# INFERENCE ON A DISTRIBUTION FROM NOISY DRAWS

Koen Jochmans*
University of Cambridge

Martin Weidner†
University College London

June 20, 2018

**Abstract**

We consider a situation where the distribution of a random variable is being estimated by the empirical distribution of noisy measurements of the random variable. This is common practice in many settings, including the evaluation of teacher value-added and the assessment of firm efficiency through stochastic-frontier models. 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. Analytical and jackknife corrections for the empirical distribution are derived that recenter the limit distribution and yield confidence intervals with correct coverage in large samples. A similar adjustment is also presented for the quantile function. These corrections are non-parametric and easy to implement. 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 the estimation of a stochastic-frontier model is also provided.

**JEL Classification:** C14, C23

**Keywords:** bias correction, estimation noise, non-parametric inference, regression to the mean, shrinkage.

---

*Address: University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. E-mail: kj345@cam.ac.uk.

†Address: University College London, Department of Economics, Drayton House, 30 Gordon Street, London WC1 H0AX, United Kingdom; and CeMMAP. E-mail: m.weidner@ucl.ac.uk.

We are grateful to Isaiah Andrews, Stéphane Bonhomme, Bo Honoré, and Ryo Okui for comments and discussion, and to Peter Schmidt for suggesting to us the empirical application.

Jochmans gratefully acknowledges financial support from the European Research Council through Starting Grant nº 715787. 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 grant ERC-2014-CoG-646917-ROMIA.

# 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 $\vartheta_1, \ldots, \vartheta_n$. In Rockoff (2004), for example, $\theta_i$ is a teacher effect, $\vartheta_i$ is an estimator of it obtained from data on student test scores, and we care about the distribution of teacher value-added (see, e.g., Jackson, Rockoff and Staiger 2014). Schmidt and Sickles (1984) recover estimates of firm inefficiency from fitting productions functions with fixed effects to panel data. 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.

In this paper we analyze the properties of the plug-in estimator of $F$ in an asymptotic embedding where the noise in $\vartheta_1, \ldots, \vartheta_n$ shrinks with the sample size ($n$). If we write the variances of $\vartheta_1, \ldots, \vartheta_n$ as $\sigma_1^2/m, \ldots, \sigma_n^2/m$ for some real number $m$, we consider double asymptotics where $n, m \to \infty$ jointly. This embedding is intuitive in settings where $\vartheta_i$ is an estimator of $\theta_i$ obtained from a sample of size $m$, as it is in the examples mentioned above. It is related to, yet different from, small measurement-error approximations as in, e.g., Chesher (1991), and has been used in the analysis of panel data models with fixed effects (see, e.g., Alvarez and Arellano 2003; Hahn and Kuersteiner 2002), although for different purposes.

We will focus on the case where

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

although our results hold more generally in situations where the

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

are random draws from some well-behaved (but unknown) distribution. While for the most part we will work under the assumption that the $\sigma_i^2$ are known our results carry over to the case where only a consistent estimator is available. Formally dealing with this would, however, require additional technical conditions that, we feel, would cloud the exposition.

The focus on the normal case helps to connect with the literature on shrinkage and selection bias as recently dealt with by Efron (2011) and to contrast our approach with one based on deconvolution.

Efron (2011) essentially entertains the homoskedastic setting where

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

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

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

where $p$ is the marginal density of the $\vartheta_i$ and $\nabla^1$ 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, non-parametric implementation would 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 $\vartheta_i$ being too thick. The same shrinkage factor is applied to each $\vartheta_i$, a consequence of the noise being homoskedastic. How to deal with heteroskedastic noise in an Empirical Bayes framework is not obvious; see, e.g., Xie, Kou and Brown (2012) and Weinstein, Ma, Brown and Zhang (2018) for discussion and recent contributions. Shrinkage is achieved by introducing a bias of order $m^{-1}$ in the individual estimators. In general, this bias order is passed-through to plug-in estimators of the distribution and its functionals. Thus, while it improves on $\vartheta_1, \ldots, \vartheta_n$ in terms of estimation risk, shrinkage 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 bias of the naive plug-in estimator of the distribution,

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

and correct for it directly. In the James-Stein problem, where $\theta_i \sim N(\eta, \psi^2)$, for example, the bias under homoskedastic noise turns out to equal

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

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 non-parametric 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 $n/m^4 \to 0$. So, bias correction is preferable to the naive plug-in approach for typical data sets 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 $F$.

Given a known distribution for the (potentially heteroskedastic) noise, recovering $F$ from noisy data is a (generalized) deconvolution problem (as in Wang, Fan and Wang 2010) and can be solved for fixed $m$. However, it is well documented that deconvolution-based estimators have a slow rate of convergence and can behave quite poorly in small samples. In response to this Efron (2016) has recently argued for a return to a more parametric approach. Our estimation approach delivers an intuitive and fully non-parametric estimator that enjoys the usual parametric convergence rate and is numerically well behaved. Our bias formulae (and subsequent bias correction) also do not require the noise distribution to be known. Bias correction further ensures that size-correct inference can be performed, provided that $n/m^4$ is small. It is not clear how to conduct inference based on deconvolution estimators.

While our estimators are straightforward to apply it should be noted that working out the leading bias of $\hat{F}$ (and of its quantile function) is mathematically challenging because $\hat{F}$ is a non-smooth function of $\vartheta_1, \ldots, \vartheta_n$. As such, the approach taken here is different from, and complementary to, recent work on estimating average marginal effects in panel data models with heterogenous coefficients, which has focused exclusively on inference on smooth functionals (Fernández-Val and Lee 2013; Dhaene and Jochmans 2015; Okui and Yanagi 2017). The impact of noise on smooth transformations of the $\vartheta_i$ can be handled using conventional methods based on Taylor-series expansions. In work contemporaneous to our own, Okui and Yanagi (2018) derive the bias of a kernel-smoothed estimator of $F$ and its derivative. Such smoothing greatly facilitates the calculation of the bias, thus allowing for weaker assumptions on the noise distribution, but it also introduces additional bias terms that require further restrictions on the relative growth rates of $n$, $m$, and the bandwidth that governs the smoothing.

Simulation evidence on the improvement of our approach over the plug-in estimator (and Empirical Bayes) is presented. We present results for both normal and non-normal noise distributions and focus on samples where $m$ is much smaller than $n$, as is typically the case in practice. In such settings the bias in the plug-in estimator dominates its sampling error and test procedures over-reject under the null. The deviation from the nominal size of the test is substansive and makes the naive estimator unsuitable as a tool for inference. Adjusting for noise through our procedures makes the bias small relative to the standard error. It yields confidence interval with broadly correct coverage and, at the same time leads to a reduction in mean squared error.

As an empirical illustration we fit a stochastic-frontier model (Aigner, Lovell and Schmidt, 1977) to a short panel on Spanish dairy farms. The object of interest in such an analysis is the distribution of firm inefficiencies. A parametric approach would specify this distribution, typically as half-normal (Pitt and Lee, 1981), and maximize the resulting integrated likelihood. A non-parametric approach is to estimate a firm's inefficiency by its fixed effect in a standard panel data regression (Schmidt and Sickles, 1984). This strategy is common practice but is subject to the bias issues tackled here. Consequently, we apply our corrections to non-parametrically estimate the distribution of firm inefficiencies in these data.

## 2 Estimation and inference

Let $F$ be a univariate distribution on the real line. Here, we are interested in estimation of and inference on $F$ and its 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 $n$. Moreover, we assume the following.

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

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

*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 $\vartheta_i$ is an estimator of $\theta_i$ obtained from a sample of size $m$ that converges at the parametric rate.[1] We allow $\theta_i$ and $\sigma_i^2$ to be correlated, implying that the noise $\vartheta_i - \theta_i$ is not independent of $\theta_i$. Recovering the distribution of $\theta_i$ 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 $\vartheta_i$ at $\theta$. As we will show below, under suitable regularity conditions, such plug-in estimators are consistent and asymptotically normal as $n \to \infty$

[1] 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 $\sigma_i^2$ as $\sigma_i^2/p_i$. When the $\vartheta_i$ 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.

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 $\varphi$. An (infeasible) plug-in estimator based on $\theta_1, \ldots, \theta_n$ would be

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

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

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

under regularity conditions provided in the Appendix, by a Taylor-series expansion we have

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

and

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

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 $\varphi$ is linear. It can be corrected for by subtracting a plug-in estimator of $b_\mu/m$ from $\hat{\mu}$. Doing so, again under regularity conditions given in the Appendix, delivers and estimator that is asymptotically unbiased as long as $n/m^4 \to 0$.

## 2.1 Estimation of the distribution function

Now consider estimation of the distribution function $F$ using the plug-in estimator $\hat{F}$. Again, the use of noisy measurements introduces bias. The machinery from above cannot be applied to deduce the bias of $\hat{F}$, however, as it is a step function and, hence, is non-differentiable.

To derive the bias we impose the following conditions.

**Assumption 2.** *The density function $f$ is three times differentiable with uniformly bounded derivatives and one of the following two sets of conditions hold.*

*A. (i) The function $E(\sigma_i^{p+1}|\theta_i = \theta)$ is $p$-times differentiable for $p = 1, 2, 3$; (ii) the joint density of $(\theta_i, \sigma_i)$ exists, and the conditional density function of $\theta_i$ given $\sigma_i$ is three times differentiable with respect to $\theta_i$ and the third derivative is bounded in absolute value by a function $e(\sigma_i)$ such that $E(e(\sigma_i)) < \infty$.*

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

Assumption 2 distinguishes between the cases where the relation between $\theta_i$ and $\sigma_i^2$ is stochastic (Assumption 2.A) and deterministic (Assumption 2.B). It requires smoothness of certain densities and conditional expectations.

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 $\min\{\theta, \theta'\}$. Our first theorem gives the leading bias and variance of $\hat{F}$. All proofs are collected in the supplementary appendix.

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

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

*where the order of the remainder terms is uniform in $\theta$.*

To illustrate the result suppose that $\sigma_i^2$ is independent of $\theta_i$ and that $\theta_i$ has density function

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

as in the James and Stein (1961) problem. Letting $\sigma^2$ denote the mean of the $\sigma_i^2$ an application of Theorem 1 yields

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

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, and gives an alternative explanation of why the James and Stein (1961) estimator shrinks toward the overall mean $\eta$.

A bias-corrected estimator based on Theorem 1 is

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

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

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

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

Parts (i) and (ii) of Assumption 3 are simple smoothness and boundedness requirements. Parts (iii) and (iv) are tail conditions on the marginal density of the $\theta_i$ and on the bias function $b_F(\theta)$.

We have the following result.

**Proposition 1.** *Let Assumptions 1, 2, and 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* $\theta$, *where* $\mathbb{G}_F(\theta)$ *is a mean zero Gaussian process with covariance function* $\sigma_F(\theta_1, \theta_2)$.

The implications of Proposition 1 are qualitatively similar to those for smooth functionals discussed above. Indeed, for any fixed $\theta$, 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 the Appendix for details. The cross-validated bandwidth then is $\check{h} := \arg\min_h v(h)$ on the interval $(0, +\infty)$.

Theorem 1 equally validates 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 of $\hat{F}$ is proportional to $m^{-1}$ and is based on re-estimating $\theta_1, \ldots, \theta_n$ from subsamples. This would require access to the data from which the $\vartheta_i$ were calculated. On the other hand, an interesting feature of such an estimator is that it does not require knowledge of (or estimation of) the $\sigma_i^2$ to be implementable. A somewhat different jackknife procedure can be constructed from the observation that, if $\vartheta_1, \ldots, \vartheta_n$ would have variance $\lambda^2\sigma_1^2, \ldots, \lambda^2\sigma_n^2$, then the bias in $\hat{F}$ would equally be multiplied by $\lambda^2$. This is apparent from the definition of $\beta$ and suggests the jackknife estimator

$$\dot{F}(\theta) := \hat{F}(\theta) - \frac{\dot{b}_F(\theta)}{m} = \frac{1+\lambda^2}{\lambda^2}\,\hat{F}(\theta) - \frac{1}{\lambda^2}\,\hat{F}_\lambda(\theta),$$

where

$$\dot{b}_F(\theta) := m\,\frac{\hat{F}_\lambda(\theta) - \hat{F}(\theta)}{\lambda^2}, \qquad \hat{F}_\lambda(\theta) := n^{-1}\sum_{i=1}^n \Phi\left(\frac{1}{\lambda}\frac{\theta - \vartheta_i}{\sigma_i/\sqrt{m}}\right).$$

Note that $\dot{F}$ can be computed without re-estimating $\theta_1, \ldots, \theta_n$. Such an approach bears similarities to the jackknife estimator of a density function introduced in Schucany and Sommers (1977). The reason this estimator is bias-reducing is as follows. By Assumption 1 and iterated expectations,

$$E(\hat{F}(\theta)) = E\left(\Phi\left(\frac{\theta - \theta_i}{\sigma_i/\sqrt{m}}\right)\right) = F(\theta) + \frac{b_F(\theta)}{m} + O(m^{-2}).$$

Further, by a standard convolution argument,

$$E(\hat{F}_\lambda(\theta)) = E\left(\Phi\left(\frac{1}{\sqrt{1+\lambda^2}}\frac{\theta - \theta_i}{\sigma_i/\sqrt{m}}\right)\right) = F(\theta) + (1+\lambda^2)\,\frac{b_F(\theta)}{m} + O(m^{-2}).$$

Thus, our $\dot{b}_F(\theta)$ is a sample version of $b_F(\theta)$. Like in Schucany and Sommers (1977), the approach exploits variation in a bandwidth parameter. However, while they address smoothing bias in non-parametric density estimation (in a similar way as would the use of a higher-order kernel), our estimator attacks bias introduced through estimation noise.

Note, finally, that the sample variance of

$$1\{\vartheta_i \le \theta\} - \frac{1}{\lambda^2}\left(\Phi\left(\frac{1}{\lambda}\frac{\theta - \vartheta_i}{\sigma_i/\sqrt{m}}\right) - 1\{\vartheta_i \le \theta\}\right)$$

can be used for inference in stead of that of only $1\{\vartheta_i \le \theta\}$ although, again, both will be valid asymptotically.

## 2.2 Estimation of the quantile function

The bias in $\hat{F}$ translates to bias in estimators of the quantile function. A natural estimator for $\tau$th-quantile $q(\tau)$ is given by $\hat{q}(\tau) := \hat{F}^{\leftarrow}(\hat{\tau})$, where we use $\hat{F}^{\leftarrow}$ to denote the left-inverse of $\hat{F}$. Moreover,

$$\hat{q}(\tau) = \hat{F}^{\leftarrow}(\hat{\tau}) = \vartheta_{(\lceil \tau n \rceil)},$$

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

The quantile estimator is an approximate solution to the empirical moment condition $\hat{F}(q) - \tau = 0$ (with respect to $q$); it is an approximate root only because $\hat{F}$ is a step function. From Theorem 1 we know that

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

uniformly in $\tau$, 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 asymptotic bias result.

**Corollary 1.** *Let the 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 $n/m^2 \to c \in [0, +\infty)$.*

As an example, when $\theta_i \sim N(\eta, \psi^2)$, independent of $\sigma_i^2$, 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 moment condition.)

Corollary 1 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 non-parametric estimator of the density $f$, which further shows up in the denominator. An alternative estimator that avoids this issue is

$$\check{q}(\tau) := \hat{F}^{\leftarrow}(\hat{\tau}^*), \qquad \hat{\tau}^* := \tau + \frac{\hat{b}_F(\hat{q}(\tau))}{m}.$$

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 $\vartheta_i$ relative to the $\theta_i$, 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 $\tau$ to $\tau^*$ adjusts the naive estimator and corrects for regression to the mean.

**Proposition 2.** *Let the assumptions stated in Proposition 1 hold. For $\tau \in (0,1)$, assume that $f > 0$ in a neighborhood of $q(\tau)$. Then,*

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

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

The corrected estimator has the same asymptotic variance as the uncorrected estimator. It is well-known that plug-in estimators of $\sigma_q^2$ can perform quite poorly in small samples (Maritz and Jarrett 1978). Typically, researchers rely on the bootstap, 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 this bootstrap procedure does not involve re-estimation of the individual $\theta_i$.

The view of correcting the moment condition that defines $\hat{q}(\tau)$ also suggests the jackknife estimator

$$\dot{q}(\tau) := \frac{1+\lambda^2}{\lambda^2}\hat{q}(\tau) - \frac{1}{\lambda^2}\,\hat{q}_\lambda(\tau),$$

where $\hat{q}_\lambda(\tau) := \min_q\{q : \hat{F}_\lambda(q) \geq \tau\}$, again for some chosen $\lambda$. The intuition behind this jackknife correction follows from the discussion on the bias-reducing nature of $\dot{F}$ and the definition of $\hat{q}$.

# 3 Numerical illustrations

## 3.1 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 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, $\sigma^2$, for fixed $m$, our focus is on between-group characteristics and the distribution of the $\theta_i$ as a whole. We estimate $\theta_i$ 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 $\sigma^2/m$. Rather than assuming this variance to be known we implement our procedure 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 $\vartheta_i$ are homoskedastic in estimating the noise or in constructing the bias correction. To iterate, our procedure is non-parametric and does not require knowledge of the noise distribution for implementation.

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 and. 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 $\psi^2$. To illustrate how this invalidates inference in typically-sized data sets 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 shows 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 $\psi = 1$. The nominal size is set to 5%. In practice, however, the test rejects in virtually each 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.

A popular approach in empirical work to deal with noise in $\vartheta_1, \ldots, \vartheta_n$ is shrinkage estimation (see, e.g., Chetty, Friedman and Rockoff 2014). This procedure is not designed

Table 1: Variance estimation under normal noise

| | | 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 |

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, $\psi^2$, on average.

The left plots in Figure 1 provide simulation results for the distribution function $F$ for the same Monte Carlo designs. The upper-, middle-, and lower plots are for the sample size $(50,3)$, $(100,4)$, $(200,5)$, respectively. Each plot contains the true curve $F$ (black; solid) together with (the average over the Monte Carlo replications of) the naive plug-in estimator (red; dashed), the empirical distribution of the Empirical-Bayes point estimates (purple; dashed-dotted), and the analytically bias-corrected estimator (blue; solid). 95% confidence bands are placed around the latter estimator. The bandwidth in the correction term in $\check{F}$ was chosen via the cross-validation procedure discussed above. Empirical Bayes was implemented non-parametrically (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. Simulations results for a jackknife correction yielded very similar corrections and are omitted here for brevity (results for the jackknife can be found in previous versions of this paper).

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 $\check{F}$ in the middle and bottom plots. The Empirical-Bayes estimator 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, towards $\hat{F}$.[2] Only $\check{F}$ is sufficiently bias-reducing. Indeed, its confidence band settles around $F$ as the same grows. We note that, while $\check{F}$ tends to be slightly more volatile than $\hat{F}$ in small samples, the bias-reduction outweighs this in terms of root mean squared error (RMSE). Indeed, the RMSE of $(\hat{F}, \check{F})$ across the designs are $(.0969, .0816)$, $(.0756, .0578)$, and $(.0620, .0424)$, respectively.

The reduction in bias is again sufficient to bring 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 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 is observed at all deciles.

The right plots in Figure 1 provide simulation results for estimators of the deciles of $F$. The presentation is constructed around a QQ plot of the standard normal, pictured as the black dashed-dotted line in each plot. Along the QQ plot the average (over the Monte Carlo replications) of the naive estimator (red), Empirical Bayes (purple), and the (analytically) bias-corrected quantiles (blue) are shown by $*$ symbols. Confidence intervals around the latter (in blue,-o) are again equally provided. Like the naive estimator, the Empirical Bayes estimators are the appropriate order statistics of $\vartheta_1, \ldots, \vartheta_n$, after shrinkage has been

[2]Recall that the Empirical-Bayes estimator is not designed for inference on $F$ but, in stead, 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, .874 for Empirical Bayes.

Figure 1: Estimation of $F$ and $q$ under normal noise

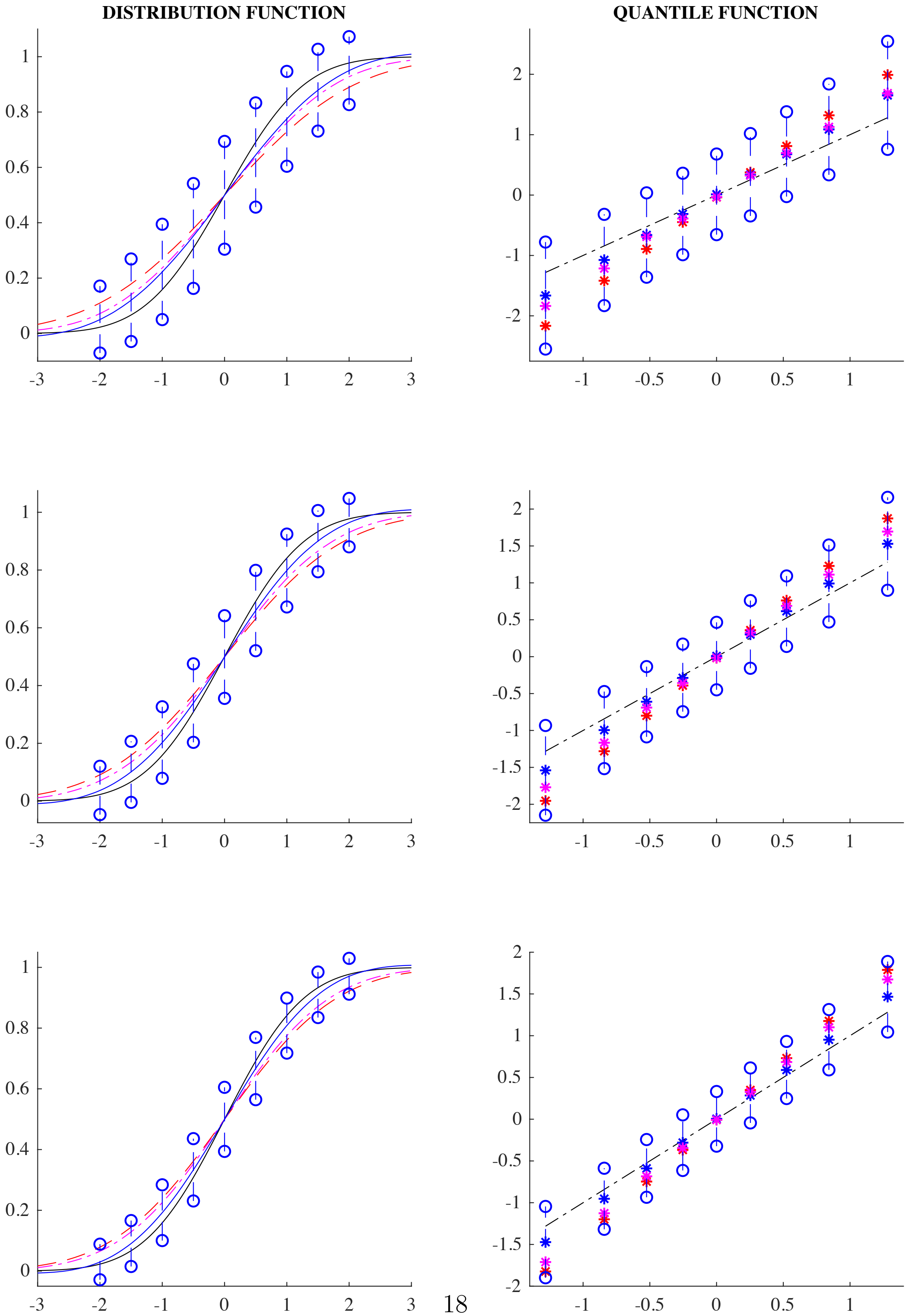


Table 2: Inference on $F$ under normal noise: empirical size

| $\tau$ | .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 |

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 adjust for bias arising from estimation noise in $\vartheta_1, \ldots, \vartheta_n$. More detailed results on inference are available in a previous version of this paper.

As said our approach does not hinge on the assumption of normal noise. To verify this we re-did the simulation exercise with logistic noise. To make all result comparable we rescale the logistic distribution so that its variance matches the one assumed prior in the normal design. The rest of the setup is unaltered. Figure 2 contains the plots for the estimators of the distribution and quantile function. Table 3 provides the empirical size of hypothesis tests on $F$ at the 5% nominal level. The layout of the figure and table fully matches those for the normal design. A glance at the output allows to verify that our corrections indeed are equally effective in this case.

## 3.2 Empirical example

As an empirical illustration we estimate a fixed-effect version of a stochastic-frontier model, as in Schmidt and Sickles (1984). We follow Belotti, Daidone, Ilardi and Atella (2013) and estimate a translog production function for milk (in liters per year) from a panel data set

Figure 2: Estimation of $F$ and $q$ under logistic noise

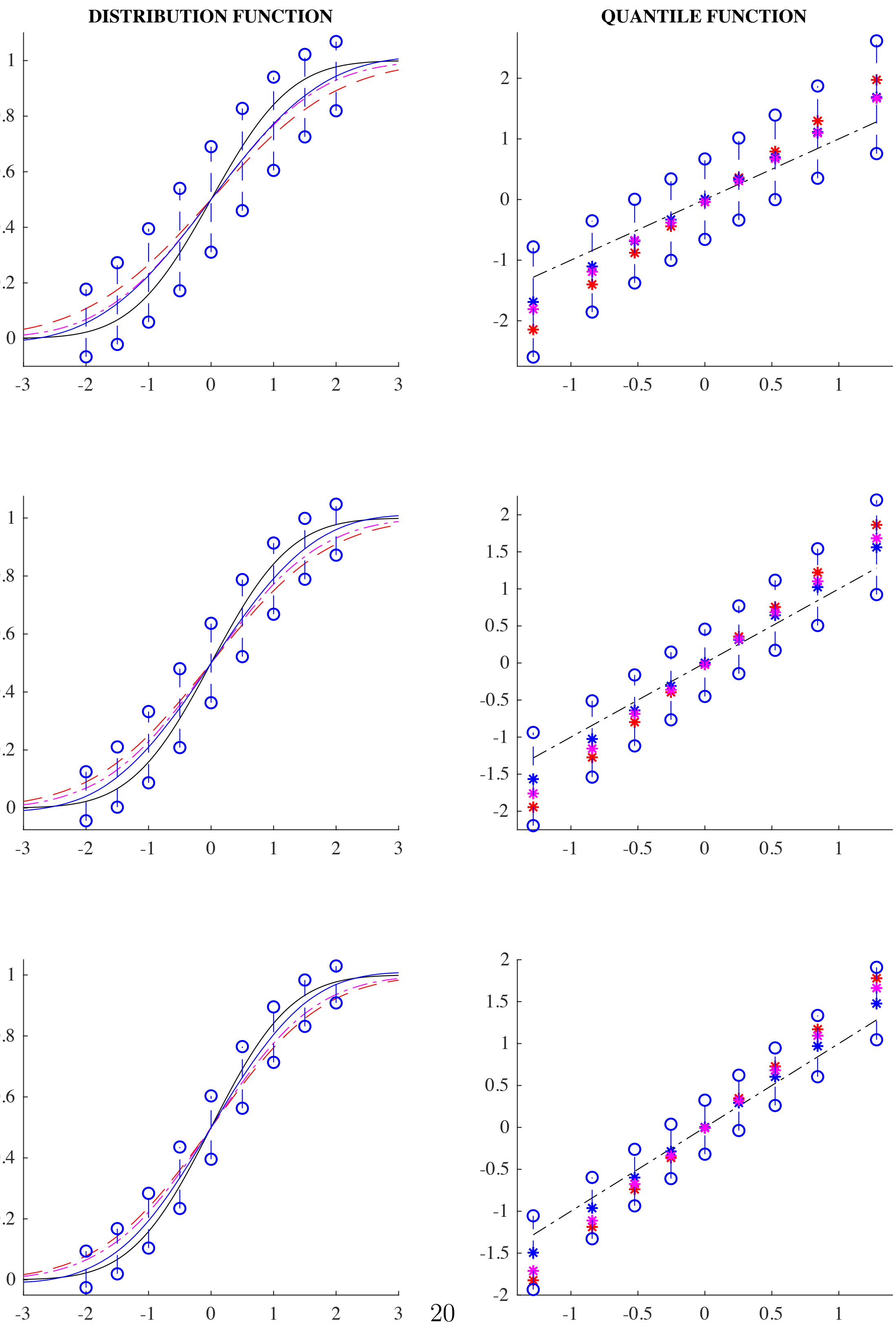


Table 3: Inference on $F$ under logistic noise: empirical size

| $\tau$ | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $(n,m)=(50,3)$ | | | | |
| $\hat{F}$ | 0.4572 | 0.5240 | 0.3474 | 0.1488 | 0.0680 | 0.1486 | 0.3460 | 0.5210 | 0.4620 |
| $\check{F}$ | 0.0694 | 0.1086 | 0.1106 | 0.0770 | 0.0582 | 0.0756 | 0.1050 | 0.1028 | 0.0692 |
| | | | | | $(n,m)=(100,4)$ | | | | |
| $\hat{F}$ | 0.6828 | 0.7223 | 0.5436 | 0.2292 | 0.0484 | 0.2268 | 0.5448 | 0.7324 | 0.6836 |
| $\check{F}$ | 0.0664 | 0.0982 | 0.1216 | 0.0796 | 0.0452 | 0.0772 | 0.1152 | 0.1041 | 0.0708 |
| | | | | | $(n,m)=(200,5)$ | | | | |
| $\hat{F}$ | 0.9176 | 0.8928 | 0.7444 | 0.3108 | 0.0564 | 0.3156 | 0.7520 | 0.8920 | 0.9124 |
| $\check{F}$ | 0.0544 | 0.0920 | 0.1124 | 0.0804 | 0.0480 | 0.0856 | 0.1292 | 0.0864 | 0.0512 |

of 247 Spanish dairy farms over the three-year period 1993-1995. The regressors are (the natural logarithm of) the number of milking cows, the number of man-equivalent units, the number of hectares devoted to pasture and crops, and the kilograms of feedstuffs fed to the dairy cows, as well as the interactions between all these inputs. Year dummies are also included to control for neutral technological change over the sampling period. Letting $y_{it}$ denote log output and $x_{it}$ the vector of all regressors the fixed-effect version of the stochastic-frontier model is

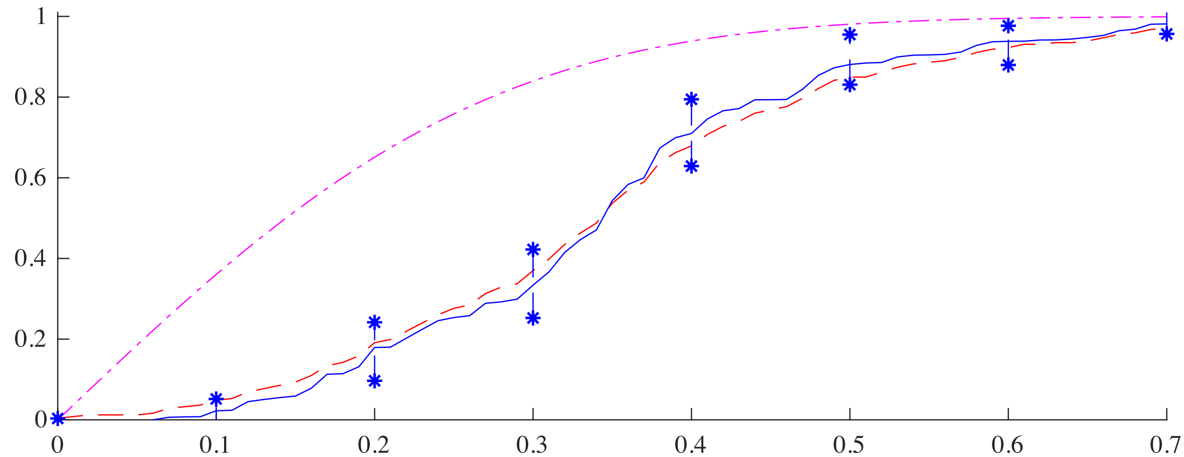
$$y_{it} = \alpha + x_{it}'\beta - \theta_i + \varepsilon_{it},$$

where $\varepsilon_{it}$ is a zero-mean normal error and $\theta_i \geq 0$ represents technical inefficiency of firm $i$. The distribution of this (in)efficiency measure is of interest. If we rewrite the above model as

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_i, \qquad \alpha_i := \alpha - \theta_i,$$

it takes the form of a standard panel data model with firm-specific effects. A common way to proceed is by taking a random-effect approach, following early work by following Pitt and Lee (1981). A default specification would assume $\alpha_i$ to follow a half-normal distribution and be independent of all the input factors in $x_{it}$. We will report the integrated-likelihood estimator for this specification below. We take a a semiparametric fixed-effect approach, as originally proposed by Schmidt and Sickles (1984). Moreover, we treat the $\alpha_i$ as parameters

Figure 3: Estimates of firm-inefficiency distribution




and estimate them by linear regression for each farm $i$. This gives the estimator $\hat{\alpha}_i$, say. We then construct the estimator

$$\vartheta_i = \max_i(\hat{\alpha}_i) - \hat{\alpha}_i$$

for the (in)efficiency parameter $\theta_i$. By doing so we are normalizing the most efficient firm in the sample as being 100% efficient. The least-squares estimator does not hinge on a normal specification for the regression errors and, for robustness, we use heteroskedasticity-robust standard errors.

Standard statistical packages report (conventional plug-in) estimates of the mean and standard deviation of the technical inefficiency measure obtained via the Schmidt and Sickles (1984) procedure. In our data, mean efficiency, $E(\theta_i)$, is estimated to be .3490 (with a standard error of .0103) and the standard deviation of $\theta_i$ is estimated as .1611 (with a standard error of .0078), respectively. Correcting the estimator of the standard deviation for the use of $\vartheta_i$ in stead of $\theta_i$ as discussed above (and allowing for cross-sectional heteroskedasticity) gives an adjusted point estimate of .1361 (with a standard error of .0092, which is slightly higher). This correction of 2.5 percentage points is subtantial relative to the standard error.

Figure 3 contains the estimated distribution of firm inefficiency. It reports the plug-in estimator (red; dashed) and its bias-adjusted version (blue; solid); the latter again comes with confidence intervals (blue; -*). As observed in the simulations, the bias-adjustment takes the form of moving-away mass from the tails of the distribution. This displacement is large relative to the estimated standard error. The figure also contains an estimate of the inefficiency distribution based on a random-effect specification with a half-normal distribution (a normal distribution folded upon itself, with its mean as turning point). The standard error of this distribution is estimated as .2136. This is much larger than the non-parametric estimates. The plot clearly shows that our non-parametric approach allows rejection of the half-normal as an appropriate parametric specification for firm inefficiency in these data.

# 4 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. These estimators are straightforward to implement. Moreover, they provide a simple and numerically stable (approximate) solution to a generalized deconvolution problem that, in addition, yields valid inference procedures.

Empirical contexts where our procedures are of direct use are regression models with fixed effects such as those in the teacher value-added literature and those used to infer stochastic production frontiers, for example. Our approach also connects to hierarchical models and, hence, can be of use in many other settings; an example is the recent literature on meta-analysis of field experiments (Vivalt 2015; Meager 2018).

To illustrate the usefulness of our work we have presented simulation results that show the vast improvement of our corrections over the commonly-used plug-in estimator and over shrinkage, which has recently been pursued in empirical work. We have equally presented an empirical application on the estimation of a stochastic frontier model for dairy farms, where our non-parametric approach allows a clear rejection of standard parametric specifications.

# References


Aigner, D. J., C. A. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*, 21–37.

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

Belotti, F., S. Daidone, G. Ilardi, and V. Atella (2013). Stochastic frontier analysis using Stata. *Stata Journal 13*, 719–758.

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

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

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

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.

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

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

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

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

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

Meager, R. (2018). Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature. Forthcoming in *American Economic Journal: Applied Economics.*

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

Okui, R. and T. Yanagi (2017). Panel data analysis with heterogeneous dynamics. Available on SSRN at http://dx.doi.org/10.2139/ssrn.2694627.

Okui, R. and T. Yanagi (2018). Kernel estimation for panel data with heterogenous dynamics. Available on arXiv.org as arXiv:1802.08825v2 [econ.EM].

Pitt, M. M. and L.-F. Lee (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics 9*, 43–64.

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

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

Schmidt, P. and R. C. Sickles (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics 2*, 367–374.

Schucany, W. and J. Sommers (1977). Improvement of kernel type density estimators. *Journal of the American Statistical Association 72*, 420–423.

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

Wang, X.-F., Z. Fan, and B. Wang (2010). Estimating smooth distribution function in the presence of heteroscedastic measurement errors. *Computational Statistics & Data Analysis 54*, 25–36.

Weinstein, A., Z. Ma, L. D. Brown, and C.-H. Zhang (2018). Group-linear Empirical Bayes estimates for a heteroscedastic normal mean. Forthcoming in *Journal of the Americal Statistical Association*.

Xie, X., S. C. Kou, and L. D. Brown (2012). SURE estimates for a heteroscedastic hierarchical model. *Journal of the American Statistical Association 107*, 1465–1479.