employer-employee data.

However, both homoskedasticity and independence can be restrictive. To relax homoskedasticity, one can introduce covariates $W = (w_i)$ (for example, some functions of the elements of Z), and model

$$\Omega(Z) = \text{diag}(\sigma_\theta^2(w_1), \dots, \sigma_\theta^2(w_n)),$$

where $\sigma_\theta^2(w_i)$ is a parametric function of w_i indexed by some parameter θ .⁶ In Example 2, Chetty and Hendren (2018b) assume that $S(Z)$ in (16) is a diagonal matrix, i.e., that the estimates $\hat{\eta}_j$ are uncorrelated across neighborhoods. Under independence, Kline, Saggio, and Sølvesten (2020) model $\Omega(Z)$ as a diagonal matrix with unrestricted diagonal elements. They propose a leave-out method that provides unbiased estimates of these diagonal variance elements.

To relax independence, one can model $\Omega(Z)$ as a parametric, matrix-valued, non-diagonal function of the covariates w_i . In panel data settings, Arellano and Bonhomme (2012) propose parametric ARMA specifications to allow for serial correlation. Relaxing independence has been shown to be important in Example 3, where assuming serial independence within employment spells is often empirically restrictive.

Lastly, it is worth noting that one cannot leave the matrix $\Omega(Z)$ fully unrestricted while at the same time identifying moments of the effects η_j . This is because η_j and v_j in (8) are both unobserved random variables. As a result, there is an essential trade-off between heterogeneity (the η_j 's) and the dependence of errors (the matrix $\Omega(Z)$).

⁶While here we assume that w_i is a function of the elements of Z , the covariates w_i could also contain additional covariates not functions of Z , such as neighborhood characteristics or worker or firm characteristics, depending on the application.

5.2 Specification of $\mu(Z)$ and $\Sigma(Z)$

Turning now to $\mu(Z)$ and $\Sigma(Z)$, a possible approach in applications is to specify $\mu(Z)$ as a function of some covariates W (e.g., as a linear function of W), and to model $\Sigma(Z)$ similarly (e.g., as a constant diagonal matrix). However, the cost of such an approach is that it implicitly imposes a *conditional independence assumption* that may be economically restrictive. We now illustrate this important point through the help of examples.

Suppose that, in Example 2, the mean and variance of neighborhood effects are specified as functions of a handful of covariates, for instance some variables measuring the racial and economic composition of the neighborhoods. In that case, the researcher is effectively assuming that the de-meaned neighborhood exposures in Z , see (13), are independent of the location-specific effects η_j conditional on those covariates. This assumption may be hard to reconcile with an economic model of location choice, and mobility across locations, where families' decisions are in part determined by neighborhood heterogeneity.

Similarly, in Example 3, assuming that the means, variances, and covariances of worker and firm effects only depend on some worker and firm characteristics restricts job mobility to be independent of the worker and firm effects η_j conditional on those characteristics. This assumption may be at odds with economic models of sorting where workers' and firms' decisions are in part determined by worker and firm heterogeneity.

To state the argument formally, let W denote some covariates. Then a specification where $\mu(Z)$ and $\Sigma(Z)$ only depend on W amounts to assuming the following:

$$\eta \mid Z, W \sim \mathcal{N}(\mu(W), \Sigma(W)),$$

which in particular imposes that:

$$Z \text{ and } \eta \text{ are independent given } W.$$

Likewise, assuming that $\mu_j(Z) = \mu_j(W_j)$, where W_j denotes the covariates of unit j , imposes that:

$$\eta_j \text{ are mean of independent } (W_1, \dots, W_{j-1}, W_{j+1}, \dots, W_p) \text{ given } W_j.$$

To illustrate that such conditional independence assumptions may be economically restrictive, let us consider two models of firm choice for Example 3. The first model assumes that workers maximize utility among all firms in a market, period-by-period (Card, Cardoso, Heining, and Kline, 2018, Lamadon, Mogstad, and Setzler, 2022). Suppose that worker k 's indirect utility in firm ℓ at time t is

$$V_{k\ell t} = \rho W_{k\ell t} + \varepsilon_{k\ell t},$$

where $\varepsilon_{k\ell t}$ are i.i.d. type I extreme value preference shocks, independent of log wages $W_{k\ell t}$, and we abstract from non-wage amenities for simplicity. Then the probability that worker k chooses firm ℓ , given all wages $W = \{W_{k'\ell't}\}$, is

$$\Pr(\ell | k, W) = \frac{\exp(\rho W_{k\ell t})}{\sum_{\ell' \in \mathcal{M}(k)} \exp(\rho W_{k\ell't})}, \quad (25)$$

where $\mathcal{M}(k)$ is the market that k considers when looking for a job. Therefore, the elements in the Z matrix depend on the log wages $W_{k\ell t}$. If, further, log wages are a function of worker heterogeneity α_k and firm heterogeneity ψ_ℓ ,⁷ then (25) implies that the η_j 's, which here are the α_k 's and ψ_ℓ 's, are *not* independent of Z , even conditional on observed characteristics.

Consider next a dynamic model of workers' mobility across firms, as proposed by [Lentz, Piyapromdee, and Robin \(2023\)](#) (see [Sorkin, 2018](#) for a related model). The probability that worker k moves between firms ℓ and ℓ' , conditional on worker heterogeneity $\alpha = \{\alpha_k\}$ and firm heterogeneity $\psi = \{\psi_\ell\}$, is specified as

$$\Pr(\ell' | \ell, k, \alpha, \psi) = \lambda_{k\ell'} \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}}, \quad (26)$$

where $\lambda_{k\ell'}$ is the probability that k meets firm ℓ' , and $\gamma_{k\ell}$ is interpreted as worker k 's value of working in firm ℓ . If the value $\gamma_{k\ell} = \gamma(\alpha_k, \psi_\ell)$ depends on worker heterogeneity α_k and firm heterogeneity ψ_ℓ , then neither α_k nor ψ_ℓ are independent of Z , even conditional on observed characteristics.

We now discuss several examples of specifications for $\mu(Z)$ and $\Sigma(Z)$ used in practice. In [Example 2](#), a common approach is to model $\mu_j(Z)$ to be a linear function of some covariates W_j , and $\Sigma(Z)$ to be a diagonal matrix, independent of Z and W . The model then assumes independence across j 's, and rules out dependence between the true effects (η_j) and location choice and mobility (Z) conditional on the covariates. Recently, [Chen \(2023\)](#) proposes an extension of this approach that allows for dependence between the true effects η_j and the precision of their estimates $\hat{\eta}_j$ (as measured by the diagonal elements of $S(Z)$ in (16)), while maintaining independence across units.

In [Example 3](#), [Woodcock \(2015\)](#) postulates a normal RC model where neither $\mu(Z)$ nor $\Sigma(Z)$ depend on Z , and $\Sigma(Z)$ is a diagonal matrix. However, these assumptions impose that workers' sorting patterns, which are encoded in Z , do not depend on the worker and firm effects η_j . [Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler \(2023\)](#) refine the [Woodcock \(2015\)](#) model by allowing $\mu(Z)$ and $\Sigma(Z)$ to depend on Z . To model the dependence,

⁷For example, if log wages are given by the additive specification $W_{k\ell t} = \alpha_k + \psi_\ell + U_{k\ell t}$ of [Abowd, Kramarz, and Margolis \(1999\)](#), in which case $\Pr(\ell | k, W)$ in (25) does not depend on k .

they cluster firms into a small number of groups using the k-means algorithm based on their wage distributions (as in [Bonhomme, Lamadon, and Manresa, 2019](#)). Given this grouping, they allow the means and variances of worker and firm effects to depend on the groups, but not on the worker and firm identities within these groups. Similarly, they allow the covariances in $\Sigma(Z)$, which they do not assume to be diagonal, to depend on the groups. Generalizing such approaches to accommodate structural economic models of workers' mobility across firms is a promising area for future investigation.

6 Estimation

In this section we describe various estimation strategies for the parameters of the normal RC model and the quantities of interest introduced in [Section 3](#).

6.1 Estimating the model's parameters

We start by providing moment conditions on the primitive parameters of the normal RC model: the error variance $\Omega(Z)$, and the mean and variance of the unit-specific effects $\mu(Z)$ and $\Sigma(Z)$.

Moment conditions. Let [Assumption 2](#) hold. Then we have

$$\mathbb{E}[YY' | Z] = Z\mathbb{E}[\eta\eta' | Z]Z' + \Omega(Z), \quad (27)$$

where the cross-product term is zero since $\mathbb{E}[U | Z, \eta] = 0$. This is a system of $n \times n$ moment conditions. Following [Arellano and Bonhomme \(2012\)](#), we can construct a suitable $n^2 \times n^2$ matrix $M(Z)$ that “differences out” the first term on the right-hand side of [\(27\)](#),⁸ to obtain

$$\mathbb{E}[M(Z) \text{vec}(YY' - \Omega(Z)) | Z] = 0. \quad (28)$$

where the vec operator stacks together the n columns of an $n \times n$ matrix into a $n^2 \times 1$ vector. [\(28\)](#) provides a system of conditional moment conditions on the elements of $\Omega(Z)$. Note that, depending on the specification of $\Omega(Z)$, the moment conditions [\(28\)](#) do not necessarily guarantee that the elements of $\Omega(Z)$ are identified. Note also that $M(Z)$ in [\(28\)](#) being singular implies that the system of equations has multiple solutions in the absence of restrictions on $\Omega(Z)$.

⁸Let $A \otimes B$ denote the Kronecker product between A and B , and let I_{n^2} be the $n^2 \times n^2$ identity matrix. A possible choice is $M(Z) = I_{n^2} - (Z(Z'Z)^{-1}Z') \otimes (Z(Z'Z)^{-1}Z')$. The key property is that $M(Z)(Z \otimes Z) = 0$, which implies that $M(Z) \text{vec}(Z\mathbb{E}[\eta\eta' | Z]Z') = M(Z)(Z \otimes Z) \text{vec}(\mathbb{E}[\eta\eta' | Z]) = 0$.

Next, taking the conditional mean in (15), we obtain

$$\mathbb{E}[\hat{\eta} - \mu(Z) | Z] = 0, \quad (29)$$

which provide conditional moment conditions on $\mu(Z)$. These moment conditions depend neither on $\Omega(Z)$ nor on $\Sigma(Z)$.

Lastly, using (15) we obtain

$$\text{Var}[\hat{\eta} | Z] = \Sigma(Z) + S(Z),$$

where $S(Z)$ given by (16) is a function of $\Omega(Z)$. This gives the following moment conditions on $\Sigma(Z)$, $\Omega(Z)$, and $\mu(Z)$,

$$\mathbb{E}[(\hat{\eta} - \mu(Z))(\hat{\eta} - \mu(Z))' - S(Z) - \Sigma(Z) | Z] = 0. \quad (30)$$

Remark 2. *In general, (28)-(29)-(30) do not exhaust all the information about $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$. The complete set of first- and second-moment conditions implied by Assumption 2 consists of (29) and*

$$\mathbb{E}[YY' | Z] = Z(\Sigma(Z) + \mu(Z)\mu(Z)')Z' + \Omega(Z). \quad (31)$$

An advantage of the moment conditions in (28) is that they are robust to possible misspecification of $\mu(Z)$ and $\Sigma(Z)$.

Parameter estimation. Given a parametric or semi-parametric specification for $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$, a possible estimation approach is to exploit the moment conditions (28)-(29)-(30) using method-of-moments or minimum-distance estimation. This strategy is used in Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2023) in Example 3, for instance. Let θ be a parameter vector indexing $\mu_\theta(Z)$, $\Sigma_\theta(Z)$, and $\Omega_\theta(Z)$, possibly adding other conditioning covariates W . This estimation step delivers an estimate $\hat{\theta}$, as well as estimates $\hat{\mu}(Z) = \mu_{\hat{\theta}}(Z)$, $\hat{\Sigma}(Z) = \Sigma_{\hat{\theta}}(Z)$, and $\hat{\Omega}(Z) = \Omega_{\hat{\theta}}(Z)$. An alternative is to perform (quasi-) maximum likelihood estimation, as in the following remark.

Remark 3. *An alternative approach to estimation of μ , Σ and Ω is to rely on the log-likelihood function conditional on Z ,*

$$\begin{aligned} L(\mu, \Sigma, \Omega) = & -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|Z\Sigma(Z)Z' + \Omega(Z)|) \\ & - \frac{1}{2} (Y - Z\mu(Z))' [Z\Sigma(Z)Z' + \Omega(Z)]^{-1} (Y - Z\mu(Z)). \end{aligned} \quad (32)$$

In the absence of normality, *i.e.*, under Assumption 2, $L(\mu, \Sigma, \Omega)$ can be interpreted as a quasi log-likelihood function.⁹

6.2 Estimating the quantities of interest: three strategies

We now present three types of estimators for the quantities of interest introduced in Section 3.

#1 Bias-corrected fixed-effects estimators. Suppose first that the researcher is only interested in estimating linear combinations of the η_j 's or quadratic forms. In that case, she does not need to estimate $\mu(Z)$ and $\Sigma(Z)$. Indeed, linear combinations $m_c = \mathbb{E}[c'\eta]$ and, given knowledge of $\Omega(Z)$, quadratic forms $v_Q = \mathbb{E}[\eta'Q\eta]$, are nonparametrically identified under Assumption 2 (without the need for normality), as

$$m_c = \mathbb{E}[c'\hat{\eta}] \quad (33)$$

and

$$v_Q = \mathbb{E}[\hat{\eta}'Q\hat{\eta}] - \mathbb{E}[\text{Trace}(QS(Z))], \quad (34)$$

respectively. By (33), linear combinations of the elements in $\hat{\eta}$ are unbiased. Moreover, (34) shows that, while quadratic forms in $\hat{\eta}$ are biased, the bias is a known function of the variance-covariance matrix $S(Z)$ of $\hat{\eta}$.

Then, a fixed-effects estimator of m_c is

$$\hat{m}_c^{\text{FE}} = c'\hat{\eta}. \quad (35)$$

Moreover, given an estimator $\hat{\Omega}(Z)$ and an associated estimator $\hat{S}(Z)$ given by

$$\hat{S}(Z) = (Z'Z)^{-1}Z'\hat{\Omega}(Z)Z(Z'Z)^{-1}, \quad (36)$$

a bias-corrected fixed-effects estimator of v_Q is

$$\hat{v}_Q^{\text{FE}} = \hat{\eta}'Q\hat{\eta} - \text{Trace}(Q\hat{S}(Z)). \quad (37)$$

Under Assumption 2, the estimator \hat{v}_Q^{FE} is unbiased whenever $\hat{\Omega}(Z)$, and hence $\hat{S}(Z)$, are themselves unbiased.

⁹Moreover, when based on $A'Y$ and $A'Z$, where $M = AA'$ is the Cholesky decomposition of the matrix M of Remark 1, (32) is the basis for restricted maximum likelihood (REML) estimation, and the resulting REML estimator does not depend on the choice of A (see, *e.g.*, Jiang and Nguyen, 2007).

In Example 3, Andrews, Gill, Schank, and Upward (2008) assume that $\Omega(Z) = \sigma^2 I_n$. In this case, (28) is equivalent to

$$\mathbb{E} \left[(I_n - Z(Z'Z)^{-1}Z') Y Y' (I_n - Z(Z'Z)^{-1}Z') \mid Z \right] = \sigma^2 (I_n - Z(Z'Z)^{-1}Z').$$

So, by taking the trace and expectation with respect to Z , it follows that

$$\sigma^2 = \mathbb{E} \left[\frac{Y' (I_n - Z(Z'Z)^{-1}Z') Y}{n - p} \right]. \quad (38)$$

The formula (38) is the well-known degree of freedom correction for variance estimation. In Example 3, Andrews, Gill, Schank, and Upward (2008) propose the estimators

$$\hat{\sigma}^2 = \frac{Y' (I_n - Z(Z'Z)^{-1}Z') Y}{n - p} \quad (39)$$

and

$$\hat{v}_Q^{\text{FE}} = \hat{\eta}' Q \hat{\eta} - \hat{\sigma}^2 \text{Trace} \left(Q (Z'Z)^{-1} \right), \quad (40)$$

which are unbiased for σ^2 and v_Q , respectively. Kline, Saggio, and Sølvesten (2020) generalize their approach to the case where $\Omega(Z)$ is a diagonal matrix with unrestricted diagonal elements.

#2 Model-based estimators. Suppose now that the researcher wishes to estimate not only linear combinations and quadratic forms but also other quantities, such as distributions, nonlinear moments, or predictors. In that case, she first needs to produce estimates $\hat{\Omega}(Z)$, $\hat{\mu}(Z)$, and $\hat{\Sigma}(Z)$. Given those, the researcher can produce estimates of all the quantities of interest we listed in Section 3, as follows:

$$\hat{m}_c = c' \hat{\mu}(Z), \quad (41)$$

$$\hat{v}_Q = \hat{\mu}(Z)' Q \hat{\mu}(Z) + \text{Trace} \left(Q \hat{\Sigma}(Z) \right), \quad (42)$$

$$\hat{w}_H = \int_{\mathbb{R}^p} H(\eta) \frac{1}{(2\pi)^{\frac{p}{2}} |\hat{\Sigma}(Z)|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\eta - \hat{\mu}(Z))' \left[\hat{\Sigma}(Z) \right]^{-1} (\eta - \hat{\mu}(Z)) \right) d\eta, \quad (43)$$

$$\hat{F}_\omega(a) = \sum_{j=1}^p \omega_j \Phi \left(\frac{a - \hat{\mu}_j(Z)}{\sqrt{\hat{\Sigma}_{j,j}(Z)}} \right), \quad (44)$$

including estimates of the posterior mean and variance of η :

$$\hat{\mathbb{E}}[\eta \mid Y, Z] = \hat{G}(Z) \left(\hat{S}(Z)^{-1} \hat{\eta} + \hat{\Sigma}(Z)^{-1} \hat{\mu}(Z) \right), \quad (45)$$

$$\hat{G}(Z) = \left(\hat{S}(Z)^{-1} + \hat{\Sigma}(Z)^{-1} \right)^{-1}, \quad (46)$$

where $\hat{S}(Z)$ is given by (36).

#3 Posterior estimators. To motivate the third type of estimators, consider a generic nonlinear moment of η , $w_H = \mathbb{E}[H(\eta)]$, and note that, by the law of iterated expectations,

$$\begin{aligned} w_H &= \mathbb{E}[\mathbb{E}(H(\eta) | Y, Z)] \\ &= \mathbb{E} \left[\int_{\mathbb{R}^p} H(\eta) \frac{1}{(2\pi)^{\frac{p}{2}} |G(Z)|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\eta - \mathbb{E}[\eta | Y, Z])' [G(Z)]^{-1} (\eta - \mathbb{E}[\eta | Y, Z]) \right) d\eta \right], \end{aligned} \quad (47)$$

where we have used the expression in (24) of the conditional density of $\eta | Y, Z$.

Equation (47) motivates the following posterior estimator of w_H :

$$\hat{w}_H^{\text{POST}} = \int_{\mathbb{R}^p} H(\eta) \frac{1}{(2\pi)^{\frac{p}{2}} |\hat{G}(Z)|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\eta - \hat{\mathbb{E}}[\eta | Y, Z])' [\hat{G}(Z)]^{-1} (\eta - \hat{\mathbb{E}}[\eta | Y, Z]) \right) d\eta, \quad (48)$$

where $\hat{\mathbb{E}}[\eta | Y, Z]$ and $\hat{G}(Z)$ are given by (45) and (46), respectively.

To provide intuition about the posterior estimator, we note that, by a change in variables,

$$\hat{w}_H^{\text{POST}} = \mathbb{E} \left[H \left(\hat{\mathbb{E}}[\eta | Y, Z] + \hat{G}(Z)^{\frac{1}{2}} \varepsilon \right) | Y, Z \right], \quad (49)$$

where $\varepsilon | Y, Z \sim \mathcal{N}(0, I_p)$. Consider two special cases in (49). When the $\hat{\eta}_j$ are poorly estimated so $\hat{S}(Z)^{-1} \approx 0$, then (49) becomes

$$\hat{w}_H^{\text{POST}} \approx \mathbb{E} \left[H \left(\hat{\mu}(Z) + \hat{\Sigma}(Z)^{\frac{1}{2}} \varepsilon \right) | Z \right], \quad (50)$$

which coincides with the model-based estimator (43). Now, when the $\hat{\eta}_j$ are well estimated so $\hat{S}(Z) \approx 0$ (and $S(Z) \approx 0$), then (49) becomes

$$\hat{w}_H^{\text{POST}} \approx H(\hat{\eta}) \approx H(\eta), \quad (51)$$

which means that the posterior estimator recovers the true quantity in this situation, irrespective of whether the normal RC model is well specified or not. This illustrates a robustness property of posterior estimators, as studied by [Arellano and Bonhomme \(2009\)](#) and [Bonhomme and Weidner \(2022\)](#), and provides a motivation for reporting such estimators in practice.

6.3 Remarks on asymptotic properties and inference

There are several challenges to deriving asymptotic properties for the estimators mentioned in this section, in the type of settings we are focusing on in this paper. A first challenge is the dimensionality of the data and model. Typically, the vector $\mu(Z)$ and the matrices $\Omega(Z)$ and

$\Sigma(Z)$ are very large (i.e., both n and p are large). It is therefore necessary to work in a high-dimensional asymptotic regime where both n and p tend to infinity. Moreover, the specifications for $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$ often depend on many parameters. Another challenge is the nature of the matrix Z . In Examples 2 and 3, Z does not have a block-diagonal structure, which further complicates the asymptotic analysis. Related to this, $\Sigma(Z)$ is often non-diagonal, hence creating complex forms of dependence among observations.

To give a simple example, suppose we are interested in the mean $m_c = \mathbb{E}[\frac{1}{p} \sum_{j=1}^p \eta_j]$. Consistency of $\hat{m}_c = \frac{1}{p} \sum_{j=1}^p \hat{\mu}_j(Z)$ for m_c is not immediate. To see this, consider first the case where there is no estimation error and one can compute $\tilde{m}_c = \frac{1}{p} \sum_{j=1}^p \mu_j(Z)$. Consistency of \tilde{m}_c for m_c requires limiting the dependence of the $\mu_j(Z)$'s across j . For instance, in Example 2, this requires limiting the dependence of mean exposure effects across neighborhoods, while in Example 3, this requires limiting the dependence of mean firm effects across different firms, for example, allowing for dependence among “similar” firms only. Next, the addition of estimation error further complicates the argument, since the $\hat{\mu}_j(Z)$'s are dependent across j conditional on Z .

Nevertheless, several results are available in the literature. For the case of fixed-effects estimators of quadratic forms, as in (37), Kline, Saggio, and Sølvesten (2020) provide conditions for consistency, as well as inference theory, in a setup that assumes $\Omega(Z)$ to be diagonal while leaving $\mu(Z)$ and $\Sigma(Z)$ unrestricted.

The mixed models literature in statistics (as reviewed by, e.g., Jiang, 2017) provides a variety of asymptotic results for the quantities of interest we have listed here, including for estimators of posterior means such as (45). However, the models studied in the theoretical literature on mixed models impose specific assumptions on $\Omega(Z)$, $\mu(Z)$, and $\Sigma(Z)$. In particular, a typical assumption is that $\Sigma(Z)$ is a diagonal matrix, as in the so-called “mixed ANOVA” model. We have argued here that such an assumption may be economically restrictive. For example, in Example 3, $\Sigma(Z)$ is not diagonal whenever the effects of two firms j and j' depend on each other given Z , as happens in the model of Lentz, Piyapromdee, and Robin (2023) mentioned in Section 5.

Under Assumption 1, which imposes normality, Heijmans and Magnus (1986b) provide conditions for consistency of the maximum likelihood estimator (as in Remark 3) in models with general forms of dependence across observations induced by the matrices $\Omega(Z)$ and $\Sigma(Z)$. Heijmans and Magnus (1986a) provide a further set of conditions under which the maximum likelihood estimator is asymptotically normal. Importantly, however, while their setup allows for general forms of dependence, it relies on the assumption that the dimension of the model's

parameter θ is kept fixed as the sample size tends to infinity.

To make the model more flexible while keeping the number of parameters to estimate moderate, a possibility is to assume that, while means and variances vary across observations, this variation is driven by a small number of groups. For example, [Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler \(2023\)](#) assume that the means and variances of worker and firm effects depend on firm groups. Grouping methods, where group membership is estimated using methods akin to the k-means algorithm, are theoretically justified under suitable conditions ([Bonhomme and Manresa, 2015](#), [Bonhomme, Lamadon, and Manresa, 2022](#)). Assuming a grouped structure reduces the dimension of the model, and can allow for a simpler characterization of asymptotic distributions.

Given the relevance of these high-dimensional, dependent settings for the applied economic literature, more research needs to be done regarding the formal analysis of asymptotic properties and inference methods, both in cases where normality is assumed to hold (as in [Assumption 1](#)) and in cases where distributions may be non-normal. We next comment on the possibility that normality may be violated.

7 Extensions

While the normal random coefficients model provides a simple and powerful tool to estimate heterogeneous effects in high-dimensional settings, it relies on functional forms assumptions that may be empirically restrictive. In this final section of the paper, we briefly explore some strategies that could be used to relax some of those assumptions.

Non-normal effects. When it is suspected that the distribution of $\eta | Z$ is not truly normal, that distribution is sometimes interpreted as a Bayesian prior. This interpretation is central to the empirical Bayes approach, and it is often invoked in applications such as [Example 2](#). In this perspective, the conditional distribution of $\eta | Y, Z$ in [\(24\)](#) can be interpreted as the posterior distribution of η . Linear shrinkage predictors as in [\(45\)](#) possess attractive robustness properties under misspecification ([James and Stein, 1961](#)).¹⁰ Moreover, estimators of nonlinear moments based on posterior distributions (see [Subsection 6.2](#)) are less sensitive to misspecification of the normal model compared to estimators based on the prior distribution $\eta | Z$. However, in many applications the variance of the noise is substantial, and it is still important to correctly model

¹⁰In particular, the James-Stein linear shrinkage estimator in the normal means model achieves asymptotically minimax expected sum-of-squares loss in Euclidean balls (see [Theorem 7.48](#) in [Wasserman, 2006](#)).

the distribution of $\eta | Z$.

Suppose that we maintain the normality of $U | Z$ in Assumption 1, while leaving the conditional distribution of $\eta | Z$ unrestricted. Then, (8) becomes a nonparametric deconvolution model with normal errors. Identification and estimation strategies exist in a variety of settings. In Example 1, [Kline, Rose, and Walters \(2022\)](#) use Efron’s deconvolution method ([Efron, 2016](#)) to estimate the density of firm-specific discrimination under an independence assumption. [Chen \(2023\)](#) proposes a location-scale model to handle situations where the precision of the estimates $\hat{\eta}_j$ predicts the true η_j , assuming independence across units. He shows that this modeling can improve the prediction performance of conditional mean estimates. Relaxing independence across units, thereby fitting the settings of Examples 2 and 3, is an important task for future work on these approaches.

Non-normal noise. One may also suspect that the distribution of $U | Z$ is not normal. In applications to Examples 1 and 2, the error V in (15) may be approximately normal even when $U | Z$ is not, due to the fact that V is an estimation error, hence asymptotically normal under standard conditions as the relevant sample size tends to infinity. Approximate normality of V will typically hold in grouped data settings as Example 1 provided that group sizes be large enough.¹¹ However, the normal approximation need not be accurate in other applications.¹²

Dealing with settings where V in (15) is not normal, and $\eta | Z$ is not normal either, is challenging. This requires using the individual outcome model (1), while allowing for a non-normal distribution of $U | Z$. In panel data settings, [Arellano and Bonhomme \(2012\)](#) show how to identify and estimate the distribution of U under the assumption that it follows a linear independent factor model (see also [Kotlarski, 1967](#), [Li and Vuong, 1998](#), and [Bonhomme and Robin, 2010](#)). We are not aware of extensions of such generalized deconvolution approaches to the settings of Examples 2 and 3, however.

Nonlinear mean. Lastly, an important assumption in the RC model is that the mean outcome is linear in η , in the sense that

$$\mathbb{E}[Y | Z, \eta] = Z\eta. \tag{52}$$

¹¹For example, V may be approximately normal even when Y is a vector of binary outcomes. In [Kline, Rose, and Walters \(2022\)](#), the outcomes y_i are binary call-back indicators, yet estimates $\hat{\eta}_j$ may still be approximately normal with variance $S(Z)$.

¹²When $U | Z$ is not normal, the right-hand side in (23) remains unbiased whenever U has zero mean given Z and η . However, it is no longer the best predictor of η in general.

Equation (52) has empirical and economic content. In Example 2, assuming that neighborhood exposures experienced by children have a separable, constant impact on outcomes may be restrictive (Chetty and Hendren, 2018b). In Example 3, assuming away interactions between worker and firm effects in wages implicitly imposes restrictive assumptions on input complementarity, which in turn drives the nature of sorting patterns in many economic models (e.g., Becker, 1973, Eeckhout and Kircher, 2011).

Within the context of Example 3, Bonhomme, Lamadon, and Manresa (2019) show how to allow for complementarity patterns between firm and worker effects, in a setup where they assume that firm heterogeneity is discrete.¹³ More work is needed in this direction.

¹³In that application, the presence of complementarity between worker and firm effects can alternatively be interpreted as reflecting individual heterogeneity in the “treatment effect” of a firm, as studied in Hull (2018).

References

- ABOWD, J. M., R. H. CREECY, AND F. KRAMARZ (2002): “Computing person and firm effects using linked longitudinal employer-employee data,” Discussion paper, Center for Economic Studies, US Census Bureau.
- ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): “High wage workers and high wage firms,” *Econometrica*, 67(2), 251–333.
- ABOWD, J. M., F. KRAMARZ, AND S. WOODCOCK (2008): “Econometric analyses of linked employer-employee data,” in *The econometrics of panel data: Fundamentals and recent developments in theory and practice*, pp. 727–760. Springer.
- ABOWD, J. M., K. L. MCKINNEY, AND I. M. SCHMUTTE (2019): “Modeling endogenous mobility in earnings determination,” *Journal of Business & Economic Statistics*, 37(3), 405–418.
- ALONI, T., AND H. AVIVI (2023): “One Land, Many Promises: The Unequal Consequences of Childhood Location for Natives and Immigrants in Israel,” .
- AMITI, M., AND D. E. WEINSTEIN (2018): “How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data,” *Journal of Political Economy*, 126(2), 525–587.
- 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 (2009): “Robust priors in nonlinear panel data models,” *Econometrica*, 77(2), 489–536.
- (2012): “Identifying distributional characteristics in random coefficients panel data models,” *The Review of Economic Studies*, 79(3), 987–1020.
- BECKER, G. S. (1973): “A theory of marriage: Part I,” *Journal of Political economy*, 81(4), 813–846.
- BERGMAN, P., R. CHETTY, S. DELUCA, N. HENDREN, L. F. KATZ, AND C. PALMER (2019): “Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice,” Discussion paper, National Bureau of Economic Research.
- BERTRAND, M., AND E. DUFLO (2017): “Field experiments on discrimination,” *Handbook of economic field experiments*, 1, 309–393.
- BONHOMME, S. (2020): “Econometric analysis of bipartite networks,” in *The econometric analysis of network data*, pp. 83–121. Elsevier.

- 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.
- BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): “A distributional framework for matched employer employee data,” *Econometrica*, 87(3), 699–739.
- (2022): “Discretizing unobserved heterogeneity,” *Econometrica*, 90(2), 625–643.
- BONHOMME, S., AND E. MANRESA (2015): “Grouped patterns of heterogeneity in panel data,” *Econometrica*, 83(3), 1147–1184.
- BONHOMME, S., AND J.-M. ROBIN (2010): “Generalized non-parametric deconvolution with an application to earnings dynamics,” *The Review of Economic Studies*, 77(2), 491–533.
- BONHOMME, S., AND M. WEIDNER (2022): “Posterior average effects,” *Journal of Business & Economic Statistics*, 40(4), 1849–1862.
- CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): “Firms and labor market inequality: Evidence and some theory,” *Journal of Labor Economics*, 36(S1), S13–S70.
- CARD, D., J. HEINING, AND P. KLINE (2013): “Workplace heterogeneity and the rise of West German wage inequality,” *The Quarterly journal of economics*, 128(3), 967–1015.
- CHAMBERLAIN, G. (1992): “Efficiency bounds for semiparametric regression,” *Econometrica: Journal of the Econometric Society*, pp. 567–596.
- (2016): “Fixed Effects, Invariance, and Spatial Variation in Intergenerational Mobility,” *American Economic Review*, 106(5), 400–404.
- CHAMBERLAIN, G., AND M. J. MOREIRA (2009): “Decision theory applied to a linear panel data model,” *Econometrica*, 77(1), 107–133.
- CHEN, J. (2023): “Empirical Bayes When Estimation Precision Predicts Parameters,” .
- CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American economic review*, 104(9), 2593–2632.
- CHETTY, R., AND N. HENDREN (2018a): “The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects,” *The Quarterly Journal of Economics*, 133(3), 1107–1162.
- (2018b): “The impacts of neighborhoods on intergenerational mobility II: County-level estimates,” *The Quarterly Journal of Economics*, 133(3), 1163–1228.

- EECKHOUT, J., AND P. KIRCHER (2011): “Identifying sorting—in theory,” *The Review of Economic Studies*, 78(3), 872–906.
- EFRON, B. (2012): *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, vol. 1. Cambridge University Press.
- (2016): “Empirical Bayes deconvolution estimates,” *Biometrika*, 103(1), 1–20.
- FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2016): “Sources of geographic variation in health care: Evidence from patient migration,” *The quarterly journal of economics*, 131(4), 1681–1726.
- GU, J., AND R. KOENKER (2023): “Invidious comparisons: Ranking and selection as compound decisions,” *Econometrica*, 91(1), 1–41.
- HEIJMANS, R. D., AND J. R. MAGNUS (1986a): “Asymptotic Normality of Maximum Likelihood Estimators Obtained from Normally Distributed but Dependent Observations,” *Econometric Theory*, 2(3), 374–412.
- (1986b): “Consistent maximum-likelihood estimation with dependent observations: the general (non-normal) case and the normal case,” *Journal of Econometrics*, 32(2), 253–285.
- HULL, P. (2018): “Estimating treatment effects in mover designs,” *arXiv preprint arXiv:1804.06721*.
- JAMES, W., AND C. STEIN (1961): “Estimation with quadratic loss,” in *Proc. 4th Berkeley Symp. on Math. Statist. and Prob., 1961*.
- JIANG, J. (2017): *Asymptotic analysis of mixed effects models: theory, applications, and open problems*. CRC press.
- JIANG, J., AND T. NGUYEN (2007): *Linear and generalized linear mixed models and their applications*, vol. 1. Springer.
- JOCHMANS, K., AND M. WEIDNER (2019): “Fixed-effect regressions on network data,” *Econometrica*, 87(5), 1543–1560.
- KANE, T. J., AND D. O. STAIGER (2008): “Estimating teacher impacts on student achievement: An experimental evaluation,” Discussion paper, National Bureau of Economic Research.
- KLINE, P., E. K. ROSE, AND C. R. WALTERS (2022): “Systemic discrimination among large US employers,” *The Quarterly Journal of Economics*, 137(4), 1963–2036.

- KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): “Leave-out estimation of variance components,” *Econometrica*, 88(5), 1859–1898.
- KOTLARSKI, I. (1967): “On characterizing the gamma and the normal distribution,” *Pacific Journal of Mathematics*, 20(1), 69–76.
- LALIBERTÉ, J.-W. (2021): “Long-term contextual effects in education: Schools and neighborhoods,” *American Economic Journal: Economic Policy*, 13(2), 336–377.
- LAMADON, T., M. MOGSTAD, AND B. SETZLER (2022): “Imperfect competition, compensating differentials, and rent sharing in the US labor market,” *American Economic Review*, 112(1), 169–212.
- LENTZ, R., S. PIYAPROMDEE, AND J.-M. ROBIN (2023): “The Anatomy of Sorting—Evidence From Danish Data,” *Econometrica*, 91(6), 2409–2455.
- LI, T., AND Q. VUONG (1998): “Nonparametric estimation of the measurement error model using multiple indicators,” *Journal of Multivariate Analysis*, 65(2), 139–165.
- MCCULLOCH, C. E., AND S. R. SEARLE (2004): *Generalized, linear, and mixed models*. John Wiley & Sons.
- ROBINSON, G. K. (1991): “That BLUP is a good thing: the estimation of random effects,” *Statistical science*, pp. 15–32.
- SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2019): “Firming up inequality,” *The Quarterly journal of economics*, 134(1), 1–50.
- SORKIN, I. (2018): “Ranking firms using revealed preference,” *The quarterly journal of economics*, 133(3), 1331–1393.
- WASSERMAN, L. (2006): *All of nonparametric statistics*. Springer Science & Business Media.
- WOODCOCK, S. D. (2015): “Match effects,” *Research in Economics*, 69(1), 100–121.