# Estimation uncertainty in repeated finite populations*

Víctor Sancibrián†

November 2024

[Updated regularly. Please click here for the latest version.]

**Abstract**

Standard errors need to be adjusted down when the sample is a large fraction of the population of interest (a *finite population* setup). I consider the empirically relevant case where a finite population coexists with a measurement problem, in that the features of interest are not necessarily observable even if the entire population is sampled. I show that conventional standard errors remain generally conservative in this context and propose Finite Population Corrections (FPCs) that guarantee non-conservative inference when repeated measurements are available. FPCs rely on weak dependence across measurements and are very simple to implement. I apply these methods to two empirical settings where uncertainty has been previously understood in different ways: predicting lethal encounters with police using data on all U.S. police departments, and studying firm misallocation with a census of large Indonesian firms. Finite-population inference leads to confidence intervals that are up to 50% shorter in the former and illustrates the need to account for measurement uncertainty in the latter.

**Keywords:** Finite populations, inference, heterogeneity, repeated measurements, prediction, misallocation.

*I have been incredibly privileged to have Manuel Arellano as my Ph.D. advisor; his unwavering support, insightful guidance and generosity are truly beyond measure. I also thank Martín Almuzara, Dmitry Arkhangelsky, Stéphane Bonhomme, Jinyong Hahn, Mikkel Plagborg-Møller, Enrique Sentana, Jeffrey M. Wooldridge and workshop participants at CEMFI. I acknowledge financial support from Grant PRE2022-000906 funded by MCIN/AEI/10.13039/501100011033 and by "ESF +", the María de Maeztu Unit of Excellence CEMFI MDM-2016-0684, funded by MCIN/AEI/10.13039/501100011033, Fundación Ramón Areces and CEMFI. All errors are my own.

†CEMFI: victor.sancibrian@cemfi.edu.es

# 1 Introduction

Empirical researchers are often interested in features of finite populations — those for which all or a non-negligible number of units are sampled: all schools in a district, most households in a village, nearly universal records on firms or workers.

When these features are directly observable upon sampling, the usual standard error formulas need to be adjusted down to reflect this abundance of information.[1] This is, however, of limited applicability in many relevant problems: school quality might not be directly observable even if all schools of reference were to be sampled; instead, we might only have access to imperfect measurements such as average test scores for different student cohorts. Similar ideas apply to learning about household-level preferences or about the frictions firms face in a particular sector.

In this paper, I propose new methods to assess estimation uncertainty in a framework where a finite population coexists with a measurement problem — where even if we observe the entire population, we may only have access to a few noisy measurements of the underlying attributes of interest. I show that conventional inference methods remain generally conservative in this context and propose Finite Population Corrections (FPCs) that lead to asymptotically correct inference for any sample-to-population fraction. FPCs rely on weak dependence across measurements and are very simple to implement.

I apply these methods to two empirical settings where uncertainty has been previously understood in different ways: predicting lethal encounters with police using data on all U.S. police departments, and studying firm misallocation with a census of large Indonesian firms. Inference is of primary interest in the former, and I show that FPCs lead to up to 50% shorter confidence intervals. Inference is usually second-order in the latter, where full-population datasets are common. When a measurement problem is nonetheless present, finite-population confidence intervals correctly reflect the dominant source of estimation uncertainty.

[1] Such results belong to a long-standing statistical literature; Cochran (1977) is a classical reference. The earlier work was done in the context of survey sampling, see for instance Neyman (1934), Hansen and Hurwitz (1943) and Horvitz and Thompson (1952).

**Setup and scope for empirical work.** The methods in this paper are relevant for applications that share two key ingredients. First, there is a well-defined population of units characterized by a set of characteristics — or attributes — and the object of interest is defined over this population: say, an average response coefficient or a regression parameter. [2] The analyst has access to a (random) sample of units from this population. Using of a sample to learn about a population introduces *sampling uncertainty*; the extent of this is determined by the sample-to-population fraction $f \in [0, 1]$. Here $f = 1$ captures full population setups, such as with data on all U.S. counties, whereas $f = 0$ is appropriate for CPS data, where it is reasonable to view it as a random sample from a much larger, "infinite" population.

Second, some of these attributes might remain unobserved even if a given unit is sampled; instead, a few error-ridden measurements of the underlying attributes are available. These are then used to estimate the parameter of interest, introducing *measurement uncertainty*. We will require that these are "good measurements" in the sense that it is possible to construct unbiased estimators of the underlying attributes. For this purpose, I consider a general class of measurement models that are affine in the underlying attributes of interest, analogous to random coefficient models in the panel data literature (Chamberlain, 1992; Arellano and Bonhomme, 2012). This is a different setup from the one considered in experimental analyses in finite populations, where uncertainty is induced by treatment randomization (Neyman, 1923/1990; Abadie, Athey, Imbens, and Wooldridge, 2020). My setup is in the model-based tradition where policy variation is not exploited for inference.

These two ingredients are prominent in the two empirical applications I consider. The first one is based on Montiel-Olea, O'Flaherty, and Sethi (2021), who draw from records on all local police departments in the U.S. to study the determinants of police use of deadly force and conduct prediction exercises involving these agencies. Some of these predictors are directly observable (such as regional laws), while others are not (such as departmental culture). The authors use a panel

[2] Sometimes it is not obvious whether one should adopt a finite-population perspective or treat the sample as drawn from a larger population that includes new, hypothetical units. This might be a useful conceptual exercise, see the discussion in Section 2 (Remark 1). The methods in this paper allow to quantify and decompose estimation uncertainty under both approaches.

of lethal encounters over 2013–2018 and a measurement system analogous to a heterogeneous Poisson model to disentangle their separate effects.

The second is in the spirit of a large literature following Hsieh and Klenow (2009). In a nutshell, firms face frictions that prevent them from choosing their inputs optimally, and these translate into firm-specific "wedges" in marginal products relative to the optimal allocation. Interest is here on investigating how these frictions relate to firm characteristics or on quantifying their cross-sectional dispersion, which is directly informative on aggregate TFP losses from misallocation. Measuring these underlying frictions is challenging: I use census panel data for manufacturing Indonesian firms from Peters (2020) and consider a persistent–transitory (fixed-effects) decomposition, following recent approaches in the literature (David and Venkateswaran, 2019; Chen, Restuccia, and Santaeulàlia-Llopis, 2022; Adamopoulos, Brandt, Leight, and Restuccia, 2022; Nigmatulina, 2023).

More generally, the analysis here is relevant for a large class of problems involving latent variables, fixed effects, factor models and random coefficient models.[3] Note that repeated measurements need not have a clear time ordering as in panel data; measurements over space or parallel measurements are also common. For instance, Kline et al. (2022) are interested in studying firm-level discrimination for a finite population of 108 Fortune 500 U.S. firms and have job-level repeated measurements for each company.

[3] Additional examples include heterogeneous earnings profiles (Guvenen, 2009), school or teacher value-added models (Gilraine, Gu, and McMillan, 2022; Hahn, Singleton, and Yildiz, 2023), modelling skill and scalability in mutual funds (Barras, Gagliardini, and Scaillet, 2022), microforecasting (Liu, Moon, and Schorfheide, 2020, 2023; Giacomini, Lee, and Sarpietro, 2023), state or country-level regressions (Villacorta, 2021), total factor productivity estimation (Klette, 1999; Combes, Duranton, Gobillon, Puga, and Roux, 2012), risk-sharing in village economies (Townsend, 1994; Schulhofer-Wohl, 2011; Chiappori, Samphantharak, Schulhofer-Wohl, and Townsend, 2014), firm-level discrimination audits (Kline, Rose, and Walters, 2022), meta-analyses (Meager, 2022), heterogeneity in returns to technology adoption in developing countries (Suri, 2011), schooling models as in Magnac, Pistolesi, and Roux (2018) or the difference-in-differences model in Bonhomme and Sauder (2011), elicitation of preferences and risk attitudes (Barsky, Juster, Kimball, and Shapiro, 1997; Andreoni and Samuelson, 2006; Ahn, Choi, Gale, and Kariv, 2014), low-rank models for time-varying treatment effects (Bonhomme and Denis, 2024), and many others.

**Finite-population inference.** I propose consistent variance estimators for a fixed number of measurements in a method-of-moments framework that incorporates these two ingredients. The parameters of interest include finite-population estimands, defined by linear instrumental-variable moment conditions for the attributes of interest, and common parameters of the measurement system.[4]

The proposed finite-population variance estimator is constructed such that it accounts for sampling-based and measurement-based uncertainty in the right proportions, that is, it is indexed by $f \in [0, 1]$. In essence, it exploits a parallel between the *measurement–sampling* decomposition in the asymptotic variance of the estimator and a (generalized) *within–between* variance decomposition. The "within" part embeds the notion of measurement uncertainty: residual variation around the underlying latent attributes of interest. The "between" part captures the idea of sampling uncertainty: differences between sample and population attributes. The finite-population variance estimator weights the latter by $(1 - f)$; this generalizes the standard Finite Population Correction (FPC) to problems where the features of interest are not directly observable upon sampling.

The within-between decomposition requires limited dependence across measurements. This is the main assumption in the paper and the point of departure relative to conventional variance estimators. I specify weak dependence as linear restrictions on the covariance matrix of the measurement errors, such as those implied by $m$-dependent processes.[5] In other words, measurement errors should be not too dependent relative to the number of measurements, and fewer restrictions are needed as more measurements become available — a common notion of weak dependence in the time-series literature. This is also natural to many problems with repeated measurements; for instance, it is the key assumption in deconvolution problems, in which one is interested in the distributional characteristics of latent variables. Importantly, all other elements in the covariance matrix

[4] I extend the framework to nonlinear transformations of the latent attributes such as variances in Remark 8. Such objects are relevant in the empirical application in Section 5.2.

[5] This is a popular approach in minimum-distance estimation of covariance structures, see Arellano (2003, Chapter 5).

remain completely unrestricted and free to vary with observable and unobservable attributes. For instance, in the context of our school quality example, it is reasonable to assume that within variation in average test scores for different cohorts is uncorrelated, while we allow for dispersion to vary over cohorts and to select on school quality. The generalization of within-between decompositions to weak dependence of this form is established by Arellano and Bonhomme (2012) in the context of estimation of distributional characteristics of random coefficient models.

The resulting variance estimator takes the form of a FPC applied to the conventional "sandwhich" estimator and is very simple to implement: FPCs only require to inversion of a selection matrix specifying the restrictions on the measurement part of the model. A drawback of the resulting estimator is that is that it is not guaranteed to be positive semi-definite in finite samples, although this is not a problem in the simulations and the empirical illustrations in this paper. The presence of other observable attributes can be exploited to construct conservative estimators following similar proposals in the design-based literature (Fogarty, 2018; Abadie et al., 2020).

The asymptotic approximations in this paper are established for a sequence of growing finite populations such that the limiting sample-to-population fraction remains representative of the sampling framework, an embedding often referred to as in the literature as finite-population asymptotics (Lehmann, 1975; Li and Ding, 2017; Abadie et al., 2020; Xu, 2021), and for a fixed number of measurements. Finite-population inference via FPCs is correct in large samples for any $f \in [0, 1]$, unlike conventional inference methods which implicitly set $f = 0$ and remain generally conservative. An exception are common parameters: intuitively, uncertainty about the measurement system itself should not depend on $f$, since the measurement problem is present regardless of the sampling framework. The same is true for finite-population estimands in the absence of heterogeneity in the underlying attributes of interest. Conversely, FPCs tend to be larger the more dispersed population attributes are, the less noise there is in the measurement system and the more measurements are available.

I complement these theoretical results with simulations for realistic designs to study the finite-sample properties of the proposed FPCs. The results show that finite-population inference maintains correct (nominal) coverage even for relatively small sample sizes ($N = 200$) and for different sample-to-population fractions, while the coverage probability for conventional confidence intervals is often one. The designs are calibrated to match reasonable signal-to-noise ratios (in the sense of relative weights of sampling and measurement components), which map to the relative width of conventional and finite-population confidence intervals in line with the discussion above.

**Empirical illustrations.** The framework developed in this paper has practical implications for a wide range of applications. To illustrate this, I revisit two very different empirical problems: a prediction exercise where uncertainty quantification is first-order and an investigation of firm-level frictions and misallocation, where empirical moments are often reported without measures of estimation uncertainty.

In the first exercise, I apply Finite Population Corrections to the results in Montiel-Olea et al. (2021). Here the population of interest are local U.S. police departments, and the data comes from the Law Enforcement Agency Identifiers Crosswalk dataset (LEAIC), which compiles information on all state and local law enforcement agencies; here we set $f = 1$. The final dataset contains 7,585 agencies and information on the number of yearly lethal encounters with police and a number of covariates including local demographics, the number of officers per thousand inhabitants and state-level dummies on the severity of laws regarding officer misconduct. The authors posit an exponential model and obtain coefficient estimates via nonlinear least squares. Next, they propose a method to obtain predictions for counterfactual-like questions of the form "what would happen to number of lethal encounters if all 10 largest agencies had the department-specific attributes of the Chicago Police Department?"

Uncertainty is here of primary interest — more so than point prediction or statistical significance — and thus the authors directly report sampling-based confidence intervals (CIs). Finite-population inference instead identifies the right source of estimation uncertainty for this problem: the fact we only observed error-ridden measures of agency-specific baseline police violence. Applying Finite Population Corrections, I find that the conventional variance estimators were overly pessimistic about prediction uncertainty: standard errors for the projection coefficients are between 20% and 60% smaller, and this in turn leads to shorter prediction intervals. For instance, the conventional 90% CI for the question above is $(545, 700)$; the finite-population 90% CI is $(565, 667)$. From a policy perspective, the finite-population CI now excludes the realized number of lethal encounters during this period, whereas the effect remains ambiguous when treating the data as a negligible sample from a hypothetical population of U.S. local police departments.

In the second exercise, I apply these methods to a measurement exercise concerned with labor "wedges" for manufacturing Indonesian (formal) firms. I follow Peters (2020), who uses census data from Statistik Industri and focuses on young firms. The final dataset is an unbalanced yearly panel covering $17,000$ firms that enter the market over 1991–1999. Following recent contributions (David and Venkateswaran, 2019; Chen et al., 2022; Adamopoulos et al., 2022; Nigmatulina, 2023), I allow for measurement error in firm-level marginal revenue products of labor and focus on fixed-effect measures of wedges. I then explore the relationship between firm-level labor wedges and firm size upon entry, which might be suggestive of size-dependent policies and regulations (Guner, Ventura, and Yi, 2008). Similar exercises are commonplace in the literature (Yeh, Macaluso, and Hershbein, 2022; Gorodnichenko, Revoltella, Svejnar, and Weiss, 2021). I also extend the framework to cover wedge dispersion statistics, an often reported measure of "allocative efficiency" that can be mapped to the TFP loss from misallocation (Hsieh and Klenow, 2009).

Finite-population inference provides again a clear recipe for uncertainty quantification: despite having data on all firms we are interested in, measurement-based uncertainty needs to be accounted for. The results point at a very imprecise relationship between firm size and labor wedges for smaller firms — even more so if one were to calculate confidence intervals as if the sample was drawn from a superpopulation of firms. The emphasis on measurement problems also has implications for allocative efficiency calculations: fixed-effects measures revise down the TFP losses from misallocation of labor to about 15% from a (biased) baseline of around 20% on average across different size groups. Finite-population confidence intervals also suggest that this difference is statistically meaningful.

In essence, finite-population inference provides a unified approach to uncertainty quantification in problems where estimation uncertainty has been previously understood in very different ways.

**Related literature.** This paper contributes to various strands of the literature.

First, it relates to the longstanding statistics literature on finite population analysis (Neyman, 1934; Hansen and Hurwitz, 1943; Horvitz and Thompson, 1952; Hájek, 1960; Erdős and Rényi, 1959; Li and Ding, 2017), which laid out the foundations of survey sampling and developed limit theorems under simple random sampling for growing sequences of finite populations; see Lehmann (1975) for a review. The focus of this literature has been on inference under sampling-based uncertainty, whereas I consider problems where sampling and measurement uncertainty coexist.

The discussion of measurement issues in this literature has revolved around the biases introduced by different forms of survey errors on estimation and prediction, see for instance Hansen, Hurwitz, and Bershad (1961) for an early contribution on (across units) interviewer bias. The literature has also noted the validity of standard variance formulas that ignore the presence of (classical) measurement error altogether as long as the Finite Population Correction is negligible, see for instance Fuller (1995). Detailed treatments of measurement issues can be found in Cochran (1977, Chapter 13) and Mukhopadhyay (2001, Chapter 7). My focus is on inference with unbiased repeated measurements, which I exploit to propose consistent standard errors for a large class of empirically relevant models.[6]

Second, this paper relates to the literature on design-based inference, which starting with Neyman (1923/1990) has been traditionally studied in a potential outcomes finite population context. These are (quasi)experimental setups where a source of uncertainty arises from randomized treatment assignment. In a context where both sampling and design uncertainty coexist, Neyman (1923/1990) noted the conservativeness of conventional variance estimators in a binary treatment setting.

[6]The use of the term "measurement error" in my setup should be understood in a broad sense; it refers to any source of random variation that contaminates the underlying features of interest.

Later contributions generalized this setup to regression models with additional covariates, general sampling frameworks and assignment mechanisms, panel experiments and nonlinear models (Freedman, 2008; Rosenbaum, 2002; Abadie, Imbens, and Zheng, 2014; Fogarty, 2018; Abadie et al., 2020; Abadie, Athey, Imbens, and Wooldridge, 2023; Bojinov, Rambachan, and Shephard, 2021; Xu, 2021). An interesting set of extensions (Deeb and de Chaisemartin, 2022; Startz and Steigerwald, 2023, 2024) allows for stochastic potential outcomes (say, due to post-randomization aggregate shocks in an RCT); this is in the spirit of measurement-based uncertainty in a cross-sectional setting. Fogarty (2018), Abadie et al. (2020) and Xu (2021) also propose conservative finite-population variance estimators exploiting the predictive power of (observable) fixed attributes in different contexts.

I regard my framework as complementary to results in this tradition, both conceptually and in practice. The conceptual difference between measurement and experimental variation is clear, and which is more appropriate is application-specific. In practice, the two setups afford different tools for estimation and inference. For instance, exact inference for sharp nulls is possible if the analyst has access to the randomization distribution in the design world (Rosenbaum, 2002; Bojinov et al., 2021). In a model-based framework, I show that the availability of repeated measurements leads to asymptotically non-conservative finite-population inference.

Third, this paper connects with the literature on fixed-$T$ panel data random coefficient models. In particular, within-group and between-group transformations to deal with permanent unobserved heterogeneity are at the heart of this literature (Chamberlain, 1992; Arellano, 2003; Arellano and Bonhomme, 2012; Graham and Powell, 2012), and weak dependence over measurements has proved useful when estimation of distributional characteristics is of interest (Kotlarski, 1967; Arellano and Bonhomme, 2012). It turns out that Finite Population Corrections can be written as a variance over heterogeneous unit-level moment conditions; I leverage these insights to study inference in a finite population context.

Finally, the framework developed in this paper allows us to reinterpret some of the existing results on inference in fixed-effects models as finite-population inference in the limit case with no sampling uncertainty. This is the case in the "many covariates" literature, which is concerned with linear regression models with a growing number of parameters (Cattaneo, Jansson, and Newey, 2018a,b; Kline, Saggio, and Sølvsten, 2020). Intuitively, removing the incidental parameters problem in this setup is analogous to removing sampling-based uncertainty. In practice, this only makes a difference for inference when the objects of interest involve functions of the large-dimensional part of the model, as in Kline et al. (2020). Another example are average marginal effects in large-$T$ nonlinear panel data models, which have been traditionally defined conditioning on the in-sample fixed effects, see Higgins and Jochmans (2024) for a recent contribution.

**Outline.** Section 2 builds intuition and illustrates the results in a simple example under simple random sampling. Section 3 generalizes the framework and presents the main results on finite-population inference. Section 4 discusses a comprehensive simulation study and Section 5 contains the two empirical applications. Proofs can be found in Appendix A.

## 2 Simple example

I first illustrate the main points of the paper in a simple example where interest is in a population average but the outcome of interest is contaminated with independent measurement noise. For reference, it might help to think of estimating average school quality in a particular district using average test scores for different cohorts.

**Setup.** Consider a population of size $n$. Unit $i$ in the population is indexed by a fixed attribute $\theta_i$, and we are interested in the population average of $\theta_i$:

$$\beta_n = E_n[\theta_i] \equiv \frac{1}{n}\sum_{i=1}^{n}\theta_i.$$

The task of the researcher is to obtain an estimate $\hat{\beta}$ of $\beta_n$ together with a quantification of estimation uncertainty, such as a standard error or a confidence interval. Randomness in $\hat{\beta}$ might arise from two sources, what I refer to as sampling-based and measurement-based uncertainty. First, we might only have access to a representative sample from the population of interest, which we indicate via the vector of inclusion indicators $R_{1:n} = (R_1, \ldots, R_n) \in \{0,1\}^n$, where $R_i = 1$ indicates that unit $i$ is sampled. Second, even if unit $i$ is sampled, we might only observe noisy measurements $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ of $\theta_i$, so that in a given sample the analyst has access to $\{R_i, R_i Y_i\}_{i=1}^n$. Additionally, this requires an equation specifying the measurement system — the way $Y_i$ relate to the underlying attributes of interest.

I formalize sampling and measurement in Assumptions S1 and S2, later generalized in Section 3.

**Assumption S1 (Simple random sampling).**

$$P\left(R_{1:n} = r_{1:n}\right) = 1 \Big/ \binom{n}{N},$$

*for each n-vector $r_{1:n}$ with $E_n[r_i] = N/n$.*

Assumption S1 describes random sampling without replacement, which leads to a sample of size $N$ from the target population. The sample-to-population fraction is thus $N/n$; the limit case where all population units are sampled corresponds to $N/n = 1$. Similarly, this formulation nests the "infinite" superpopulation framework where the sample represents a negligible fraction of the population if we let $n \to \infty$ for fixed $N$.

The attributes of interest for the sampled units are not directly observed. Instead, we have access to noisy measurements according to

$$Y_{it} = \theta_i + \varepsilon_{it}, \quad \text{for } t = 1, \ldots, T. \tag{1}$$

We assume that $E[\varepsilon_{it}] = 0$, a reasonable requirement that ensures that $Y_{it}$ are "good measurements" in the sense of being unbiased for $\theta_i$ for each unit. Note that while here we index measurements by $t$, these need not have a time ordering.

On top of this, we also assume below that there is limited dependence across measurement errors, a necessary condition in order to gauge the extent of measurement uncertainty.

**Assumption S2 (Weakly dependent measurements in the simple model).** *The measurement errors $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ in (1) satisfy $E[\varepsilon_i \varepsilon_i'] = I_T \sigma^2$; $\sigma^2 < \infty$.*

Assumption S2 implies that randomness around the attribute of interest is uncorrelated over measurements; that is, $E\left[\varepsilon_{it}\varepsilon_{is}\right] = 0$ for $s \neq t$. Throug the lens of the school quality example where $Y_{it}$ are average test scores for the $t$th student cohort, this is a natural starting point: it implies that cohort-specific variation in test scores is uncorrelated across cohorts. Similar assumptions are common in measurement systems such as (1); see for instance the discussion in Gilraine et al. (2022) in the context of teacher value-added models. The homoskedasticity assumption is made for simplicity.

In Section 3, I generalize this assumption to allow for unrestricted heteroskedasticity and different forms of dependence across measurements such as moving average errors, and I discuss the extent to which weak dependence assumptions are testable. These notions are at the heart of measurement models, as they reflect the intrinsic trade-off between unobserved heterogeneity and persistence: in an error-component model such as (1), it is $\theta_i$ that induces strong dependence in $Y_{it}$ across measurements. The suitability of a given set of restrictions should be discussed jointly with that of a given measurement model.

**Estimation and estimation uncertainty.** Let $\bar{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$ and $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$. A natural estimator for $\beta_n$ is

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i\bar{Y}_i = \frac{1}{N}\sum_{i=1}^{n} R_i\theta_i + \frac{1}{N}\sum_{i=1}^{n} R_i\bar{\varepsilon}_i,$$

where note that we average over all units, but only those with $R_i = 1$ effectively enter the sums. Using $E\left[R_i\right] = N/n$, it is easy to see that the estimator is unbiased: $E\left[\hat{\beta}\right] = \beta_n$.

The above expression also shows that the estimator decomposes into two different terms, which are at the core of much that follows. In particular, they represent orthogonal sources of estimation uncertainty: sampling and measurement. This can be read off directly from the variance of the estimator:

$$\text{(2)}\quad \text{Var}\left(\hat{\beta}\right) = \text{Var}\left(\frac{1}{N}\sum_{i=1}^{n} R_i\theta_i\right) + \text{Var}\left(\frac{1}{N}\sum_{i=1}^{n} R_i\bar{\varepsilon}_i\right) = \left(1 - \frac{N}{n}\right)\frac{\text{Var}_n\left(\theta_i\right)}{N} + \frac{\text{Var}\left(\varepsilon_{it}\right)/T}{N}.$$

where $\mathrm{Var}_n(\theta_i) = (n-1)^{-1}\sum_{i=1}^{n}(\theta_i - \beta_n)^2$ and where we have used Assumptions S1 and S2.[7] These two terms embed the notion of sampling and measurement uncertainty, respectively. The first term is the variance of the ideal estimator of $\beta_n$ if $\theta_i$ were directly observable; sampling is the only source of randomness here. The second term captures uncertainty induced by the measurement problem.

For our purposes, the most important feature in the expression above is that sampling-based uncertainty is indexed by the sample-to-population fraction $f = N/n$. It is helpful to view $\mathrm{Var}\left(\hat{\beta}\right)$ as a function of $f$, let it be denoted by $\mathrm{V}(f)$. When $f = 1$, there is no sampling uncertainty: in the absence of a measurement problem, the ideal estimator of $\beta_n$ would be $\beta_n$ itself. On the other extreme, sampling uncertainty is largest when we regard the sample as a negligible fraction of the population. This is captured by $\mathrm{V}(0) = \lim_{f\to 0}\mathrm{V}(f)$.

At the same time, measurement uncertainty does not depend on $f$: our ability to obtain more accurate measurements of $\theta_i$ for each unit is not related to the sampling framework. The relative weight of these two components in estimation uncertainty is modulated by signal-to-noise in the data: sampling uncertainty is relatively larger the more dispersed the underlying attributes are in the population (signal) and the less noise there is in the measurement system, captured by the size of the measurement errors and the number of measurements.

**Remark 1 (External validity).** An advantage of the sampling–measurement framework is that it sheds light on the relevant sources of uncertainty for a given question of interest. One notion of external validity researchers might be concerned with is that of extrapolation beyond the specific circumstances that occurred during measurement. For instance, this might involve prediction exercises or "parallel universes" where a different sequence of shocks could have realized. Appropriately accounting for measurement-based uncertainty implies that $\beta_n$ is directly

[7] In particular, we use basic results on simple random sampling without replacement repeatedly. Assumption S1 implies $E[R_i] = N/n$ and $E\left[R_i R_j\right] = N(N-1)/n(n-1)$ for $j \neq i$. It is this dependence across sampling indicators that induces the form of the first term in (2).

informative for these questions.[8] Another notion of external validity in the literature is that of generalizability of results to an exchangeable population of interest; here it is adequately accounting for sampling-based uncertainty what guarantees external validity. When that population is the one over which $\beta_n$ is defined, this follows from Assumption S1. Alternatively, we can think of extrapolability of the results to new hypothetical units drawn from a superpopulation where the original and the new "target" units are exchangeable; say with size $\tilde{n} \geq n$. We then just need to redefine $\beta_{\tilde{n}}$ as the parameter of interest. Such conceptual exercises are often useful in meta-analyses (Meager, 2022). A more meaningful question for policy purposes is that of transferability of results to a new, different population. This requires additional tools that are independent of the sampling framework, see Jin and Rothenhäusler (2024) for a discussion in a finite population context.

**Conventional inference.** The conventional variance estimator for $\hat{\beta} = N^{-1} \sum_{i=1}^{n} R_i \bar{Y}_i$ would be

$$\hat{\mathrm{V}}^{\mathrm{cluster}} = \frac{1}{N(N-1)} \sum_{i=1}^{n} R_i \left(\bar{Y}_i - \hat{\beta}\right)^2 ;$$

a cluster-robust variance estimator (Liang and Zeger, 1986; Arellano, 1987). This is the natural choice here: clustering within units accounts for the presence of the persistent component $\theta_i$ in the measurement equation in (1); this is true regardless of the degree of dependence across measurement errors. In Appendix B.1, I show that

$$E\left[\hat{\mathrm{V}}^{\mathrm{cluster}}\right] = \mathrm{V}(0) \geq \mathrm{V}(f),$$

for any sample-to-population fraction $f$. That is, the conventional variance estimator implicitly treats the sample as a random draw from a much larger population, and using $\hat{\mathrm{V}}^{\mathrm{cluster}}$ for inference introduces superfluous sampling uncertainty when this is not the case.

[8] This notion is often present in discussions about external validity when certain shocks are not accounted for; see for instance Hahn, Kuersteiner, and Mazzocco (2020) and Deeb and de Chaisemartin (2022) in the context of aggregate shocks.

By how much $\hat{\mathrm{V}}^{\text{cluster}}$ exaggerates estimation uncertainty is a matter of signal-to-noise. For instance, letting $\mathrm{Var}_n(\theta_i) = 1$ and $\mathrm{Var}(\varepsilon_{it})/T = 1$, the variance is on average twice as large as it should be when the sample is also the population. An exception is the limit case where $\theta_i = \theta$ for all units: since all underlying attributes are equal to each other, which population units are sampled and which ones are not is irrelevant.

**Finite Population Corrections.** It turns out that we can make progress when repeated measurements are available. In particular, the standard within-between variance decomposition gives

$$\widehat{\mathrm{Var}(\varepsilon_{it})} = \frac{1}{N(T-1)} \sum_{i=1}^{n} R_i \sum_{t=1}^{T} (Y_{it} - \bar{Y}_i)^2,$$
$$\widehat{\mathrm{Var}_n(\theta_i)} = \frac{1}{N-1} \sum_{i=1}^{n} R_i \left(\bar{Y}_i - \hat{\beta}\right)^2 - \frac{\widehat{\mathrm{Var}(\varepsilon_{it})}}{T},$$

which rely on weak dependence for their validity (Assumption S2). The finite-population variance estimator is then constructed via a simple adjustment to the conventional estimator — a Finite Population Correction:

$$\hat{\mathrm{V}}(f) = \hat{\mathrm{V}}^{\text{cluster}} - f\, \frac{\widehat{\mathrm{Var}_n(\theta_i)}}{N}.$$

It is not difficult to show that for any $f$ we have

$$E\left[\hat{\mathrm{V}}(f)\right] = \mathrm{V}(f),$$

so that the finite-population variance estimator reflects the right amount of sampling and measurement uncertainty in the problem. Under the regularity conditions in Section 3, $\hat{\mathrm{V}}(f)$ can then be used to perform asymptotically correct inference for any $f$ and a fixed number of measurements.

# 3 General case

In this section, I study estimation and inference for finite-population estimands in a general framework where sampling-based and measurement-based uncertainty coexist and provide Finite Population Corrections that guarantee non-conservative inference for any sample-to-population fraction.

I introduce the setup in Section 3.1 and characterize estimation uncertainty for moment-based estimators of the parameters of interest in Section 3.2. I introduce Finite Population Corrections and state the main result on non-conservative inference in Section 3.3. Proofs and derivations and relegated to Appendix A.

## 3.1 Setup

Consider a population of size $n$. Unit $i$ in the population is characterized by a set of fixed attributes $\{\theta_i, W_i\}$, and the researcher is interested in a summary measure $\beta_n$ of outcome $\theta_i$, say, an average over the population or a coefficient on a regression involving characteristics $W_i$. Probability statements are understood to hold conditional on these fixed attributes. Population averages are denoted as $E_n[f(x_i)] := n^{-1}\sum_{i=1}^n f(x_i)$ for a function $f$ applied to an array $(x_i)_{i=1}^n$.

Given a sampling framework, we use $R = (R_1, \dots, R_n) \in \{0,1\}^n$ to denote the vector of inclusion indicators; $R_i = 1$ indicates that unit $i$ is sampled. The observed data for each sampled unit is a vector of noisy measurements $Y_i = (Y_{i1}, \dots, Y_{iT})'$. In a given sample, the analyst has access to $\{R_i, R_iY_i, R_iW_i\}_{i=1}^n$. We now define the objects of interest and formalize each dimension of uncertainty.

### 3.1.1 Estimands

Let $\dim \theta_i = 1$ and $\dim \beta_n = p$. That $\theta_i$ are scalar outcomes is for simplicity and all results extend with minor modifications to the multivariate case; I will point those out throughout the exposition. The target objects $\beta_n$ solve population moment conditions $h(\theta, W, b_n)$ affine in $\theta$:

$$E_n\left[h_1(W_i;\beta_n)\left(\theta_i - h_0(W_i;\beta_n)\right)\right] = 0, \tag{3}$$

where $h_0$ (scalar-valued) and $h_1$ (of size $p \times 1$) are known functions continuously differentiable in $\beta_n$. I assume that $h_1(W_i; \beta_n) \neq 0$ for each unit in the population.[9]

The moment conditions in (3) define a broad class that includes moment-based methods such as linear regression models, IV-like estimands or nonlinear least squares. Setting $p = 1$, $h_1(W_i; \beta_n) = 1$ and $h_0(W_i; \beta_n) = \beta_n$ recovers $\beta_n = E_n[\theta_i]$ as in Section 2, a population average over heterogeneous attributes.

**Remark 2 (Prediction example, revisited).** In this application the population of interest comprises all local police agencies in the U.S., and $\theta_i$ is the baseline level of police violence associated to each department over the panel horizon. Montiel-Olea et al. (2021) are interested in observable determinants $Z_i$ of $\theta_i$ such as state-level laws or demographics and specify an exponential model of the form

$$\theta_i = \exp(Z_i'\beta_n + \alpha_i), \tag{4}$$

where $\alpha_i$ is the unexplained component and the target coefficients $\beta_n$ are defined via a nonlinear least squares problem. Through the lens of our framework, $Z_i$ are observable attributes included in $W_i$, $h_1(W_i; \beta_n) = Z_i \exp(Z_i'\beta_n)$ and $h_0(W_i; \beta_n) = \exp(Z_i'\beta_n)$; equation (10) in Montiel-Olea et al. (2021) is then exactly (3). The ultimate objects of interest (predictions and counterfactuals) then involve known transformations of these $\beta_n$ parameters.

**Remark 3 (Misallocation example, revisited).** In the exercise in Section 5.2 we are interested in characterizing the extent of resource misallocation in the formal manufacturing sector in Indonesia, and $\theta_i$ are "wedges" that measure firm-level deviations from optimal allocation of labor. A popular approach is to explore whether wedges relate systematically to observable firm-level characteristics $Z_i$, such as measures of firm size. In this context, $\beta_n$ are least-squares projection coefficients.

Equation (3), on the other hand, excludes nonlinear transformations of $\theta_i$. Generally speaking, nonlinear transformations of unbiased measurements are not unbiased, and identification and estimation require additional assumptions. It is

[9] This is trivially satisfied by redefining the subpopulation to the set of units that satisfy this condition. In other words, this condition is definitional.

nonetheless possible to extend this framework to cover certain nonlinear transformations. Some of these are of great relevance in the empirical illustration in Section 5.2, where the extent of cross-sectional dispersion in "wedges" $\theta_i$ can be directly mapped to macroeconomic aggregates. I discuss extensions to this case in Remark 8 below and in the empirical application.

#### 3.1.2 Measurement

I specify the following measurement equation for $\theta_i$:

$$Y_i = g_0(X_i; \delta) + g_1(X_i; \delta)\theta_i + \varepsilon_i, \quad E[\varepsilon_i] = 0, \tag{5}$$

where $X_i = (X'_{i1}, \ldots, X'_{iT})'$ is a $T \times \dim X_{it}$ matrix of fixed attributes contained in $W_i$,[10] $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ are independent but not necessarily identically distributed measurement errors and $\delta$ is a fixed $k$-vector of unknown parameters ($k \leq \dim X_{it}$). These might be of direct or auxiliary interest to the researcher and do not depend on population size $n$. Finally, $g_0$ and $g_1$ (of size $T \times 1$) are known functions continuously differentiable in $\delta$. This formulation allows the model to be nonlinear in observable attributes $X_i$ and common parameters and to be known only up to the latter; $T > 1$ is needed to estimate the system. Allowing for nonlinear terms in the measurement equation substantially broadens the applicability of the methods developed here; for instance, (5) covers factor models and multiplicative models with unobserved components such as the Poisson regression model in Section 5.1 (see the introduction section for examples). We also assume that $\det g_1(X_i; \delta)' g_1(X_i; \delta) \neq 0$ for all $i$, which essentially amounts to requiring that the data are informative of attribute $\theta_i$ for the population of interest.[11]

[10] Attributes in $W_i$ but not in $X_i$ are excluded instruments from the point of view of the measurement equation (5); these might help describe $\theta_i$ and enter the moment condition for $\beta_n$.

[11] This condition is analogous to that on $h_1$ in equation (3), in that it implies that the results apply for the subpopulation of units that satisfy this requirement. For instance, for a difference-in-difference measurement model where $X_{it} \in \{0, 1\}$ and $T = 2$, the restriction that $X_{i1} + X_{i2} \neq 0$ redefines the population of interest to treated units and reflects the familiar result that $\beta_n = E_n[\theta_i]$ is the ATT rather than the ATE (in the absence of additional assumptions). See Graham and Powell (2012) for a discussion of irregular models where $\det g_1(X_i; \delta)' g_1(X_i; \delta)$ might be close to zero for some units.

The zero mean assumption $E[\varepsilon_i] = 0$ ensures that the repeated measurements $Y_{it}$ unbiased for the nonstochastic component of the model. In a panel data context defined in a superpopulation where $\{W_i, \theta_i\}$ are treated as random quantities, $E[\varepsilon_i] = 0$ is a zero conditional mean assumption as in the random coefficients model in Chamberlain (1992) and Arellano and Bonhomme (2012). Assumption 1 below is the main necessary assumption of the paper and formalizes the notion of limited dependence in measurements.

**Assumption 1 (Weakly dependent errors).**

*Let $S_{(m)}$ be a $T^2 \times m$ full column rank selection matrix such that $E[\varepsilon_i \otimes \varepsilon_i] = S_{(m)}\omega_i$ for an m-vector of parameters $\omega_i$ and the measurement system defined in* (5)*. Then*

$$m \leq \frac{T(T+1)}{2} - 1. \tag{6}$$

Assumption 1 rules out fully unrestricted covariance matrices, but allows for arbitrary patterns of dependence and heteroskedasticity in the non-restricted elements $\omega_i$. Assumption 1 operationalizes the notion of weak dependence over repeated measurements via the choice of selection matrix $S_{(m)}$, which imposes linear restrictions on $\Omega_i = E[\varepsilon_i \varepsilon_i']$.[12,13]

Limited dependence is particularly appealing in a repeated measurements context, where it is expected that randomness in those is (partly) non-systematic. When measurements are drawn in parallel, independence might be reasonable; when measurements have a natural time or spatial ordering, a stronger association might be expected between closer errors than between those far apart. Moving

[12] For instance, with $T = 2$ measurements, the restriction $E[\varepsilon_{i1}\varepsilon_{i2}] = 0$ can be represented as

$$S_{(2)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}'.$$

Note that this leaves $E\left[\varepsilon_{it}^2\right]$ for $t \in \{1, 2\}$ completely unrestricted.

[13] If $\dim \theta_i > 1$, (6) is modified to

$$m \leq \frac{T(T+1)}{2} - \frac{\dim \theta_i(\dim \theta_i + 1)}{2}.$$

This result is established in Arellano and Bonhomme (2012) and reflects a fundamental trade-off between unobserved heterogeneity and error persistence in panel data models with unit-level coefficients.

average or $m$-dependent processes are convenient implementations of this idea. Similar notions of weak dependence also underpin much of the work in time series econometrics.[14]

**Remark 4 (Testable restrictions.).** When the order condition (6) is strict, Assumption 1 is testable. A standard $J$-test can be constructed following the long-standing panel data literature on testing covariance structures (Abowd and Card, 1989; Arellano, 2003; Arellano and Bonhomme, 2012). The informative content of the data for Assumption 1 and its plausibility is evident in the empirical applications that are discussed below.

#### 3.1.3 Sampling

Assumption 2 formalizes random sampling and embeds the population into a sequence of finite populations of growing sizes $n \to \infty$.

**Assumption 2 (Random sampling).**

*(i) Unit i is independently sampled with probability* $f_n > 0$.

*(ii)* $f_n$ *satisfies* $nf_n \to \infty$ *and* $f_n \to f \in [0, 1]$.

Assumption 2(i) implies that the sample size $N = \sum_{i=1}^{n} R_i$ is random and is a convenience device to avoid dealing with dependence across inclusion indicators $R_i$, moving beyond the exact results in Section 2. Note that $\hat{f} = N/n$ is a natural estimator of $f_n$; this is inconsequential for the large-sample results presented here as long as $\hat{f}/f_n \xrightarrow{\text{p}} 1$ as $n \to \infty$, which will be the case. Assumption 2(ii) ensures that as $n \to \infty$ the expected sample size $nf_n$ also increases and that the sampling

[14] Classical references are Hansen and Singleton (1982) and Newey and West (1987). The literature of repeated measurements has often favored moving average processes since they imply linear restrictions on $\Omega_i$, see for instance Bonhomme and Robin (2009) and Bonhomme and Robin (2010) in the context of factor models and the discussion in Arellano (2003, Chapter 5). Autoregressive processes, on the other hand, are not covered by Assumption 1. Still, the methods in this paper can be extended to cover such forms of dependence over measurements via quasi-differencing; see also Arellano and Bonhomme (2012, Section 3.2) for the counterpart to Assumption 1.

fraction $f_n$ has a well-defined limit. In essence, this is a way of relying on large-sample approximations for finite populations while ensuring that these remain representative of the sampling framework; for instance, choosing a sequence such that $\lim f_n = 0$ allows us to capture the standard environment where the sample becomes negligible relative to the population.[15]

#### 3.1.4 Estimator

Let $\gamma_n = (\delta', \beta_n')'$ be the $(k+p)$ parameters of interest, including both finite-population estimands and parameters of the measurement system. For a generic $\tilde{\gamma}$, let

$$u\left(Y_i, W_i, \tilde{\gamma}\right) = Y_i - g_0\left(X_i; \tilde{\delta}\right) - g_1\left(X_i; \tilde{\delta}\right) h_0\left(W_i; \tilde{\beta}\right) \tag{7}$$

and let $Q_i\left(\tilde{\delta}\right) = I_T - g_1\left(X_i; \tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right)^{\dagger}$ denote the projection on the orthogonal of the span of $g_1\left(X_i; \tilde{\delta}\right)$.[16] Consider a (non-redundant) set of instruments $A\left(W_i, \tilde{\delta}\right)$ for $\delta$, assumed to be continuously differentiable in $\delta$. I consider a method-of-moments approach with moment function

$$\psi(Y_i, W_i, \tilde{\gamma}) = \begin{pmatrix} \psi_{\delta}(Y_i, W_i, \tilde{\gamma}) \\ \psi_{\beta}(Y_i, W_i, \tilde{\gamma}) \end{pmatrix} = \begin{pmatrix} A\left(W_i, \tilde{\delta}\right) Q_i\left(\tilde{\delta}\right) \\ h_1(W_i; \tilde{\beta}) g_1\left(X_i; \tilde{\delta}\right)^{\dagger} \end{pmatrix} u\left(Y_i, W_i, \tilde{\gamma}\right). \tag{8}$$

Some intuition is as follows. Unobserved attributes $\theta_i$ are incidental parameters from the point of view of estimation of the parameters of the measurement system (5). The role of $Q_i\left(\tilde{\delta}\right)$ is to induce a transformation of the system that does not depend on $\theta_i$; note that $Q_i\left(\tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right) = 0_T$. The moment conditions for $\beta_n$, on the other hand, rescale the system so that $g_1\left(X_i; \tilde{\delta}\right)^{\dagger} g_1\left(X_i; \tilde{\delta}\right) = 1$ and then bring in the population moment conditions for $\beta_n$ defined in (3). The method-of-moments

[15] These embeddings are referred to as finite-population asymptotics in the literature, see Lehmann (1975), Aronow, Green, and Lee (2014), Li and Ding (2017), Abadie et al. (2020) and Xu (2021) for applications.

[16] For a matrix $B$, $B^{\dagger}$ denotes its Moore–Penrose pseudoinverse. In the context of panel data models, $g_1\left(X_i; \tilde{\delta}\right)^{\dagger}$ and $Q_i\left(\tilde{\delta}\right)$ are often referred to as generalized between- and within-group operators, respectively (see, for instance, Chamberlain, 1992; Arellano and Bonhomme, 2012).

estimator $\hat{\gamma}$ solves:

$$\sum_{i=1}^{n} R_i \psi(Y_i, W_i, \hat{\gamma}) = 0. \tag{9}$$

## 3.2 Characterizing estimation uncertainty

Here, I study the large-sample properties of $\hat{\gamma}$ as an estimator of $\gamma_n$, characterize its asymptotic variance and discuss estimation uncertainty in finite populations.

First, note that $\gamma_n$ solves

$$E_n \left[ E \left[ \psi(Y_i, W_i, \gamma_n) \right] \right] = 0. \tag{10}$$

This can be verified by noting that the moment function at $\gamma_n$ satisfies:

$$\psi_\delta(Y_i, W_i, \gamma_n) = A\left(W_i, \delta\right) Q_i\left(\delta\right) \varepsilon_i,$$
$$\psi_\beta(Y_i, W_i, \gamma_n) = h_1(W_i, \beta_n)\left(\theta_i - h_0\left(W_i; \beta_n\right)\right) + h_1(W_i, \beta_n) g_1\left(X_i; \delta\right)^\dagger \varepsilon_i.$$

The result follows from $E\left[\varepsilon_i\right] = 0$ and averaging over population attributes.[17] Note that setting $h_1 = 1$, $h_0 = \beta_n$ and $g_1 = 1_T$, $\psi_\beta(Y_i, W_i, \gamma_n)$ reduces to the estimation error in the simple example in Section 2. Let

$$\begin{aligned} \mathrm{V}_{\psi,n}(f_n) = {} & E_n \left[ E \left[ \psi(Y_i, W_i, \gamma_n) \psi(Y_i, W_i, \gamma_n)' \right] \right] \\ & - f_n E_n \left[ E \left[ \psi(Y_i, W_i, \gamma_n) \right] E \left[ \psi(Y_i, W_i, \gamma_n) \right]' \right]. \end{aligned} \tag{11}$$

Below we establish that the limit of $\mathrm{V}_{\psi,n}(f_n)$ as $n \to \infty$ is the inner term of the asymptotic variance; the second term in (11) is the Finite Population Correction. Finally, let $H_n = E_n \left[ E \left[ \nabla_{\tilde{\gamma}} \psi(Y_i, W_i, \gamma_n) \right] \right]$. Proposition 1 characterizes the asymptotic distribution of the (rescaled) estimation error; the following limits are assumed to exist as part of the regularity conditions.

**Proposition 1 (Asymptotic distribution).** *Under the measurement system in* (5)*, Assumption* 2 *and the regularity conditions in Assumption* 3 *in Appendix* A*, as $n \to \infty$ and*

[17] I take as given that $\delta$ is identified from $E\left[\psi_\gamma(Y_i, W_i, \gamma_n\right] = 0$. A necessary requirement is that A contains $\dim \delta = k$ valid instruments.

*for given* $T > 1$*:*

$$\sqrt{N}\,(\hat{\gamma} - \gamma_n) \xrightarrow{\text{d}} \mathrm{N}\,(0, V(f))\,,$$

*where*

$$V(f) = \left(\lim_{n\to\infty} H_n\right)^{-1} \lim_{n\to\infty} V_{\psi,n}(f_n) \left(\lim_{n\to\infty} H_n'\right)^{-1}. \tag{12}$$

*Proof.* See Appendix A. □

Proposition 1 is the finite-population counterpart to standard results for problems with repeated measurements.[18] While the asymptotic variance has the usual "sandwich" form, it is indexed by $f$. In other words, estimation uncertainty depends on the sample-to-population fraction — even if the measurement problem does not. The reason for this is simple: these two sources of uncertainty are orthogonal to each other, and randomness in $\hat{\gamma}$ reflects both. As $f \to 1$, randomness due to sampling disappears and the asymptotic variance of $\hat{\gamma}$ adjusts proportionally via the Finite Population Correction. Note that the FPC is positive-semidefinite; it follows that for $f' \geq f$, $\mathrm{V}(f') \leq \mathrm{V}(f)$ in the matrix sense.

Two particular cases are worth highlighting. First, when $f = 0$ the standard sandwich formula recovers. This is the basis for the standard, superpopulation-based approach to inference; let $\hat{\mathrm{V}}(0)$ denote any such estimator. It then follows that $\hat{\mathrm{V}}(0)$ is generally inconsistent for the finite-population variance when $f > 0$, and that conventional standard errors tend to exaggerate estimation uncertainty.

Second, the FPC is zero when $E\left[\psi(Y_i, W_i, \gamma_n)\right] = 0$, that is, when the population moment condition (10) holds *for each unit* $i$. When it holds only on average, the FPC is precisely equal to the variance of these heterogeneous unit-level moment conditions over the population, and is larger the more dispersed these are. In the simple example in Section 2, $E\left[\psi(Y_i, W_i, \gamma_n)\right] = \theta_i - \beta_n$ and the FPC equals $E_n\left[(\theta_i - \beta_n)^2\right]$, the variance over population heterogeneous responses. It is this

[18] This can be established with our sampling assumption (Assumption 2) and regularity conditions for moment-based estimators (Newey and McFadden, 1994), but does not rely on assumptions of dependence across measurements. See also Xu (2021) for a similar result in a world with design-based uncertainty and no repeated measurements.

insight and the connection to random coefficient models that I exploit to propose Finite Population Correction estimators.

One relevant case in which FPCs are zero are parameters of the measurement model, denoted here $\delta$. This follows from $E\left[\psi_\delta(Y_i, W_i, \gamma_n)\right] = 0$; it can be verified that the upper-left $k \times k$ block of $\Sigma_n$ is a zero matrix. This is intuitive: since the measurement problem is present regardless of how much of the population we sample, uncertainty about the measurement system itself should not depend on $f$. When population attributes $\{\theta_i\}_{i=1}^n$ are actually homogeneous, $\beta_n$ becomes a common parameter. Through the lens of the causal inference literature, this is the classical result that conventional standard errors with randomized treatments are not conservative under constant treatment effects (Neyman, 1923/1990; Abadie et al., 2020).

**Remark 5 (Perfect measurements).** Suppose that outcome attributes are observed without error; for simplicity, set $Y_{it} = \theta_i$ and $T = 1$. Then the moment function $\psi_\beta$ is nonstochastic and the variance of the moment condition in (11) adapts to reflect so:

$$\mathrm{V}_{\psi,n}(f_n) = (1 - f_n) E_n\left[\psi_\beta(Y_i, W_i, \gamma_n)\psi_\beta(Y_i, W_i, \gamma_n)'\right].$$

This is analogous to the classical finite-populations literature where sampling-based uncertainty is the only source of randomness in $\hat{\beta}$. The usual variance estimator (say, heteroskedasticity-robust) $\hat{\mathrm{V}}(0)$ is conservative, but an adjustment is here straightforward: $\hat{\mathrm{V}}(f) = (1 - f)\hat{\mathrm{V}}(0)$ will do.

Whichever the setup, the dominant paradigm in empirical work is to interpret uncertainty as-if derived from an infinite population. In the next section, I show that the discussion above is not only about the interpretation of uncertainty but has practical consequences: estimating FPCs is possible when repeated measurements are available.

## 3.3 Finite Population Corrections

Our goal in this section is to propose consistent finite-population standard errors and confidence intervals for $\hat{\gamma}$.

Let $\hat{u}_i \equiv u(Y_i, W_i, \hat{\gamma})$, where $u(Y_i, W_i, \tilde{\gamma})$ is defined in (7). This is a compound residual term that includes both types of unobservables in the measurement equation: attributes and measurement errors. In a nutshell, the idea is to mimic the approach in the simple example in Section 2 and propose variance estimators that weight these two elements according to some $\tilde{f} \in [0,1]$. Let $Q_i^*\left(\tilde{\delta}\right)$ denote the projection on the orthogonal of the span of $g_1\left(X_i; \tilde{\delta}\right) \otimes g_1\left(X_i; \tilde{\delta}\right)$.[19] The following are weighted unit-level contributions to the finite-population variance:[20]

$$\hat{\Lambda}_i\left(\tilde{f}\right) = \text{vec}^{-1}\left[\left(1-\tilde{f}\right) I_{T^2} + \tilde{f} S_{(m)}\left(Q_i^*\left(\hat{\delta}\right) S_{(m)}\right)^{\dagger} Q_i^*\left(\hat{\delta}\right)\right](\hat{u}_i \otimes \hat{u}_i)\,. \tag{13}$$

The cross-products $\hat{u}_i \otimes \hat{u}_i$ weighted by $\left(1-\tilde{f}\right)$ include both attributes and measurement errors; those weighted by $\tilde{f}$ include only measurement errors. The latter are constructed by imposing the covariance structure in Assumption 1 and then projecting out the part involving the attributes via $Q_i\left(\hat{\delta}\right)$. This estimator is based on the constructive identification proof in Arellano and Bonhomme (2012) for covariance structures in panel data random coefficient models. Importantly, this is the only modification that finite-population standard errors will require relative to conventional approaches: implementation only requires defining a selection matrix and a projection matrix.

Now, the finite-population variance estimator of the score is

$$\hat{V}_{\psi}\left(\hat{f}\right) = \frac{1}{N}\sum_{i=1}^{n} R_i \begin{pmatrix} A\left(W_i, \hat{\delta}\right) Q_i\left(\hat{\delta}\right) \\ h_1(W_i; \hat{\beta}) g_1\left(X_i; \hat{\delta}\right)^{\dagger} \end{pmatrix} \hat{\Lambda}_i\left(\hat{f}\right) \begin{pmatrix} A\left(W_i, \hat{\delta}\right) Q_i\left(\hat{\delta}\right) \\ h_1(W_i; \hat{\beta}) g_1\left(X_i; \hat{\delta}\right)^{\dagger} \end{pmatrix}',$$

where recall that $\hat{f} = N/n$. Note that using $\hat{\Lambda}_i(0)$ instead yields the conventional estimator of the variance of the scores for repeated measurement models, which

[19] This is the counterpart of $Q_i\left(\tilde{\delta}\right)$ for cross-products of the data:

$$Q_i^*\left(\tilde{\delta}\right) = I_{T^2} - g_1\left(X_i; \tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right)^{\dagger} \otimes g_1\left(X_i; \tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right)^{\dagger}.$$

[20] $\text{vec}_{m,n}^{-1} : \mathbb{R}^{mn} \to \mathbb{R}^{m\times n}$ is the inverse vec operator. For an $m \times n$ matrix $B$, we have $= \text{vec}^{-1}\,\text{vec}\,B = B$. I omit the subscripts in the text since I only use $\text{vec}^{-1}$ here to reconstruct $T \times T$ matrices. This is readily available in commercial software, such as via `reshape` in MATLAB.

averages over both dispersion in attributes and dispersion in measurement errors: the presence of the former reminds us that this is in the class of cluster-robust variance estimators.

Let $\hat{H} = N^{-1} \sum_{i=1}^{n} R_i \nabla_{\tilde{\gamma}} \psi(Y_i, W_i, \hat{\gamma})$. The estimator of the finite-population variance in (12) is given by

$$\hat{\mathrm{V}}\left(\hat{f}\right) = \hat{H}^{-1} \hat{\mathrm{V}}_{\psi}\left(\hat{f}\right) \hat{H}'^{-1}. \tag{14}$$

For $\hat{\mathrm{V}}\left(\hat{f}\right) \geq 0$ and an arbitrary column vector $\lambda \neq 0_{(k+p)\times 1}$, the finite-population standard error is $\hat{\sigma}_{\lambda}\left(\hat{f}\right) = \sqrt{\lambda' \hat{\mathrm{V}}\left(\hat{f}\right) \lambda / N}$. Finally, the $(1-\alpha)$ confidence interval for $\lambda' \gamma_n$ is

$$\hat{C}_{\lambda,\alpha}\left(\hat{f}\right) = \left[\lambda' \hat{\gamma} \pm z_{1-\alpha/2} \hat{\sigma}_{\lambda}\left(\hat{f}\right)\right],$$

where $z_q$ is the $q$-quantile of the standard normal distribution. Proposition 2 below states that this leads to non-conservative inference for any $f_n$ that satisfies Assumption 2.

**Proposition 2 (Asymptotically correct inference).** *Under the measurement system in* (5), *Assumption* 1, *Assumption* 2 *and the regularity conditions in Assumption* 3 *in Appendix* A, *if* $\operatorname{rank} Q_i^* (\delta) S_{(m)} = m$ *then as* $n \to \infty$ *and for given* $T > 1$

$$\lim_{n\to\infty} P\left(\lambda' \gamma_n \in \hat{C}_{\lambda,\alpha}\left(\hat{f}\right)\right) = 1 - \alpha.$$

*Proof.* See Appendix A. □

The most relevant assumption behind Proposition 2 is that of weak dependence across measurements. Assumption 1, however, is not sufficient. We also require the more primitive condition that $Q_i^* (\delta) S_{(m)}$ has linearly independent columns and thus the left inverse in (13) is well-defined. This rank condition rules out cases where it is not possible to distinguish attributes from dependence in measurement errors from the second-order moments of the data even if restrictions are such that there are sufficient free parameters in the "reduced-form" covariance matrix.[21]

[21] One such example is provided by the panel data literature on distinguishing unobserved heterogeneity from genuine dependence, where a measurement model with uncorrelated measurements is

Finally, note that all elements of $\hat{\mathrm{V}}\left(\hat{f}\right)$ are generally speaking a function of $\hat{f}$, despite our discussion following Proposition 1 that estimation uncertainty for common parameters is independent of the sampling fraction. Under Assumption 1, the proposed variance is nonetheless valid: as we move along $\tilde{f} \in [0,1]$, we are only changing the relative weight of the attributes component in the compound residual term, but the upper-left $k \times k$ block of $\hat{V}_{\psi}(\tilde{f})$ is constructed such that it projects out this component.[22] For $\tilde{f} = 1$ and from the point of view of common parameters, this can be seen as a generalization of the approach in Stock and Watson (2008).

Note that the variance estimator in (14) is not guaranteed to be positive semidefinite for all $\tilde{f} \in [0,1]$ as written, although there always exists some $\tilde{f}$ for which this is the case. A natural alternative is to use a conservative estimator, say $\hat{\mathrm{V}}(0)$.[23] This is not an issue neither in our empirical applications nor in the simulation evidence presented in Section 4.

**Remark 6 (Finite Population Corrections).** Note that (14) can be written as:

$$\hat{\mathrm{V}}\left(\hat{f}\right) = \hat{\mathrm{V}}(0) - \hat{f}\left(\hat{\mathrm{V}}(0) - \hat{\mathrm{V}}(1)\right),$$

which has the intuitive form "conventional estimator − FPC." The above representation is most useful when different conventional estimators might be available, such as when $\delta$ are controls and estimation proceeds in two steps. It is then common to resort to bootstrap methods for inference on $\hat{\beta}$; an example of this is the first of my empirical illustrations. Let us focus on the $j$th entry of $\hat{\beta}$ and denote by

observationally equivalent to one with common attributes ($\theta_i = \theta$ for all units) and serial correlation in very short panels ($T = 2$), see Arellano (2003, pp. 58–60) for additional details. More generally, the rank condition fails when covariance restrictions do not restrict dependence across measurements but only impose homogeneity assumptions such as equal diagonal entries.

[22]This follows from the definition of common parameters themselves, see again the discussion in Section 3.1.4. Of course, if Assumption 1 fails, only the estimator that sets $f = 0$ would be consistent.

[23]This is a common drawback of estimators that are constructed by subtracting terms, as it becomes clear in the remark below. As such, asymptotically valid estimators can also be constructed in a number of standard ways in the literature, such as rotating the eigenvalues in the eigendecomposition of $\hat{\mathrm{V}}\left(\hat{f}\right)$; see for instance the discussion for two-way clustering in Cameron, Gelbach, and Miller (2011).

$\tilde{\mathrm{V}}_{\beta,j}(0)$ the bootstrap variance.[24] A finite-population variance estimator for $\hat{\beta}_j - \beta_{n,j}$ that is valid in the sense of Proposition 2 is then

$$\tilde{\mathrm{V}}_{\beta,j}(0) - \hat{f} e_j' \hat{H}_\beta^{-1} \left( \hat{V}_{\psi_\beta}(0) - \hat{V}_{\psi_\beta}(1) \right) \hat{H}_\beta'^{-1} e_j,$$

where $e_j$ is the basis vector of size $p$, $\hat{H}_\beta$ indexes the corresponding $p \times p$ block of $\hat{H}$ and $\hat{V}_{\psi_\beta}\left(\hat{f}\right)$ is analogous to $\hat{V}_\psi\left(\hat{f}\right)$ but only involves the $\psi_\beta$. In other words, the analyst just needs an estimate of FPC, and $\tilde{\mathrm{V}}_{\beta,j}(0)$ automatically takes care of two-step uncertainty (see also Newey and McFadden, 1994, Chapter 6).

**Remark 7 (Conservative finite-population inference).** Asymptotically correct inference requires limited dependence in measurements. When Assumption 1 is not attractive, it is nonetheless possible to compute partial FPCs by leveraging the predictive content of covariates $W_i$ for the attributes of interest. In particular, building on similar ideas from the causal inference literature (Fogarty, 2018; Abadie et al., 2020; Xu, 2021), one can regress scores $\psi_\beta(Y_i, W_i, \hat{\gamma})$ on observable attributes and use the variance of the predicted values to form partial FPCs.

**Remark 8 (Extensions to higher order moments).** Model (5) is a measurement equation for $\theta_i$. The results presented here extend to a nonlinear transformation of $\theta_i$, say $\theta_i^2$, if a measurement equation for $\theta_i^2$ is available. A different question is whether this is also the case if we maintain (5) as a baseline equation. One possibility is as follows; suppose for simplicity that $\delta$ are known. Let $Y_i^* = (Y_i - g_0(X_i;\delta)) \otimes (Y_i - g_0(X_i;\delta))$, $g_1^*(X_i;\delta) = g_1(X_i;\delta) \otimes g_1(X_i;\delta)$ and define

$$\tilde{Y}_i^* = \left[ I_{T^2} - S_{(m)} \left( Q_i^*(\delta)\, S_{(m)} \right)^\dagger Q_i^*(\delta) \right] Y_i^*. \tag{15}$$

It can then be shown that:

$$\tilde{Y}_i^* = g_1^*(X_i;\delta)\theta_i^2 + \tilde{\varepsilon}_i^*, \quad E\left[\tilde{\varepsilon}_i^*\right] = 0,$$

which is a measurement equation for $\theta_i^2$ of the form (5). We can then make progress by characterizing the covariance structure of $\tilde{\varepsilon}_i^*$ in parallel to the exposition above.

[24] Similar ot the cluster-robust case, it can be shown that nonparametric (block) bootstrap estimators are also only consistent for V(0), regardless of the sampling fraction $f$.

A limitation of this approach is that it imposes more stringent conditions on the number of available measurements relative to those needed for estimation. Still, objects such as the population-level dispersion of $\theta_i$ are of great interest in the context of the application in Section 5.2. I take a slightly different route and propose there a non-conservative variance estimator based on higher-order cumulants that directly estimates the FPC. This also illustrates the discussion in Remark 6.

# 4 Simulation study

Here I discuss a simulation study intended to illustrate the discussion so far and verify the finite-sample properties of the inference procedures proposed in the previous section.

**Design.** The design here considers a relatively simple measurement system that is additive in a scalar attribute of interest $\theta_i$, and thus in the spirit of Section 2 and the empirical illustration in Section 5.2. I augment it with some additional ingredients as follows.

First, we consider population outcome attributes $\{\theta_i\}_{i=1}^n$ that are drawn from a superpopulation $\theta_i \sim \mathrm{N}\left(1, \sigma_\theta^2\right)$; interest is on the population average $\beta_n = E_n\left[\theta_i\right]$. We also define a $T$-vector of observable attributes $X_i$ such that

$$X_{i0} = (1 - 0.25\theta_i) + |\theta_i| U_{i0},$$
$$X_{it} = 0.8 X_{i,t-1} + U_{it},$$

and $U_{it} \sim \mathrm{t}_{(\kappa)}$ independently for $t = 0, \ldots, T$. This allows for persistence and non-normal features in attributes $X_i$, and induces dependence or "fixed-effects endogeneity" in $\theta_i$. The population is thus characterized by $\{\theta_i, X_i\}_{i=1}^n$. The measurement equation for $\theta_i$ is specified as

$$Y_{it} = \theta_i + \delta X_{it} + \epsilon_{it}, \tag{16}$$

where $\epsilon_{it} \sim \mathrm{N}\left(0, X_{it}^2\right)$ independently over measurements. This augments the simple measurement model with a common parameter $\delta$, which has to be estimated in a first step, and heteroskedasticity in measurement errors.

The design sets $(\sigma_\theta, \kappa)$ to control signal-to-noise. That is, we consider different relative weights of sampling-to-measurement uncertainty in the variance of the estimator. Note that the presence of $\delta$ adds two-step uncertainty to the problem, which in practice is equivalent to measurement uncertainty in the sense that the measurement system is not fully known. I also consider relatively small sample sizes ($N = 200$) and $T = 3$ measurements, and vary population size according to a grid of sample-to-population fractions $f \in \{0, 0.1, \ldots, 0.9, 1\}$. For instance, $f = 0.1$ is associated to a population of $n = 2,000$ units. The results for $f = 0$ correspond to the superpopulation data generating process (that is, the estimand equals one).

**Results.** Figures 1, 2 and 3 report coverage and width of finite-population confidence intervals over different sample-to-population fractions and for three signal-to-noise regimes (low, moderate and large, respectively). I also report conservative (or superpopulation) confidence intervals Liang and Zeger (1986); Arellano (1987) that impose $f = 0$ regardless of the actual sampling framework. I use critical values based on $t_{n-1}$ as recommended in (Hansen, 2007) for cluster-robust estimators.

The results suggest an excellent performance of finite-population inference even for relatively small sample sizes, maintaining coverage close to nominal for all sample-to-population fractions and signal-to-noise regimes considered. Similarly, the figures also illustrate the conservativeness of conventional estimators for $f > 0$. In particular, actual coverage increases monotonically as $f \to 1$, and is one or close to one for cases where the sample is a large fraction of the population.

The extent of conservativeness is better captured by looking at the relative width of confidence intervals, and so is the size of Finite Population Corrections as a result. In line with the discussion above, these are larger the more dispersed the underlying attributes are relative to the size of measurement errors. In particular, for the limit case $f = 1$ the relative width is around 0.85 in the low signal-to-noise regime, 0.6 in the moderate one and around 0.5 in the high signal-to-noise one. As expected and in parallel, the actual coverage probability tends to increase for conservative confidence intervals as signal dominates, while remaining close to 0.95 for finite-population ones.

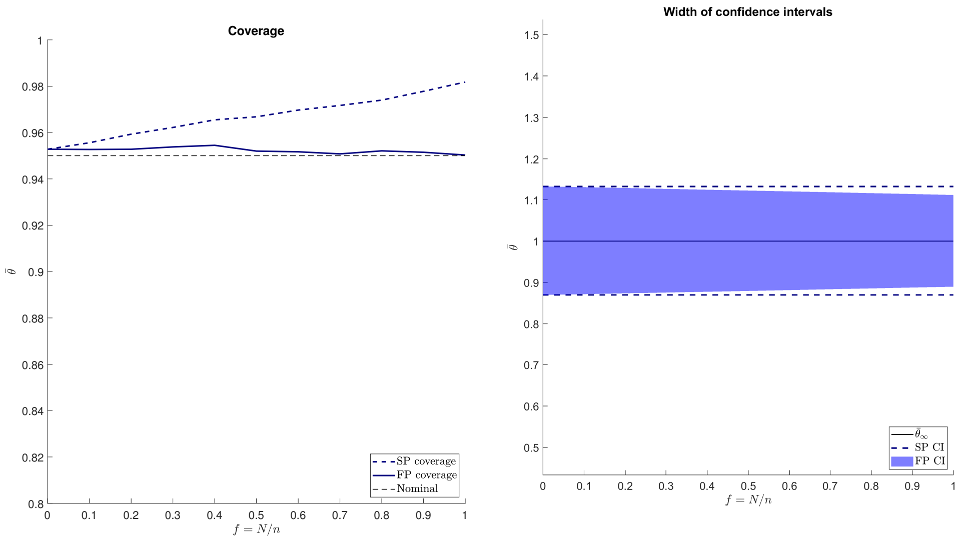



FIGURE 1. Results for $\beta_n$ and the measurement model in equation (16): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise $\approx$ 0.5.

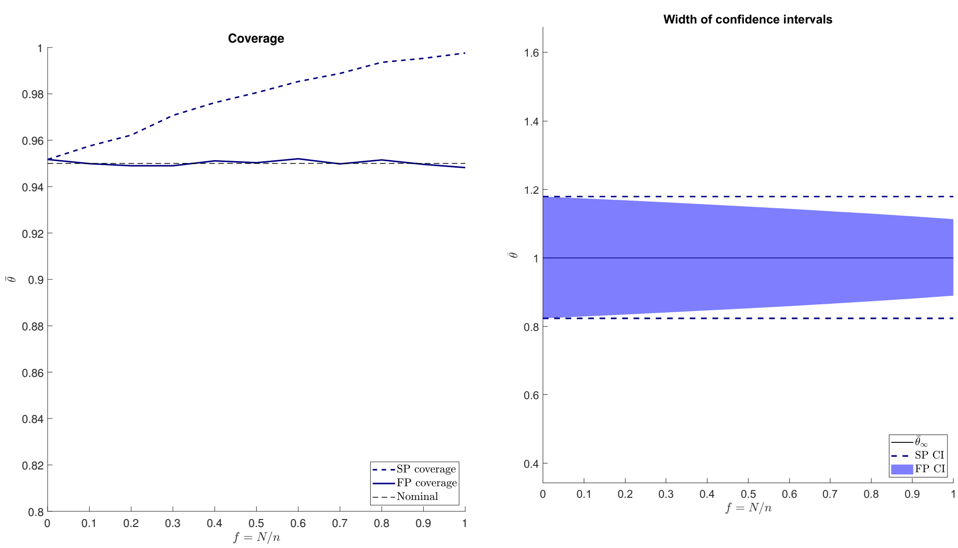



FIGURE 2. Results for $\beta_n$ and the measurement model in equation (16): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise $\approx$ 1.5.

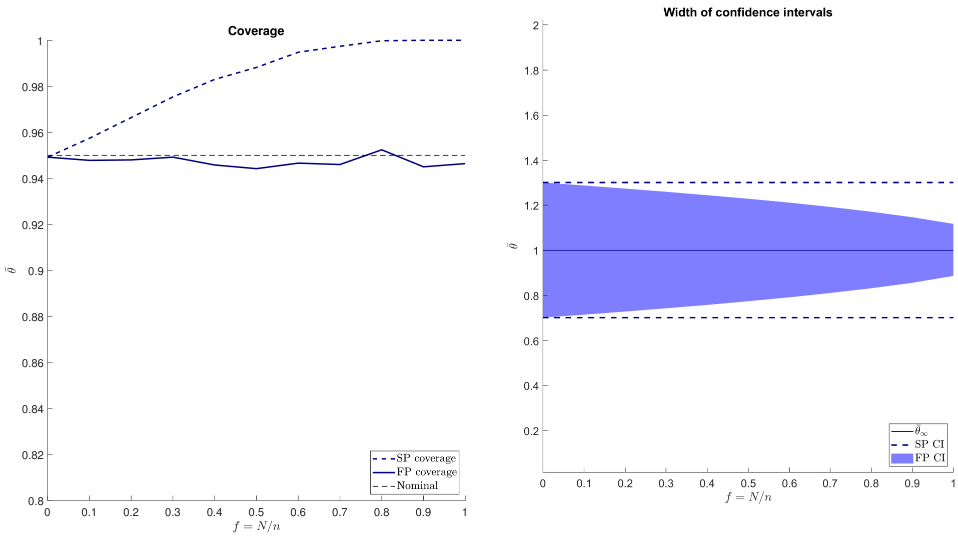



FIGURE 3. Results for $\beta_n$ and the measurement model in equation (16): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise $\approx$ 2.5.

# 5 Empirical illustrations

In this section, I illustrate how the methods in the previous sections allow for a systematic approach to uncertainty quantification in finite populations by considering two setups that span the wide range of empirical applications for which this paper is relevant.

## 5.1 Predicting police violence

The first exercise is based on Montiel-Olea et al. (2021), which are interested in the determinants of the use of deadly force by police officers in the United States. More generally, it is aimed at illustrating finite-population inference in the microforecasting literature (Liu et al., 2020, 2023; Giacomini et al., 2023), which is concerned with prediction of individual outcomes in short panels.

**Data and background.** The authors collect data on all local police departments in the United States, defined as those that serve a well-defined population. They use census records from the Law Enforcement Agency Identifiers Crosswalk dataset (LEAIC), and the final dataset contains $N = 7,585$ agencies.[25] The authors are interested in characterizing the determinants of police use of deadly force in the U.S. at the department level and aim at retrieving comprehensive records for all such departments: $f = 1$ is arguably a reasonable description of the sampling framework.

Here interest is a on a composite index $\theta_i$ for the agency-specific baseline level of lethal encounters, which is specified as an exponential model including observed, candidate determinants $Z_i$ and unobserved, residual attributes $\alpha_i$; see again equation (4). The measurement system is specified as a multiplicative (Poisson) model for the number of lethal encounters, which can be recovered by setting $g_0 = 0_T$ and $g_1(X_i; \delta) = \exp(X_i\delta)$ in equation (5). In a slightly rearranged form, we have

$$Y_{it} = \alpha_i \exp(Z_i'\beta_n + X_{it}\delta) + \varepsilon_{it} \tag{17}$$

for $t = 1, \ldots, 6$, corresponding to yearly measurements over 2013–2018. In the main specification, $\dim X_{it} = 1$ and $X_{it}$ are murders per 100,000 population served. Throughout the period, 1,179 agencies have at least one lethal encounter, for a total number of 3,504 homicides. The parameters of interest are $\gamma_n = (\delta, \beta_n')'$. The method-of-moments estimator $\hat{\gamma}$ solves a moment condition of the form (9).[26]

The ultimate objects of interest are counterfactual lethal encounters for agency $i$ that would obtain if we were to replace some of its observed and unobserved intrinsic characteristics — encapsulated in $Z_i$ and $\alpha_i$, respectively — with those of agency $j$. Since the latter remain unobserved, the authors propose an estimator based on an Empirical Bayes approach, the Poisson model and the assumption of weakly dependent measurements. For our purposes, what matters is that this is a known mapping of estimated coefficients to the predicted number of lethal

[25] The authors have made the data and code publicly available at https://github.com/jm4474/EmpiricalBayesCounterfactuals.

[26] In particular, we have discussed how to write $\beta_n$ as a finite-population estimand in the sense of equation (5) in Remark 2. This leads to moment conditions for $\beta_n$ (given $\delta$) as in eq. (10) in Montiel-Olea et al. (2021). For the common parameter $\delta$, I follow the authors and set $\mathrm{A}(X_i, \delta) = X_i'$ in the moment function in (8); see eq. (8) in Montiel-Olea et al. (2021).

encounters:

$$\hat{Y}^*_t(i, j, z) = \frac{\bar{Y}_j + 1}{\exp\left\{Z'_j\hat{\beta}\right\}\sum_{t=1}^{T}\exp\left\{X_{jt}\hat{\delta}\right\}}\exp\left\{z'\hat{\beta} + X_{it}\hat{\delta}\right\}, \tag{18}$$

where $(i, j, z)$ denotes a counterfactual for agency $i$ if it had the unobserved characteristics of agency $j$ and the observed characteristics $z$. For instance, $\hat{Y}^*_t(i, i, z_j)$ are estimated counterfactuals for agency $i$ with its own unobserved determinants (interpreted as selection and training practices and departmental culture) and with the observed characteristics of agency $j$.

**Results.** Table 1 reports the point estimates and standard errors for $\hat{\gamma}$; the first row corresponds to $X_{it}$ and the subsequent entries correspond to the predictors $Z_i$. The second and third columns report conventional and finite-population standard errors, respectively. Conventional standard errors are based on standard method-of-moments variance estimators, which correspond to the diagonal entries of the square root of $\hat{\mathrm{V}}(0)$ in equation (14). Finite-population standard errors are constructed by calculating $\hat{\mathrm{V}}(1)$ assuming (conditionally) uncorrelated measurements, a maintained assumption in Montiel-Olea et al. (2021).

Table 1 shows that ignoring the finite-population dimension of the problem leads to standard errors that are between 1.2 and 2.5 times larger than the finite-population ones, a byproduct of introducing sampling-based uncertainty. For example, the standard error on the estimated coefficient associated to the poverty share goes from 0.003 to 0.007. Differences in the magnitude of the change can be traced back to how each particular predictor loads on signal-to-noise. Note that the standard errors on the coefficient associated to murders per 100,000 population served are unchanged: the Finite Population Correction is zero for common parameters, along the lines of our discussion in Section 3.

Importantly, this is just a first step towards computing counterfactuals. Statistical significance is not necessarily of interest here; instead, Table 1 is relevant in that uncertainty in the estimated counterfactuals in equation (18) stems directly from the covariance matrix of these estimated coefficients. The authors consider different types of counterfactuals; here we focus on those of the form $\hat{Y}^*_t(i, j, z_j)$, where both

TABLE 1. Estimates of the parameters in equation (17).

| | Coefficient | Conventional s.e. | FP s.e. |
|---|---|---|---|
| Murders per pop. (in hund. ths.) | 0.005 | 0.003 | 0.003 |
| Log of avg. pop. (in m.) | 1.192 | 0.049 | 0.036 |
| Officers per pop. (in ths.) | 0.012 | 0.004 | 0.004 |
| Gun death rate (%) | 0.049 | 0.01 | 0.004 |
| Share in poverty (%) | 0.04 | 0.007 | 0.003 |
| Share black (%) | -0.024 | 0.004 | 0.002 |
| Garner | -0.031 | 0.127 | 0.102 |
| LEOBR | -0.05 | 0.113 | 0.066 |
| Land area (sq. km. per m.) | 1.0231e-05 | 1.1511e-06 | 7.4896e-07 |

*Notes:* The first row corresponds to the time-varying variable $X_{it}$ in equation (17); the rest are time-invariant predictors $Z_i$. "Garner" are dummy variables indicating the severity of state laws on the use of deadly force and "LEOBR" are dummy variables for state laws protecting police from misconduct allegations, see Montiel-Olea et al. (2021) for additional details. "*hund. ths.*" stands for "hundred thousands", "*m.*" stands for 'millions' and "*sq. km.*" for "square kilometers". The second and third columns report baseline standard errors (as in Montiel-Olea et al. (2021)) and finite-population standard errors for $f = 1$, respectively.

observed and unobserved characteristics of agency $i$ are replaced with those of agency $j$.[27]

Table 2 reports the results for four of the ten largest departments according to population served (Phoenix, Chicago, Philadelphia and New York) and for these ten combined ("Totals"). A full list of counterfactuals is reported in Appendix C.1. Note that Table 2 directly reports prediction intervals rather than point estimates, calculated by drawing from the estimated asymptotic distribution. This is in line with the authors' emphasis on quantifying estimation uncertainty, something that makes this application particularly interesting for our purposes. Rows correspond to agency $i$ and columns to agency $j$; the diagonal elements are the actual number

[27]In particular, we consider the counterfactual values of *Officers per pop.*, *Gun death rate*, *Share in poverty*, *Garner* and *LEOBR* (see Table 1).

of realized lethal encounters during the period. For instance, we might ask the following question:

*"What would happen to the number of lethal encounters if all ten largest agencies had the department-specific attributes of the Chicago Police Department?"*

We can read this off Table 2: the 90% finite-population prediction interval is $(565, 667)$, and the number of lethal encounters is thus expected to increase from a (realized) baseline of 548 encounters during the period. The answer to this question is however inconclusive if we were to calculate these prediction intervals as if the U.S. local police departments were a small subset of a much larger superpopulation: the 90% finite-population prediction interval increases to $[545, 700]$. Not only is the finite-population interval 34% smaller than the conventional one, it also leads to substantively different policy directions for the questions that the authors seek to answer.

The discussion here illustrates that finite-population inference identifies the right source of estimation uncertainty for this problem — the fact that we only observe error-ridden measurements of agency-specific baseline police violence — and that Finite Population Corrections can lead to substantially more precise and meaningful uncertainty assessments.

TABLE 2. Counterfactual homicides: observed and unobserved determinants (selection of departments)

| | Phoenix | Chicago | Philadelphia | New York |
|---|---|---|---|---|
| Phoenix | **93** | [28,32] | [21,30] | [5,9] |
| | | (28,32) | (23,28) | (6,8) |
| Chicago | [189,216] | **63** | [46,64] | [12,19] |
| | (190,214) | | (49,61) | (13,18) |
| Philadelphia | [89,130] | [29,40] | **28** | [7,9] |
| | (96,120) | (30,38) | | (7,9) |
| New York | [568,1013] | [184,310] | [175,236] | **55** |
| | (643,870) | (204,275) | (180,228) | |
| Totals | [1689,2279] | [545,700] | [481,567] | [125,166] |
| (**548**) | (1791,2094) | (565,667) | (481,564) | (134,155) |

*Note*: The agencies above are a selection of those in Table 3 in Appendix C.1, which cover the ten largest departments by population served. Diagonal entries are observed lethal encounters (totalling 548 encounters for the top ten departments). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, j, z_j)$ in equation (18), which replace characteristics of agency $i$ in the rows with those of agency $j$ in the columns; see the text for additional details. Baseline prediction intervals (as in Montiel-Olea et al. (2021)) are reported in brackets and finite-population prediction intervals for $f = 1$ are reported in parenthesis.

## 5.2 Misallocation

The second exercise is motivated by the large literature on resource misallocation, which is based on the observation that differences in aggregate TFP might not be driven solely by technology but also by allocative efficiency. Following Hsieh and Klenow (2009), an extensive body of work has provided evidence of substantial heterogeneity in revenue productivity within industries, which under appropriate conditions can be used to quantify the extent of misallocation. See Restuccia and Rogerson (2017) for a review.

Exploring the sources of misallocation and obtaining aggregate summary statistics requires a combination of rich microdata and careful measurement, which makes this an appealing framework to illustrate the methods in this paper. In remarkable contrast to the previous exercise, here the literature has often understated or ignored estimation uncertainty.

**Data and background.** For illustration, consider the monopolistic competition framework in Hsieh and Klenow (2009), where firms hire labor and capital in competitive markets, have Cobb-Douglas production functions and might face output, capital or labor distortions such as output subsidies, differential access to credit or labor market regulations. These create "wedges" relative to the efficient allocation, which manifests in heterogeneity in the marginal revenue product of capital and labor (MRPK and MRPL, respectively) within a given industry.

Measuring these firm-level wedges is challenging. Even if marginal revenue products can be measured in the data, within firm variation over short periods of time might reflect measurement errors, adjustment costs or transitory shocks. A popular approach is to focus on persistent–transitory decompositions such as firm fixed-effects in marginal revenue products as measures of these underlying wedges (David and Venkateswaran, 2019; Chen et al., 2022; Adamopoulos et al., 2022; Chen, Restuccia, and Santaeulàlia-Llopis, 2023; Nigmatulina, 2023).

For this exercise, I use data from the Statistik Industri, an annual census of all formal manufacturing firms in Indonesia with more than 20 employees. I follow Peters (2020), who focuses on firms that enter after 1990 and is interested in heterogeneous markups — a particular form of misallocation — to motivate a model of firm dynamics and market power.[28]

This leads to an unbalanced panel of about $N = 17,000$ firms, which also comprise the population of interest, for the period 1991–2000. Motivated by the literature

[28]Many firm surveys in developing countries have such a size-based/formal employer cutoff. This qualifies the population of interest and complicates measuring the extensive margin. Peters (2020) argues that a new firm in the census is also an entrant to the relevant product markets to the extent those are the ones formal firms compete in; see Section 3.1 in the paper for additional discussion. The data and replication files are available online at https://onlinelibrary.wiley.com/doi/full/10.3982/ECTA15565.

above, I consider the following measurement model for log labor wedges:

$$\log \widetilde{\mathrm{MRPL}}_{it} = \theta_i + \varepsilon_{it}, \tag{19}$$

where $\log \widetilde{\mathrm{MRPL}}_{it}$ is log MRPL demeaned with respect to industry averages and where $\theta_i$ are firm-level wedges.[29] This is a natural formulation in a context where the distortions of interest are persistent market features such as frictions or regulations.

I then use this framework to compute popular misallocation statistics. First, I explore the relationship between labor distortions and firm size (labor force) in line with similar exercises in the literature (Gorodnichenko et al., 2021; Yeh et al., 2022). The finite-population estimands $\beta_n$ are then least-squares coefficients from the projection of $\theta_i$ on firm size bins $Z_i$; I group firms into ventiles according to their position in the size distribution at entry.

Second, I calculate measures of allocative efficiency, or the aggregate TFP loss associated to the extent of misallocation. An often-used formula that has a closed form expression under normality (Hsieh and Klenow, 2009; Gorodnichenko et al., 2021) is

$$d \log \mathrm{TFP} = -\left(\frac{\alpha(1-\alpha)}{2} + \frac{(1-\alpha)^2\sigma}{2}\right)\mathrm{Var}_n\left(\theta_i\right), \tag{20}$$

where $1-\alpha$ is the labor share and $\sigma$ is the elasticity of substitution. (I follow Hsieh and Klenow (2009) and set $\sigma = 3$ and $\alpha = 0.33$.) The finite-population estimand here is $\beta_n = \mathrm{Var}_n\left(\theta_i\right)$, the dispersion of labor wedges across firms in the economy.

As discussed in Remark 8, our baseline setup does not allow for such objects without further assumptions: while conceptually the problem is identical, the class of estimands considered rules out nonlinear transformations of the latent

[29] In particular, under Hsieh and Klenow (2009) marginal revenue products can be measured up to scale via average revenue products, which are directly available in most datasets. As usual, labor wedges are here identified up to a normalization with respect to other firm-level frictions. Here labor is measured via the wage bill instead of the number of employees and log MRPL is demeaned with respect to narrowly-defined industry indicators and time dummies, following Peters (2020). Finally, note that model (19) is a representative specification, but more general formulations are possible along the lines of Section 3.

attributes. In Appendix B.2, I extend the framework to cover $\text{Var}_n(\theta_i)$ in the context of this application. The corresponding Finite Population Corrections rely on the same notion of weak dependence across measurements as in Assumption 1 and do not require additional measurements. We do need to limit the higher order dependence of measurement errors on latent attributes. This is not surprising: similar assumptions are needed in any deconvolution-like exercise when interest is in nonlinear features or higher-order moments, see Arellano and Bonhomme (2012) for further discussion.

**Results.** Figure 4 show the estimated relationship between labor wedges and firm size at entry, together with finite-population confidence intervals under the benchmark of conditionally uncorrelated measurements.[30] I also report the confidence intervals that would obtain if we were to treat the population of young formal Indonesian firms as a negligible fraction of some hypothetical superpopulation.

Overall, the results suggest a positive relationship between labor-related distortions and firm size, which might be indicative of size-dependent regulations that tend to distort the optimal allocation of labor (Guner et al., 2008). Estimation uncertainty is however not negligible: the relationship is quite noisy overall, and not statistically significant up to the 35% percentile. Ignoring uncertainty altogether would seem to suggest a stronger positive relationship; treating the population as a small sample from an infinite superpopulation would rule out much of a relationship in the bottom half of the distribution. Instead, finite-population inference correctly identifies the nature of estimation uncertainty in this context — the measurement problem in model (19).

Consider now $\beta_n = \text{Var}_n(\theta_i)$ and let $Y_{it} = \log \widetilde{\text{MRPL}}_{it}$ and $\bar{Y} = N^{-1} \sum_{i=1}^{n} R_i \bar{Y}_i$. In empirical work, an often-reported object is the variance of the estimated firm fixed

[30]This exercise also illustrates the applicability of our methods to unbalanced panels, with data missing at random. In particular, note that Assumption 1 applies unit-by-unit, and that the finite-population adjustments only appear in the unit-level weighted contributions to the variance in equation (13). As such, a simple modification to our framework allowing for unit-specific selection matrices $S_{i,(m)}$ according to the number of available measurements (and similarly for projection matrices) would do.

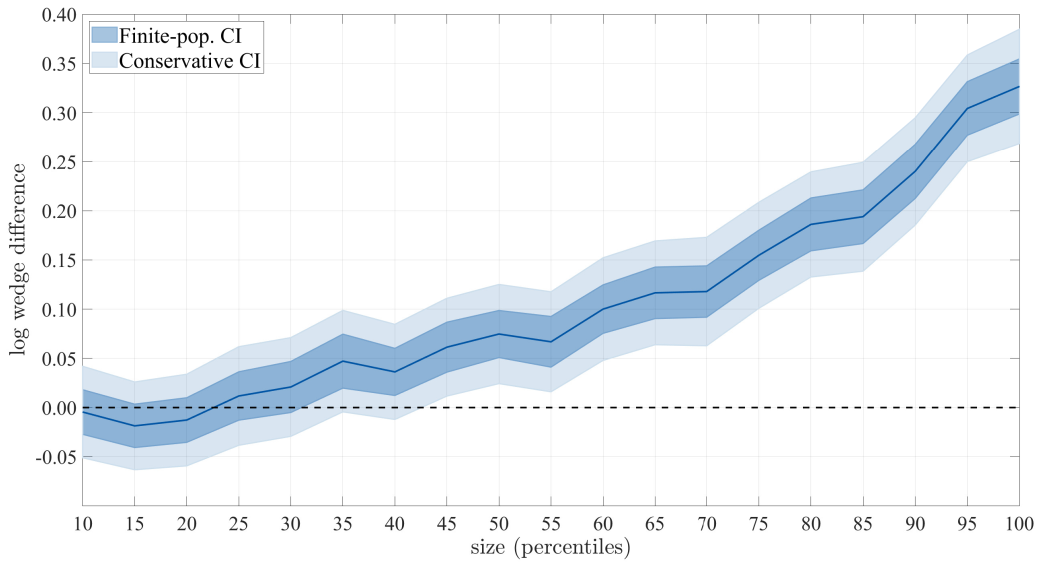



FIGURE 4. Labor wedges across the distribution of firm size at entry (relative to 5th percentile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates.

effects, which ignores the measurement problem:

$$\tilde{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i \left( T^{-1} \sum_{t=1}^{T} Y_{it} - \bar{Y} \right)^2 .$$

Consider the following alternative. Let $Q_i^* = I_{T^2} - T^{-2} 1_{T^2 \times T^2}$ and define $S_{(T)}$ as the selection matrix that has zeros everywhere but at positions $(1,1), (T+2,2), \ldots, (T^2, T)$. Letting $\hat{Y}_i^* = (Y_i - 1_T \bar{Y}) \otimes (Y_i - 1_T \bar{Y})$, a consistent estimator of the population-wide dispersion in $\theta_i$ is

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i T^{-2} 1'_{T^2} \left[ I_{T^2} - S_{(T)} \left( Q_i^* S_{(T)} \right)^{\dagger} Q_i^* \right] \hat{Y}_i^* . \tag{21}$$

Note that we are now imposing independence over measurements at the estimation step — a form of Assumption 1 (Arellano and Bonhomme, 2012). See again Appendix B.2 for additional details. Given this, an estimate of $d \log \text{TFP}$ in equation (20) is readily available. I explore the evolution of this measure of allocative efficiency at entry over 1991–1999, analogous to similar exercises in empirical work (García-Santana, Moral-Benito, Pijoan-Mas, and Ramos, 2020; Bils, Klenow, and Ruane, 2021).[31] Through the lens of this framework, I characterize the second moments of a sequence of evolving finite populations.

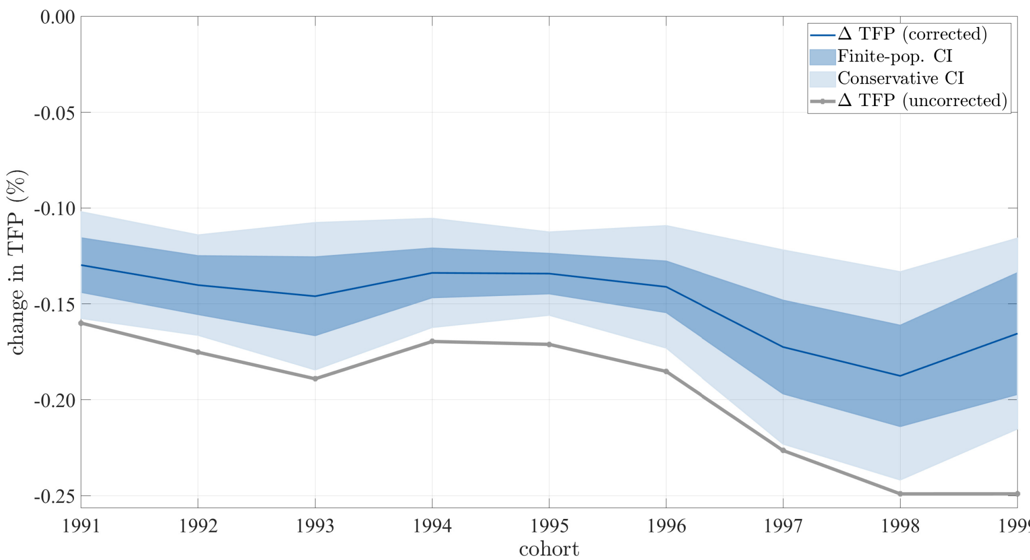



FIGURE 5. Evolution of allocative efficiency as in equation (20) for each cohort (within firms in the bottom size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates.

Figure 5 shows the results for firms in the bottom quartile of the size distribution, a group for which labor-related distortions do not seem to differ systematically based on the number of employees.

The figure shows that the aggregate productivity losses from misallocation (if the economy-wide distortions were like those of entering firms) are of the order of 12–15%, with a slight upward trend over time. Importantly, this is a revised down estimate of around five percentage points in every cohort relative to the standard

[31] Specifically, I calculate (21) for each entry cohort over this period. Note that we still use all measurements for each firm in estimation — this is what allows us to separate the persistent component from measurement errors. Finally, note that in order to calculate (21) at least two (independent) measurements are needed, which means that we cannot report results for firms in the 2000 cohort.

calculation that does not take the measurement problem into account (reported in gray in the figure). This is also lower than the observed dispersion in MRPL in the data (around 40% in terms of equation (20)). These magnitudes are broadly consistent with conventional wisdom that a large part of observed dispersion in total factor revenue productivity is in the firm fixed effect, but emphasize the role of measurement error and transitory shocks. Furthermore, these differences are statistically significant, but estimation uncertainty is here far from negligible: as an example, the change in TFP is within a confidence band of 14–20% for the 1999 cohort. In fact, the difference relative to the uncorrected estimates is only marginally significant if one were to add sampling uncertainty on top of the finite-population confidence bands.

In Appendix C.2, I report additional results allowing for dependence over measurements, with similar implications. If anything, finite-population confidence intervals tend to be wider. This stresses the importance of finite-population inference in this context — where sampling uncertainty is indeed small (even if one treats the population as a negligible fraction of some hypothetical superpopulation) while measurement uncertainty remains sizeable.

All in all, these results illustrate that the methods presented in this paper provide guidance on the relevant sources of estimation uncertainty yet again — this time in a context where the contrast between sampling and measurement is particularly salient and where the conventional approach to inference has been to understate rather than exaggerate estimation uncertainty.

# 6 Conclusion

Finite population problems, where the sample at hand is a relevant fraction of the population of interest, are ubiquitous in empirical work. Despite its salience, the standard treatments of inference in finite populations assume that the features of interest are observable upon sampling, which limits their adoption in applications.

In this paper, I propose new methods to assess estimation uncertainty in problems where a finite population coexists with a measurement problem. I propose finite-population variance estimators that guarantee non-conservative inference and apply these methods to two different empirical applications: on predicting police violence and on studying firm misallocation with census data. Finite-population inference allows for a systematic approach to uncertainty quantification in setups where uncertainty has been previously understood in different ways and leads to large gains in precision in setups where routine practice has been to report standard errors as if the sample was negligible relative to the population.

I also leave some interesting dimensions for future work. Extending the finite-populations framework to a "many measurements" context presents no conceptual difficulty. If anything, some tasks are simplified: whereas weak dependence continues to be a key assumption, more agnostic approaches to dependence are possible, as those in the time series tradition. Having access to many measurements also allows extensions to more general nonlinear models and facilitates constructing Finite Population Corrections. Further formalizing these ideas also seems a promising direction for future work. Similarly, the framework in this paper can be extended to more general persistent–transitory measurement models and dynamic panel data problems in short panels.

# A Proofs

The derivations here are a proof sketch of Propositions 1 and 2; the former is here integrated in the proof of the latter.

Let $\lambda$ denote a fixed column vector $\lambda \neq 0_{(k+p)\times 1}$ and let $\hat{f} = N/n$. The finite-population variance is $\hat{\mathrm{V}}(\hat{f})$ in (14). Throughout, we condition on $\hat{\mathrm{V}}(\hat{f}) \geq 0$ (in the matrix sense), which is a measure-one event in the limit. At all times, we maintain Assumption 2 and the regularity conditions in Assumption 3. For (B) below we also invoke Assumption 1 and assume rank $Q_i^*(\delta)\, S_{(m)} = m$. Then, as $n \to \infty$:

(A) $(\lambda'\mathrm{V}(f)\lambda)^{-1/2}\, \sqrt{N}\lambda'(\hat{\gamma} - \gamma_n) \xrightarrow{\ \mathrm{d}\ } \mathrm{N}(0,1)$,

(B) $\left(\lambda'\hat{\mathrm{V}}(\hat{f})\lambda\right) / (\lambda'\mathrm{V}(f)\lambda) \xrightarrow{\ \mathrm{p}\ } 1$.

where $\mathrm{V}(f)$ is defined in (12). (A) and (B) are established in Lemmas 1 and 2, respectively. Since $\lambda$ can be chosen arbitrarily, (A) and the Cramér-Wold device imply Proposition 1. (A) and (B) imply that

$$\left(\lambda'\hat{\mathrm{V}}(\hat{f})\lambda/N\right)^{-1/2} \lambda'(\hat{\gamma} - \gamma_n) \xrightarrow{\ \mathrm{d}\ } \mathrm{N}(0,1),$$

and thus Proposition 2 follows.

**Assumption 3 (Regularity conditions for limit theorems).** *[To be completed.] These include the existence of limits in the sequence of finite populations, bounds on population attributes and on moments of measurement errors, regularity conditions for M-estimators as in Newey and McFadden (1994) and regularity conditions for limit theorems of i.n.i.d. random elements.*

**Lemma 1 (Asymptotic normality of the rescaled estimation error).** *For an arbitrary column vector $\lambda \neq 0_{(k+p)\times 1}$,*

$$(\lambda' V(f)\lambda)^{-1/2}\, \sqrt{N}\lambda'(\hat{\gamma} - \gamma_n) \xrightarrow{\ \mathrm{d}\ } \mathrm{N}(0,1).$$

*Proof.* Under regularity conditions in Assumption 3, the sample moment condition (9) admits an expansion

$$\sqrt{N}(\hat{\gamma} - \gamma_n) = H_n^{-1} n^{-1/2} \sum_{i=1}^{n} \frac{R_i}{\sqrt{f_n}} \psi(Y_i, W_i, \gamma_n) + o_p(1),$$

where we have also used that $\hat{f}/f_n \xrightarrow{\text{p}} 1$ by Assumption (2). Now, note that from independent random sampling and repeatedly using $E[R_i] = f_n$,

$$\text{Var}\left(\frac{R_i}{\sqrt{f_n}}\psi(Y_i, W_i, \gamma_n)\right) = E\left[\psi(Y_i, W_i, \gamma_n)\psi(Y_i, W_i, \gamma_n)'\right] - f_n E\left[\psi(Y_i, W_i, \gamma_n)\right] E\left[\psi(Y_i, W_i, \gamma_n)\right]', \tag{22}$$

and that averaging over the population yields $\mathrm{V}_{\psi,n}(f_n)$ in (11). For an arbitrary vector $\lambda \neq 0_{(k+p)\times 1}$, $\{\lambda'\psi(Y_i, W_i, \gamma_n)\}_n$ is a row-wise independent triangular array and note that $a'E_n\left[\psi(Y_i, W_i, \gamma_n)\right] = 0$ from (10). Asymptotic normality of these averages follows by a Lyapunov-type condition; here we invoke Lemma A.1 in Abadie et al. (2020). Letting $\mathrm{V}_\psi(f) = \lim_{n\to\infty} \mathrm{V}_{\psi,n}(f_n)$, the asymptotic variance is given by $\lambda'\mathrm{V}_\psi(f)\lambda$, and the result follows via the the Cramér-Wold device. □

**Lemma 2 (Consistency of the finite-population standard error).** *For an arbitrary column vector $\lambda \neq 0_{(k+p)\times 1}$,*

$$\frac{\lambda'\hat{V}(\hat{f})\lambda}{\lambda'V(f)\lambda} \xrightarrow{\text{p}} 1.$$

*Proof.* I focus on $\hat{\mathrm{V}}_\psi(\hat{f})$; the regularity conditions in Assumption 3 immediately imply convergence of $\hat{H}$ to its limits. We first characterize $\mathrm{V}_{\psi,n}(f_n)$. It is immediate from the expressions in (10), $E\left[\varepsilon_i\varepsilon_i'\right] = \Omega_i$ and the result in (22) that

$$\mathrm{V}_{\psi,n}(f_n) = \begin{pmatrix} E_n\left[\Psi_{\delta\delta,i}\right] & E_n\left[\Psi_{\delta\beta,i}\right] \\ E_n\left[\Psi'_{\delta\beta,i}\right] & (1-f_n)E_n\left[\Psi_{\beta\beta,i}\right] + f_n E_n\left[\tilde{\Psi}_{\beta\beta,i}\right] \end{pmatrix}.$$

where $\psi_{\beta,i} = h_1(W_i, \beta_n)\left(\theta_i - h_0\left(W_i; \beta_n\right)\right)$ and

$$\begin{aligned} \Psi_{\delta\delta,i} &= \mathrm{A}(W_i, \delta)Q_i(\delta)\Omega_i Q_i(\delta)\mathrm{A}(W_i, \delta)', \\ \Psi_{\delta\beta,i} &= \mathrm{A}(W_i, \delta)Q_i(\delta)\Omega_i\left(g_1\left(X_i; \delta\right)^\dagger h_1(W_i, \beta_n)\right)', \\ \Psi_{\beta\beta,i} &= \psi_{\beta,i}\psi'_{\beta,i} + \underbrace{h_1(W_i, \beta_n)g_1\left(X_i; \delta\right)^\dagger \Omega_i\left(g_1\left(X_i; \delta\right)^\dagger h_1(W_i, \beta_n)\right)'}_{\equiv\tilde{\Psi}_{\beta\beta,i}}. \end{aligned}$$

Next, define

$$\Lambda_i\left(f_n\right) = \mathrm{vec}^{-1}\left[\left(1-f_n\right)I_{T^2} + f_n S_{(m)}\left(Q_i^*\left(\delta\right)S_{(m)}\right)^\dagger Q_i^*\left(\delta\right)\right]\left(u\left(Y_i, W_i, \gamma_n\right)\otimes u\left(Y_i, W_i, \gamma_n\right)\right),$$

and note that

$$u(Y_i, W_i, \gamma_n) = g_1(X_i; \delta)(\theta_i - h_0(W_i; \beta_n)) + \varepsilon_i.$$

Furthermore, let $g_1^*(X_i; \delta) = g_1(X_i; \delta) \otimes g_1(X_i; \delta)$ and note that using Assumption 1,

$$E[u(Y_i, W_i, \gamma_n) \otimes u(Y_i, W_i, \gamma_n)] = g_1^*(X_i; \delta)(\theta_i - h_0(W_i; \beta_n))^2 + S_{(m)}\omega_i$$

and $Q_i^*(\delta) E[u(Y_i, W_i, \gamma_n) \otimes u(Y_i, W_i, \gamma_n)] = Q_i^*(\delta) S_{(m)}\omega_i$. Further, using that $Q_i^*(\delta) S_{(m)}$ has column rank, it follows that

$$E[\Lambda_i(f_n)] = (1 - f_n)\left[g_1(X_i; \delta)(\theta_i - h_0(W_i; \beta_n))^2 g_1(X_i; \delta)' + \Omega_i\right] + f_n\Omega_i.$$

The above shows unbiasedness of $\Lambda_i(f_n)$ precisely for the term in the (finite-population) score. We can then bound the variance of this term . One can proceed similarly for the residual term (using $\hat{\gamma}$ instead of $\gamma_n$). Note that $Q_i(\delta)g_1(X_i; \delta) = 0_T$ by construction, which shows why the finite-population variance is independent of $f_n$ for common parameters despite the way it is contructed.

□

# B Additional derivations

## B.1 Section 2: conservativeness of the cluster-robust variance

Here we show that $E\left[\hat{\mathrm{V}}^{\text{cluster}}\right] = \mathrm{V}(0)$, where

$$\hat{\mathrm{V}}^{\text{cluster}} = \frac{1}{N(N-1)} \sum_{i=1}^{n} R_i \left(\bar{Y}_i - \hat{\beta}\right)^2.$$

In order to see this, it is helpful to rewrite the expression as

$$\hat{\mathrm{V}}^{\text{cluster}} = \frac{1}{N^2} \sum_{i=1}^{n} R_i \bar{Y}_i^2 - \frac{1}{N^2(N-1)} \sum_{i=1}^{n} \sum_{j \neq i} R_i R_j \bar{Y}_i \bar{Y}_j.$$

Taking expectations, we have

$$E\left[\hat{\mathrm{V}}^{\text{cluster}}\right] = \frac{E_n\left[E\left[\bar{Y}_i^2\right]\right]}{N} - \frac{\frac{1}{n-1} E_n\left[\sum_{j \neq i} E\left[\bar{Y}_i \bar{Y}_j\right]\right]}{N}$$

$$= \frac{E_n\left[\theta_i^2\right] + \sigma^2/T}{N} - \frac{\frac{1}{n-1}E_n\left[\sum_{j\neq i}\theta_i\theta_j\right]}{N}$$

$$= \frac{\frac{n}{n-1}E_n\left[(\theta_i - \beta_n)^2\right]}{N} + \frac{\sigma^2/T}{N} = \mathrm{V}(0),$$

where we have used Assumption S1 in the first line[32] and Assumption S2 in the second one.

## B.2 Finite-population inference for variances

Here I extend the results in Section 3 to cover $\beta_n = \mathrm{Var}_n(\theta_i) = (n-1)^{-1}\sum_{i=1}^n (\theta_i - \bar{\theta}_n)^2$ (where $\bar{\theta}_n$ is the average $\theta_i$ in the population) in the context of the misallocation empirical application in Section 5.2 and the measurement model in equation (19).

In particular, consider the variance estimator in equation (21), which we can extend to allow for weak dependence as in Assumption 1:

$$(23)\qquad \hat{\beta} = \frac{1}{N}\sum_{i=1}^n R_i T^{-2} 1'_{T^2}\left[I_{T^2} - S_{(m)}\left(Q_i^* S_{(m)}\right)^{\dagger} Q_i^*\right]\hat{Y}_i^* = \frac{1}{N}\sum_{i=1}^n R_i T^{-2} 1'_{T^2}\tilde{Y}_i^*,$$

with an obvious definition of $\tilde{Y}_i^*$. The motivation for this estimator can be traced back to equation (15) in Remark 8, which recasts the measurement system for $\theta_i$ as a measurement system for $\theta_i^2$. A valid, conservative (large-sample) variance estimator for $\hat{\beta}$ is given by

$$\hat{\mathrm{V}}(0) = \frac{1}{N}\sum_{i=1}^n R_i\left(T^{-2}1'_{T^2}\tilde{Y}_i^* - \hat{\beta}\right)^2.$$

Again through the lens of equation (15), it can be seen that the corresponding FPC is given by

$$\mathrm{FPC} = \lim_{n\to\infty} E_n\left[(\theta_i - \bar{\theta}_n)^4\right] - \beta_n^2.$$

We can then leverage Remark 6, which shows that we can construct finite-population variance estimators if we have access to a conservative estimator that

[32]In particular, note that $E[R_i] = N/n$ and $E\left[R_iR_j\right] = N(N-1)/n(n-1)$ for $j \neq i$ for simple random sampling without replacement.

is consistent for V(0) and a valid estimator of the FPC.A candidate for the latter follows by noting that the FPC is also equal to (the limit of) $\kappa_{4n} + 2\beta_n^2$, where

$$\kappa_{4n}(\theta_i) = E_n\left[(\theta_i - \bar{\theta}_n)^4\right] - 3E_n\left[(\theta_i - \bar{\theta}_n)^2\right]^2.$$

Arellano and Bonhomme (2012, Appendix A) propose estimators of fourth-order cumulants. This approach allows us to obtain valid finite-population estimators under the same notion of dependence over measurements used in estimation, but we do need to restrict the higher-order dependence between measurement errors and unobserved attributes; statistical independence would be a sufficient condition.[33]

# C Empirics: additional results

## C.1 Additional results from Section 5.1

Tables 3, 4 and 5 report finite-population confidence intervals for all results reported in the main empirical section in Montiel-Olea et al. (2021) (section 5.2 of the paper). The entries correspond to the ten largest police departments by population served. Table 3 considers counterfactuals based on both observed and unobserved determinants, Table 4 considers only counterfactual unobserved determinants and Table 5 only observed ones.

[33]One way to operationalize independence is to define a probability distribution over $\theta_i$ and measurement errors in the limit over sequences of growing finite populations and then impose these restrictions. Weaker conditions are possible imposing restrictions on limits of certain sums over the finite population. Similarly, independence can be relaxed to zero cross-cumulants up to fourth order.

TABLE 3. Counterfactual homicides for 2013-2018: observed and unobserved determinants

(a) Conventional inference

| | Phoenix | Las Vegas | Dallas | San Antonio | Los Angeles | Houston | San Diego | Chicago | Philadelphia | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [52,53] | [33,47] | [31,45] | [34,40] | [31,34] | [31,32] | [28,32] | [21,30] | [5,9] |
| Las Vegas | [93,93] | **51** | [33,46] | [31,44] | [34,40] | [31,33] | [30,32] | [28,32] | [21,30] | [5,9] |
| Dallas | [68,97] | [38,54] | **33** | [32,33] | [27,37] | [24,33] | [23,33] | [22,30] | [21,23] | [5,7] |
| San Antonio | [76,109] | [42,61] | [37,39] | **35** | [30,41] | [26,37] | [25,37] | [24,34] | [23,25] | [6,8] |
| Los Angeles | [269,314] | [150,176] | [105,141] | [99,135] | **113** | [95,106] | [89,106] | [86,98] | [66,91] | [17,27] |
| Houston | [145,156] | [81,87] | [54,74] | [50,71] | [56,62] | **51** | [48,53] | [46,50] | [34,48] | [9,14] |
| San Diego | [79,83] | [44,46] | [28,40] | [27,39] | [29,35] | [26,29] | **26** | [23,28] | [18,26] | [4,8] |
| Chicago | [189,216] | [105,121] | [72,100] | [68,96] | [75,85] | [67,73] | [62,74] | **63** | [46,64] | [12,19] |
| Philadelphia | [89,130] | [50,73] | [44,47] | [41,45] | [36,50] | [31,44] | [30,44] | [29,40] | **28** | [7,9] |
| New York | [568,1013] | [317,566] | [270,371] | [259,348] | [237,375] | [201,342] | [189,342] | [184,310] | [175,236] | **55** |
| Totals | [1689,2279] | [942,1273] | [753,882] | [717,836] | [699,848] | [596,769] | [562,776] | [545,700] | [481,567] | [125,166] |

(b) Finite-population inference

| | Phoenix | Las Vegas | Dallas | San Antonio | Los Angeles | Houston | San Diego | Chicago | Philadelphia | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [52,53] | [35,44] | [34,42] | [35,39] | [32,33] | [31,32] | [28,32] | [23,28] | [6,8] |
| Las Vegas | [93,93] | **51** | [35,43] | [33,41] | [35,39] | [31,33] | [30,32] | [28,31] | [23,28] | [6,8] |
| Dallas | [73,90] | [41,50] | **33** | [32,33] | [28,36] | [25,31] | [24,30] | [23,29] | [21,23] | [5,7] |
| San Antonio | [81,101] | [45,56] | [37,39] | **35** | [32,40] | [28,35] | [27,34] | [25,33] | [23,25] | [6,7] |
| Los Angeles | [274,307] | [153,172] | [107,137] | [102,130] | **113** | [96,104] | [91,104] | [87,97] | [69,88] | [19,24] |
| Houston | [146,155] | [82,86] | [56,71] | [53,67] | [57,62] | **51** | [48,53] | [46,50] | [36,45] | [10,13] |
| San Diego | [79,83] | [44,46] | [30,38] | [29,36] | [30,34] | [27,29] | **26** | [24,28] | [19,24] | [5,7] |
| Chicago | [190,214] | [106,120] | [74,96] | [71,92] | [75,84] | [67,72] | [62,73] | **63** | [49,61] | [13,18] |
| Philadelphia | [96,120] | [54,67] | [44,47] | [41,45] | [38,48] | [33,42] | [32,41] | [30,38] | **28** | [7,9] |
| New York | [643,870] | [359,486] | [281,355] | [269,333] | [261,332] | [225,298] | [214,292] | [204,275] | [180,228] | **55** |
| Totals | [1791,2094] | [1000,1170] | [751,882] | [716,833] | [723,809] | [626,717] | [597,704] | [565,667] | [481,564] | [134,155] |

*Note*: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}^*_t(i, j, z_j)$ in equation (18), which replace characteristics of agency $i$ in the rows with that of agency $j$ in the columns; see the text in Section 5.1 for additional details. In this case, we replace both observed and unobserved determinants of police use of deadly force.

TABLE 4. Counterfactual homicides for 2013-2018: unobserved determinants

(a) Conventional inference

| | Phoenix | Las Vegas | Dallas | Philadelphia | San Diego | Chicago | Los Angeles | Houston | San Antonio | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [64,70] | [52,74] | [46,78] | [51,66] | [48,64] | [44,54] | [44,53] | [40,57] | [13,25] |
| Las Vegas | [70,77] | **51** | [40,58] | [36,60] | [40,51] | [38,49] | [34,43] | [34,41] | [30,45] | [10,20] |
| Dallas | [43,61] | [31,44] | **33** | [28,37] | [26,38] | [25,36] | [23,31] | [22,31] | [24,28] | [8,12] |
| Philadelphia | [35,60] | [25,42] | [26,35] | **28** | [22,35] | [23,32] | [19,30] | [19,29] | [19,29] | [7,11] |
| San Diego | [39,50] | [28,35] | [24,36] | [22,35] | **26** | [23,29] | [21,26] | [21,25] | [18,28] | [6,11] |
| Chicago | [93,125] | [67,87] | [60,86] | [57,81] | [60,74] | **63** | [50,64] | [53,59] | [44,68] | [16,28] |
| Los Angeles | [198,244] | [139,175] | [125,170] | [111,172] | [121,150] | [115,146] | **113** | [103,125] | [95,130] | [34,54] |
| Houston | [92,112] | [66,79] | [57,79] | [53,78] | [56,68] | [57,63] | [47,58] | **51** | [42,62] | [15,26] |
| San Antonio | [60,85] | [42,62] | [43,52] | [36,56] | [35,54] | [34,53] | [32,43] | [30,45] | **35** | [11,17] |
| New York | [207,402] | [147,284] | [152,234] | [148,222] | [133,233] | [130,224] | [117,189] | [111,197] | [116,178] | **55** |
| Totals | [952,1279] | [677,909] | [656,802] | [591,809] | [595,762] | [577,731] | [531,614] | [510,630] | [489,630] | [177,257] |

(b) Finite-population inference

| | Phoenix | Las Vegas | Dallas | Philadelphia | San Diego | Chicago | Los Angeles | Houston | San Antonio | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [65,69] | [55,70] | [51,70] | [54,61] | [51,61] | [46,51] | [46,51] | [43,53] | [16,21] |
| Las Vegas | [71,75] | **51** | [43,54] | [40,53] | [43,47] | [40,47] | [35,41] | [36,39] | [34,41] | [13,16] |
| Dallas | [45,58] | [33,41] | **33** | [30,35] | [28,35] | [27,34] | [23,31] | [24,30] | [25,27] | [9,11] |
| Philadelphia | [39,54] | [28,38] | [28,33] | **28** | [24,32] | [24,31] | [20,29] | [21,27] | [20,26] | [8,10] |
| San Diego | [41,47] | [30,33] | [26,33] | [24,32] | **26** | [24,28] | [21,25] | [22,24] | [20,25] | [8,10] |
| Chicago | [99,118] | [71,83] | [63,82] | [60,77] | [62,71] | **63** | [51,62] | [53,58] | [48,63] | [19,24] |
| Los Angeles | [208,232] | [146,168] | [126,168] | [116,165] | [124,146] | [119,142] | **113** | [106,121] | [98,125] | [39,49] |
| Houston | [96,107] | [69,76] | [60,75] | [56,72] | [59,65] | [58,62] | [49,56] | **51** | [46,58] | [18,22] |
| San Antonio | [64,78] | [46,56] | [45,50] | [39,51] | [38,49] | [36,48] | [33,42] | [32,41] | **35** | [12,16] |
| New York | [248,326] | [177,232] | [169,212] | [159,204] | [152,199] | [147,190] | [131,165] | [130,164] | [129,162] | **55** |
| Totals | [1026,1165] | [732,828] | [662,793] | [611,777] | [626,717] | [600,691] | [535,603] | [536,585] | [508,602] | [198,232] |

*Note*: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, j, z_i)$ in equation (18), which replace characteristics of agency $i$ in the rows with that of agency $j$ in the columns; see the text in Section 5.1 for additional details. In this case, we replace only unobserved determinants of police use of deadly force.

TABLE 5. Counterfactual homicides for 2013-2018: observed determinants

(a) Conventional inference

| | Phoenix | San Antonio | Las Vegas | Los Angeles | Houston | Dallas | San Diego | Chicago | Philadelphia | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [69,79] | [70,78] | [65,77] | [57,70] | [53,67] | [45,59] | [43,59] | [31,52] | [27,46] |
| San Antonio | [43,49] | **35** | [33,39] | [34,35] | [29,33] | [27,32] | [23,27] | [22,27] | [16,24] | [14,21] |
| Las Vegas | [63,69] | [48,56] | **51** | [46,55] | [40,49] | [37,47] | [32,41] | [31,41] | [22,37] | [19,32] |
| Los Angeles | [139,164] | [116,121] | [109,130] | **113** | [95,108] | [89,103] | [77,88] | [73,89] | [52,79] | [47,68] |
| Houston | [70,86] | [56,65] | [55,67] | [55,63] | **51** | [48,50] | [38,47] | [39,44] | [28,39] | [24,35] |
| Dallas | [48,61] | [39,46] | [38,47] | [38,44] | [36,37] | **33** | [26,33] | [28,30] | [20,27] | [17,24] |
| San Diego | [43,56] | [36,42] | [35,44] | [35,40] | [30,37] | [28,35] | **26** | [23,30] | [17,26] | [16,22] |
| Chicago | [103,140] | [84,105] | [81,108] | [82,100] | [76,84] | [73,78] | [58,74] | **63** | [45,58] | [39,52] |
| Philadelphia | [52,89] | [43,66] | [41,69] | [42,63] | [39,54] | [37,50] | [30,46] | [32,41] | **28** | [23,29] |
| New York | [114,195] | [96,143] | [91,151] | [94,135] | [84,119] | [80,110] | [69,94] | [69,91] | [56,71] | **55** |
| Totals | [774,996] | [642,738] | [610,773] | [626,699] | [564,611] | [532,572] | [442,516] | [445,493] | [320,435] | [286,378] |

(b) Finite-population inference

| | Phoenix | San Antonio | Las Vegas | Los Angeles | Houston | Dallas | San Diego | Chicago | Philadelphia | New York |
|---|---|---|---|---|---|---|---|---|---|---|
| Phoenix | **93** | [72,76] | [72,76] | [69,74] | [59,68] | [55,64] | [48,55] | [46,56] | [34,47] | [31,41] |
| San Antonio | [44,47] | **35** | [35,37] | [34,35] | [29,32] | [27,31] | [24,26] | [23,27] | [17,22] | [16,19] |
| Las Vegas | [64,68] | [51,54] | **51** | [48,52] | [42,47] | [40,44] | [35,38] | [33,39] | [25,33] | [22,28] |
| Los Angeles | [145,156] | [117,120] | [114,122] | **113** | [97,106] | [91,100] | [80,85] | [75,87] | [57,73] | [52,63] |
| Houston | [72,83] | [58,64] | [58,64] | [56,61] | **51** | [48,49] | [40,44] | [40,43] | [30,36] | [27,32] |
| Dallas | [50,58] | [40,45] | [40,45] | [39,43] | [36,36] | **33** | [28,31] | [28,30] | [21,25] | [19,22] |
| San Diego | [46,52] | [38,40] | [37,41] | [36,39] | [32,35] | [30,33] | **26** | [25,28] | [19,24] | [17,20] |
| Chicago | [108,132] | [87,102] | [86,101] | [84,97] | [78,83] | [73,78] | [61,70] | **63** | [48,54] | [43,48] |
| Philadelphia | [58,79] | [47,61] | [46,61] | [45,58] | [42,50] | [40,47] | [33,42] | [34,39] | **28** | [24,27] |
| New York | [129,169] | [105,130] | [102,130] | [102,123] | [92,107] | [87,100] | [74,88] | [75,84] | [60,66] | **55** |
| Totals | [813,937] | [655,719] | [648,720] | [636,686] | [578,595] | [543,560] | [459,493] | [450,484] | [341,407] | [309,353] |

*Note*: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}^*_t(i,i,z_j)$ in equation (18), which replace characteristics of agency $i$ in the rows with that of agency $j$ in the columns; see the text in Section 5.1 for additional details. In this case, we replace only observed determinants of police use of deadly force.

## C.2 Additional results from Section 5.2

In this section, I report additional results from the application to firm misallocation in Section 5.2.

First, for completeness, Figures 8, 9 and 10 report results for allocative efficiency in the sense of $d \log \text{TFP}$ in equation (20) for the remaining size quartiles. Second, whereas in the text I have focused on a baseline measurement model with uncorrelated measurements (but unrestricted heteroskedasticity), I report here results further allowing for weak dependence.

In particular, given the unbalanced nature of the panel, I allow for richer forms of $m$-dependence as a larger number of measurements becomes available. Let $T_i$ denote the number of such periods for firm $i$. I choose $m = 1$ for $T_i = 3$, $m = 2$ if $T_i \in \{4, 5\}$, $m = 3$ if $T_i \in \{5, 6, 7, 8\}$ and $m = 4$ if $T_i \in \{9, 10\}$. In other words, for firms that enter in 1991 and remain active throughout the panel, we allow for unrestricted dependence in measurement errors over up to four year horizons. It is easy to see that these choices satisfy the order condition in Assumption 1, sometimes with equality. Note that for figures relative to allocative efficiency not only the confidence intervals but also the point estimates might change as a consequence, see equation (23).

Figures 6 and 7 are the counterparts to Figures 4 and 5. In qualitative terms, the results are similar to those reported in the body of the paper, although the finite-population confidence intervals tend to be wider. This reinforces the main message in Section 5.2 emphasizing the need to account for measurement-based uncertainty while leaving a minor role for sampling-based uncertainty — even if the analyst treats the population as a negligible fraction of an infinite superpopulation.

It is also important to stress that sensitivity to the baseline assumption of uncorrelated measurements might suggest either a restrictive notion of weak dependence or misspecification of the underlying measurement system. Regarding the former, the $m$-dependence restrictions above are only marginally rejected when I implement a test of covariance structures along the lines of Remark 4. An alternative is to consider richer measurement equations for $\theta_i$ beyond the benchmark model. This seems particularly promising in this context, where there might be a time-varying

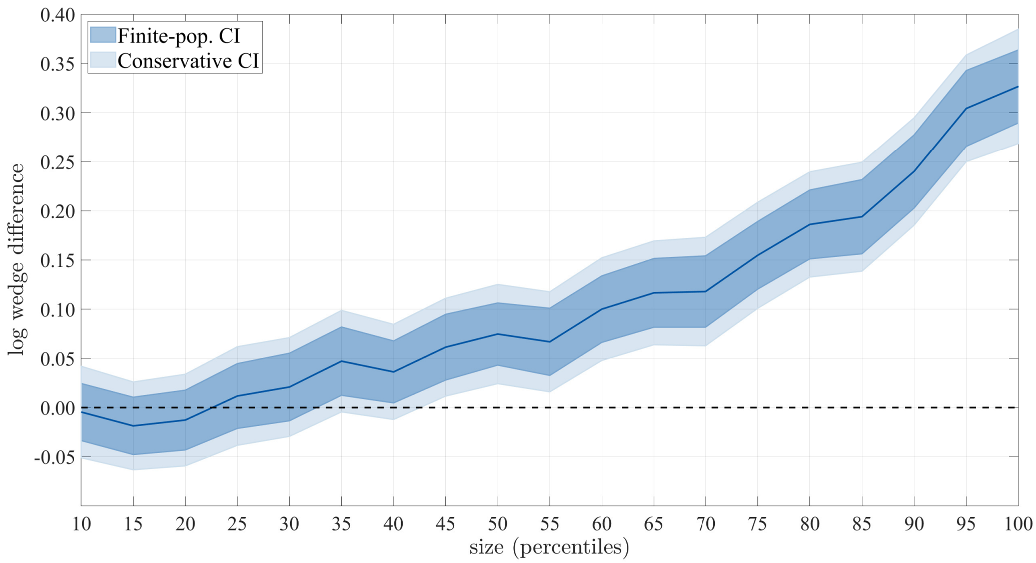



FIGURE 6. Labor wedges across the distribution of firm size at entry (relative to 5th percentile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. Results allowing for $m$-dependence in measurements, see the text for additional details.

systematic component in labor-related distortions or persistent, predictable variation in MRPL beyond what is captured in equation (19). All of these can be framed within the class of models discussed in Section 3.

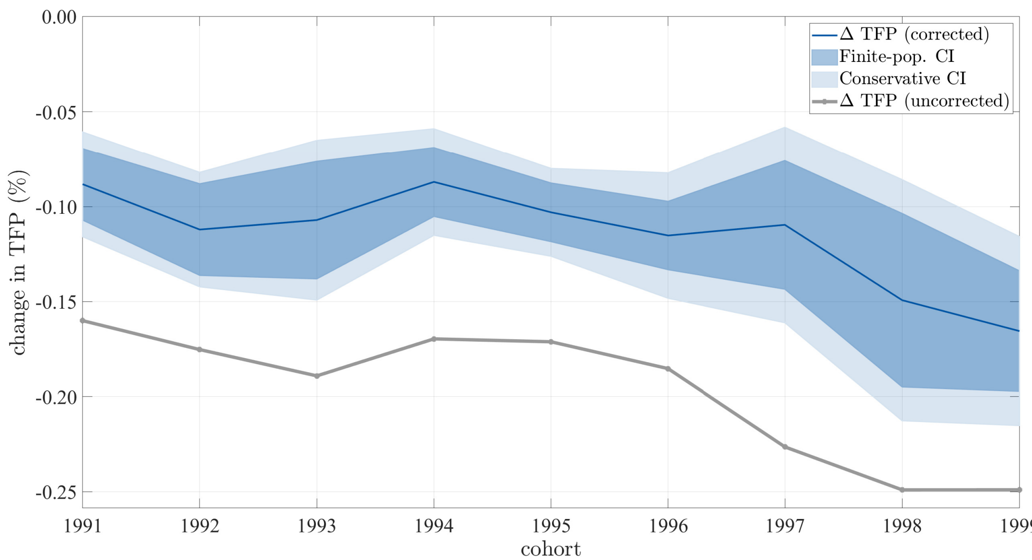
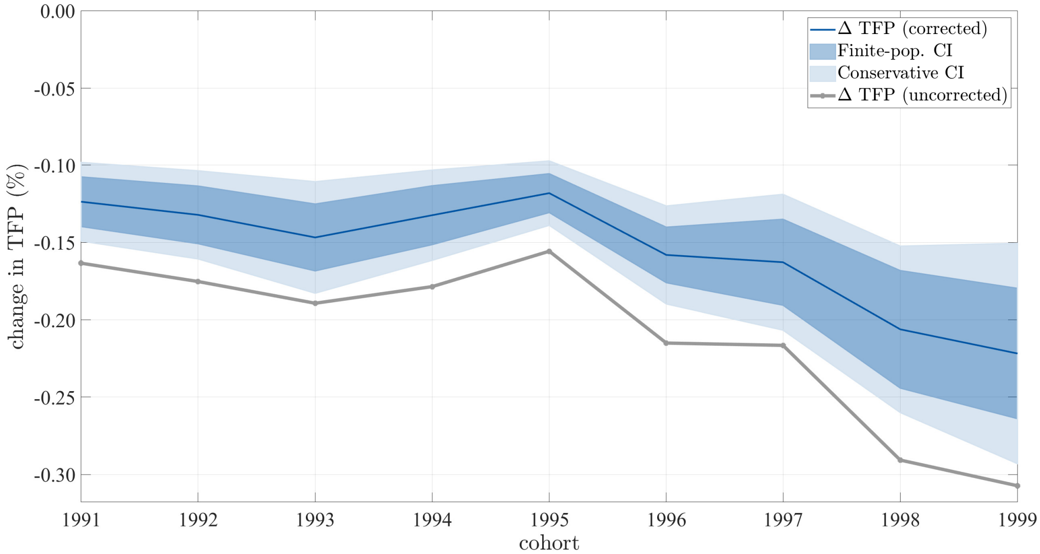
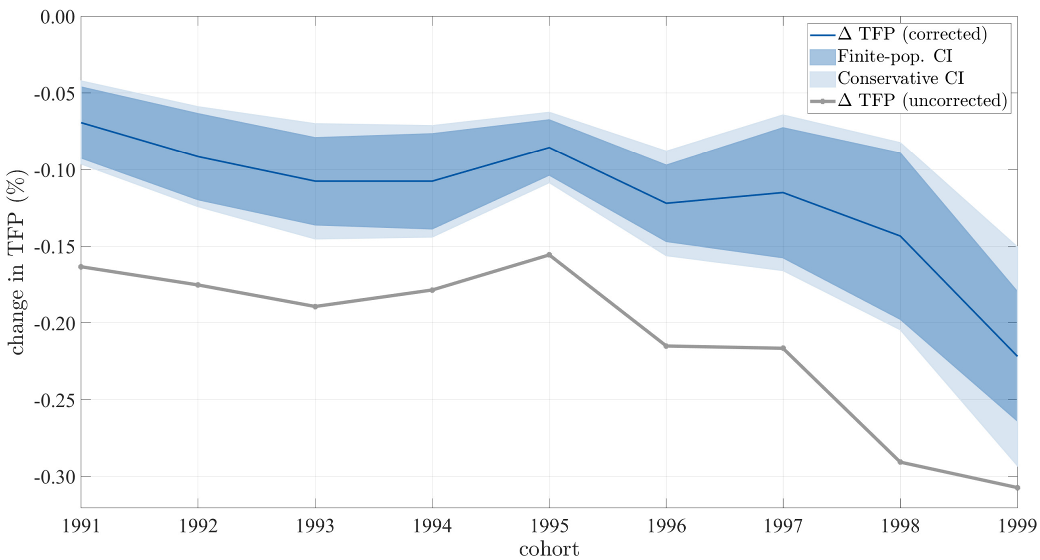



FIGURE 7. Evolution of allocative efficiency as in equation (20) for each cohort (within firms in the bottom size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. Results allowing for $m$-dependence in measurements, see the text for additional details.

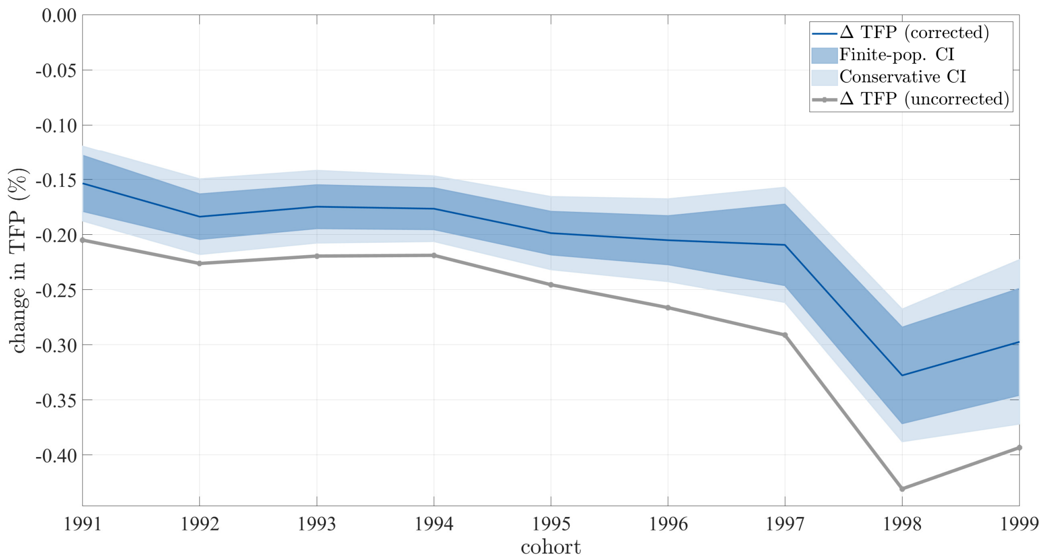



(a) Uncorrelated measurements.

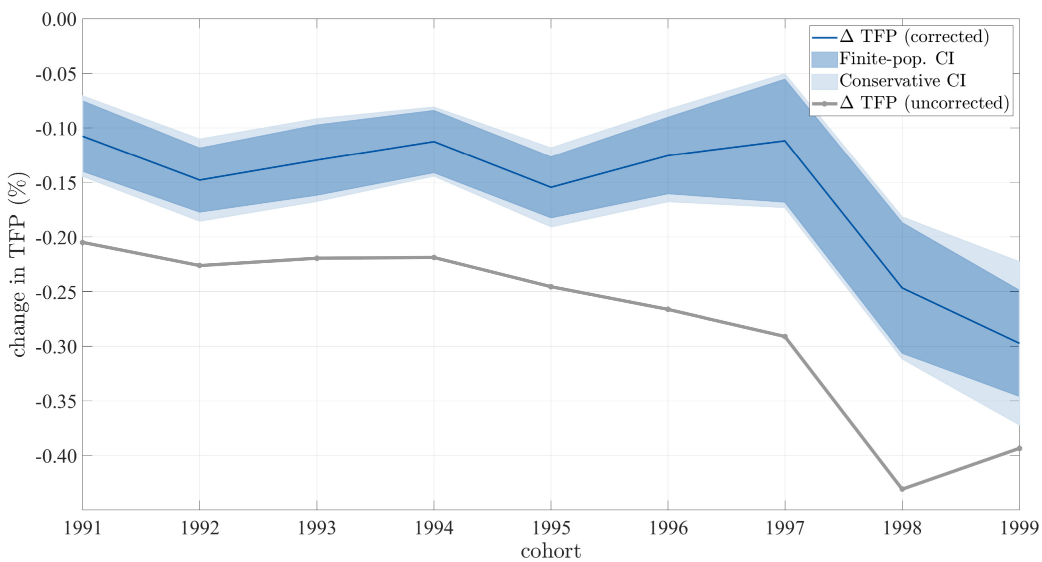



(b) *m*-dependent measurements.

FIGURE 8. Evolution of allocative efficiency as in equation (20) for each cohort (within firms in the second size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.

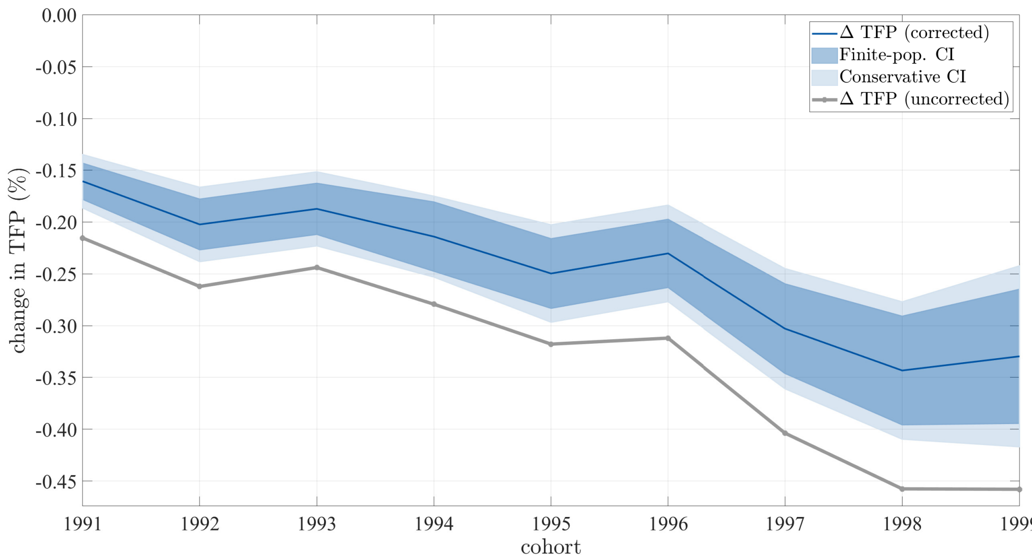



(a) Uncorrelated measurements.

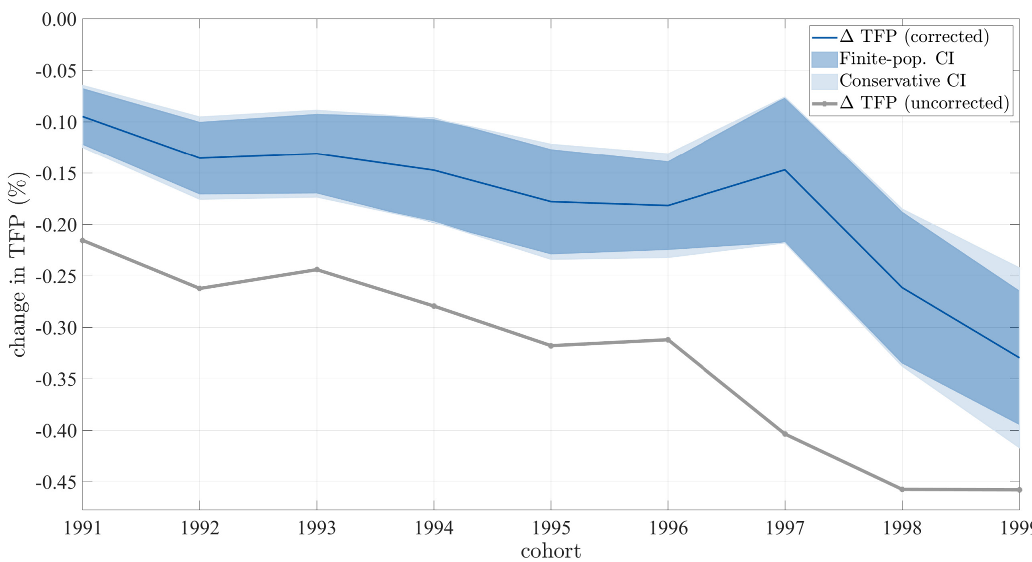



(b) $m$-dependent measurements.

FIGURE 9. Evolution of allocative efficiency as in equation (20) for each cohort (within firms in the third size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.




(a) Uncorrelated measurements.




(b) $m$-dependent measurements.

FIGURE 10. Evolution of allocative efficiency as in equation (20) for each cohort (within firms in the upper size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.

# References


Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based vs design-based uncertainty in regression analysis." *Econometrica*, 88, 265–296.

——— (2023): "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics*, 138, 1–35.

Abadie, A., G. W. Imbens, and F. Zheng (2014): "Inference for misspecified models with fixed regressors." *Journal of the American Statistical Association*, 109, 1601–1614.

Abowd, J. M. and D. Card (1989): "On the covariance structure of earnings and hours changes." *Econometrica*, 57, 411–445.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2022): "Misallocation, Selection, and Productivity: A Quantitative Analysis With Panel Data From China." *Econometrica*, 90, 1261–1282.

Ahn, D., S. Choi, D. Gale, and S. Kariv (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment." *Quantitative Economics*, 5, 195–223.

Andreoni, J. and L. Samuelson (2006): "Building Rational Cooperation." *Journal of Economic Theory*, 127, 117–154.

Arellano, M. (1987): "Computing Robust Standard Errors for Within-Group Estimators." *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

——— (2003): *Panel Data Econometrics*, Oxford University Press.

Arellano, M. and S. Bonhomme (2009): "Identifying distributional characteristics in random coefficients panel data models." CEMFI Working Paper 0904.

——— (2012): "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies*, 79, 987–1020.

Aronow, P. M., D. P. Green, and D. K. K. Lee (2014): "Sharp bounds on the variance in randomized experiments," *The Annals of Statistics*, 42, 850–871.

Barras, L., P. Gagliardini, and O. Scaillet (2022): "Skill, Scale, and Value Creation in the Mutual Fund Industry." *Journal of Finance*, 77, 601–638.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112, 537–579.

Bils, M., P. J. Klenow, and C. Ruane (2021): "Misallocation or Mismeasurement?" *Journal of Monetary Economics*, 124S.

Bojinov, I., A. Rambachan, and N. Shephard (2021): "Panel experiments and dynamic causal effects: A finite population perspective," *Quantitative Economics*, 12, 1171–1196.

Bonhomme, S. and A. Denis (2024): "Fixed Effects and Beyond: Bias Reduction, Groups, Shrinkage, and Factors in Panel Data." Working paper.

Bonhomme, S. and J.-M. Robin (2009): "Consistent noisy independent component analysis." *Journal of Econometrics*, 149, 12–25.

Bonhomme, S. and J. M. Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics." *Review of Economic Studies*, 77, 491–533.

Bonhomme, S. and U. Sauder (2011): "Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling." *Review of Economics and Statistics*, 93, 479–494.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering." *Journal of Business and Economic Statistics*, 29, 238–249.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018a): "Alternative asymptotics and the partially linear model with many regressors," *Econometric Theory*, 34, 277–301.

——— (2018b): "Inference in Linear Regression Models with Many Covariates and Heteroscedasticity." *Journal of the American Statistical Association*, 113, 1350–1361.

Chamberlain, G. (1992): "Efficiency bounds for semiparametric regression." *Econometrica*, 60, 567–596.

Chen, C., D. Restuccia, and R. Santaeulàlia-Llopis (2022): "The effects of land markets on resource allocation and agricultural productivity." *Review of Economic Dynamics*, 45, 41–54.

——— (2023): "Land Misallocation and Productivity." *American Economic Journal: Macroeconomics*, 15, 441–465.

Chiappori, P.-A., K. Samphantharak, S. Schulhofer-Wohl, and R. M. Townsend (2014): "Heterogeneity and Risk Sharing in Village Economies." *Quantitative Economics*, 5, 1–27.

Cochran, W. G. (1977): *Sampling Techniques*, John Wiley & Sons, 3rd ed.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica*, 80, 2543–2594.

David, J. M. and V. Venkateswaran (2019): "The Sources of Capital Misallocation." *American Economic Review*, 109, 2531–2567.

Deeb, A. and C. de Chaisemartin (2022): "Clustering and External Validity in Randomized Controlled Trials." Working Paper arxiv:1912.01052.

Erdős, P. and A. Rényi (1959): "On the Central Limit Theorem for Samples from a Finite Population," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 4, 49–61.

Fogarty, C. B. (2018): "Regression-assisted inference for the average treatment effect in paired experiments." *Biometrika*, 105, 994–1000.

Freedman, D. (2008): "On regression adjustments in experiments with several treatments." *The Annals of Applied Statistics*, 2, 176–196.

Fuller, W. A. (1995): "Estimation in the presence of measurement error." *International Statistical Review*, 63, 121–141.

García-Santana, M., E. Moral-Benito, J. Pijoan-Mas, and R. Ramos (2020): "Growing Like Spain: 1995–2007." *International Economic Review*, 61, 1–468.

Giacomini, R., S. Lee, and S. Sarpietro (2023): "A Robust Method for Microforecasting and Estimation of Random Effects." Working Paper arXiv:2308.01596.

Gilraine, M., J. Gu, and R. McMillan (2022): "A Nonparametric Approach for Studying Teacher Impacts." Working paper.

Gorodnichenko, Y., D. Revoltella, J. Svejnar, and C. T. Weiss (2021): "Resource Misallocation in European Firms: The Role of Constraints, Firm Characteristics and Managerial Decisions." Working paper.

Graham, B. S. and J. L. Powell (2012): "Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models." *Econometrica*, 80, 2105–2152.

Guner, N., G. Ventura, and X. Yi (2008): "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics*, 11, 721–744.

Guvenen, F. (2009): "An empirical investigation of labor income processes." *Review of Economic Dynamics*, 12, 58–79.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020): "Estimation with Aggregate Shocks." *Review of Economic Studies*, 87, 1365–1398.

Hahn, J., J. D. Singleton, and N. Yildiz (2023): "Identification of Non-Additive Fixed Effects Models: Is the Return to Teacher Quality Homogeneous?" NBER Working Paper 31384, National Bureau of Economic Research.

Hansen, C. B. (2007): "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics*, 141, 597–620.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hansen, M. H. and W. N. Hurwitz (1943): "On the determination of optimum probabilities in sampling." *Annals of Mathematical Statistics*, 14, 333–362.

Hansen, M. H., W. N. Hurwitz, and M. A. Bershad (1961): "Measurement errors in censuses and surveys." *Bulletin of the International Statistical Institute*, 38, 359–374.

Higgins, A. and K. Jochmans (2024): "Bootstrap inference for fixed-effect models." *Econometrica*, 92, 411–427.

Horvitz, D. G. and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association*, 47, 663–685.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124, 1403–1448.

Hájek, J. (1960): "Limiting Distributions in Simple Random Sampling From a Finite Population," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 361–374.

Jin, Y. and D. Rothenhäusler (2024): "Tailored inference for finite populations: conditional validity and transfer across distributions." *Biometrika*, 111, 215–230.

Klette, T. J. (1999): "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data." *Journal of Industrial Economics*, 47, 451–476.

Kline, P., E. K. Rose, and C. R. Walters (2022): "Systemic Discrimination Among Large U.S. Employers." *Quarterly Journal of Economics*, 137, 1963–2036.

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

Kotlarski, I. (1967): "On characterizing the gamma and the normal distribution." *Pacific Journal of Mathematics*, 20, 69–76.

Lehmann, E. L. (1975): *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day.

Li, X. and P. Ding (2017): "General forms of finite population central limit theorems with applications to causal inference." *Journal of the American Statistical Association*, 112, 1759–1769.

Liang, K. and S. Zeger (1986): "Longitudinal data analysis using generalized linear models." *Biometrika*, 73, 13–22.

Liu, L., H. R. Moon, and F. Schorfheide (2020): "Forecasting with Dynamic Panel Data Models." *Econometrica*, 88, 171–201.

——— (2023): "Forecasting with a panel Tobit model." *Quantitative Economics*, 14, 117–159.

Magnac, T., N. Pistolesi, and S. Roux (2018): "Post-Schooling Human Capital Investments and the Life Cycle of Earnings." *Journal of Political Economy*, 126, 1219–1249.

Meager, R. (2022): "Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature." *American Economic Review*, 112, 1818–1847.

MONTIEL-OLEA, J. L., B. O'FLAHERTY, AND R. SETHI (2021): "Empirical Bayes Counterfactuals in Poisson Regression with an Application to Police Use of Deadly Force." Working paper.

MUKHOPADHYAY, P. (2001): *Topics in Survey Sampling*, vol. 153 of *Lecture Notes in Statistics*, Springer.

NEWEY, W. K. AND D. L. MCFADDEN (1994): "Large sample estimation and hypothesis testing." in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, chap. 36.

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55, 703–708.

NEYMAN, J. (1923/1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Roczniki Nauk Rolniczych*, 10, 1–51, reprinted in English in Statistical Science, 1990, 5, 463-480.

——— (1934): "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society, Series A*, 97, 558–606.

NIGMATULINA, D. (2023): "Sanctions and Misallocation: How Sanctioned Firms Won and Russians Lost," Working paper.

PETERS, M. (2020): "Heterogeneous Markups, Growth, and Endogenous Misallocation." *Econometrica*, 88, 2037–2073.

RESTUCCIA, D. AND R. ROGERSON (2017): "The Causes and Costs of Misallocation." *Journal of Economic Perspectives*, 31, 151–174.

ROSENBAUM, P. R. (2002): "Covariance adjustment in randomized experiments and observational studies." *Statistical Science*, 17, 286–327.

SCHULHOFER-WOHL, S. (2011): "Heterogeneity and Tests of Risk Sharing." *Journal of Political Economy*, 119, 925–958.

STARTZ, R. AND D. G. STEIGERWALD (2023): "Inference and extrapolation in finite populations with special attention to clustering." *Econometric Reviews*, 42, 343–357.

——— (2024): "The variance of regression coefficients when the population is finite." *Journal of Econometrics*, 240.

Stock, J. H. and M. W. Watson (2008): "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression." *Econometrica*, 76, 155–174.

Suri, T. (2011): "Selection and Comparative Advantage in Technology Adoption." *Econometrica*, 79, 159–209.

Townsend, R. M. (1994): "Risk and Insurance in Village India." *Econometrica*, 62, 539–591.

Villacorta, L. (2021): "Estimating Unobserved Country Heterogeneity in Capital-Labor Substitution: A Nonlinear Panel Data Framework." Working paper.

Xu, R. (2021): "Potential outcomes and finite-population inference for M-estimators." *Econometrics Journal*, 24, 162–176.

Yeh, C., C. Macaluso, and B. Hershbein (2022): "Monopsony in the US Labor Market." *American Economic Review*, 112, 2099–2138.