Econometric Theory, **40**, 2024, 60–97.
doi:10.1017/S0266466622000378

INFERENCE ON A DISTRIBUTION FROM NOISY DRAWS

KOEN JOCHMANS
Université Toulouse 1 Capitole

MARTIN WEIDNER
University of Oxford

We consider a situation where the distribution of a random variable is being estimated by the empirical distribution of noisy measurements of that variable. This is common practice in, for example, teacher value-added models and other fixed-effect models for panel data. We use an asymptotic embedding where the noise shrinks with the sample size to calculate the leading bias in the empirical distribution arising from the presence of noise. The leading bias in the empirical quantile function is equally obtained. These calculations are new in the literature, where only results on smooth functionals such as the mean and variance have been derived. We provide both analytical and jackknife corrections that recenter the limit distribution and yield confidence intervals with correct coverage in large samples. Our approach can be connected to corrections for selection bias and shrinkage estimation and is to be contrasted with deconvolution. Simulation results confirm the much-improved sampling behavior of the corrected estimators. An empirical illustration on heterogeneity in deviations from the law of one price is equally provided.

1. INTRODUCTION

Let $\theta_1, \ldots, \theta_n$ be a random sample from a distribution F that is of interest. Suppose that we only observe noisy measurements of these variables, say $\vartheta_1, \ldots, \vartheta_n$. A popular approach is to do inference on F and its functionals using the empirical distribution of $\vartheta_1, \ldots, \vartheta_n$. This is common practice when analyzing panel data with heterogenous coefficients. In the literature on student achievement, for example, θ_i is a teacher effect, ϑ_i is an estimator of it obtained from data on student test scores, and we care about the distribution of teacher value-added

We are grateful to Isaiah Andrews, Stéphane Bonhomme, Bo Honoré, Ryo Okui, and Peter Schmidt for comments, and to Arthur Lewbel and three referees for feedback on an earlier version. We also greatly appreciate the help of Ryo Okui and Mototsugu Shintani in providing us with the data used in our empirical illustration. Jochmans gratefully acknowledges financial support from the European Research Council through grant ERC-2016-StG-715787-MiMo and from the French Government and the ANR under the Investissements d' Avenir program, grant ANR-17-EURE-0010. Weidner gratefully acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 and from the European Research Council grants ERC-2014-CoG-646917-ROMIA and ERC-2018-CoG-819086-PANEDA. Address correspondence to Koen Jochmans, Toulouse School of Economics, Université Toulouse 1 Capitole, 1 esplanade de l'Université, 31080 Toulouse, France; e-mail: koen.jochmans@tse-fr.eu.

(see, e.g., Jackson, Rockoff, and Staiger, 2014 for an overview). In the same vein, Guvenen (2009), Browning, Ejrnæs, and Alvarez (2010), and Magnac and Roux (2021) estimate heterogenous earning profiles, whereas Ahn et al. (2014) find substantial heterogeneity in ambiguity aversion. In a nonlinear fixed-effect model, the marginal effect is heterogenous across units and interest lies in the distribution of these effects as well as its functionals (Chamberlain, 1984; Hahn and Newey, 2004). Although the plug-in approach is popular, using $\vartheta_1,\ldots,\vartheta_n$ rather than $\theta_1,\ldots,\theta_n$ introduces bias that is almost entirely ignored in practice. Barras, Gagliardini, and Scaillet (2021)), who are interested in the distribution of the skill of fund managers, find that not accounting for bias leads to substantial overestimation of tail mass and fails to pick up the substantial asymmetry in the skill distribution.

We analyze the properties of the plug-in estimator of F in a location-scale setting where

$$\vartheta_i = \theta_i + \frac{\sigma_i}{\sqrt{m}}\,\varepsilon_i, \qquad \varepsilon_i\,|\,(\theta_i,\sigma_i^2) \sim \text{i.i.d. } (0,1),$$

where m is a parameter that grows with n. As the variance of the (heteroskedastic) noise is σ_i^2/m, this device shrinks the noise as the sample size grows. This is a very natural asymptotic embedding in settings where ϑ_i is an estimator of θ_i obtained from a sample of size m, as in a panel data setting or meta-analysis (Vivalt, 2015). It is related to, yet different from, an approach based on small measurement-error approximations as in Chesher (1991, 2017),[1] and has precedent in the analysis of fixed-effect models for panel data, although for different purposes, as discussed in more detail below (see, e.g., Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003).

Efron (2011) essentially entertains the homoskedastic setting with normal noise, where

$$\vartheta_i|\,\theta_i \sim N(\theta_i, \sigma^2/m),$$

and defines selection bias as the tendency of the ϑ_i's associated with the (in magnitude) largest θ_i's to be larger than their corresponding θ_i. He proposes to deal with selection bias by using the well-known Empirical-Bayes estimator of Robbins (1956), which here is equal to

$$\vartheta_i + \frac{\sigma^2}{m}\,\nabla^1 \log p(\vartheta_i),$$

[1] Chesher (1991) provides expansions for densities, whereas we focus on distribution and quantile functions. Chesher (2017) discusses the impact of noise in the explanatory variables in a quantile-regression model; this is a different setup than the one considered here. Evdokimov and Zeleneev (2020) use our device of measurement error that shrinks with the sample size to correct inference in generalized method-of-moment problems.

where p is the marginal density of the ϑ_i and ∇^1 denotes the first-derivative operator. For example, when $\theta_i \sim N(0,\psi^2)$, this expression then yields the (infeasible) shrinkage estimator

$$\left(1-\frac{\sigma^2/m}{\sigma^2/m+\psi^2}\right)\vartheta_i,$$

a parametric plug-in estimator of which would be the James and Stein (1961) estimator. More generally, nonparametric implementation would also require estimation of p and its first derivative. Shrinkage to the overall mean (in this case zero) is intuitive, as selection bias essentially manifests itself through the tails of the empirical distribution of the ϑ_i being too thick.[2] Shrinkage is commonly applied in empirical work (see, e.g., Rockoff, 2004; Chetty, Friedman, and Rockoff, 2014). It should be stressed, though, that, while shrinkage improves on $\vartheta_1,\dots,\vartheta_n$ in terms of estimation risk, it does not lead to preferable estimators of the distribution F or its moments.

The approach taken here is different from Efron (2011). Without making parametric assumptions on F, we calculate the (leading) bias of the naive plug-in estimator of the distribution,

$$\hat{F}(\theta) := n^{-1}\sum_{i=1}^{n} 1\{\vartheta_i \le \theta\}.$$

This calculation allows to construct estimators that correct for the bias directly. In the James–Stein problem, where $\theta_i \sim N(\eta,\psi^2)$, for example, the bias under homoskedastic noise equals

$$-\frac{\theta-\eta}{2}\,\frac{\sigma^2/\psi^2}{m}\phi\left(\frac{\theta-\eta}{\psi}\right)+o(m^{-1}).$$

Thus, the empirical distribution is indeed upward biased in the left tail and downward biased in the right tail. A bias order of m^{-1} implies incorrect coverage of confidence intervals unless $n/m^2 \to 0$. We present plug-in and jackknife estimators of the leading bias and show that the bias-corrected estimators are asymptotically normal with zero mean and variance $F(\theta)\,(1-F(\theta))$ as long as $n/m^3 \to 0$. So, bias correction is preferable to the naive plug-in approach for typical data sizes encountered in practice, where m tends to be quite small relative to n. We also provide corresponding bias-corrected estimators of the quantile function of F.

If the distribution of $\sigma_i\,\varepsilon_i$ is fully known, recovering F is a (generalized) deconvolution problem that can be solved for fixed m. Deconvolution-based estimators

[2]The same shrinkage factor is applied to each ϑ_i, a consequence of the noise being homoskedastic. How to deal with heteroskedastic noise in an Empirical-Bayes framework is not obvious. Discussion and a recent contribution can be found in Weinstein et al. (2018).

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

are well studied (see, e.g., Carroll and Hall, 1988; Delaigle and Meister, 2008). However, they have a very slow rate of convergence, and it is well documented that they can behave quite poorly in small samples.[3] In response to this, Efron (2016) has recently argued for a return to a more parametric approach. Our approach delivers intuitive estimators that enjoy the usual parametric convergence rate and are numerically well behaved. Although it does not deliver a fixed-m consistent estimator, bias correction further ensures that size-correct inference can be performed, provided that n/m^3 is small. It is not clear how to conduct inference based on deconvolution estimators.

Working out the statistical properties of $\hat{F}$ (and of its quantile function) is nontrivial because $\hat{F}$ is a nonsmooth function of the data $\vartheta_1, \ldots, \vartheta_n$. As such, the approach taken here is different from, and complementary to, recent work on estimating average marginal effects in panel data models, which only looks at smooth functionals such as the mean and variance (see, e.g., Fernández-Val and Lee, 2013; Okui and Yanagi, 2019). The impact of noise on smooth transformations of the ϑ_i can be handled using conventional methods based on Taylor-series expansions. We contrast such an approach with our derivations below. How to perform inference on the quantiles of marginal effects in nonlinear panel models is a long-standing open question (Dhaene and Jochmans, 2015), and the current work can be seen as a first step in that direction.

In work contemporaneous to our own, Okui and Yanagi (2020) derive the bias of a kernel-smoothed estimator of F and its derivative. Such smoothing greatly facilitates the calculation of the bias, making it amenable to conventional analysis. However, it also introduces additional bias terms that require much stronger moment conditions as well as further restrictions on the relative growth rates of n, m, and the bandwidth that governs the smoothing. Nevertheless, the (leading) bias term obtained in Okui and Yanagi (2020, Thm. 3) coincides with ours in Proposition 1. Additional discussion on and comparison between the two different approaches is given in Okui and Yanagi (2020, pp. 169–170).

2. LARGE-SAMPLE PROPERTIES OF PLUG-IN ESTIMATORS

Let F be a univariate distribution on the real line. We are interested in estimation of and inference on F and its quantile function $q(\tau) := \inf_\theta \{\theta : F(\theta) \geq \tau\}$. If a random sample $\theta_1, \ldots, \theta_n$ from F would be available, this would be a standard problem. We instead consider the situation where $\theta_1, \ldots, \theta_n$ themselves are unobserved and we observe noisy measurements $\vartheta_1, \ldots, \vartheta_n$, with variances $\sigma_1^2/m, \ldots, \sigma_n^2/m$ for a positive real number m which, in our asymptotic analysis below, will be required to grow with n. We assume the following.

[3]There are also solutions to the measurement-error problem based on repeated measurements (or instrumental variables), coupled with suitable independent restrictions (see, for example, Horowitz and Markatou, 1996; Li and Vuong, 1998; Hu, 2008; Hu and Schennach, 2008; Bonhomme, Jochmans, and Robin, 2016a, 2016b). These can be useful alternatives in static models for panel data, where the object of interest is the distribution of the random intercept, as in the work of Horowitz and Markatou (1996), for example.

Assumption 1. The variables $(\theta_i, \sigma_i^2, \vartheta_i)$ are i.i.d. across i, with

$$E(\vartheta_i \,|\, \theta_i, \sigma_i^2) = \theta_i, \quad E((\vartheta_i - \theta_i)^2 \,|\, \theta_i, \sigma_i^2) = \frac{\sigma_i^2}{m},$$

and $\sigma_i^2 \in [\underline{\sigma}^2, \overline{\sigma}^2] \subset (0, \infty)$ for all i.

Our setup reflects a situation where the noisy measurements $\vartheta_1, \ldots, \vartheta_n$ converge in squared mean to $\theta_1, \ldots, \theta_n$ at the rate m^{-1}. A leading case is the situation where ϑ_i is an estimator of θ_i obtained from a sample of size m that converges at the parametric rate.[4] We allow θ_i and σ_i^2 to be correlated, implying that the noise $\vartheta_i - \theta_i$ is not independent of θ_i. Hence, we allow for measurement error to be nonclassical. Recovering the distribution of θ_i from a sample of $(\vartheta_i, \sigma_i^2)$ is, therefore, not a standard deconvolution problem.

It is common to estimate $F(\theta)$ by

$$\hat{F}(\theta) := n^{-1} \sum_{i=1}^{n} 1\{\vartheta_i \leq \theta\},$$

the empirical distribution of the ϑ_i at θ. As we will show below, under suitable regularity conditions, that such plug-in estimators are consistent and asymptotically normal as $n \to \infty$ provided that m grows with n so that n/m^2 converges to a finite constant. The use of $\vartheta_1, \ldots, \vartheta_n$ rather than $\theta_1, \ldots, \theta_n$ introduces bias of the order m^{-1}, in general. This bias implies that test statistics are size distorted and the coverage of confidence sets is incorrect unless n/m^2 converges to zero.

The bias problem is easy to see (and fix) when interest lies in smooth functionals of F,

$$\mu := E(\varphi(\theta_i)),$$

for a (multiple-times) differentiable function φ. An (infeasible) plug-in estimator based on $\theta_1, \ldots, \theta_n$ would be

$$\tilde{\mu} := n^{-1} \sum_{i=1}^{n} \varphi(\theta_i).$$

[4]Everything to follow can be readily modified to different convergence rates as well as to the case where

$$\mathrm{var}(\vartheta_i \,|\, \theta_i, \sigma_i^2) = \sigma_i^2 / m_i,$$

with $m_i := p_i m$ for a random variable $p_i \in (0,1]$. It suffices to redefine σ_i^2 as σ_i^2/p_i. When the ϑ_i represent estimators, this device allows for the sample size to vary with i. For example, in a panel data setting, it would cover unbalanced panels under a missing-at-random assumption. Furthermore, the requirement that ϑ_i is unbiased can be relaxed to allow for standard nonlinearity bias of order m^{-1}. We do not do this here as it is possible quite generally to reduce the bias down to $O(m^{-3/2})$, for example, via a jackknife or bootstrap correction, making it negligible in our analysis below. Furthermore, the split-sample jackknife approach to bias correction that we discuss below would automatically take care of this additional m^{-1} bias without modification.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Clearly, this estimator is unbiased and satisfies $\tilde{\mu} \overset{a}{\sim} N(\mu, \sigma_\mu^2/n)$ as soon as $\sigma_\mu^2 := \operatorname{var}(\varphi(\theta_i))$ exists. For the feasible plug-in estimator of μ,

$$\hat{\mu} := n^{-1} \sum_{i=1}^{n} \varphi(\vartheta_i),$$

under standard regularity conditions, a Taylor-series expansion of $\varphi_i(\vartheta_i)$ around θ_i yields

$$E(\hat{\mu} - \mu) = \frac{b_\mu}{m} + O(m^{-3/2}), \qquad b_\mu := \frac{E(\nabla^2 \varphi(\theta_i)\, \sigma_i^2)}{2},$$

and

$$\operatorname{var}(\hat{\mu}) = \frac{\sigma_\mu^2}{n} + O\left(n^{-1} m^{-1}\right).$$

Hence, letting $z \sim N(0,1)$, we have

$$\frac{\hat{\mu} - \mu}{\sigma_\mu / \sqrt{n}} \overset{a}{\sim} z + \sqrt{\frac{n}{m^2}} \frac{b_\mu}{\sigma_\mu} \sim N(c\, b_\mu / \sigma_\mu, \sigma_\mu^2),$$

as $n/m^2 \to c^2 < \infty$ when $n, m \to \infty$. The noise in $\vartheta_1, \ldots, \vartheta_n$ introduces bias unless φ is linear. It can be corrected for by subtracting a plug-in estimator of b_μ / m from $\hat{\mu}$. Doing so, again under regularity conditions, delivers an estimator that is asymptotically unbiased as long as $n/m^3 \to 0$.

2.1. Distribution function

The machinery from above cannot be applied to deduce the bias of $\hat{F}$ as it is a step function and, hence, nondifferentiable. We will derive its leading bias under the following conditions. To state them, we let

$$\varepsilon_i := \frac{\vartheta_i - \theta_i}{\sigma_i / \sqrt{m}}$$

and write f for the density function of F.

Assumption 2. The variables ε_i are independent of (θ_i, σ_i^2), and their distribution is absolutely continuous and has finite fourth-order moment. The function f is three times differentiable with uniformly bounded derivatives, and one of the following two sets of conditions holds:

A. The function $E(\sigma_i^{p+1} | \theta_i = \theta)$ is p-times differentiable for $p = 1, 2$, the joint density of (θ_i, σ_i) exists, the conditional density function of θ_i given σ_i is twice differentiable with respect to θ_i, and the derivatives are bounded in absolute value by a function $e(\sigma_i)$ such that $E(e(\sigma_i)) < \infty$.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

B. There exists a deterministic function σ so that $\sigma_i = \sigma(\theta_i)$ for all i; and σ is three times differentiable and has uniformly bounded derivatives.

Assumption 2 imposes smoothness on certain densities and conditional expectations but not on the estimator of F.

Define the function

$$\beta(\theta) := \frac{E(\sigma_i^2|\theta_i = \theta)f(\theta)}{2},$$

which is well behaved under Assumption 2, and let

$$b_F(\theta) := \beta'(\theta)$$

be its derivative. We also introduce the covariance function

$$\sigma_F(\theta,\theta') := F(\theta \wedge \theta') - F(\theta)\,F(\theta'),$$

where we use $\theta \wedge \theta'$ to denote $\min\{\theta,\theta'\}$. Proposition 1 summarizes the large-sample properties of $\hat{F}$.

PROPOSITION 1. *Let Assumptions 1 and 2 hold. Then, as $n,m \to \infty$,*

$$E(\hat{F}(\theta)) - F(\theta) = \frac{b_F(\theta)}{m} + O(m^{-3/2}), \quad \operatorname{cov}\big(\hat{F}(\theta),\hat{F}(\theta')\big) = \frac{\sigma_F(\theta,\theta')}{n} + O(n^{-1}m^{-1}),$$

where the order of the remainder terms is uniform in θ. If furthermore $n/m^2 \to c \in [0,+\infty)$, then

$$\sqrt{n}\left(\hat{F}(\theta) - F(\theta) - \frac{b_F(\theta)}{m}\right) \rightsquigarrow \mathbb{G}_F(\theta),$$

where $\mathbb{G}_F(\theta)$ is a mean-zero Gaussian process with covariance function $\sigma_F(\theta_1,\theta_2)$.

Proof. The proof is in Appendix A. □

To illustrate the result, suppose that σ_i^2 is independent of θ_i and that θ_i has density function

$$f(\theta) = \frac{1}{\psi}\phi\left(\frac{\theta-\eta}{\psi}\right),$$

as in the James and Stein (1961) problem. Letting σ^2 denote the mean of the σ_i^2, an application of Proposition 1 yields

$$b_F(\theta) = -\frac{\theta-\eta}{2}\,\frac{\sigma^2}{\psi^2}\,\phi\left(\frac{\theta-\eta}{\psi}\right).$$

Thus, $\hat{F}(\theta)$ is upward biased when $\theta < \eta$ and is downward biased when $\theta > \eta$. This finding is a manifestation of the phenomenon of regression to the mean (or selection bias, or the winner's curse; see Efron, 2011). It implies that the empirical distribution tends to be too disperse.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

2.2. Quantile function

The bias in $\hat{F}$ translates to bias in estimators of the quantile function. A natural estimator of the quantile function is the left inverse of $\hat{F}$. With this definition, the plug-in estimator of the τth quantile is

$$\hat{q}(\tau) := \vartheta_{(\lceil \tau n \rceil)},$$

where $\vartheta_{(\lceil \tau n \rceil)}$ is the $\lceil \tau n \rceil$th-order statistic of our sample, where $\lceil a \rceil$ delivers the smallest integer at least as large as a.

To calculate the leading bias in $\hat{q}(\tau)$, observe that it is an approximate solution to the empirical moment condition

$$\hat{F}(q) - \tau = 0$$

(with respect to q). From Proposition 1, we know that

$$E(\hat{F}(q(\tau))) - \tau = \frac{b_F(q(\tau))}{m} + O(m^{-3/2}),$$

uniformly in τ, so the moment condition that defines the estimator $\hat{q}(\tau)$ is biased. Letting

$$b_q(\tau) := -\frac{b_F(q(\tau))}{f(q(\tau))}, \qquad \sigma_q^2(\tau) := \frac{\tau(1-\tau)}{f(q(\tau))^2},$$

we obtain the following result.

PROPOSITION 2. *Let Assumptions 1 and 2 hold. For $\tau \in (0,1)$, assume that $f > 0$ in a neighborhood of $q(\tau)$. Then,*

$$\sqrt{n}\left(\hat{q}(\tau) - q(\tau) - \frac{b_q(\tau)}{m}\right) \xrightarrow{d} N(0, \sigma_q^2(\tau)),$$

as $n, m \to \infty$ with $n/m^2 \to c \in [0, +\infty)$.

Proof. The proof is in Appendix A. □

As an example, when $\theta_i \sim N(\eta, \psi^2)$, independent of σ_i^2, we have

$$b_q(\tau) = \frac{\sigma^2/\psi^2}{2}(q(\tau) - \eta),$$

which, in line with our discussion on regression to the mean above, is positive for all quantiles below the median and negative for all quantiles above the median. The median itself is, in this particular case, estimated without plug-in bias of order m^{-1}. It will, of course, still be subject to the usual n^{-1} bias arising from the nonlinear nature of the estimating equation.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

3. ESTIMATION AND INFERENCE

Propositions 1 and 2 complement the existing results on the bias in smooth functionals (Fernández-Val and Lee, 2013; Okui and Yanagi, 2019) of the distribution of heterogenous parameters in panel data models. Our calculations confirm that the order of the bias in the empirical distribution and in the quantile function is of the same order as in the smooth case, m^{-1}.

3.1. Split-panel jackknife estimation

Importantly, our results validate a traditional jackknife approach to bias correction as in Hahn and Newey (2004) and Dhaene and Jochmans (2015). Such an approach exploits the fact that the bias is proportional to m^{-1} and is based on re-estimating $\theta_1, \ldots, \theta_n$ from subsamples. The simplicity of such a method makes it very useful in panel data applications, for example.

To illustrate how the jackknife would work here, consider a stationary (balanced) $n \times m$ panel. Let ϑ_{i,m_1} be an estimator of θ_i constructed from the $n \times m_1$ subpanel consisting of the first m_1 cross sections only. Then

$$\hat{F}_{m_1}(\theta) := n^{-1} \sum_{i=1}^{n} 1\{\vartheta_{i,m_1} \leq \theta\}$$

is the plug-in estimator of $F(\theta)$ based on this subpanel alone. From Proposition 1, it follows that

$$E(\hat{F}_{m_1}(\theta)) = F(\theta) + \frac{b_F(\theta)}{m_1} + O(m_1^{-3/2}).$$

Using the remaining $m_2 := m - m_1$ cross sections from the full panel, we can equally calculate estimators ϑ_{i,m_2} and subsequently construct

$$\hat{F}_{m_2}(\theta) := n^{-1} \sum_{i=1}^{n} 1\{\vartheta_{i,m_2} \leq \theta\},$$

for which

$$E(\hat{F}_{m_2}(\theta)) = F(\theta) + \frac{b_F(\theta)}{m_2} + O(m_2^{-3/2})$$

follows in the same way. Consequently,

$$\tilde{b}_F(\theta) := m_1 \hat{F}_{m_1}(\theta) + m_2 \hat{F}_{m_2}(\theta) - m\hat{F}(\theta)$$

is a split-panel jackknife estimator of the leading bias term $b_F(\theta)$. Hence,

$$\tilde{F}(\theta) := \hat{F}(\theta) - \frac{\tilde{b}_F(\theta)}{m}$$

is a nonparametric bias-corrected estimator.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

A jackknife estimator of the quantile function can be defined in the same way. Moreover, let $\vartheta_{(\lceil \tau n \rceil),m_1}$ and $\vartheta_{(\lceil \tau n \rceil),m_2}$ be the $\lceil \tau n \rceil$th-order statistic of the re-estimated quantities in the first and second subsamples, respectively. Recall that $\vartheta_{(\lceil \tau n \rceil),m_1}$ is the (approximate) solution to $\hat{F}_{m_1}(q) - \tau = 0$, and so is our estimator of $q(\tau)$ as obtained from the information in the $n \times m_1$ subpanel only. As before,

$$\tilde{b}_q(\tau) := m_1 \vartheta_{(\lceil \tau n \rceil),m_1} + m_2 \vartheta_{(\lceil \tau n \rceil),m_2} - m \vartheta_{(\lceil \tau n \rceil)}$$

is a nonparametric estimator of $b_q(\tau)$ that gives rise to a jackknife bias-corrected estimator of the quantile function.

The large-sample behavior of these jackknife estimators is the same as for the analytic corrections in Propositions 3 and 4. The split-sample jackknife is simple to implement but requires access to the original data from which $\vartheta_1, \ldots, \vartheta_n$ were computed. This can be infeasible in meta-analysis problems, where each of the ϑ_i is an estimator constructed from a different dataset that need not all be accessible. It can also be complicated in structural econometric models, where ϑ_i may be the solution to a cumbersome optimization program that can be time-consuming to solve. We discuss an alternative bias-correction estimator next.

3.2. Analytic bias correction

We will formulate regularity conditions for a plug-in estimator of the bias to be consistent under the maintained assumption that the $\sigma_1^2, \ldots, \sigma_m^2$ are known. We conjecture that, under suitable conditions, the results below will continue to go through when the σ_i^2 are replaced by estimators.

A bias-corrected estimator based on Proposition 1 takes the form

$$\check{F}(\theta) := \hat{F}(\theta) - \frac{\hat{b}_F(\theta)}{m}, \qquad \hat{b}_F(\theta) := -\frac{(nh^2)^{-1} \sum_{i=1}^n \sigma_i^2 \, k'\left(\frac{\vartheta_i - \theta}{h}\right)}{2},$$

where k' is the derivative of kernel function k and h is a nonnegative bandwidth parameter. Thus, we estimate the bias using standard kernel methods. For simplicity, we will use a Gaussian kernel throughout, so $k'(\eta) := -\eta \phi(\eta)$.

We establish the asymptotic behavior of $\check{F}$ under the following conditions.

Assumption 3.
(i) The conditional density of θ_i given σ_i is five times differentiable with respect to θ_i and the derivatives are bounded in absolute value by a function $e(\sigma_i)$ such that $E(e(\sigma_i)) < \infty$.
(ii) There exists an integer $\omega > 2$, and real numbers $\kappa > 1 + (1-\omega^{-1})^{-1}$ and $\eta > 0$ so that $\sup_\theta (1 + |\theta|^\kappa) f(\theta) = O(1)$ and $\sup_\theta (1 + |\theta|^{1+\eta}) |\nabla^1 b_F(\theta)| = O(1)$, and $\sup_\theta |b_F(\theta)| = O(1)$.
(iii) The density of ε, g, satisfies $g(\varepsilon) \leq C(1+|\varepsilon|)^{-\alpha}$ for finite constant C and $\alpha \geq \kappa + 1$.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Assumption 3 contains simple smoothness and boundedness requirements on the conditional density of θ_i given σ_i^2, as well as tail conditions on the marginal density of the θ_i and on the bias function $b_F(\theta)$.

We have the following result.

PROPOSITION 3. *Let Assumptions 1–3 hold, and let* $\varepsilon := (3-\omega^{-1})\omega^{-1} > 0$. *If* $h = O(m^{-1/2})$, $h^{-1} = O(m^{2/3-4/9\varepsilon})$, *and* $h^{-1} = O(n)$, *as* $n \to \infty$ *and* $m \to \infty$ *with* $n/m^4 \to 0$, *then*

$$\sqrt{n}(\check{F}(\theta) - F(\theta)) \rightsquigarrow \mathbb{G}_F(\theta)$$

as a stochastic process indexed by θ, *where* $\mathbb{G}_F(\theta)$ *is a mean-zero Gaussian process with covariance function* $\sigma_F(\theta_1, \theta_2)$.

Proof. The proof is in Appendix A.5. □

The implications of Proposition 3 are qualitatively similar to those for smooth functionals discussed above. Indeed, for any fixed θ, it implies that

$$\check{F}(\theta) \overset{a}{\sim} N(F(\theta), F(\theta)(1-F(\theta))/n)$$

as $n \to \infty$ and $m \to \infty$ with $n/m^4 \to 0$. Thus, the leading bias is removed from $\hat{F}$ without incurring any cost in terms of (asymptotic) precision. Given the correction term, the sample variance of

$$1\{\vartheta_i \leq \theta\} + \frac{1}{2}\frac{1}{mh^2}\sigma_i^2 \kappa'\left(\frac{\vartheta_i - \theta}{h}\right)$$

is a more natural basis for inference in small samples than is that of $1\{\vartheta_i \leq \theta\}$.

A data-driven way of choosing h is by cross validation. A plug-in estimator of the integrated squared error $\int_{-\infty}^{+\infty}(\check{F}(\theta) - F(\theta))^2\, d\theta$ (up to multiplicative and additive constants) is

$$v(h) := \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\sigma_i^2\sigma_j^2}{h^2}\,\underline{\phi}'(\vartheta_i, \vartheta_j; h) + \sum_{i=1}^{n}\sum_{j\neq i}\frac{\sigma_i^2}{h}\left(m\phi'\left(\frac{\vartheta_i - \vartheta_j}{h}\right) - \frac{nm}{n-1}\phi\left(\frac{\vartheta_i - \vartheta_j}{h}\right)\right),$$

where we use the shorthand

$$\underline{\phi}'(\vartheta_i, \vartheta_j; h) := \frac{1}{4}\frac{1}{\sqrt{2}h}\phi\left(\frac{\vartheta_i - \vartheta_j}{\sqrt{2}h}\right)\left(\frac{1}{2} - \frac{(\vartheta_i + \vartheta_j)^2}{4h^2} + \frac{\vartheta_i\vartheta_j}{h^2}\right).$$

See Appendix B for details on the derivation. The cross-validated bandwidth then is $\check{h} := \arg\min_h v(h)$ on the interval $(0, +\infty)$.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Now, turn the bias-corrected estimation of the quantile function. Proposition 2 readily suggests a bias-corrected estimator of the form

$$\hat{q}(\tau) - \frac{\hat{b}_q(\tau)}{m}, \qquad \hat{b}_q(\tau) := -\frac{\hat{b}_F(\hat{q}(\tau))}{\hat{f}(\hat{q}(\tau))},$$

using obvious notation. While (under suitable regularity conditions) such an estimator successfully reduces bias, it has the unattractive property that it requires a nonparametric estimator of the density f, which further shows up in the denominator.

An alternative estimator that avoids this issue is

$$\check{q}(\tau) := \vartheta_{(\lceil \hat{\tau}^* n \rceil)}, \qquad \hat{\tau}^* := \tau + \frac{\hat{b}_F(\hat{q}(\tau))}{m}.$$

The justification for this estimator comes from the fact that $E(\hat{F}(q(\tau))) - \tau^* = O(m^{-2})$, where $\tau^* = \tau + b_F(q(\tau))/m$, and its interpretation is intuitive. Given the noise in the ϑ_i relative to the θ_i, the empirical distribution of the former is too heavy-tailed relative to the latter, and so $\hat{q}(\tau)$ estimates a quantile that is too extreme, on average. Changing the quantile of interest from τ to τ^* adjusts the naive estimator and corrects for regression to the mean.

PROPOSITION 4. *Let the assumptions stated in Proposition 3 hold. For $\tau \in (0,1)$, assume that $f > 0$ in a neighborhood of $q(\tau)$. Then,*

$$\sqrt{n}\left(\check{q}(\tau) - q(\tau)\right) \xrightarrow{d} N(0, \sigma_q^2(\tau)),$$

as $n, m \to \infty$ with $n/m^4 \to 0$.

Proof. The proof is in Appendix A.5. □

The corrected estimator has the same asymptotic variance as the uncorrected estimator. It is well known that plug-in estimators of σ_q^2 can perform quite poorly in small samples (Maritz and Jarrett, 1978). Typically, researchers rely on the bootstrap, and we suggest doing so here. Moreover, draw (many) random samples of size n from the original sample $\vartheta_1, \ldots, \vartheta_n$ and re-estimate $q(\tau)$ by the bias-corrected estimator for each such sample. Then construct confidence intervals for $q(\tau)$ using the percentiles of the empirical distribution of these estimates. Note that, again, this bootstrap procedure does not involve re-estimation of the individual θ_i.

4. NUMERICAL ILLUSTRATIONS

4.1. Simulated data

To support our theory, we provide simulation results for a James and Stein (1961) problem where $\theta_i \sim N(0, \psi^2)$ and we have access to an $n \times m$ panel on independent

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

realizations of the random variable

$$x_{it}|\theta_i \sim N(\theta_i, \sigma^2).$$

This setup is a simple random-coefficient model. It is similar to the classic many normal means problem of Neyman and Scott (1948). While their focus was on consistent estimation of the within-group variance, σ^2, for fixed m, our focus is on between-group characteristics and the distribution of the θ_i as a whole. We estimate θ_i by the fixed-effect estimator, i.e.,

$$\vartheta_i = m^{-1} \sum_{t=1}^{m} x_{it}.$$

The sampling variance of $\vartheta_i|\theta_i$ is σ^2/m. Rather than assuming this variance to be known, we implement our analytical bias correction using the estimator

$$s_i^2 := (m-1)^{-1} \sum_{t=1}^{m} (x_{it} - \vartheta_i)^2.$$

We do not make use of the fact that the ϑ_i are homoskedastic in estimating the noise or in constructing the bias correction. Moreover, the implementation of our procedure is nonparametric in the noise distribution.

A deconvolution argument implies that

$$\vartheta_i \sim N(0, \psi^2 + \sigma^2/m).$$

Thus, indeed, the empirical distribution of the fixed-effect estimator is too fat-tailed. In particular, the sample variance of $\vartheta_1, \ldots, \vartheta_n$,

$$\hat{\psi}^2 := \frac{1}{n-1} \sum_{i=1}^{n} (\vartheta_i - \overline{\vartheta})^2, \qquad \overline{\vartheta} := n^{-1} \sum_{i=1}^{n} \vartheta_i,$$

is a biased estimator of ψ^2. To illustrate how this invalidates inference in typically sized datasets, we simulated data for $\psi^2 = 1$ (so F is standard normal) and $\sigma^2 = 5$. The panel dimensions (n, m) reported on are $(50, 3)$, $(100, 4)$, and $(200, 5)$. Table 1 reports the bias and standard deviation of $\hat{\psi}^2$ as well as the empirical rejection frequency of the usual two-sided t-test for the null that $\psi^2 = 1$. The nominal size is set to 5%. In practice, however, the test rejects in virtually all of the 10,000 replications. The table provides the same summary statistics for the bias-corrected estimator

$$\check{\psi}^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left((\vartheta_i - \overline{\vartheta})^2 - \frac{s_i^2}{m} \right).$$

The adjustment reduces the estimator's bias relative to its standard error and brings down the empirical rejection frequencies to just over their nominal value for the sample sizes considered.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

TABLE 1. Inference on ψ^2 in the James–Stein problem from $n \times m$ panel data

		Bias		Std		SE/Std		Size (5%)	
n	m	$\hat{\psi}^2$	$\check{\psi}^2$	$\hat{\psi}^2$	$\check{\psi}^2$	$\hat{\psi}^2$	$\check{\psi}^2$	$\hat{\psi}^2$	$\check{\psi}^2$
50	3	1.616	−0.054	0.525	0.577	0.964	0.971	0.973	0.082
100	4	1.224	−0.028	0.321	0.337	0.966	0.969	0.997	0.073
200	5	0.989	−0.010	0.199	0.205	0.985	0.985	1.000	0.062

Notes: $\hat{\psi}^2$ is the plug-in estimator of ψ^2. $\check{\psi}^2$ is its (analytically) bias-corrected version constructed using estimators of the variance of the noise distributions. The table reports the bias and standard deviation of these estimators, along with the ratio of the average standard error to the standard deviation and empirical rejection frequencies of a two-sided t-test for the null that $\psi^2 = 1$, which is the value with which the data were generated.

A popular approach in empirical work to deal with noise in $\vartheta_1, \ldots, \vartheta_n$ is shrinkage estimation (see, e.g., Chetty et al., 2014). This procedure is not designed to improve estimation and inference of F or its moments, however. In the current setting, the (infeasible, parametric) shrinkage estimator is simply

$$\left(1 - \frac{\sigma^2/m}{\sigma^2/m + \psi^2}\right) \vartheta_i.$$

Its exact sampling variance is

$$\left(\frac{\psi^2}{\sigma^2/m + \psi^2}\right) \psi^2 = \psi^2 - \frac{\sigma^2/\psi^2}{m} + o(m^{-1}).$$

It follows that the sample variance of the shrunken $\vartheta_1, \ldots, \vartheta_n$ has a bias that is of the same order as that in the sample variance of $\vartheta_1, \ldots, \vartheta_n$. Interestingly, note that, here, this estimator overcorrects for the presence of noise, and so will be underestimating the true variance, ψ^2, on average.

The upper two plots in Figures 1–3 provide simulation results for the distribution function F for the same Monte Carlo designs (see the online version of the paper for color figures). The figures deal with the sample sizes $(50, 3)$, $(100, 4)$, $(200, 5)$, respectively. The left plots contain (the average over the Monte Carlo replications of) the analytically bias-corrected estimator (solid blue line), with the bandwidth chosen according to a cross-validation procedure, together with 95% confidence bands placed around in. Each of the plots also provides the average of the naive plug-in estimator (dashed red line), the empirical distribution of the Empirical-Bayes point estimates (dash-dotted purple line), and the actual standard-normal distribution that is being estimated (solid black line).[5] The upper right plots in

[5]Empirical Bayes was implemented nonparametrically (and correctly assuming homoskedasticity) based on the formula stated in the Introduction using a kernel estimator and the optimal bandwidth that assumes knowledge of the normality of the target distribution.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

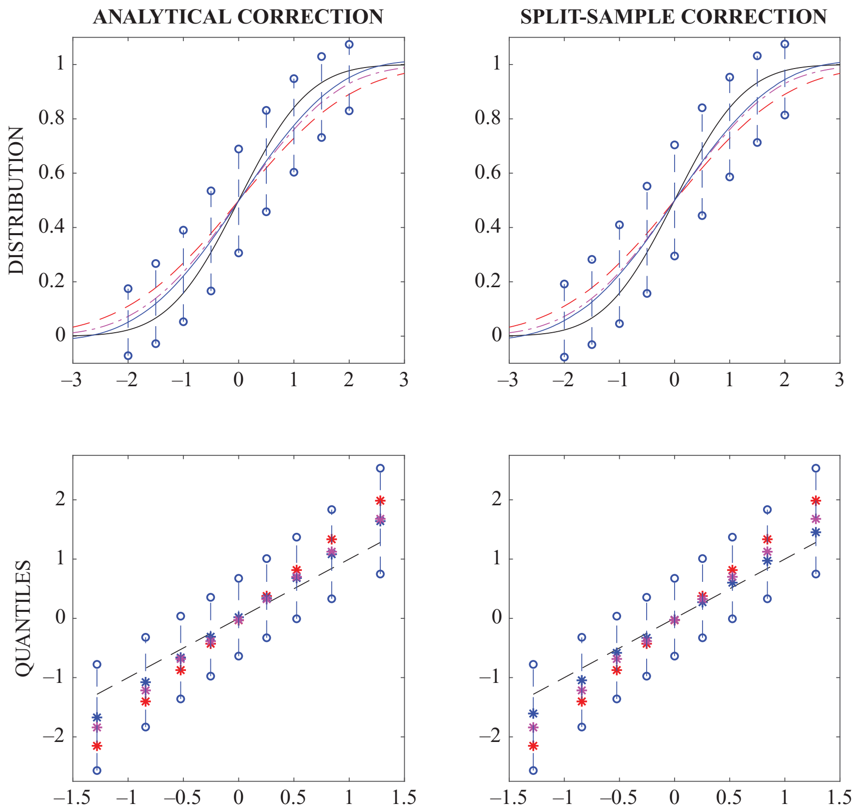

FIGURE 1. Estimation of F and q in the James–Stein problem from 50×3 panel data. *Notes:* The upper plots contain the average (over the Monte Carlo replications) distribution function (full blue line) obtained via analytical bias correction (left plot) and split-sample jackknife estimation (right plot) along with 95% confidence intervals around them at each of the quantiles of F (blue $\circ$). Each plot also contains the true curve (full black line) and the average of the empirical distribution function of the estimated θ_i (dashed red line) and of their Empirical-Bayes adjustments (dash-dotted purple line). The lower plots contain corresponding QQ-plots of the average bias-corrected quantile function (blue $*$) at each of the deciles of F together with 95% confidence intervals. The 45° line (dashed black line) corresponds to the truth. Average estimates for the naive (red $*$) and Empirical-Bayes (purple $*$) estimators are equally pictured.

Figures 1–3 have the same structure, and only now the bias-corrected estimator being plotted is the split-sample jackknife.

The simulations clearly show the substantial bias in the naive estimator. This bias becomes more pronounced relative to its standard error as the sample size grows and, indeed, $\hat{F}$ starts falling outside of the confidence bands (of the bias corrected estimator) as the sample size increases. The Empirical-Bayes estimator

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

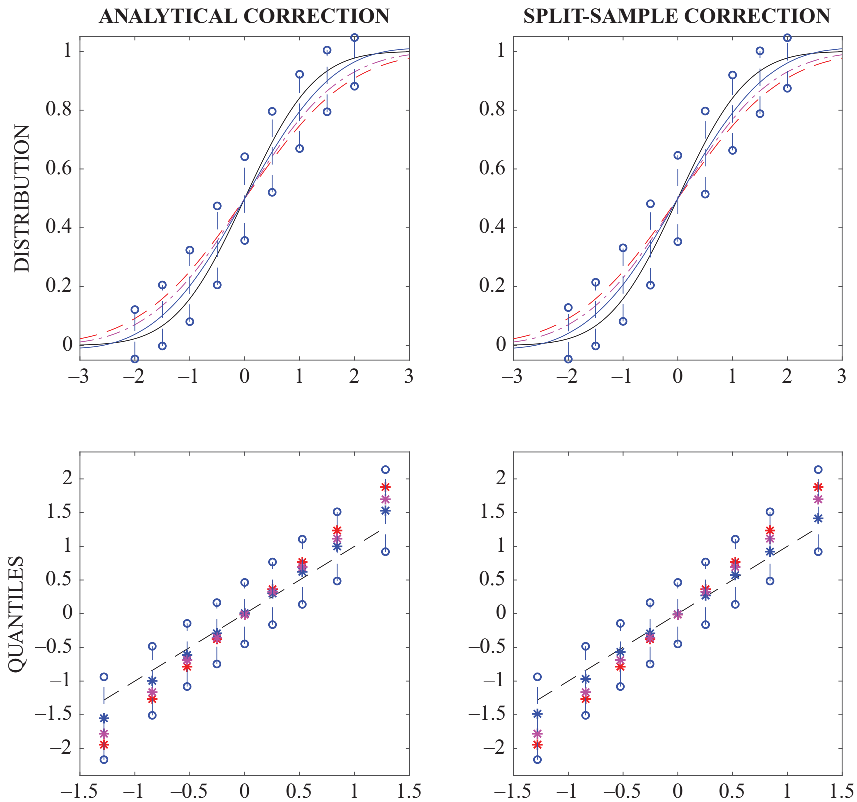

FIGURE 2. Estimation of F and q in the James–Stein problem from 100×4 panel data. *Notes:* The upper plots contain the average (over the Monte Carlo replications) distribution function (full blue line) obtained via analytical bias correction (left plot) and split-sample jackknife estimation (right plot) along with 95% confidence intervals around them at each of the quantiles of F (blue $\circ$). Each plot also contains the true curve (full black line) and the average of the empirical distribution function of the estimated θ_i (dashed red line) and of their Empirical-Bayes adjustments (dash-dotted purple line). The lower plots contain corresponding QQ-plots of the average bias-corrected quantile function (blue $*$) at each of the deciles of F together with 95% confidence intervals. The 45° line (dashed black line) corresponds to the truth. Average estimates for the naive (red $*$) and Empirical-Bayes (purple $*$) estimators are equally pictured.

is less biased than $\hat{F}$. However, its bias is of the same order and so, as the sample size grows, it does not move toward F but, rather, toward $\hat{F}$.[6] The confidence bands of $\check{F}$ and $\tilde{F}$ settle around F as the sample grows. The results also show near

[6]Recall that the Empirical-Bayes estimator is not designed for inference on F but, instead, aims to minimize risk in estimating $\theta_1, \ldots, \theta_n$. In terms of RMSE, it dominates $\vartheta_1, \ldots, \vartheta_n$. For the three sample sizes considered here, the RMSEs are 1.667, 1.246, and 1.000 for the plug-in estimators and 1.233, 1.018, 0.874 for Empirical Bayes.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

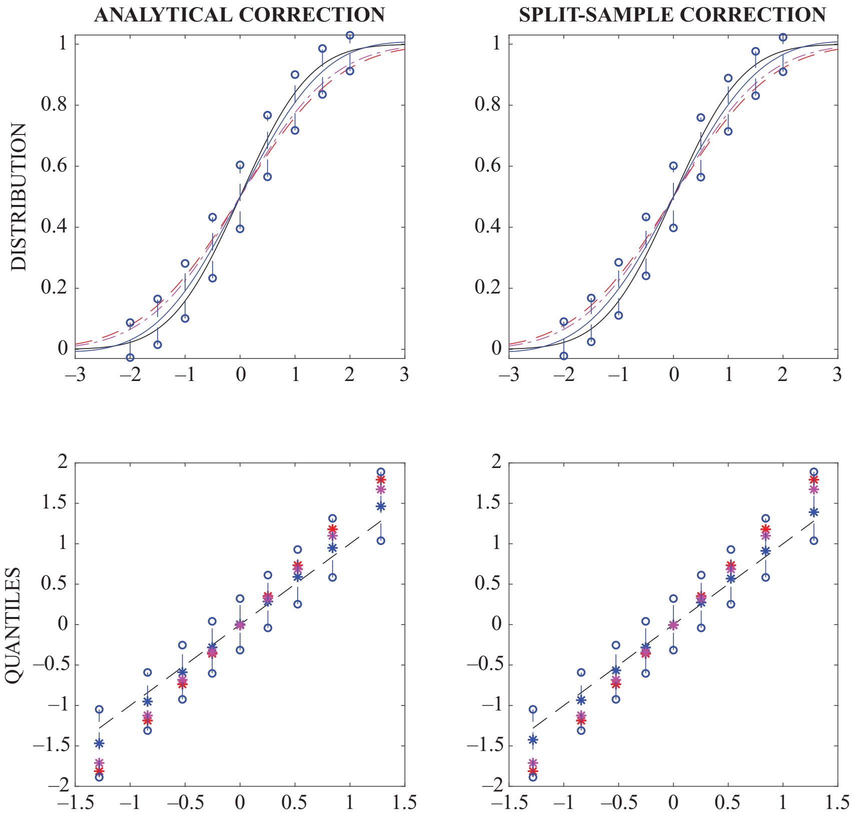

Figure 3. Estimation of F and q in the James–Stein problem from 200×5 panel data. *Notes:* The upper plots contain the average (over the Monte Carlo replications) distribution function (full blue line) obtained via analytical bias correction (left plot) and split-sample jackknife estimation (right plot) along with 95% confidence intervals around them at each of the quantiles of F (blue $\circ$). Each plot also contains the true curve (full black line) and the average of the empirical distribution function of the estimated θ_i (dashed red line) and of their Empirical-Bayes adjustments (dash-dotted purple line). The lower plots contain corresponding QQ-plots of the average bias-corrected quantile function (blue $*$) at each of the deciles of F together with 95% confidence intervals. The 45° line (dashed black line) corresponds to the truth. Average estimates for the naive (red $*$) and Empirical-Bayes (purple $*$) estimators are equally pictured.

identical performance of the split-sample approach and the analytical approach based on our bias formula. Indeed, the curves in the left and right plots are virtually indistinguishable.

The reduction in bias in our estimators of F is again sufficient to bring the empirical size of tests in line with their nominal size. To see this, Table 2 provides empirical rejection frequencies of two-sided tests at the 5% level for F at each of its

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

TABLE 2. Inference on F in the James–Stein problem from $n \times m$ panel data

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9
					$(n,m) = (50,3)$				
$\hat{F}$	0.4814	0.5518	0.3695	0.1530	0.0681	0.1598	0.3801	0.5610	0.4828
$\check{F}$	0.0600	0.0928	0.1039	0.0785	0.0563	0.0745	0.1029	0.0891	0.0628
					$(n,m) = (100,4)$				
$\hat{F}$	0.6962	0.7304	0.5564	0.2280	0.0566	0.2312	0.5586	0.7352	0.7034
$\check{F}$	0.0608	0.0848	0.0920	0.0664	0.0494	0.0734	0.0932	0.0782	0.0532
					$(n,m) = (200,5)$				
$\hat{F}$	0.926	0.902	0.7634	0.3288	0.0576	0.3212	0.7646	0.903	0.9146
$\check{F}$	0.0536	0.0828	0.0996	0.0770	0.0496	0.0792	0.0978	0.0780	0.0554

Notes: $\hat{F}$ is the empirical distribution of the ϑ_i. $\check{F}$ is its (analytically) bias-corrected version constructed using estimators of the variance of the noise distributions. The table provides, for several combinations of (n,m), rejection frequencies of the associated two-sided tests of the null that $F(\Phi^{-1}(\tau)) = \tau$ for a range of different quantiles τ; the data were generated with F set to the standard-normal distribution function.

deciles using both $\hat{F}$ and $\check{F}$. The rejection frequencies based on the naive estimator are much too high for all sample sizes and deciles and get worse as the sample gets larger. Empirical size is much closer to nominal size after adjusting for noise, and this improvement is observed at all deciles of the distribution.

The lower two plots in Figures 1–3 provide corresponding simulation results for estimators of the deciles of F. The presentation is constructed around a QQ-plot of the standard normal, pictured as the dash-dotted black line in each plot. Along the QQ-plot, the averages (over the Monte Carlo replications) of the naive estimator (red), Empirical Bayes (purple), and the bias-corrected quantiles (blue) are shown by $*$ symbols. Again, the left plots deal with the analytical correction and the right plots show results for the split-sample approach. Confidence intervals around the corrected estimators (in blue, -o) are also again provided. Like the naive estimator, the Empirical-Bayes estimators reported are the appropriate order statistics of $\vartheta_1, \ldots, \vartheta_n$, after shrinkage has been applied to each. Visual inspection reveals that the results are in line with those obtained for the distribution function. As the sample size grows, only $\check{q}$ successfully adjusts for bias arising from estimation noise in $\vartheta_1, \ldots, \vartheta_n$. Here, the split-sample correction is slightly more effective than our analytical approach.

4.2. Empirical illustration

We use quarterly panel data on a set of 48 consumer price index items in 52 U.S. cities. The data span the period of 1990–2007, yielding 72 time series observations. They were used by Parsley and Wei (2001), Crucini, Shintani, and Tsuruga (2015),

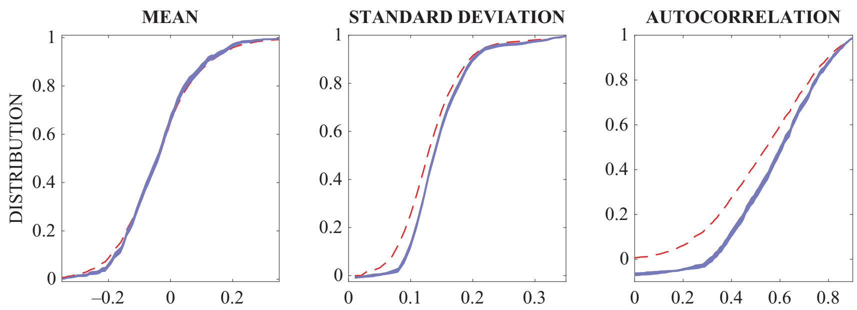

FIGURE 4. Deviations from the law of one price. *Notes:* The empirical distribution functions of the means (left), standard deviations (middle), and autocorrelations (right) of the time series of $x_{cit} = \log(p_{cit}) - \log(p_{1it})$ for all city/item combinations (dashed red line) along with 95% confidence bands constructed from the split-sample jackknife estimator of each of these distributions (shaded blue region).

and Okui and Yanagi (2019, 2020) to investigate the cross-sectional heterogeneity in deviations from the law of one price. Let p_{cit} be the price of item i in city c at time t and define the random variable

$$x_{cit} = \log\left(\frac{p_{cit}}{p_{1it}}\right) = \log(p_{cit}) - \log(p_{1it})$$

for all $(52-1) \times 48 = 2,448$ city/item combinations apart from the reference city (which here is Albuquerque, New Mexico). For each city/item combination, we estimate the mean, standard deviation, and first-order autocorrelation of x_{cit} nonparametrically from the time dimension of our panel. Our interest lies in the distribution functions of their population counterparts. We estimate these three distributions by the empirical distributions of the cross-sectional estimates, and then correct for plug-in bias via the split-sample jackknife procedure. Our results complement the analysis of Okui and Yanagi (2020, Figure 1), which gives corresponding estimates of the associated density functions.

The results are collected in Figure 4. The plots contain the empirical distribution functions (dashed red lines) together with 95% confidence bands based on the split-sample jackknife (shaded blue regions). The correction for regression to the mean to the empirical distribution is clearly visible for the mean (left plot). It is also statistically significant, with the tails of the empirical distribution falling out of the confidence region. The sample standard deviation and autocorrelation obtained from the time series are biased estimators and so the empirical distribution function for these parameters (middle and right plots, respectively) suffer from an additional bias that is of the same order of magnitude as is the bias due to estimation noise (see the discussion on Footnote 4). The split-sample jackknife corrects for both these sources of bias automatically. Here, the bias adjustment leads to a pronounced shift of the empirical distribution; the corrected distribution functions all but

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

stochastically dominate the naive plug-in estimators. The differences between the corrected and uncorrected functions are quantitatively large, and, given the small standard error, they are also statistically significant.

5. CONCLUSIONS

In this paper, we have considered inference on the distribution of latent variables from noisy measurements. In an asymptotic embedding where the variance of the noise shrinks with the sample size, we have derived the leading bias in the empirical distribution function of the noisy measurements and suggested both an analytical and a jackknife correction. They provide a simple and numerically stable (approximate) solution to a generalized deconvolution problem that, in addition, yields valid inference procedures. The split-sample jackknife is particularly straightforward to implement, and we recommend its use whenever possible.

APPENDIX A:

Notational Convention: We let $\nabla_p^q \varphi$ denote the qth derivative of φ with respect to its pth argument. We omit the subscript for univariate φ.

Proof of Proposition 1

The following result is useful to prove Proposition 1.

LEMMA A.1 (Komlós, Major, and Tusnády, 1975). *Let $\mathbb{G}_n$ denote the empirical cumulative distribution of an i.i.d. sample of size n from a uniform distribution on* [0,1]. *Let $\mathbb{B}_n$ denote a sequence of Brownian bridges. Then,*

$$\sup_{u\in[0,1]} \left|\sqrt{n}\,(\mathbb{G}_n(u)-u)-\mathbb{B}_n(u)\right| = O_p(\log(n)/\sqrt{n}).$$

Proof of Proposition 1. We begin with the bias calculation. Suppose, first, that Part A of Assumption 2 holds. Then, (θ_i,σ_i) have a joint density, $h(\theta_i,\sigma_i)$, say. We will denote the marginal density of σ_i by $h(\sigma_i)$ and the conditional density of θ_i given σ_i by $h(\theta_i|\sigma_i)$. For any real number δ, let

$$G(\theta,\delta) := E(1\{\theta_i+\delta\sigma_i \leq \theta\}) = \int_{\underline{\sigma}}^{\overline{\sigma}} \int_{-\infty}^{\theta-\delta\sigma} h(\vartheta,\sigma)\, d\vartheta\, d\sigma.$$

Note that $G(\theta,0)=F(\theta)$ and that

$$E(\hat{F}(\theta)) = E(1\{\vartheta_i \leq \theta\}) = E\left(1\left\{\theta_i + \frac{\varepsilon_i}{\sqrt{m}}\sigma_i \leq \theta\right\}\right) = E\left(G(\theta,\varepsilon_i/\sqrt{m})\right). \tag{A.1}$$

Assumption 2 implies that G is smooth and differentiable in its second argument. By definition of the function $e(\sigma_i)$,

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

$$\sup_{\theta}\sup_{\delta}|\nabla_2^3 G(\theta,\delta)| = \sup_{\theta}\sup_{\delta}\left|\int_{\underline{\sigma}}^{\overline{\sigma}}\sigma^3\,\nabla_1^2 h(\theta-\delta\sigma|\sigma)\,h(\sigma)\,d\sigma\right| \leq \int_{\underline{\sigma}}^{\overline{\sigma}}\sigma^3\,e(\sigma)\,h(\sigma)\,d\sigma, \tag{A.2}$$

which equals $E(\sigma_i^3 e(\sigma_i))$ and is finite by assumption. Therefore, by (A.1) and a third-order expansion of $G(\theta,\varepsilon_i/\sqrt{m})$ in its second argument around zero, we find that

$$E(\hat{F}(\theta)) = F(\theta) + \frac{1}{2}\frac{\nabla_2^2 G(\theta,0)}{m} + \frac{1}{6}\frac{E(\varepsilon_i^3\,\nabla_2^3 G(\theta,\varepsilon_i^*/\sqrt{m}))}{m^{3/2}} = \frac{1}{2}\frac{\nabla_2^2 G(\theta,0)}{m} + O(m^{-3/2}),$$

where ε_i^* is some value between zero and ε_i, and where, in addition to (A.2), we have used that $E(\varepsilon_i)=0$ and $E(\varepsilon_i^2)=1$ by construction and that $E(|\varepsilon_i|^3)<\infty$ by assumption. By direct calculation,

$$\nabla_2^2 G(\theta,0) = 2\,b_F(\theta).$$

Therefore,

$$E(\hat{F}(\theta)) = F(\theta) + \frac{b_F(\theta)}{m} + O(m^{-3/2}),$$

as claimed.

Suppose, next, that Part B of Assumption 2 holds. Then, we have a deterministic relationship between θ_i and σ_i. We may define $G(\theta,\delta)$ as above but have to take care when Taylor expanding in δ, as the function may be noncontinuous. A noncontinuity occurs whenever the number of solutions t (on the real line) to the equation $t+\delta\sigma(t)=\theta$ changes. However, at $\delta=0$, the only solution to this equation is $t=\theta$, and because we assume that the function $\sigma(\theta)$ has uniformly bounded derivative σ', there always exists $\eta>0$ such that, for all $\delta\in(-\eta,\eta)$ and all real θ, the equation $t+\delta\sigma(t)=\theta$ has a unique solution in t on the real line. We denote this solution by $t^*(\theta,\delta)$, that is, we have $t^*(\theta,\delta)+\delta\sigma(t^*(\theta,\delta))=\theta$. Using this, we find that, for $\delta\in(-\eta,\eta)$, we have

$$G(\theta,\delta) = F(t^*(\theta,\delta)), \quad \nabla_2^1 t^*(\theta,\delta) = -\frac{\sigma(t^*(\theta,\delta))}{1+\delta\,\sigma'(t^*(\theta,\delta))},$$

where the last equation is obtained by taking derivatives of $t^*(\theta,\delta)+\delta\sigma(t^*(\theta,\delta))=\theta$ with respect to δ and then solving for the derivative. Because we have that $t^*(\theta,0)=\theta$ we then find

$$G(\theta,0) = F(\theta), \quad \nabla_2^1 G(\theta,0) = -\sigma(\theta)f(\theta), \quad \nabla_2^2 G(\theta,0) = 2b_F(\theta).$$

Differentiating further, we see that $\nabla_2^3 G(\theta,0)$, and $\nabla_2^4 G(\theta,0)$ are functions of the derivatives of f and σ up to third order. Our assumption that these derivatives are uniformly bounded implies that

$$\sup_{\theta}\sup_{\delta\in[-\eta,\eta]}\left|\nabla_2^4 G(\theta,\delta)\right| < \infty. \tag{A.3}$$

The only obstacle that now prevents us from proceeding with an expansion as we did under Assumption 2.A is that the bound (A.3) is restricted to a neighborhood around zero. To complete the derivation of the bias, we argue that the restriction that $\delta\in(-\eta,\eta)$ relaxes

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

sufficiently fast as m grows. We do so as follows. Note, first, that, by Markov's inequality,

$$P(|\varepsilon_i| > \eta\sqrt{m}) \leq m^{-2}\frac{E(\varepsilon_i^4)}{\eta^4} = O(m^{-2}).$$

Then,

$$\begin{aligned}E(\hat{F}(\theta)) &= E\left(G(\theta,\varepsilon/\sqrt{m})\right)\\ &= E\left(\{|\varepsilon_i| \leq \eta\sqrt{m}\}\, G(\theta,\varepsilon/\sqrt{m})\right) + E\left(\{|\varepsilon_i| > \eta\sqrt{m}\}\, G(\theta,\varepsilon/\sqrt{m})\right)\\ &= E\left(\{|\varepsilon_i| \leq \eta\sqrt{m}\}\, G(\theta,\varepsilon/\sqrt{m})\right) + O(m^{-2}),\end{aligned}$$

uniformly in θ, because

$$\sup_\theta E(\{|\varepsilon_i| > \eta\,\sqrt{m}\}\, G(\theta,\varepsilon_i/\sqrt{m})) \leq P(|\varepsilon_i| > \eta\,\sqrt{m}) = O(m^{-2}),$$

noting that $\sup_\theta \sup_\delta G(\theta,\delta) \leq 1$ by definition of the function G. Next, a Taylor expansion of G around $\delta = 0$ gives

$$E(\hat{F}(\theta)) = E(G(\theta,\varepsilon_i/\sqrt{m})) = F(\theta) + \frac{1}{2}\frac{\nabla_2^2 G(\theta,0)}{m} + \frac{1}{6}\frac{\nabla_2^3 G(\theta,0)}{m^{3/2}} + R(\theta) + O(m^{-2}),$$

where we have used that $F(\theta) = G(\theta,0)$, that $E(\varepsilon_i) = 0$, and that $E(\varepsilon_i^2) = 1$, and have introduced the notational shorthand

$$R(\theta) := R_2(\theta) - R_1(\theta)$$

for

$$\begin{aligned}R_1(\theta) :=\; & P(|\varepsilon_i| > \eta\,\sqrt{m})\, F(\theta)\\ +\; & E(\{|\varepsilon_i| > \eta\,\sqrt{m}\}\,\varepsilon_i)\,\frac{\nabla_2^1 G(\theta,0)}{\sqrt{m}}\\ +\; & \frac{1}{2}E(\{|\varepsilon_i| > \eta\,\sqrt{m}\}\,\varepsilon_i^2)\,\frac{\nabla_2^2 G(\theta,0)}{m}\\ +\; & \frac{1}{6}E(\{|\varepsilon_i| > \eta\,\sqrt{m}\}\,\varepsilon_i^3)\,\frac{\nabla_3^2 G(\theta,0)}{m^{3/2}}\end{aligned}$$

and

$$R_2(\theta) := \frac{1}{24}\frac{E(\{|\varepsilon_i| \leq \eta\,\sqrt{m}\}\,\varepsilon_i^4\,\nabla_2^4 G(\theta,\varepsilon_i^*/\sqrt{m}))}{m^2};$$

here, ε_i^* lies in between zero and ε_i. To validate our bias expression, it remains only to establish that $\sup_\theta |R(\theta)| = O(m^{-3/2})$. To do so, we show that $\sup_\theta |R_1(\theta)| = O(m^{-2})$, and that $\sup_\theta |R_2(\theta)| = O(m^{-2})$, in turn. By Hölder's inequality,

$$|E(\{|\varepsilon_i| > \eta\,\sqrt{m}\}\,\varepsilon_i)| \leq E(\{|\varepsilon_i| > \eta\,\sqrt{m}\})^{3/4} E(\varepsilon_i^4)^{1/4} = O(P(|\varepsilon_i| > \eta\,\sqrt{m})^{3/4}) = O(m^{-3/2}).$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

In the same way,

$$|E(\{|\varepsilon_i| > \eta\sqrt{m}\}\varepsilon_i^2) = O(m^{-1}), \qquad |E(\{|\varepsilon_i| > \eta\sqrt{m}\}\varepsilon_i^3) = O(m^{-1/2}),$$

follow. Consequently,

$$\sup_\theta |R_1(\theta)| = O(m^{-2}) \sup_\theta (1 + \nabla_2^1 G(\theta,0) + \nabla_2^2 G(\theta,0) + \nabla_2^3 G(\theta,0)) = O(m^{-2}),$$

using that all relevant derivatives on the right-hand side are bounded. Next, noting that, as $|\varepsilon_i^*| \leq |\varepsilon_i|$, the event $|\varepsilon_i|/\sqrt{m} \leq \eta$ implies that $|\varepsilon_i^*|/\sqrt{m} \leq \eta$, we have

$$\begin{aligned}
\sup_\theta |R_2(\theta)| &= \frac{1}{24}\frac{\sup_\theta E(\{|\varepsilon_i|/\sqrt{m} \leq \eta\}\varepsilon_i^4 \nabla_2^4 G(\theta, \varepsilon_i^*/\sqrt{m}))}{m^2} \\
&\leq \frac{1}{24}\frac{\sup_\theta \sup_{\delta\in[-\eta,\eta]} |\nabla_2^4 G(\theta,\delta)|\, E(\varepsilon_i^4)}{m^2} \\
&= O(m^{-2}),
\end{aligned}$$

because of (A.3). Therefore, $\sup_\theta |R(\theta)| = O(m^{-2})$, and so

$$E(\hat{F}(\theta)) = F(\theta) + \frac{b_F(\theta)}{m} + O(m^{-3/2}),$$

as before.

Now, turning to the result on the covariance, note that

$$\operatorname{cov}(\hat{F}(\theta_1), \hat{F}(\theta_2)) = \frac{E(\hat{F}(\theta_1 \wedge \theta_2)) - E(\hat{F}(\theta_1))\, E(\hat{F}(\theta_2))}{n}$$

depends only on $E(\hat{F}(\theta))$ which, up to $O(m^{-3/2})$ and uniformly in θ, has been calculated above. Moreover,

$$\begin{aligned}
\operatorname{cov}(\hat{F}(\theta_1), \hat{F}(\theta_2)) &= \frac{\Big(F(\theta_1 \wedge \theta_2) + O(m^{-1})\Big) - \Big(F(\theta_1) + O(m^{-1})\Big)\Big(F(\theta_2) + O(m^{-1})\Big)}{n} \\
&= \frac{F(\theta_1 \wedge \theta_2) - F(\theta_1)F(\theta_2)}{n} + O(n^{-1}m^{-1}) \\
&= \frac{\sigma_F(\theta_1, \theta_2)}{n} + O(n^{-1}m^{-1}),
\end{aligned}$$

as stated in the proposition.

To complete the proof, it remains only to verify the limit distribution of the scaled empirical distribution function. Let $F_m(\theta) := E(1\{\vartheta_i \leq \theta\})$, the distribution function of ϑ_i. Our assumptions imply that F_m is continuous and that it has no mass points. With $u_i := F_m(\vartheta_i)$, we therefore have that u_i is i.i.d. uniformly distributed on $[0,1]$ by the probability integral transform. An application of Lemma A.1 with $u = F_m(\theta)$ and exploiting monotonicity of distribution functions then gives

$$\sup_\theta \left| \sqrt{n}(\hat{F}(\theta) - F_m(\theta)) - \mathbb{B}_n(F_m(\theta)) \right| = O_p(\log(n)/\sqrt{n}).$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

We have already shown that, uniformly in θ,

$$F_m(\theta) = F(\theta) + \frac{b_F(\theta)}{m} + O(m^{-3/2}).$$

Therefore, using that $n/m^3 \to 0$ if $n/m^2 \to c \in [0, +\infty)$ as $n, m \to \infty$,

$$\sqrt{n}(\hat{F}(\theta) - F_m(\theta)) = \sqrt{n}\left(\hat{F}(\theta) - F(\theta) - \frac{b_F(\theta)}{m}\right) + o(1),$$

holds uniformly in θ. Furthermore, our bias calculation implies that $F_m(\theta) - F(\theta)$ converges to zero uniformly in θ as $m \to 0$, so that applying Lévy's modulus-of-continuity theorem, that is,

$$\lim_{\epsilon \to 0} \sup_{t \in [0, 1-\epsilon]} \frac{|\mathbb{B}_n(t) - \mathbb{B}_n(t+\epsilon)|}{\sqrt{\epsilon \log(1/\epsilon)}} = O(1), \qquad \epsilon > 0,$$

to our problem yields $\sup_\theta |\mathbb{B}_n(F_m(\theta)) - \mathbb{B}_n(F(\theta))| \xrightarrow{p} 0$ as $m \to \infty$. We thus have that $\mathbb{B}_n(F_m(\theta)) \rightsquigarrow \mathbb{B}_n(F(\theta))$. Putting everything together and noting that, by definition, $\mathbb{B}_n(F(\theta)) = \mathbb{G}_F(\theta)$, we obtain

$$\sup_\theta \left| \sqrt{n}\left(\hat{F}(\theta) - F(\theta) - \frac{b_F(\theta)}{m}\right) - \mathbb{G}_F(\theta) \right| = o_p(1),$$

which completes the proof of the proposition. □

Proof of Proposition 2

LEMMA A.2. *Let Assumptions 1 and 2 hold. Let f_m denote the density function of ϑ_i. Then:*

(i) $\sup_\theta |f_m(\theta) - f(\theta)| = O(m^{-1})$,
(ii) $\sup_\theta |\nabla^1 f_m(\theta) - \nabla^1 f(\theta)| = O(m^{-1})$,
(iii) $\sup_\theta |\nabla^2 f_m(\theta) - \nabla^2 f(\theta)| = O(1)$,
(iv) $\sup_\theta |\nabla^3 f_m(\theta) - \nabla^3 f(\theta)| = O(1)$.

Proof. From the argument in the proof of Proposition 1, we have

$$F_m(\theta) - F(\theta) = \frac{1}{2} \frac{E(\varepsilon_i^2 H(\theta, \varepsilon_i^*/\sqrt{m}))}{m}$$

by a second-order expansion, where ε_i^* is a value between zero and ε_i and we introduce the function

$$H(\theta, \delta) := \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma^2 \nabla_1^1 h(\theta - \delta\sigma | \sigma) \, h(\sigma) \, d\sigma,$$

where $h(\theta_i|\sigma_i)$ and $h(\sigma_i)$ are the density functions of θ_i given σ_i and of σ_i, respectively. Differentiating with respect to θ yields the first conclusion of the lemma as

$$\sup_\theta |f_m(\theta) - f(\theta)| = \sup_\theta \left| \frac{1}{2} \frac{E(\varepsilon_i^2 \nabla_1^1 H(\theta, \varepsilon_i^*/\sqrt{m}))}{m} \right| \leq \frac{E(\sigma_i^2)}{m} \frac{\sup_\theta \sup_\delta |\nabla_1^1 H(\theta, \delta)|}{2} = O(m^{-1}),$$

which follows from the inequality

$$\sup_\theta \sup_\delta |\nabla_1^1 H(\theta,\delta)| = \sup_\theta \sup_\delta \left| \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma^3 \, \nabla_1^2 h(\theta - \delta\sigma|\sigma) \, h(\sigma) \, d\sigma \right| \leq \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma^3 \, e(\sigma) \, h(\sigma) \, d\sigma < \infty$$

and the definition of the function $e(\sigma)$ in Assumption 2. The second conclusion of the lemma follows in the same manner, differentiating once more. Finally, the third and fourth conclusions are obtained similarly. The point of departure is now the following identity, which is derived in the proof of Proposition 1,

$$F_m(\theta) = E\left(G(\theta, \varepsilon_i^*/\sqrt{m})\right),$$

where

$$G(\theta,\delta) := \int_{\underline{\sigma}}^{\overline{\sigma}} \int_{-\infty}^{\theta-\delta\sigma} h(\vartheta|\sigma) \, h(\sigma) \, d\vartheta \, d\sigma.$$

Repeated differentiation shows that

$$\sup_\theta \sup_\delta |\nabla_1^3 G(\theta,\delta)| = \sup_\theta \sup_\delta |\int_{\underline{\sigma}}^{\overline{\sigma}} \nabla_1^2 h(\theta - \delta\sigma|\sigma) \, h(\sigma) \, d\sigma| \leq |\int_{\underline{\sigma}}^{\overline{\sigma}} e(\sigma) \, h(\sigma) \, d\sigma| < \infty,$$
$$\sup_\theta \sup_\delta |\nabla_1^4 G(\theta,\delta)| = \sup_\theta \sup_\delta |\int_{\underline{\sigma}}^{\overline{\sigma}} \nabla_1^3 h(\theta - \delta\sigma|\sigma) \, h(\sigma) \, d\sigma| \leq |\int_{\underline{\sigma}}^{\overline{\sigma}} e(\sigma) \, h(\sigma) \, d\sigma| < \infty,$$

and so $\sup_\theta |\nabla^3 F_m(\theta)| = O(1)$ and $\sup_\theta |\nabla^4 F_m(\theta)| = O(1)$ follow. Furthermore,

$$\sup_\theta |\nabla^2 f_m(\theta) - \nabla^2 f(\theta)| \leq \sup_\theta |\nabla^2 f_m(\theta)| + \sup_\theta |\nabla^2 f(\theta)| = O(1),$$
$$\sup_\theta |\nabla^3 f_m(\theta) - \nabla^3 f(\theta)| \leq \sup_\theta |\nabla^3 f_m(\theta)| + \sup_\theta |\nabla^3 f(\theta)| = O(1),$$

follows because f has uniformly bounded derivatives up to third order by assumption. This completes the proof. □

Proof of Proposition 2. The ϑ_i are i.i.d. draws from the distribution F_m which according to Lemma A.2 has nondegenerate density f_m, that is, the ϑ_i are continuously distributed. Thus,

$$u_{(k)} := F_m(\vartheta_{(k)})$$

is the kth-order statistic of a uniform sample. We set $k = \lceil \tau n \rceil$ for the rest of the proof. Then, $\hat{q}(\tau) = \vartheta_{(k)}$. Since $k/n \to \tau$ by construction, it is well known that

$$\sqrt{n}(u_{(k)} - \tau) \xrightarrow{d} N(0, \tau(1-\tau)). \tag{A.4}$$

Let $q_m(\tau) := F_m^{-1}(\tau)$, the τth-quantile of F_m. By expanding the function F_m^{-1} around τ, we find that

$$\hat{q}(\tau) = F_m^{-1}(u_{(k)}) = q_m(\tau) + \frac{u_{(k)} - \tau}{f_m(q_m(\tau))} + r_{(k)}$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

for the remainder term

$$r_{(k)} := -\frac{f_m'(\xi_{(k)})}{f_m(\xi_{(k)})^3}\left(u_{(k)} - \tau\right)^2,$$

where $\xi_{(k)}$ is a value between $F_m^{-1}(\tau)$ and $F_m^{-1}(u_{(k)})$. From (A.4), we have $u_{(k)} - \tau = O_P(n^{-1/2})$. This implies that $\xi_{(k)} \xrightarrow{p} \tau$. Using Lemma A.2, we may conclude that $f_m(\xi_{(k)}) \xrightarrow{p} f_m(\tau) \to f(\tau) > 0$, and, therefore, that $r_{(k)} = O_p(n^{-1})$. We thus have

$$\hat{q}(\tau) = q_m(\tau) + \frac{u_{(k)} - \tau}{f_m(q_m(\tau))} + O_p(n^{-1}).$$

Again using Lemma A.2 and our assumption that $f(\theta) > 0$ in a neighborhood of $q(\tau) = F^{-1}(\tau)$, we have $f_m(q_m(\tau))^{-1} = f(q(\tau))^{-1} + O(m^{-1})$, and therefore

$$\hat{q}(\tau) = q_m(\tau) + \frac{u_{(k)} - \tau}{f(q(\tau))} + O_p(n^{-1} + n^{-1/2}m^{-1}). \tag{A.5}$$

From Proposition 1, we know $F_m(\theta) = E(\hat{F}(\theta)) = F(\theta) + b_F(\theta)/m + O(m^{-3/2})$, and therefore

$$q_m(\tau) = q(\tau) - \frac{b_F(q(\tau))/f(q(\tau))}{m} + O(m^{-3/2}). \tag{A.6}$$

Combining (A.4)–(A.6) gives the statement of the theorem. □

Proof of Proposition 3

LEMMA A.3. *Let the assumptions of Proposition 3 hold. Then:*

(i) $\sup_\theta E(\hat{b}_F(\theta) - b_F(\theta)) = O(m^{-1}) + O(h^2)$,

(ii) $\sup_\theta \mathrm{var}(\hat{b}_F(\theta)) = O(n^{-1}h^3)$,

(iii) $\sup_\theta (1 + |\theta|^{1+\eta})\,|\nabla^1 \hat{b}_F(\theta) - \nabla^1 b_F(\theta)| = O_p(h^{-(\omega+1)/\omega})$.

LEMMA A.4. *Let Assumption 1 hold and define*

$$b_i(\theta) := -\frac{\sigma_i^2}{h^2}\frac{\phi'\left(\frac{\vartheta_i - \theta}{h}\right)}{2}.$$

If f is bounded, then, for any $\epsilon > 0$,

$$\sup_\theta E(|b_i(\theta) - E(b_i(\theta))|^\epsilon)^{1/\epsilon} = O(h^{-2+\epsilon^{-1}}).$$

The proofs of those two lemmas are provided below, after the proof of the main text results.

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Proof of Proposition 3. We first show that

$$\sup_{\theta\in\mathbb{R}}\left|\hat{b}_F(\theta)-b_F(\theta)\right| = O(m^{-1})+O(h^2)+O(n^{-1/2}\,h^{-3/2-\varepsilon}).$$

The result of the proposition then follows readily. For a finite ν, introduce the function

$$t(\theta) := \operatorname{sgn}(\theta)\,\frac{1-(1+|\theta|)^{-\nu}}{\nu}.$$

Note that t maps to the finite interval $(-\nu^{-1},\nu^{-1})$ and is monotone increasing; moreover, $\nabla^1 t(\theta) = (1+|\theta|)^{-(1+\nu)}$. Now, consider the reparameterization $\tau = t(\theta)$; note that τ lives in a bounded interval. From Lemma A.3(iii), using the chain rule of differentiation, it follows that

$$\sup_{\tau\in(-\nu^{-1},\nu^{-1})}\left|\nabla^1_\tau\hat{b}_F(t^{-1}(\tau))-\nabla^1_\tau b_F(t^{-1}(\tau))\right| = O_p(h^{-(1+\omega^{-1})}), \tag{A.7}$$

where we use the notation ∇_τ to indicate derivatives with respect to τ. We therefore have that $\hat{b}_F(t^{-1}(\tau))-b_F(t^{-1}(\tau))$, as a function τ, has a uniformly bounded Lipschitz constant. Now, let I_h be a partition of $(-\nu,-\nu^{-1})$ with subintervals that are (approximately) of length $l_h := h^{3-\omega^{-1}}$. Then, (A.7) implies that

$$\sup_{\theta}|\hat{b}_F(\theta)-b_F(\theta)| = \sup_{\tau\in(-\nu,\nu)}|\hat{b}_F(t^{-1}(\tau))-b_F(t^{-1}(\tau))|$$

is equal to

$$\max_{\tau\in I_h}|\hat{b}_F(t^{-1}(\tau))-b_F(t^{-1}(\tau))|+O_p(h^2). \tag{A.8}$$

Here, the order of the remainder terms follows from the choice of l_h. Now, introduce the shorthand

$$\hat{\Delta}(\theta) := \hat{b}_F(\theta)-E(\hat{b}_F(\theta)).$$

Then,

$$\max_{\tau\in I_h}|\hat{b}_F(t^{-1}(\tau))-b_F(t^{-1}(\tau))| \le \max_{\tau\in I_h}|\hat{\Delta}(t^{-1}(\tau))|+\sup_{\theta}|E(\hat{b}_F(\theta))-b_F(\theta)|,$$

and so Lemma A.3(i) implies that

$$\max_{\tau\in I_h}|\hat{b}_F(t^{-1}(\tau))-b_F(t^{-1}(\tau))| \le \max_{\tau\in I_h}|\hat{\Delta}(t^{-1}(\tau))|+O(m^{-1}+h^2).$$

Moving on, observe that the number of subintervals making up I_h is equal to $\lceil l_h^{-1}\rceil = \lceil h^{-3+\omega^{-1}}\rceil$, where $\lceil a\rceil$ delivers the smallest integer at least as large as a. We therefore have

$$
\begin{aligned}
E\left(\left(\max_{\tau\in I_h}\left|\hat{\Delta}(t^{-1}(\tau))\right|\right)^{\omega}\right) &= E\left(\max_{\tau\in I_h}\left|\hat{\Delta}(t^{-1}(\tau))\right|^{\omega}\right)\\
&\leq E\left(\sum_{\tau\in I_h}\left|\hat{\Delta}(t^{-1}(\tau))\right|^{\omega}\right)\\
&= \sum_{\tau\in I_h} E\left(\left|\hat{\Delta}(t^{-1}(\tau))\right|^{\omega}\right) \leq \left\lceil h^{-3+1/\omega}\right\rceil \sup_{\theta\in\mathbb{R}} E\left|\hat{\Delta}(\theta)\right|^{\omega}.
\end{aligned}
\tag{A.9}
$$

Let $b_i(\theta) := -\frac{1}{2}h^{-2}\sigma_i^2\phi'\left(\frac{\vartheta_i-\theta}{h}\right)$ and $\Delta_i(\theta) := b_i(\theta) - Eb_i(\theta)$. We may then write $\hat{\Delta}(\theta) = n^{-1}\sum_{i=1}^{n}\Delta_i(\theta)$. Notice that $\Delta_i(\theta)$ are independent and mean zero. By Rosenthal (1970, Thm. 3), we therefore have that

$$
\left(E\left(\left|n^{-1/2}\sum_{i=1}^{n}\Delta_i(\theta)\right|^{\omega}\right)\right)^{1/\omega}
$$

is bounded from above by

$$
c\max\left\{\left(n^{-1}\sum_{i=1}^{n}E\left(\Delta_i(\theta)^2\right)\right)^{1/2},\ n^{-1/2}\left(\sum_{i=1}^{n}E\left(|\Delta_i(\theta)|^{\omega}\right)\right)^{1/\omega}\right\},
$$

where the constant c only depends on ω. Using Lemma A.3(ii), we obtain

$$
\sup_{\theta\in\mathbb{R}}\left(n^{-1}\sum_{i=1}^{n}E(\Delta_i(\theta)^2)\right)^{1/2} = \sup_{\theta\in\mathbb{R}}\left(n\operatorname{var}\hat{b}_F(\theta)\right)^{1/2} = O(h^{-3/2}).
$$

Using Lemma A.4, we obtain

$$
\begin{aligned}
n^{-1/2}\sup_{\theta\in\mathbb{R}}\left(\sum_{i=1}^{n}E\left(|\Delta_i(\theta)|^{\omega}\right)^{1/\omega}\right) &= n^{-1/2+1/\omega}\sup_{\theta\in\mathbb{R}}\left(E|\Delta_i(\theta)|^{\omega}\right)^{1/\omega}\\
&= O(n^{-1/2+1/\omega}h^{-2+1/\omega}) = O(h^{-3/2}),
\end{aligned}
$$

where in the last step we used the condition that $h^{-1} = O(n)$. We can therefore conclude from Rosenthal's inequality above that

$$
\left(\sup_{\theta\in\mathbb{R}}E\left(|\hat{\Delta}(\theta)|^{\omega}\right)\right)^{1/\omega} = n^{-1/2}\left(E\left(\left|n^{-1/2}\sum_{i=1}^{n}\Delta_i(\theta)\right|^{\omega}\right)\right)^{1/\omega} = O(n^{-1/2}h^{-3/2}).
$$

Using this and (A.9), we obtain

$$
\max_{\tau\in I_h}\left|\hat{\Delta}(t^{-1}(\tau))\right| = O(h^{(-3+1/\omega)/\omega}n^{-1/2}h^{-3/2}) = O(n^{-1/2}h^{-3/2-\varepsilon}),
$$

where $\varepsilon = 3/\omega - 1/\omega^2$. Combining this with (A.8) and (A.9), we thus conclude

$$\sup_{\theta\in\mathbb{R}} \left|\hat{b}_F(\theta) - b_F(\theta)\right| = O(m^{-1}) + O(h^2) + O(n^{-1/2} h^{-3/2-\varepsilon}),$$

as claimed.

Now, with $h = O(m^{-1/2})$ and $h^{-1} = O(n^{1-2\omega^{-1}})$, we find

$$\begin{aligned}\sup_{\theta\in\mathbb{R}} \frac{\sqrt{n}}{m} \left|\hat{b}_F(\theta) - b_F(\theta)\right| &= O_P(n^{1/2} m^{-1} h^2 + n^{1/2} m^{-2} + m^{-1} h^{-3/2-\varepsilon}) \\ &= O_P(n^{1/2} m^{-2} + m^{-4/9\epsilon^2}) \\ &= o_P(1),\end{aligned}$$

where in the last step we also used that $n/m^4 \to 0$ and that $m \to \infty$. The result of Proposition 3 now follows immediately from Proposition 1. □

Proof of Proposition 4

Let $\mathbb{G}_n(u) := \hat{F}(F_m^{-1}(u))$ be the empirical distribution function of the i.i.d. sample $u_i = F_m(\vartheta_i)$. Lemma A.1 and Theorem 1 in Doss and Gill (1992) give

$$\sup_{\tau\in[0,1]} \left|\sqrt{n}\left(\mathbb{G}_n^{\leftarrow}(\tau) - \tau\right) + \mathbb{B}_n(\tau)\right| = o_P(1), \tag{A.10}$$

where $\mathbb{G}_n^{\leftarrow}$ again denotes the left inverse of $\mathbb{G}_n \mathbb{B}_n(\tau)$ is the sequence of Brownian bridges that previously appeared in Lemma A.1.

Equation (A.10) yields

$$\mathbb{G}_n^{\leftarrow}(\hat{\tau}^*) - \mathbb{G}_n^{\leftarrow}(\tau) = (\hat{\tau}^* - \tau) - n^{-1/2}\left[\mathbb{B}_n(\hat{\tau}^*) - \mathbb{B}_n(\tau)\right] + o_p(n^{-1/2}).$$

Furthermore, $\hat{\tau}^* - \tau = O_p(m^{-1})$ follows from the results above. Lévy's modulus-of-continuity theorem then implies that $\mathbb{B}_n(\hat{\tau}^*) - \mathbb{B}_n(\tau) = o_P(1)$. Therefore,

$$\mathbb{G}_n^{\leftarrow}(\hat{\tau}^*) - \mathbb{G}_n^{\leftarrow}(\tau) = O_p(m^{-1}) + o_p(n^{-1/2}).$$

By definition, we have $\check{q}(\tau) = \hat{F}^{\leftarrow}(\hat{\tau}^*)$ and $\hat{q}(\tau) = \hat{F}^{\leftarrow}(\tau)$, and also that $\mathbb{G}_n^{\leftarrow}(\tau) = F_m(\hat{F}^{\leftarrow}(\tau))$. Substituting this into the last displayed equation yields

$$F_m(\check{q}(\tau)) - F_m(\hat{q}(\tau)) = O_p(m^{-1}) + o_p(n^{-1/2}).$$

Lemma A.2 and our assumptions guarantee that $F_m(\tau)$ has a density $f_m(\tau)$ that is bounded from below in a neighborhood of $q(\tau)$ for the quantile of interest τ. The last result therefore also implies that

$$\check{q}(\tau) - \hat{q}(\tau) = O_p(m^{-1}) + o_p(n^{-1/2}). \tag{A.11}$$

Next, The result (A.10) implies $\sqrt{n}(\mathbb{G}_n^{\leftarrow}(\tau) - \tau) \rightsquigarrow \mathbb{B}(\tau)$ for a Brownian bridge $\mathbb{B}$. For $\check{q}(\tau) = \hat{F}^{\leftarrow}(\hat{\tau}^*)$, we have $F_m(\check{q}(\tau)) = \mathbb{G}_n^{\leftarrow}(\hat{\tau}^*)$, and therefore

$$\sqrt{n}(F_m(\check{q}(\tau)) - \hat{\tau}^*) \rightsquigarrow \mathbb{B}(\tau).$$

From Proposition 1, we know that $F_m(\theta) = E(\hat{F}(\theta)) = F(\theta) + b_F(\theta)/m + O(m^{-2})$, uniformly in θ. We then find

$$\sqrt{n}\left(F(\check{q}(\tau)) - \tau + \frac{b_F(\check{q}(\tau)) - \hat{b}_F(\hat{q}(\tau))}{m} + O(m^{-2})\right) \xrightarrow{d} N(0, \tau(1-\tau)).$$

From the proof of Proposition 3, we also know that $\sup_\theta (\sqrt{n}/m)\left|\hat{b}_F(\theta) - b_F(\theta)\right| = o_p(1)$, and therefore

$$\sqrt{n}\left(F(\check{q}(\tau)) - \tau + \frac{b_F(\check{q}(\tau)) - b_F(\hat{q}(\tau))}{m} + O(m^{-2})\right) \xrightarrow{d} N(0, \tau(1-\tau)).$$

Smoothness of the function b_F and (A.11) imply $b_F(\check{q}(\tau)) - b_F(\hat{q}(\tau)) = O(m^{-1}) + o_p(n^{-1/2})$. We thus obtain $\sqrt{n}\left(F(\check{q}(\tau)) - \tau\right) \xrightarrow{d} N(0, \tau(1-\tau))$ An application of the delta method with transformation F^{-1} then gives the result. This completes the proof. ■

APPENDIX B: Proofs of Lemmas A.3 and A.4

Before proving Lemmas A.3 and A.4, we first state one known result and establish two further intermediate lemmas.

LEMMA B.1 (Mason, 1981). *Let $\mathbb{G}_n$ be the empirical cumulative distribution of an i.i.d. sample of size n from a uniform distribution on* [0,1]. *Then, as $n \to \infty$,*

$$\sup_{u\in(0,1)} [u(1-u)]^{-1+\epsilon} \, |\mathbb{G}_n(u) - u| \to 0,$$

almost surely, for any $0 < \epsilon \leq 1/2$.

LEMMA B.2. *Let Assumption 1 hold. Then, if* $\sup_\theta (1 + |\theta|^\kappa) f(\theta) < \infty$,

$$\sup_\theta (1 + |\theta|^\kappa) f_m(\theta) = O_p(1)$$

holds.

Proof. The conditional density of $\vartheta_i - \theta_i$ given θ_i evaluated in ε is

$$p(\varepsilon|\theta) := E\left(\frac{1}{\sigma_i/\sqrt{m}} g\left(\frac{\varepsilon}{\sigma_i/\sqrt{m}}\right) \middle| \theta_i = \theta\right).$$

We thus have

$$f_m(\vartheta) = \int_{-\infty}^{\infty} p(\vartheta - \theta|\theta) f(\theta)\, d\theta = \int_{-\infty}^{\vartheta/2} p(\vartheta - \theta|\theta) f(\theta)\, d\theta + \int_{\vartheta/2}^{\infty} p(\vartheta - \theta|\theta) f(\theta)\, d\theta.$$

Without loss of generality, we will take the value ϑ to be positive throughout. We have the bound

$$f_m(\vartheta) \leq \sup_\theta f(\theta) \int_{-\infty}^{\vartheta/2} p(\vartheta - \theta|\theta)\, d\theta + \sup_{\theta \geq \vartheta/2} f(\theta) \int_{\vartheta/2}^{\infty} p(\vartheta - \theta|\theta)\, d\theta. \qquad \textbf{(B.1)}$$

Consider the second term on the right-hand side in (B.1). $\sup_{\theta \geq \vartheta/2} f(\theta) = O(1 + |\vartheta/2|^{-\kappa})$ by assumption and so it suffices to show that the integral is finite for all ϑ. To see that this is so, observe that

$$\int_{\vartheta/2}^{\infty} p(\vartheta-\theta|\theta)\,d\theta = \int_{-\infty}^{\vartheta/2} p(\varepsilon|\vartheta-\varepsilon)\,d\varepsilon = \int_{-\infty}^{\vartheta/2} E\left(\frac{1}{\sigma_i/\sqrt{m}}\,g\left(\frac{\varepsilon}{\sigma_i/\sqrt{m}}\right)\middle|\theta_i=\vartheta-\varepsilon\right)d\varepsilon$$

and use the change of variable $\epsilon^* = \sqrt{m}\,\varepsilon$

$$\begin{aligned}\int_{\vartheta/2}^{\infty} p(\vartheta-\theta|\theta)\,d\theta &\le \int_{-\infty}^{\infty} \max_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\left\{\frac{1}{\sigma/\sqrt{m}}\,g\left(\frac{\varepsilon}{\sigma/\sqrt{m}}\right)\right\}d\varepsilon\\ &= \int_{-\infty}^{\infty} \max_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\left\{\frac{1}{\sigma}\,g\left(\frac{\epsilon}{\sigma}\right)\right\}d\epsilon\\ &\le C\int_{-\infty}^{\infty} \max_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\left\{\frac{1}{\sigma}\,\left(1+\left|\frac{\epsilon}{\sigma}\right|\right)^{-\alpha}\right\}d\epsilon\\ &\le C\int_{-\infty}^{\infty} \frac{1}{\underline{\sigma}}\left(1+\left|\frac{\epsilon}{\underline{\sigma}}\right|\right)^{-\alpha}d\epsilon = C/(\alpha-1) = O(1).\end{aligned}$$

Next, for the first right-hand side term in (B.1), recall that $\sup_\theta f(\theta) < \infty$, and so we need to show that the integral vanishes sufficiently fast as $\vartheta \to \infty$. To see that this is the case, we proceed as before by observing that

$$\int_{-\infty}^{\vartheta/2} p(\vartheta-\theta|\theta)\,d\theta = \int_{\vartheta/2}^{\infty} E\left(\frac{1}{\sigma_i/\sqrt{m}}\,g\left(\frac{\varepsilon}{\sigma_i/\sqrt{m}}\right)\middle|\theta_i=\vartheta-\varepsilon\right)d\varepsilon$$

to obtain

$$\begin{aligned}\int_{-\infty}^{\vartheta/2} p(\vartheta-\theta|\theta)\,d\theta &\le \int_{\vartheta/2}^{\infty} \max_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\left\{\frac{1}{\sigma/\sqrt{m}}\,g\left(\frac{\varepsilon}{\sigma/\sqrt{m}}\right)\right\}d\varepsilon\\ &\le \int_{\sqrt{m}\vartheta/2}^{\infty} \max_{\sigma\in[\underline{\sigma},\overline{\sigma}]}\left\{\frac{1}{\sigma}\,g\left(\frac{\epsilon}{\sigma}\right)\right\}d\epsilon\\ &\le C\int_{\sqrt{m}\vartheta/2}^{\infty} \frac{1}{\underline{\sigma}}\left(1+\frac{\epsilon}{\underline{\sigma}}\right)^{-\alpha}d\epsilon = O(1+(\sqrt{m}\,\vartheta/2)^{1-\alpha}).\end{aligned}$$

Thus, as long as $\alpha > 1$ and $\alpha \ge \kappa+1$, we have

$$f_m(\vartheta) = O(1+|\vartheta/2|^{-\kappa})$$

uniformly in ϑ, as claimed. This completes the proof of the lemma. □

LEMMA B.3. *Let Assumptions 1 and 2 hold, and let*

$$\gamma_m^r(\theta) := E(\sigma_i^r|\vartheta_i=\theta)f_m(\theta), \qquad \gamma^r(\theta) := E(\sigma_i^r|\theta_i=\theta)f(\theta).$$

Then, for any integer r,

$$\sup_\theta |\nabla^q\gamma_m^r(\theta) - \nabla^q\gamma^r(\theta)| = O(m^{-1})$$

provided that the conditional density $h(\theta|\sigma)$ is $(q+2)$ times differentiable with respect to θ and that there exists a function e so that $|\nabla_1^{q+2}h(\theta|\sigma)| \le e(\sigma)$ and $E(e(\sigma_i)) < \infty$.

Proof. Fix r throughout the proof. First note that, by Bayes' rule and Assumption 1, we may write

$$\gamma_m^r(\vartheta) = \int_{\underline{\sigma}}^{\overline{\sigma}}\int_{-\infty}^{\infty} \sigma^r\,\frac{1}{\sigma/\sqrt{m}}\,g\left(\frac{\vartheta-\theta}{\sigma/\sqrt{m}}\right)h(\theta,\sigma)\,d\sigma\,d\theta.$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

A change of variable from θ to $\varepsilon := (\vartheta - \theta)/(\sigma/\sqrt{m})$ then allows to write

$$\gamma_m^r(\vartheta) = E\left(B_r(\vartheta, \varepsilon_i/\sqrt{m})\right), \qquad B_r(\theta,\delta) := \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma^r\, h(\theta - \delta\sigma, \sigma)\, d\sigma.$$

Observe that $B_r(\vartheta, 0) = \gamma^r(\vartheta)$. Now, by a Taylor expansion,

$$\nabla^q \gamma_m^r(\vartheta) - \nabla^q \gamma^r(\vartheta) = \frac{E\left(\varepsilon_i^2\, \nabla_1^q \nabla_2^2 B_r(\vartheta, \varepsilon_i^*/\sqrt{m})\right)}{m}.$$

Furthermore, as

$$\nabla_1^p \nabla_2^q B_r(\theta,\delta) = (-1)^q \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma^{r+q}\, \nabla_1^{p+q} h(\theta - \delta\sigma, \sigma)\, d\sigma$$

for any pair of integers (p,q), we have that

$$\sup_\theta \sup_\delta |\nabla_1^q \nabla_2^2 B_r(\theta,\delta)| \le \overline{\sigma}^{r+q} \sup_\theta \sup_\delta \left|\int_{\underline{\sigma}}^{\overline{\sigma}} \nabla_1^{2+q} h(\theta - \delta\sigma|\sigma)\, h(\sigma)\, d\sigma\right| \le \overline{\sigma}^{r+q} \int_{\underline{\sigma}}^{\overline{\sigma}} e(\sigma)\, h(\sigma)\, d\sigma,$$

which is finite. Therefore, uniformly in θ,

$$\nabla^q \gamma_m^r(\theta) - \nabla^q \gamma^r(\theta) = O(m^{-1}),$$

as claimed. This completes the proof. □

Proof of Lemma A.3

Part (i): With

$$\beta_m(\theta) := \frac{E(\sigma_i^2 | \vartheta_i = \theta) f_m(\theta)}{2},$$

a change of variable and integration by parts yield

$$E(\hat{b}_F(\theta)) = -\int_{-\infty}^{\infty} \frac{\beta_m(\vartheta)}{h^2}\, \phi'\left(\frac{\vartheta - \theta}{h}\right) d\vartheta = \int_{-\infty}^{\infty} \nabla^1 \beta_m(\theta + h\varepsilon)\, \phi(\varepsilon)\, d\varepsilon.$$

Taylor expanding $\nabla^1 \beta_m$ around $\varepsilon = 0$ and using our assumptions of the distribution of ε, we obtain

$$E(\hat{b}_F(\theta)) = \nabla^1 \beta_m(\theta) + h^2\, \frac{\int_{-\infty}^{\infty} \nabla^3 \beta_m(\theta + h\varepsilon^*)\, \varepsilon^2\, \phi(\varepsilon)\, d\varepsilon}{2},$$

where ε^* lies between ε and zero. From Lemma B.3, we have

$$\nabla^1 \beta_m(\theta) = \nabla^1 \beta(\theta) + O(m^{-1}) = b_F(\theta) + O(m^{-1}),$$

uniformly in θ, and $\sup_\theta |\nabla^3 \beta_m(\theta)| < \infty$. Therefore,

$$E(\hat{b}_F(\theta)) = b_F(\theta) + O(m^{-1}) + O(h^2),$$

as claimed.

Part (ii): Note that

$$\operatorname{var}(\hat{b}_F(\theta)) = E(\hat{b}_F(\theta)^2) - E(\hat{b}_F(\theta))^2 = \frac{n^{-1}}{4} E\left(\frac{\sigma_i^4}{h^4} \phi'\left(\frac{\vartheta - \theta}{h}\right)^2\right) - b_F(\theta)^2 + o(n^{-1}).$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Now, with

$$\beta_m^2(\theta) := \frac{E(\sigma_i^4|\vartheta_i=\theta)f_m(\theta)}{4},$$

we have

$$\frac{n^{-1}}{4}E\left(\frac{\sigma_i^4}{h^4}\phi'\left(\frac{\vartheta-\theta}{h}\right)^2\right) = \int_{-\infty}^{\infty}\frac{\beta_m^2(\vartheta)}{h^4}\phi'\left(\frac{\vartheta-\theta}{h}\right)^2 d\vartheta \leq \frac{\sup_\theta|\beta_m^2(\theta)|}{n}\,\frac{\int_{-\infty}^{\infty}\phi'\left(\frac{\vartheta-\theta}{h}\right)^2 d\vartheta}{h^4},$$

which is $O(n^{-1}h^3)$ uniformly in θ as $\sup_\theta|\beta_m^2(\theta)| < \infty$ because σ_i is finite and f_m is bounded, and

$$\int_{-\infty}^{\infty}\phi'\left(\frac{\vartheta-\theta}{h}\right)^2 d\vartheta = \frac{h}{4\sqrt{\pi}},$$

independent of θ. This completes the proof.

Part (iii): First observe that

$$\nabla^1 b_F(\theta) = \nabla^2\beta(\theta)/2,$$

so that $(1+|\theta|^{1+\eta})\,|\nabla^1 b_F(\theta)| < \infty$ follows directly from Assumption 3. What is left to show is that

$$\sup_\theta(1+|\theta|^{1+\eta})\,|\nabla^1\hat{b}_F(\theta)| = O_p(-(1+\omega^{-1})).$$

Note that

$$\nabla^1\hat{b}_F(\theta) = \frac{(nh^2)^{-1}}{2}\sum_{i=1}^n \sigma_i^2\,\phi''\left(\frac{\vartheta_i-\theta}{h}\right).$$

By Hölder's inequality,

$$|\nabla^1\hat{b}_F(\theta)| \leq h^{-2}\left\{\left(n^{-1}\sum_{i=1}^n(\sigma_i^2/2)^\omega\right)^{\omega^{-1}}\right\}\times\left\{\left(n^{-1}\sum_{i=1}^n\left|\phi''\left(\frac{\vartheta_i-\theta}{h}\right)\right|^\psi\right)^{\psi^{-1}}\right\},$$

where $\psi := (1-\omega^{-1})^{-1}$. The first term in braces is bounded in probability because the σ_i^2 are finite. For the second term in braces, write $\mathbb{G}_n$ for the empirical cumulative distribution of an i.i.d. sample of size n from the uniform distribution on $[0,1]$ and let $\mathbb{G}_n'(u) := n^{-1}\sum_{i=1}^n \delta_{u_i-u}$, where δ_a is Dirac's delta at a. Then, writing ∇_u for the derivative with respect to u, we get

$$\begin{aligned} n^{-1}\sum_{i=1}^n\left|\phi''\left(\frac{\vartheta_i-\theta}{h}\right)\right|^\psi &= \int_0^1\left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^\psi \mathbb{G}_n'(u)\,du \\ &= -\int_0^1\nabla_u^1\left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^\psi \mathbb{G}_n(u)\,du \end{aligned}$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

$$
\begin{aligned}
&= -\int_0^1 \nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi} u\, du \\
&\quad - \int_0^1 \nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi} (\mathbb{G}_n(u)-u)\, du, \qquad \textbf{(B.2)}
\end{aligned}
$$

where we have used integration by parts in the first step and replaced $\mathbb{G}_n(u)$ by $u+(\mathbb{G}_n(u)-u)$ in the second step. We now consider each of the integrals on the right-hand side in turn. First, integrating by parts,

$$
-\int_0^1 \nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi} u\, du = E\left(\left|\phi''\left(\frac{\vartheta_i-\theta}{h}\right)\right|^{\psi}\right). \qquad \textbf{(B.3)}
$$

Clearly, this term is bounded uniformly on any finite interval. To evaluate it for large values of θ, observe that

$$
\begin{aligned}
\frac{1}{h}E\left(\left|\phi''\left(\frac{\vartheta_i-\theta}{h}\right)\right|^{\psi}\right) &= \int_{-\infty}^{+\infty} \frac{1}{h}\left|\phi''\left(\frac{\vartheta-\theta}{h}\right)\right|^{\psi} f_m(\vartheta)\, d\vartheta \\
&= \int_{\theta-h\log(1+|\theta|)}^{\theta+h\log(1+|\theta|)} \frac{1}{h}\left|\phi''\left(\frac{\vartheta-\theta}{h}\right)\right|^{\psi} f_m(\vartheta)\, d\vartheta \\
&\quad + \int_{\log(1+|\theta|)}^{\infty} \left|\phi''(z)\right|^{\psi} f_m(\theta+zh)\, dz \\
&\quad + \int_{\log(1+|\theta|)}^{\infty} \left|\phi''(z)\right|^{\psi} f_m(\theta-zh)\, dz.
\end{aligned}
$$

Here,

$$
\int_{\theta-h\log(1+|\theta|)}^{\theta+h\log(1+|\theta|)} \frac{1}{h}\left|\phi''\left(\frac{\vartheta-\theta}{h}\right)\right|^{\psi} f_m(\vartheta)\, d\vartheta \le O(\log(1+|\theta|)) \sup_{\theta} |f_m(\theta)| = O(\log(1+|\theta|)),
$$

because $\sup_\theta |\phi''(\theta)|^{\psi} = O(1)$ and f_m is bounded. Furthermore, because

$$
\textstyle\int_x^{\infty} |\phi''(z)|^{\psi}\, dz = O(x^{2\psi-1} e^{-\psi x^2/2}), \qquad \text{as } x \to \infty,
$$

and $f_m(\theta) = O(|\theta|^{-\kappa})$ as $|\theta| \to \infty$ by Lemma B.2, we have

$$
\begin{aligned}
&\textstyle\int_{\log(1+|\theta|)}^{\infty} \left|\phi''(z)\right|^{\psi} f_m(\theta+zh)\, dz = O\left(\log(1+|\theta|)^{2\psi-1} e^{-\psi \log(1+|\theta|)^2/2}\right), \\
&\textstyle\int_{\log(1+|\theta|)}^{\infty} \left|\phi''(z)\right|^{\psi} f_m(\theta-zh)\, dz = O\left(\log(1+|\theta|)^{2\psi-1} e^{-\psi \log(1+|\theta|)^2/2}\right).
\end{aligned}
$$

Then, as

$$
e^{-\psi \log(1+|\theta|)^2/2} = o(|\theta|^{a}) \text{ for any } a > 0 \text{ as } |\theta| \to \infty,
$$

we may conclude that the term in (B.3) is $O(h|\theta|^{-\kappa}\log(1+|\theta|))$ uniformly in θ. Next, for the second term in (B.2), we use Lemma B.1 to establish that, for any $\epsilon \in (0, 1/2]$, we have

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

$$\left|\int_0^1 \nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi} (\mathbb{G}_n(u)-u)\,du\right|$$
$$\leq o_p(1)\left|\int_0^1 \left|\nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi}\right| \left(u^{1-\epsilon}(1-u)^{1-\epsilon}\right) du\right|$$
$$= o_p(1)\left|\int_{-\infty}^{+\infty} \left|\nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi}\right| \left(F_m(\vartheta)^{1-\epsilon}(1-F_m(\vartheta))^{1-\epsilon}\right) d\vartheta\right|,$$

where the $o_p(1)$ term is independent of θ. The integral term can be bounded in the same way as (B.3). Hence,

$$\left|\int_0^1 \nabla_u^1 \left|\phi''\left(\frac{F_m^{-1}(u)-\theta}{h}\right)\right|^{\psi} (\mathbb{G}_n(u)-u)\,du\right| = o_p(h|\theta|^{(1-\epsilon)(1-\kappa)}\log(1+|\theta|))$$

uniformly in θ. We therefore have that

$$\sup_\theta |\hat{b}_F(\theta)| \leq h^{-2} O_p(1)\left\{(O(h|\theta|^{-\kappa}\log(1+|\theta|)) + o_p(h|\theta|^{(1-\epsilon)(1-\kappa)}\log(1+|\theta|))^{\psi^{-1}}\right\}.$$

For any $\eta > (\kappa-1)(1-\epsilon)(1-1/\omega)-1 > 0$, it then follows that

$$\sup_\theta \left(1+|\theta|^{1+\eta}\right)|\hat{b}_F(\theta)| = O_P\left(h^{-(1+\omega^{-1})}\right).$$

Here, our assumption $\kappa > 1+(1-1/\omega)^{-1}$ guarantees that we can find $\epsilon > 0$ such that $\eta > (\kappa-1)(1-\epsilon)(1-1/\omega)-1 > 0$ holds. This concludes the proof. □

Proof of Lemma A.4. First observe that, for any $\epsilon > 0$,

$$\sup_\theta E(|b_i(\theta)-E(b_i(\theta))|^{\epsilon}) \leq \sup_\theta \sum_{p=0}^{\epsilon}\binom{\epsilon}{p} E(|b_i(\theta)|^p)\,E(|b_i(\theta)|^{\epsilon-p}) \leq 2^{\epsilon}\sup_\theta E(|b_i(\theta)|^{\epsilon}).$$

Therefore,

$$\begin{aligned}
\sup_\theta E(|b_i(\theta)-E(b_i(\theta))|^{\epsilon})^{\epsilon^{-1}} &\leq 2\sup_\theta (E(|b_i(\theta)|^{\epsilon}))^{\epsilon^{-1}} \\
&= \sup_\theta\left(\int_{-\infty}^{\infty}\frac{E(\sigma_i^{2\epsilon}|\vartheta_i=\vartheta)f_m(\vartheta)}{h^2}\left|\phi'\left(\tfrac{\vartheta-\theta}{h}\right)\right|^{\epsilon} d\vartheta\right)^{\epsilon^{-1}} \\
&\leq \sup_\vartheta (E(\sigma_i^{2\epsilon}|\vartheta_i=\vartheta)f_m(\vartheta))^{\epsilon^{-1}}\frac{\left(\sup_\theta\int_{-\infty}^{\infty}\left|\phi'\left(\tfrac{\vartheta-\theta}{h}\right)\right|^{\epsilon} d\vartheta\right)^{\epsilon^{-1}}}{h^2} \\
&= O(h^{\epsilon^{-1}-2}),
\end{aligned}$$

where we have used the definition of $b_i(\theta)$ in the first step, boundedness of the σ_i and f_m in the second step, and the fact that

$$\int_{-\infty}^{\infty}\left|\phi'\left(\frac{\vartheta-\theta}{h}\right)\right|^{\epsilon} d\vartheta = O(h),$$

independent of θ, in the final step. This completes the proof. □

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

APPENDIX C: Least-Squares Cross Validation

The integrated squared error of

$$\check{F}(\theta) = \hat{F}(\theta) - \frac{\hat{b}_F(\theta)}{m}$$

is

$$\int (\check{F}(\theta) - F(\theta))^2 \, d\theta = \frac{\int \hat{b}_F(\theta)^2 \, d\theta}{m^2} - \frac{2\int (\hat{F}(\theta) - F(\theta))\, \hat{b}_F(\theta)\, d\theta}{m} + \text{term independent of } h.$$

Using the definition of $\hat{b}_F$ and expanding the square, the first right-hand side term can be written as

$$\frac{\int \hat{b}_F(\theta)^2 \, d\theta}{m^2} = \frac{m^{-2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sigma_i^2 \sigma_j^2}{h^2} \frac{1}{4} \int \frac{1}{h} \phi' \left(\frac{\vartheta_i - \theta}{h} \right) \frac{1}{h} \phi' \left(\frac{\vartheta_j - \theta}{h} \right) d\theta,$$

and using properties of the normal distribution, we calculate

$$\int \phi' \left(\frac{\vartheta_i - \theta}{h} \right) \phi' \left(\frac{\vartheta_j - \theta}{h} \right) d\theta = \frac{1}{\sqrt{2}h} \phi \left(\frac{\vartheta_i - \vartheta_j}{\sqrt{2}h} \right) \left(\frac{h^2}{2} - \frac{(\vartheta_i + \vartheta_j)^2}{4} + \vartheta_i \vartheta_j \right).$$

Next, exploiting that $\phi'(\eta) = -\eta\, \phi(\eta)$ and using well-known results on the truncated normal distribution,

$$\begin{aligned}
-\frac{2\int \hat{F}(\theta)\, \hat{b}_F(\theta)\, d\theta}{m} &= \frac{m^{-1}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sigma_j^2}{h^2} \int_{\vartheta_i}^{+\infty} \phi' \left(\frac{\vartheta_j - \theta}{h} \right) d\theta \\
&= \frac{m^{-1}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sigma_j^2}{h^2} \int_{\vartheta_i}^{+\infty} \left(\frac{\theta - \vartheta_j}{h} \right) \phi \left(\frac{\theta - \vartheta_j}{h} \right) d\theta \\
&= \frac{m^{-1}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sigma_j^2}{h} \left(\frac{\vartheta_i - \vartheta_j}{h} \right) \phi \left(\frac{\vartheta_i - \vartheta_j}{h} \right) \\
&= \frac{m^{-1}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \frac{\sigma_i^2}{h} \phi' \left(\frac{\vartheta_i - \vartheta_j}{h} \right).
\end{aligned}$$

Omitting terms for which $j = i$ in the last expression is justified by the fact that $\phi'(0) = 0$. Finally, for the last term, integrating by parts shows that

$$\frac{2\int F(\theta)\, \hat{b}_F(\theta)\, d\theta}{m} = -\frac{m^{-1}}{n} \sum_{i=1}^{n} \frac{\sigma_i^2}{h} \int \phi \left(\frac{\vartheta_i - \theta}{h} \right) f(\theta)\, d\theta.$$

The integral in the right-hand side expression represents an expectation taken with respect to f. A leave-one-out estimator of the entire term is

$$-\frac{m^{-1}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{\sigma_i^2}{h} \phi \left(\frac{\vartheta_i - \vartheta_j}{h} \right).$$

https://doi.org/10.1017/S0266466622000378 Published online by Cambridge University Press

Combining results and multiplying the entire expression through with n^2m^2 yields the cross-validation objective function stated in the main text.

REFERENCES

Ahn, D., S. Choi, D. Gale, & S. Kariv (2014) Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics* 5, 195–223.

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

Barras, L., P. Gagliardini, & O. Scaillet (2021) Skill, scale, and value creation in the mutual fund industry. *Journal of Finance* 77(1), 601–638.

Bonhomme, S., K. Jochmans, & J.-M. Robin (2016a) Estimating multivariate latent-structure models. *Annals of Statistics* 44, 540–563.

Bonhomme, S., K. Jochmans, & J.-M. Robin (2016b) Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society, Series B* 78, 211–229.

Browning, M., M. Ejrnæs, & J. Alvarez (2010) Modeling income processes with lots of heterogeneity. *Review of Economic Studies* 77, 1353–1381.

Carroll, R.J. & P. Hall (1988) Optimal rates of convergence for deconvoluting a density. *Journal of the American Statistical Association* 83, 1184–1186.

Chamberlain, G. (1984) Chapter 22: Panel data. In Z. Griliches & M. Intriligator (eds.), *Handbook of Econometrics*. Handbooks in Economics, 2, pp. 1247–1315. Elsevier.

Chesher, A. (1991) The effect of measurement error. *Biometrika* 78, 451–462.

Chesher, A. (2017) Understanding the effect of measurement error on quantile regressions. *Journal of Econometrics* 200, 223–237.

Chetty, R., J.N. Friedman, & J.E. Rockoff (2014) Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104, 2593–2632.

Crucini, M.J., M. Shintani, & T. Tsuruga (2015) Noisy information, distance and law of one price dynamics across US cities. *Journal of Monetary Economics* 74, 52–66.

Delaigle, A. & A. Meister (2008) Density estimation with heteroscedastic error. *Bernoulli* 14, 562–579.

Dhaene, G. & K. Jochmans (2015) Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82, 991–1030.

Doss, H. & R.D. Gill (1992) An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *Journal of the American Statistical Association* 87(419), 869–877.

Efron, B. (2011) Tweedie's formula and selection bias. *Journal of the American Statistical Association* 106, 1602–1614.

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

Evdokimov, K. & A. Zeleneev (2020) Simple Estimation of Semiparametric Models with Measurement Errors. CeMMAP Working paper 08/22, Mimeo.

Fernández-Val, I. & J. Lee (2013) Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics* 4, 453–481.

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

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both *n* and *T* are large. *Econometrica* 70, 1639–1657.

Hahn, J. & W.K. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.

Horowitz, J.L. & M. Markatou (1996) Semiparametric estimation of regression models from panel data. *Review of Economic Studies* 63, 145–168.

Hu, Y. (2008) Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* 144, 27–61.

Hu, Y. & S.M. Schennach (2008) Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.

Jackson, C.K., J.E. Rockoff, & D.O. Staiger (2014) Teacher effects and teacher related policies. *Annual Review of Economics* 6, 801–825.

James, W. & C. Stein (1961) Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 361–379. University of California Press.

Komlós, J., P. Major, & G. Tusnády (1975) An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 32, 111–131.

Li, T. & Q. Vuong (1998) Nonparametric estimation of measurement error models using multiple indicators. *Journal of Multivariate Analysis* 65, 139–165.

Magnac, T. & S. Roux (2021) Heterogeneity and wage inequalities over the life cycle. *European Economic Review* 134, 103715.

Maritz, J.S. & R.G. Jarrett (1978) A note on estimating the variance of the sample median. *Journal of the American Statistical Association* 73, 194–196.

Mason, D.M. (1981) Bounds for weighted empirical distribution functions. *Annals of Probability* 9, 881–884.

Neyman, J. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.

Okui, R. & T. Yanagi (2019) Panel data analysis with heterogeneous dynamics. *Journal of Econometrics* 212, 451–475.

Okui, R. & T. Yanagi (2020) Kernel estimation for panel data with heterogenous dynamics. *The Econometrics Journal* 23, 156–175.

Parsley, D.C. & S.-J. Wei (2001) Convergence to the law of one price without trade barriers or currency fluctuations. *Quarterly Journal of Economics* 111, 1211–1236.

Robbins, H. (1956) An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 157–163. University of California Press.

Rockoff, J.E. (2004) The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review: Papers & Proceedings* 94, 247–252.

Rosenthal, H.P. (1970) On the subspaces of $L_p (p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics* 8, 273–303.

Vivalt, E. (2015) Heterogeneous treatment effects in impact evaluation. *American Economic Review: Papers & Proceedings* 105, 467–470.

Weinstein, A., Z. Ma, L.D. Brown, & C.-H. Zhang (2018) Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *Journal of the American Statistical Association* 113, 698–710.