4-2008

Semiparametric Deconvolution with Unknown Error Variance

William C. Horrace
Syracuse University

Christopher F. Parmeter
University of Miami

This is an author-produced, peer-reviewed version of this article. The published version of this document can be found online in the Journal of Productivity Analysis (doi: 10.1007/s11123-010-0193-z) published by *Springer Link*.

Semiparametric deconvolution with unknown error variance

William C. Horrace, Syracuse University
Christopher F. Parmeter, University of Miami
Keywords: Error component, Ordinary smooth, Semi-uniform consistency

Abstract

Deconvolution is a useful statistical technique for recovering an unknown density in the presence of measurement error. Typically, the method hinges on stringent assumptions about the nature of the measurement error, more specifically, that the distribution is *entirely* known. We relax this assumption in the context of a regression error component model and develop an estimator for the unknown density. We show semi-uniform consistency of the estimator and provide an application to the stochastic frontier model.

1 Introduction

Kernel deconvolution methods are used to estimate the density of a random variate (u) when contaminated (convoluted) with an independent and additive measurement error (v). Most methods have been developed for the scenario where a random sample of observations from the contaminated variate is available ($\varepsilon = u + v$). An early treatment is Stefanski and Carroll (1990), who consider kernel estimation of a continuous and bounded target density convolved with errors from a fully-known normal density.[1] They show that convergence rates of ln(n) for the target density estimates are typical. Other kernel deconvolution treatments consider errors drawn from a fully known Laplace density; these cases generally exhibit better convergence rates. With normal errors, Meister (2006) relaxes the assumption that the variance of the normal component is known, and consistently estimates both the target density *and* the unknown variance of the normal error. His identifying assumption is that the target density is from the *ordinary smooth* family of distributions (Fan 1991a), which places a lower bound on the rate of decay of its characteristic function's tails. Examples of ordinary smooth distributions are the Laplace, exponential and gamma; a precise definition and discussion of ordinary smooth densities are provided in the sequel.

An alternative, yet rarely studied, scenario is where the contaminated variate (ε) is not directly observed, but is an additive error in a regression model (e.g., $y = \alpha + x\beta + v + u$). In econometrics the contaminated variate is a ''composed error'' and the regression a ''composed error model''. Horowitz and Markatou (1996) consider the case where panel data (repeated observations) are available and neither error component density is known.[2] Essentially, the information contained in the time-dimension of the panel replaces the normal error assumption to identify uniformly consistent estimators of *both* densities in the composed error. They first estimate slope parameters from two regression transformations (''within group'' and ''first-difference''). Second, they treat the regression residuals from the first step as if they were observations of the unobserved errors in the transformed regression models. Using standard kernel deconvolution techniques, they recover the densities of the error components from these residuals. Since their regression residuals converge in distribution to that of the composed errors at a much faster rate than ln(n), deconvolution using residuals is asymptotically equivalent to deconvolution of the regression errors. They use the results to estimate an earnings mobility model where u is time-invariant worker ability. Using data from the Current Population Survey, they show that ability appears to be normally distributed and the density of v is non-normal.

This paper considers kernel deconvolution in the cross-sectional regression model with a composed error. If a panel is not available, what can be identified? It seems intuitive that if the density of v were normal and fully-known, and if the target density of u were continuous and bounded, then it could be consistently estimated at the ln(n) rate. In a cross-section, this would amount to using the Horowitz and Markatou deconvolution estimator but with Stefanski and Carroll's assumption of the fully-known normal density replacing the information lost along the time dimension of the panel. Again, the regression *residuals* converge in distribution to the composed regression *errors*, which can be decomposed into the density of the known error and the target density. Unfortunately, in a regression model the density of the error is never fully-known. That is, the usual Gauss–Markov assumption is that the density of the error is from a zero-mean, normal family with *unknown* variance. In this paper we show that if the

[1] Stefanski and Carroll derive results for a more general convolution problem, but treat the normal error case in greatest detail.
[2] Partial knowledge of the densities is typically necessary for estimation of the regression parameters. A leading case is the random-effects model for panel data.

All code used in this paper is available upon request. The usual disclaimer applies.
W. C. Horrace Center for Policy Research, Syracuse University, Syracuse, NY 13244-1020, USA e-mail: whorrace@maxwell.syr.edu
C. F. Parmeter Department of Economics, University of Miami, Coral Gables, FL 33124-6520, USA e-mail: cparmeter@bus.miami.edu

density of v is known to be normal *up to its variance*, σ^2, then the target density can be semi-uniformly consistently estimated, if it is assumed that the density of *u* is *ordinary smooth*. Hence, our deconvolution estimator is a regression generalization of the estimator of Meister (2006) and the cross-section complement of Horowitz and Markatou (1996). Our proof of semi-uniform consistency of the density estimator involves bounding an additional variance component arising from the regression residuals.

There are myriad situations in economics where these cross-sectional deconvolution techniques are useful. For example, Postel-Vinay and Robin (2002, 2003) use deconvolution to separate the distribution of productivity levels of workers in an equilibrium search model. Cost-specific factors in auction models have also benefited from deconvolution techniques, see Li et al. (2000) and Krusnatskaya (2009). We also note that research geared towards recovering the distribution of unobservable heterogeneity (e.g., hedonic models) may benefit from deconvolution methods, see Benkard and Bajari (2005). Finally, a leading case of an error component setting is the stochastic production/ cost frontier model (Aigner et al. 1977), which is the focus of our empirical example. These models are wide-spread in studies of firm productivity and efficiency and have been used in a variety of contexts. However, one of the limitations of a cross-sectional stochastic frontier analysis (SFA) is the dependency of the methods on distributional assumptions for *both* noise (v) and inefficiency (*u*). The appeal of deconvolution methods is the ability to dispense with distributional assumptions on the inefficiency term which is the object of interest in a cross-sectional SFA. To illustrate our methods, we re-analyze the technical inefficiency distribution of a random sample of US banks previously used in Kumbhakar and Tsionas (2005).

The paper is organized as follows. In Sect. 2 we provide a brief tour of deconvolution methods in statistics. Section 3 discusses issues with deconvolution inherent to our problem, as well as the assumptions needed to show semiuniform consistency of the estimator. Section 4 contains some Monte Carlo results, followed by a stochastic frontier application. Directions for future research and conclusions are in Sect. 5.

2 Deconvolution: the state of the art

While estimation strategies relating to measurement error have existed for quite some time, deconvolution techniques were introduced by Mendelsohn and Rice (1982). They use B-splines to deconvolve the number of live cells from the number of dead cells in DNA content. Kernel estimation techniques, or deconvolving kernels, were proposed by Stefanski and Carroll (1990). Even before their paper was published, Carroll and Hall (1988) showed that for the class of deconvolving estimators, the best possible rate of convergence in the presence of normal measurement error is logarithmic.

Other contributions to the asymptotic theory of deconvolution estimators from a kernel perspective are due to Devroye (1989), Liu and Taylor (1989) and Fan (1991a, b, c, 1992, 1993). Bandwidth selection issues are considered in Barry and Diggle (1995), Hesse (1999) and Delaigle and Gijbels (2002, 2004a, b). Other practical issues relating to kernel deconvolution are in Zhang and Karunamuni (2000) for boundary corrections, in Neumann (1997) for the estimation of the unknown measurement error as opposed to assuming its family, in Hesse (1995) for the case when the data of interest are dependent, and in Hesse (1996) for the case when only some of the data are measured with error.

A recent spate of papers on theoretical and computational aspects of deconvolution estimators has rekindled interest in the area. Matias (2002), Butucea and Matias (2005) and Meister (2006) propose estimators for the unknown variance of the known noise distribution under various sets of assumptions and develop the corresponding rates of convergence of the deconvolution estimator. Meister (2004a, b) develops procedures for testing whether the assumed measurement error density is correct while Delaigle and Meister (2007, 2008) extend the homoscedastic variance setting to allow for heteroscedasticity (in both errors-in-variable regression and density settings). Carroll and Hall (2004) develop a deconvolution estimator that involves information on the moments of the noise distribution (as opposed to assuming the family explicitly) which greatly improves accuracy, and Hall and Qui 2005) consider a trigonometric expansion for deconvolution that is simpler to implement than kernel methods. Delaigle and Gijbels (2007) discuss key issues with calculating the integrals arising in deconvolution settings. Delaigle and Hall (2007) and Delaigle (2008) discuss issues associated with optimal kernel choice and the appropriateness of assuming a diminishing error variance as the sample size grows. Also, Delaigle et al. (2008) develop an estimation strategy when replication copies of the noise are present. Taken as a whole these series of papers represent the state of the art in the statistics literature.

3 The convolution problem

Consider the specification:

$$\varepsilon_j = v_j + u_j, \tag{1}$$

$$y_j = x_j'\beta + \varepsilon_j, \quad j = 1, \ldots, n, \tag{2}$$

and β is a parameter vector of dimension k to be estimated. Here j indexes observations. Equation 1 is the classic deconvolution problem where ε is observed. Equation 2 is complicated by the fact that the errors are unobserved and have to be estimated. This specification appears in a variety of econometric settings. For example, if $x_j'\beta$ characterizes a production function and $u_j < 0$, then this would be a standard stochastic production frontier model. In a panel setting if it is assumed that x_j is uncorrelated with u_j then (2) would be the usual random-effects model. The key difference between (1) and (2) is that we have direct observations on ε_j in (1), while we must estimate ε_j in (2). We make the following assumptions on the random components of the model and the covariates when present.

Assumption 3.1

The x_j, v_j and u_j are pairwise independent for all $j = 1, \ldots, n$. Let the probability densities of the error components be $f_v(z)$, $f_u(z)$ and $f_\varepsilon(z)$ with corresponding characteristic functions $h_v(\tau)$, $h_u(\tau)$, and $h_\varepsilon(\tau)$. Based on the independence between v_j and u_j in Assumption 3.1,

$$h_\varepsilon(\tau) = h_v(\tau)h_u(\tau). \tag{3}$$

We restrict our attention to densities that satisfy the following two assumptions.

Assumption 3.2 The distribution of v is a member of the normal family with zero mean and unknown variance, i.e. $\mathcal{F} = \{N(0, \sigma^2); \sigma^2 > 0\}$.

Assumption 3.3 The distribution of u is a member of the family of ordinary smooth densities, i.e. $\mathcal{F}_u = \{u \text{ density}; C_1|\tau|^{-\delta} \le |h_u(\tau)| \le C_2|\tau|^{-\delta}, \forall |\tau| \ge T > 0\}$. Here we have $C_2 > C_1 > 0$ and $\delta > 1$.

Assumption 3.2 is standard and restricts v to the class of normally distributed random variates with mean 0 and unknown variance σ^2. Assumption 3.3 dictates tail behavior of the characteristic function of u. The class of ordinary smooth densities was first defined by Fan (1991a) and implies that the density of u is absolutely continuous. The upper bound is used when examining uniform consistency while the lower bound ensures the rate of decay of the tails of the characteristic function does not approach zero too rapidly and is needed for identification. The constants C_1 and C_2 are irrelevant for large T while δ ensures polynomial tail behavior and includes a wide array of densities. Polynomial tails of a characteristic function decay slower than exponential tails, thus precluding a normal target density in this class. This ensures unique identification of the variance of the normal noise distribution.

Examples of distributions that fall within the ordinary smooth family are the Laplace, gamma and exponential. Both the gamma and exponential densities have been employed in stochastic frontier applications. The most common distribution assumed for inefficiency in stochastic frontier model is the truncated normal distribution (however, this largely appears to be an assumption of convenience). This class of densities is not ordinary smooth and is excluded by Assumption 3.3. However, our requirement for the density to belong to the class of ordinary smooth densities is for excluding the non-identified case of a normal– normal composed error.

Assuming that the v random variates are from the normal family guarantees that they possess a nonzero characteristic function everywhere. Under Assumptions 3.2 and 3.3, the Fourier inversion formula identifies the first derivative of the distribution of u, which equals the density of u,

$$f_u(z) = \frac{1}{2\pi}\int e^{-i\tau z + \frac{1}{2}\sigma^2\tau^2} h_\varepsilon(\tau)d\tau, \tag{4}$$

Where $i = \sqrt{-1}$, see Lukacs (1968, p. 14). Meister (2006) has shown that to estimate σ^2 one loses the ability to estimate $f_u(z)$ uniformly consistently. He shows that one can estimate $f_u(z)$semi-uniformly consistently in the sense that for a given density in $\mathcal{F}_v$ a deconvolution estimator is uniformly consistent, but not uniformly consistent over all densities within $\mathcal{F}_v$. This is the price one pays by not knowing the variance. See Meister (2003) for more on concepts related to semi-uniform consistency.

If he were known we could, using Eq. (4), recover the density of u, but it is not, so we rely on its empirical characteristic function,

$$\widehat{h}_\varepsilon(\tau) = \frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j}. \tag{5}$$

Unfortunately, ε_j is unobserved in (2), but we can estimate it by consistently estimating β. That is, for some consistent estimate β_n, define residuals $\hat{\varepsilon}_j = y_j - x_j'\beta_n$. The precise nature of the consistent estimate, β , is discussed later. We will use the empirical characteristic function of the residuals which is defined as

$$\widehat{h}_{\hat{\varepsilon}}(\tau) = \frac{1}{n}\sum_{j=1}^{n} e^{i\tau\hat{\varepsilon}_j}. \tag{6}$$

Replacing h_ε with $\widehat{h}_\varepsilon$ or $\widehat{h}_{\hat{\varepsilon}}$ in Eq. (4), does not ensure that the integration will exist, so we convolute the integrand with a smoothing kernel.[3] Define a random variable Z with the usual Parzen (1962) kernel density $K(z)$ and corresponding (invertible) characteristic function $h_K(\tau)$. The characteristic function, $h_K(\tau)$, must have finite support to ensure that the integration exists and that the resulting estimate represents a density function.

Using $K(z) = (\pi z)^{-1}\sin(z)$, $(h_K(\tau) = 1\{|\tau| \le 1\})$, our estimator of the density of u is,

$$\widehat{f}_u(z) = \frac{1}{2\pi}\int_{-1/\lambda_n}^{1/\lambda_n} e^{-i\tau z + \frac{1}{2}\sigma_n^2\tau^2}\widehat{h}_{\hat{\varepsilon}}(\tau)d\tau, \tag{7}$$

where the limits of integration are a function of a sequence of positive constants $\lambda_n = \frac{\ln k_n}{k_n}$ which represent the degree of smoothing, while $k_n = \sqrt{\frac{\ln n}{\ln(\ln n)}}$, also a sequence of constants. The variance estimator is defined as

$$\hat{\sigma}_n^2 = \begin{cases} ll0, & \text{if } \tilde{\sigma}_n^2 < 0 \\ \tilde{\sigma}_n^2, & \text{if } \tilde{\sigma}_n^2 \in [0, \sigma_n^2] \\ \sigma_n^2, & \text{if } \tilde{\sigma}_n^2 > \sigma_n^2, \end{cases} \tag{8}$$

where $\tilde{\sigma}_n^2 = -2k_n^{-2}\ln\left(\frac{|\hat{h}_{\hat{\varepsilon}}(k_n)|}{C_1 k_n^{\delta}}\right)$ and $\sigma_n^2 = \ln(\ln(n))$,[4] Here $\delta > 1$ and $C_1 > 0$ are arbitrary. They should correspond to the parameters of the true density in 3.3, however, Meister (2006) shows that inappropriate choices of these constants have negligible effect on the performance of the estimator.

To show that the unknown variance deconvolution estimator retains its asymptotic properties when the composed error is estimated we provide two additional conditions that will be useful in the Lemmas and Theorem to follow.

Assumption 3.4 The distribution of *x* has bounded support.

Assumption 3.5 Our estimator of β, β_n is $\sqrt{n}$-consistent. That is, $\sqrt{n}(\beta_n - \beta) = O_p(1)$ as $n \to \infty$, for an estimator β_n.

Assumption 3.4 follows Horowitz and Markatou (1996) while Assumption 3.5 guarantees that the random sampling error between the composed errors and the estimated ones is asymptotically negligible. Equation (2) can be $\sqrt{n}$-consistently estimated with Ordinary Least Squares (OLS) if *u* has zero mean. If not, then any intercept in the parameter vector, β, cannot be consistently estimated, and the target density, *u*, can only be estimated up to location. We discuss this at length in the sequel.

The following lemmas will be used to establish semi-uniform consistency of $\hat{f}_u$.

Lemma 3.1 *For Assumptions 3.1, 3.3–3.5 and* $\mathcal{F}_v = \{N(0,\sigma^2) : \sigma^2 \in (0, \sigma_n^2]\}$, *the Mean Integrated Squared Error (MISE) of 7 is*

$$\sup_{f_v\in\mathcal{F}_v}\sup_{f_u\in\mathcal{F}_u} E_{f_v f_u}\left\|\hat{f}_u - f_u\right\|_{L_2}^2 \le B + V_1 + V_2 + E, \tag{9}$$

where $B \le const.\lambda_n^{2\delta-1}$, $V_1 \le const.(n\lambda_n)^{-1}e^{\sigma_n^2/\lambda_n^2}$, $V_2 \le const.(n\lambda_n^3)^{-1}e^{\sigma_n^2/\lambda_n^2}$ and

$$E \le const.\sup_{f_v\in\mathcal{F}_v}\sup_{f_u\in\mathcal{F}_u}\lambda_n^{-1}\left(\int_{-1}^{1}|h_u(s/\lambda_n)|^2 s^4 ds + \exp(\sigma_n^2/\lambda_n^2)\int_{-1}^{1}|h_u(s/\lambda_n)|^2\cdot P_{f_v f_u}(|\hat{\sigma}_n^2 - \sigma^2| > d_n)ds\right),$$

where $d_n = 2\ln(2C_2/C_1)\lambda_n^2$.

[3] See Stefanski and Carroll (1990).

[4] Meister (2006) introduces this truncation device on the variance estimator, however, as the sample size grows this truncation becomes irrelevant.

Notice the distinction between $\mathcal{F}$ in Assumption 3.2 and $\mathcal{F}_v$ above. The latter is the family of normal distributions that involves an upper bound on the variance and is a subset of the former.[5] It turns out that the bound on B and the first integral of the bound on E converge slowly and determine the convergence rates of the estimator. Since these bounds are identical to those of Meister (2006), the convergence rates are also identical. The other components of the bound (V_1, V_2, and the second integral in the bound on E) all involve an inverse power of n and converge relatively quickly. The V_2 component of the bound does not appear in the bound of Meister (2006) and arises from the estimation of the regression function. Clearly, its bound converges faster than that of V_1. The second integral on the bound of E is identical to that of Meister (2006), but the estimation of the regression function causes ours to converge more slowly. Ultimately, this is unimportant as there is an inverse n that dominates this component as we shall see in the next lemma.

B is a bias component of the estimator, and V_1 and V_2 are variance components, the bounds of which exhibit the usual bias-variance trade-off in non-parametric density estimation. As the bandwidth goes to zero, the bound on the bias (B) is decreasing, while those on V_1 and V_2 are increasing. Of course, the inverse n in the bounds on V_1 and V_2 dominate and cause these terms to go to zero. The E is a hybrid bias-variance term. The second integral of its bound behaves like a variance, while the first integral behaves like a bias-variance hybrid. It is ultimately decreasing in the bandwidth like a variance, but it does not rely specifically on an inverse n for its convergence, rather it depends on the tails of the characteristic function as well shall see.

Lemma 3.2 *Let d_n and $\mathcal{F}_v$ be as in Lemma 3.1. Then*

$$\sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} P_{f_v f_u}\left(|\hat{\sigma}_n^2 - \sigma^2| > d_n\right) \leq const.k_n^{2\delta} \exp\left(\sigma_n^2 k_n^2\right) \times \left(1 + k_n^2\right)/n. \qquad (10)$$

Given that we have to replace e with an estimate, our lemma differs from Meister (2006) through the addition of $k^2{}_n$. However, the presence of the inverse n dominates the logarithmic structure of k_n. Therefore, the increase in this bound over Meister's will not affect the optimal rates of convergence in the following theorem. This is to be expected since the convergence rates for deconvolution estimators are extremely slow. In fact, it is possible to take our analysis a step further and allow for our unknown regression function to be estimated nonparametrically since the convergence rates here are still polynomial in n.

Lemmas 3.1 and 3.2 can be used to show

Theorem 3.1 *Our deconvolved kernel density estimate (7), for any $f_v \in \mathcal{F}_v$ has*

$$\sup_{f_u \in \mathcal{F}_u} E_{f_v f_u}\left\|\hat{f}_u - f_u\right\|_{L_2}^2 \leq \begin{cases} (\ln(\ln(n)))^{3\delta-1.5}(\ln(n))^{0.5-\delta}, & \text{if } \delta < 2.5 \\ (\ln(\ln(n)))^{6}(\ln(n))^{-2}, & \text{if } \delta = 2.5 \\ (\ln(\ln(n)))^{7}(\ln(n))^{-2}, & \text{if } \delta > 2.5, \end{cases} \qquad (11)$$

The rates here are identical to those found in Meister (2006) and show the difficulty of deconvolution with normal error.[6] All of the rates, regardless of δ, are powers of logarithmic and iterated logarithmic terms of the sample size. These rates are, however, optimal for deconvolution with normal measurement error (see Carroll and Hall 1988).

It is worth mentioning again that the reason we achieve the Meister (2006) rates is that the residuals converge in distribution faster than the density deconvolution estimator converges. This happens because the terms affecting the speed of convergence of our estimator depend on B and the first integral of the bound on E in Lemma 3.1. However, the residuals only show up in the additional component of the bound on V and the second integral of the bound on E. Both of these terms contain a $1/n$ which allows them to go to zero faster than the remaining terms on the upper bound of the *MISE*. Thus, even though we have to estimate the errors, $\sqrt{n}$-consistent estimation

[5] This bounding of the variance in the class of normal distributions is what leads to semi-uniform consistency as opposed to uniform consistency. That is, uniform consistency only holds for this bounded class.

[6] Our rates are slightly different that those in Meister (2006), Theorem 2 to correct for a typo there.

guarantees a semi-uniformly consistent estimator. This is essentially what occurs in the deconvolution estimator of Horowitz and Markatou (1996) who also consider deconvolution in a regression context.

If one were to assume an ordinary smooth distribution for *v* (such as Laplace or twice convolved Laplace) then the rates would be polynomials of the sample size which are noticeably faster. As Meister (2004b) has shown, it may prove fruitful to perform deconvolution under the assumption of Laplacian error as the loss associated with assuming normal measurement error, when in fact it should be Laplacian, is finite.[7] However, assuming normal measurement error in the presence of Laplace error results in *infinite* loss. That is, if the true measurement error is Laplace, but one erroneously performs deconvolution assuming that the measurement error is normal, then the *MISE* goes to infinity as the sample size increases.

4 Performance of the method

4.1 Small sample properties

We draw from several other deconvolution simulation studies to examine the small sample properties of our estimator.[8] We consider sample sizes of 200 and 1,000.

Our model is

$$y_i = 4 + 3x_i + v_i + u_i \tag{12}$$

The x_is are generated from a standard normal. The u_is are generated from the two times self-convoluted, zero-mean Laplace density.[9] To determine the impact of the noise variance we generate vi from a zero-mean normal density with variance equal to 1 or 4. This implies that our signal to noise ratio is either 1 or 0.25.

Figures 1, 2, 3, 4 show the results for four simulations under each setting. The dotted line is the true twice convolved Laplacian density labelled Actual in the legends), the solid line is the Meister (2006) estimator using ε (labelled M in the legends), and the dashed line is the estimator discussed in this paper using $\hat{\varepsilon}$ (labelled ME in the legends). We can see that both our estimator and the Meister estimator provide a similar estimate of the density.

Figure 1 shows four individual runs for n = 200 and variance equal to 1. The estimated variances using the known variances are {0,0,0,0} and {0,0,0,0.229} using the estimated residuals. In Fig. 2 we have n = 200 but with measurement error variance equal to 4. Again, we see the fit of both estimates is similar but poor relative to those in Fig. 1 due to a decrease in the signal to noise ratio. The estimated variances in this setting are {0,0.390,0.417,0} and {0,0.417,0.258,0} for the known and estimated residuals, respectively.

Performing this analysis with a sample size of 1000 further illustrates the superiority of the $\sqrt{n}$ prate of convergence of the residuals as opposed to the logarithmic rates for the deconvolution density estimator. In Fig. 3 we have measurement error variance of 1 and the estimated variances are {0,0.227, 0.141, 0} and {0,0.330,0.186,0} for the known and estimated errors, respectively. The estimators are indistinguishable. Moving to the lower signal to noise ratio setting with measurement error variance equal to 4, we see that the fit of both estimators has degraded and yet they remain almost identical throughout the range of the simulated data. Here our estimated sets of variances are {0, 0.295, 0, 0.042} and {0,0.418,0,0} for the known and estimated errors, respectively.

Notice that there is a high occurrence of zero estimates for the unknown variance. This is an unresolved issue for these estimators. Also, the density estimates can be negative in certain regions and tend to fit the target density better in the tails than in the center of the distribution where a majority of the mass is present. These are unavoidable characteristics of deconvolution estimators. Also, these pictures depict a set of four runs and therefore are likely to be impacted by random sampling.

One point worth mentioning is that the variance estimate is impacted by the sample size as well as specification of C_1 and δ Given a zero estimate of the variance, a procedure to make it minimally positive would be to change C_1 and/or δ until it becomes nonzero. Since these choices do not affect the asymptotic performance of the estimator, this seems a reasonable strategy. However, it is also worth mentioning that interest centers on the unknown density and not consistent estimation of the noise variance. Thus, the occurrence of a zero variance estimate is not too troubling. We mention that developing a positive variance estimate is a fruitful avenue for further research.

4.2 Stochastic frontier application

[7] Here loss is taken to be MISE.

[8] Meister (2006) and Stefanski and Carroll (1990).

[9] The standard Laplace density has the form $L(x) = (2b)^{-1} e^{-|x|/b}$, where b is the scale parameter, while the twice convolved Laplace density iS $\bar{L}(x) = (4b^2)^{-1} e^{-|x|/b}(|x| + b)$ We choose b so that this density has variance 1, which corresponds to b = 1. In this setting it is known that C_1 = 1/4 and δ = 2. We are not concerned with C_2 as it has no bearing on any calculations for the estimator.

A typical stochastic frontier model (e.g., Aigner et al. 1977) is given in Eq. (2), but with $u < 0$ for a production function or u [0 for a cost function. Given distributional assumptions on inefficiency, u (e.g., exponential or gamma), and on noise, v (e.g., normal up to parameters), β can be consistently estimated and used to calculate the conditional distribution of firm-level inefficiency, which is typically characterized by the empirical distribution of u conditional on ε (e.g., Jondrow et al. 1982). While normality of v is widely accepted as plausible, the distribution of u has been debated in the literature (Greene 1990). Perhaps surprisingly little has been published on hypothesis testing of these distributional of choices (e.g., Wang et al. 2008). Our semi-uniformly consistent estimator of the distribution of u can provide insight into the shape of the distribution of u and help guide parametric choices. What follows is intended for readers familiar with the literature, a textbook treatment of which is provided in Kumbhakar and Lovell (2000). Applying the deconvolution estimator to the stochastic frontier model requires only some modest extensions to our main results. First, Assumptions 3.1 and 3.2 are standard in any parametric stochastic frontier model. However, if the distribution of u is truncated normal in the population (a typical assumption), then Assumption 3.3 is violated, and a consistent estimate of the distribution is not forthcoming. If the distribution is gamma or exponential (among others),

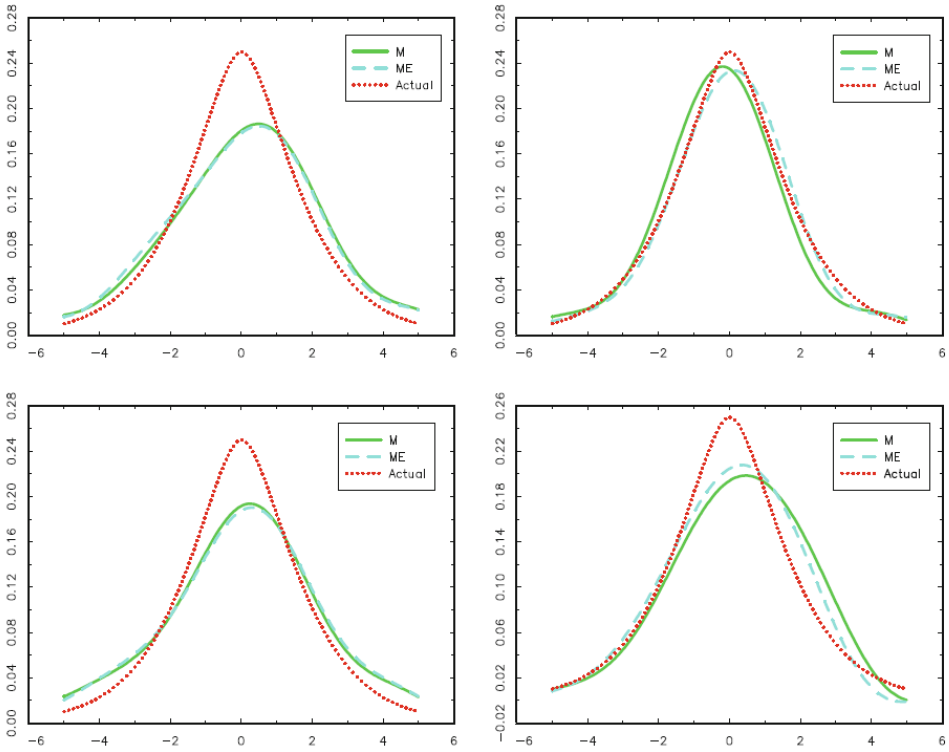


Fig. 1 Single runs for $n = 200$ and $\sigma^2 = 1$

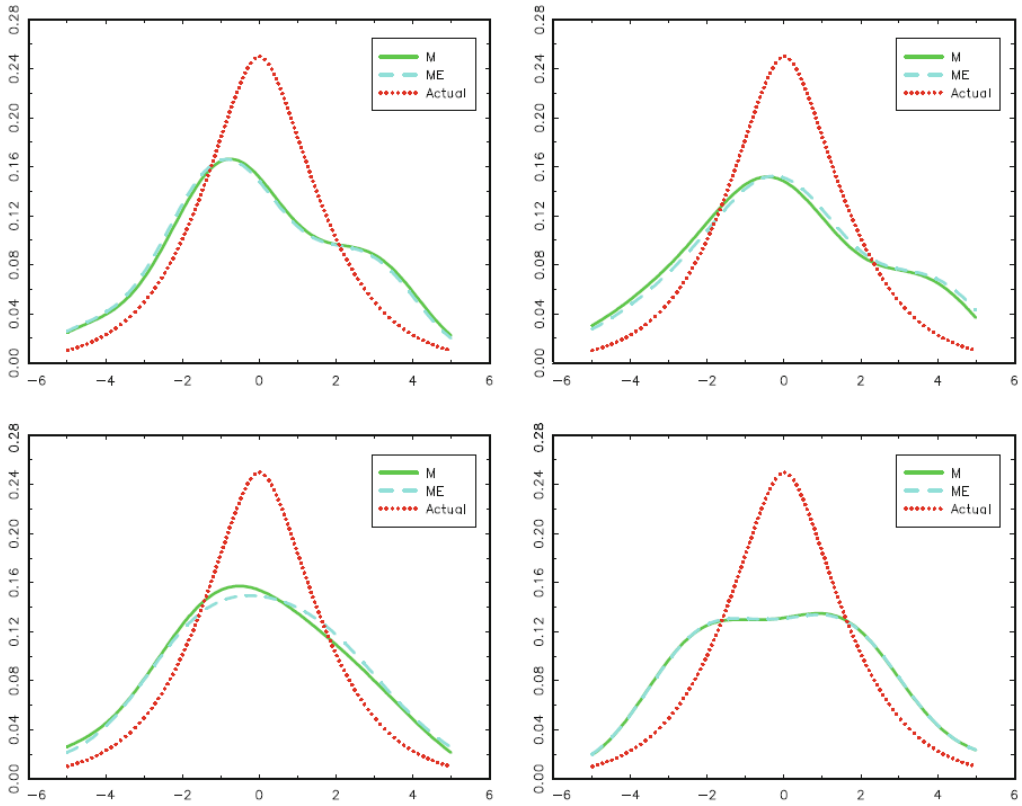


Fig. 2 Single runs for $n = 200$ and $\sigma^2 = 4$

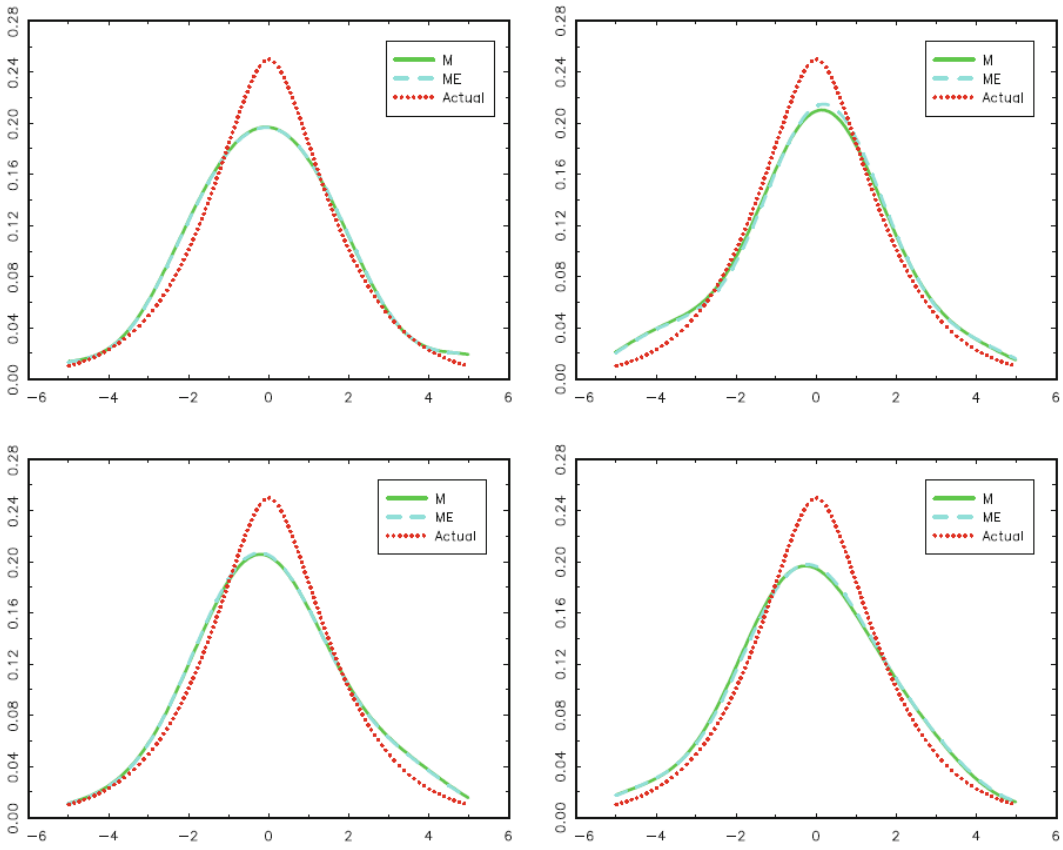


Fig. 3 Single runs for $n = 1{,}000$ and $\sigma^2 = 1$

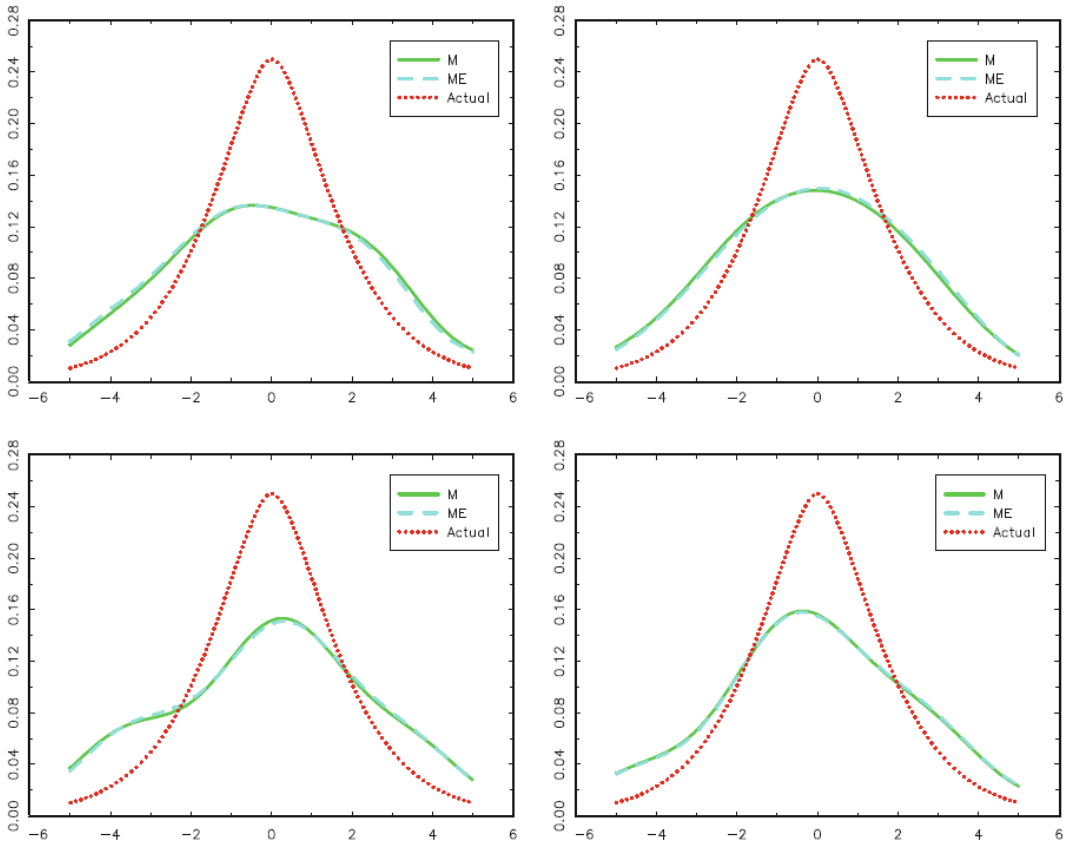


Fig. 4 Single runs for $n = 1{,}000$ and $\sigma^2 = 4$

then it is ordinary smooth and the assumption holds. Second, in a typical stochastic frontier model u has a non-zero mean, so OLS on Eq. 2 will be inconsistent for the regression intercept (violating Assumption 3.5) and will produce residuals: $\hat{\varepsilon}_j^* = \hat{\varepsilon}_j - E(u)$, where E($u$) ≠ 0. Substituting $\hat{\varepsilon}_j^*$ for $\hat{\varepsilon}_j$ everywhere in Sect. 3, yields a consistent estimate of the distribution of $\hat{u}_j^* = \hat{u}_j - E(u)$ so the distribution of u can be estimated up to location. This is fine for purposes of trying to understand the shape of the distribution of u. Third, the deconvolution estimator assumes the distribution of u has support everywhere, but the assumptions of the stochastic frontier model imply that the distribution of u will have a jump discontinuity at u = 0. However, Hall and Simar (2002) develop a procedure for detecting this jump discontinuity, which occurs at -$E(u)$ in the support of the distribution of u*. Therefore, their procedure can be used to estimate $E(u)$, which allows relocation of the estimated distribution of u* to that of u.

The estimator proposed by Hall and Simar (2002) is quite intuitive. If we observed data generated by a convolution where one random variable has a jump discontinuity (as does our inefficiency variable),[10] say μ, while the other is continuous almost surely, then the point of this discontinuity causes a severe change in the derivative of the convoluted density. Define the ordered regression residuals, $\hat{\varepsilon}_{(1)} \leq \cdots \leq \hat{\varepsilon}_{(n)}$ then Hall and Simar propose to estimate the jump discontinuity by

$$\hat{\mu} = \underset{x \in N(\hat{\varepsilon}_{(\ell)})}{argmax} \left| \widehat{f'}_{\hat{\varepsilon}}(x) \right|, \tag{13}$$

where $N(\hat{\varepsilon}_{(\ell)})$ is a neighborhood around either the left $(\ell = 1)$or right tail $(\ell = n)$of the distribution where the jump discontinuity exists. That is, while a local maximum of the derivative of the kernel density estimate can occur anywhere, Hall and Simar (2002) suggest explicitly looking in a region where the jump discontinuity is likely to

[10] The jump discontinuity occurs due to the fact that no probability is assigned to points to the left of the boundary, assuming the boundary is a lower boundary

occur. So, for a jump discontinuity appearing on the right hand side of the distribution (as in the production frontier setting) we search near the *n*th order statistic $\hat{\varepsilon}_{(n)}$, whereas if the jump discontinuity appears on the left hand side of the distribution (as in the cost frontier setting) we search around the first order statistic $\hat{\varepsilon}_{(1)}$. A kernel density estimator and its associated derivative is used to construct $\widehat{f'}_{\hat{\varepsilon}}(x)$.

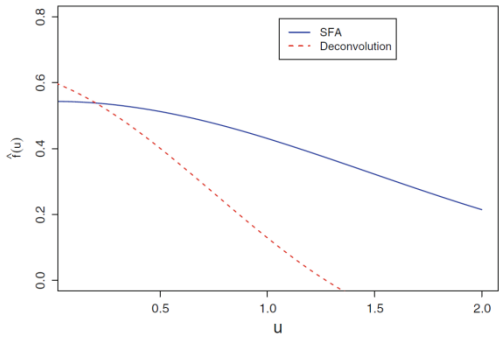
Hall and Simar (2002) show that as $n \to \infty, \hat{\mu} = \mu + o(\sigma_v^2)$. Notice that this bias does not diminish as the sample size increases. They provide conditions to further reduce the bias, however, unless one is willing to assume the the variance of the noise is diminishing as the sample increases, one can do no better than a biased estimate of the jump discontinuity. In simulations they show that the bias is actually quite small and so this estimator shows promise as a viable means of estimating the boundary of the distribution of *u*. Bandwidth selection for the kernel density estimate and the selection of the neighborhood are discussed in Hall and Simar (2002) and Delaigle and Gijbels (2006a, b).

The data come from Kumbhakar and Tsionas (2005) and are a random sample of 500 commercial banks taken from the commercial bank and bank holding company database managed by the Chicago Fed.[11] The banks are observed over the 1996–2000 time period giving us 2,500 total observations (we ignore panel aspects of the data for our application).[12] We have data on five outputs and five input prices. Our outputs are installment loans (*Y1*), real estate loans (*Y2*), business loans (*Y3*), federal funds sold (*Y4*) and other assets (*Y5*). The input variables are labor (*W1*), capital (*W2*), purchased funds (*W3*), interest-bearing deposits in total transaction accounts (*W4*) and interest-bearing deposits in total non-transaction accounts (*W5*). We impose linear homogeneity by normalizing cost and input prices by dividing by the price of input five. For simplicity we estimate a Cobb Douglas cost function as

$$\ln(C/Y5)_{jt} = \alpha_0 + \sum_{k=1}^{5} \alpha_j \ln(Wk/Y5)_{jt} + \sum_{k=1}^{4} \beta_k \ln(Yk/Y5)_{jt} + \alpha_t T + v_{jt} + u_{jt}, \qquad (14)$$

where *T* is a time trend and *j* indexes observations.

Ignoring the panel structure of the model and assuming that *u* is positive half normal and *v* is normal, we estimate Eq. (14) using maximum likelihood, yielding estimates for σ_v and σ_u of 0.730 and 1.840, respectively (σ_u is the standard deviation of u prior to truncation below zero). The estimate of σ_u implies an estimate for E(*u*) of 1.468. Other than ignoring the panel structure, this is a typical stochastic frontier analysis (SFA). A plot of the assumed half normal density is the solid line (SFA) in Fig. 5. To implement the deconvolution estimator we estimate the model in (14) by OLS, which imposes no shape assumptions on the distributions of *u* or *v*. Selecting δ = 1.5 and C_1 = 3 and using the rule-of-thumb bandwidth of Meister (2006), the deconvolution estimator yields an estimate for σ_v of 0.333 and returns the density of $u^* = u$ - E(*u*). Using the procedure of Hall and Simar (2002, p. 412) with a bandwidth of 0.256 we detect a jump discontinuity at u^*= -0.3037, implying an estimate $\hat{\mu} = \hat{E}(u) = 0.3037$.[13]



[11] The data are publicly available on Bill Greene's webpage at http://pages.stern.nyu.edu/*wgreene/Econometrics/banks.xl.
[12] This is nearly as large as the sample of long-term daily saturated fat intake in women used in Stefanski and Carroll (1990), which had 2,888 total observations.
[13] Hall and Simar (2002) point out that their estimator works well in low noise settings. An alternative would be to use the boundary estimator of Delaigle and Gijbels (2006a, b).

The jump discontinuity estimate is used to relocate the density of u^* to that of u. Our estimate of the density of u is the dashed line (Deconvolution) is Fig. 5. Notice that it has a much thinner right-tail than the positive half normal density (solid line), and this is reflected in its lower estimated mean (compare to 0.3037 to 1.468 of the half normal).[14] Visual inspection of the densities in Fig. 5 indicates that they are quite different (obviously a Kolmogorov-Smirnoff test could be preformed, but our purpose is simply to demonstrate the feasibility of the deconvolution estimator, so we did not do it). The observed differences in the densities suggests that a half normal assumption for this particular data set and model, may be incorrect.

5 Conclusions

This paper proposes a semiparametric estimator for crosssectional error components models. Our estimator is Semiparametric as it hinges on a distributional law for one of the components. This assumption is tempered by allowing for an unknown variance using the recent methods proposed by Meister (2006). Our finite sample results show that a $\sqrt{n}$ estimator of the convolved errors does not degrade the consistency or the rates of convergence of the density estimate when compared to deconvolution based on direct observation. This is intuitive given that the errors are estimated at the parametric $\sqrt{n}$ rate while deconvolution estimators typically possess a logarithmic rate. We have provided an application to the stochastic frontier model, a natural candidate for future empirical exercises. The utility of the density estimator to the stochastic frontier literature should be clear. Overall the possibilities for this estimator are multifarious. A test against known parametric densities and extensions to calculating conditional densities and expectations are worthwhile extensions of our results. Research in this area has been somewhat limited. See Wang, Amsler and Schmidt (2008) and Horrace (2009) for a test of correct specification in SFA models and Wang and Schmidt (2009) for insights into the distribution of inefficiency. As previously stated, developing a positive variance estimate is also an interesting extension of this work.

Acknowledgments The authors thank Joel Horowitz, Nicolai Kuminoff, Hailong Qian, Peter Schmidt and Leonard Stefanski.
Comments from the associate editor and two anonymous referees as well as seminar participants at Binghamton University, Syracuse
University, the University of Richmond and Virginia Tech as well as EWEPA X, European Meeting of the Econometric Society, Milan,
Italy 2008 and NAPW 2008 are gratefully acknowledged.

Appendix: Technical proofs

Proof of Lemma 3.1:

Following Meister (2006, Proof of Lemma 1) we have:

$$\begin{aligned}
&\sup_{f_v\in\mathcal{F}_v f_u\in\mathcal{F}_u} E_{f_v,f_u}\left\|\widehat{f}_u - f_u\right\|^2_{L_2(\mathbb{R})} \\
&\le (2\pi)^{-1}\left\{ \sup_{f_v\in\mathcal{F}_v f_u\in\mathcal{F}_u} 2\int_{1/\lambda_n}^{\infty} |h_u(\tau)|^2 d\tau \right. \\
&+ \sup_{f_v\in\mathcal{F}_v f_u\in\mathcal{F}_u} 2\int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u}\left|\exp\left(\widehat{\sigma}_n^2\tau^2/2\right)\left[\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau)\right]\right|^2 d\tau \\
&+ \sup_{f_v\in\mathcal{F}_v f_u\in\mathcal{F}_u} 2\int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u}\left|h_\varepsilon(\tau)/\left(\exp\left(-\widehat{\sigma}_n^2\tau^2/2\right)\right)\right. \\
&\left.\left. - h_u(\tau)\right|^2 d\tau \right\}.
\end{aligned}$$

The first addend represents the bias, does not depend on the fact that the convoluted errors are estimated and can be bound as in Lemma 1 of Meister (2006). The second term can be split into two pieces, *V1* and *V2*, where *V1* is identical to *V* in Lemma 1 of Meister (2006) while *V2* is the additional component of variance due to estimating the

[14] The careful reader will recognize that our density estimate of u does not appear to integrate to one. This occurs for two reasons. First, there is a boundary bias which we are not correcting and second, the deconvolution density estimator is not guaranteed to be positive everywhere.

composed errors. Our third term, which we call *E* can be bound almost as found in Lemma 1 of Meister (2006) but the form of the bound is more complicated due to the fact that the empirical characteristic function used to construct the variance of the normal contamination is constructed with $\hat{\varepsilon}$ instead of ε.

We split the second addend into two parts (V1 and V2) and let

$$A(\hat{h}_\varepsilon) = \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v f_u} \left| \left[\hat{h}_{\hat{\varepsilon}}(\tau) - h_\varepsilon(\tau) \right] \right|^2 d\tau. \tag{15}$$

Thus,

$$\begin{aligned}
&\sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} 2 \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v f_u} \left| \exp\left(\hat{\sigma}_n^2 \tau^2/2\right) \left[\hat{h}_{\hat{\varepsilon}}(\tau) - h_\varepsilon(\tau) \right] \right|^2 d\tau \\
&\leq 4e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} \left[E\left| \hat{h}_{\hat{\varepsilon}}(\tau) - \hat{h}_\varepsilon(\tau) \right|^2 + E\left| \hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau) \right|^2 \right] d\tau \\
&= 4e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} \left[E\left| \hat{h}_{\hat{\varepsilon}}(\tau) - \hat{h}_\varepsilon(\tau) \right|^2 \right] d\tau + 4e^{\sigma_n^2/\lambda_n^2} A(\hat{h}_\varepsilon) \\
&\leq 4e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} \tau^2 E\left(n^{-1} \sum_{j=1}^{n} \left| \hat{\varepsilon}_j - \varepsilon_j \right| \right)^2 d\tau + 4e^{\sigma_n^2/\lambda_n^2} A(\hat{h}_\varepsilon).
\end{aligned}$$

The second addend is identical to Meister's V and is bound above by:

$$V_1 \leq const.(n\lambda_n)^{-1} \exp\left(\sigma_n^2/\lambda_n^2\right). \tag{16}$$

All that is left to consider is the first addend, which is the important difference between these results and Meister's. Call this V_2. Under Assumption 3.5 it has bound:

$$V_2 \leq const.\left(n\lambda_n^3\right)^{-1} \exp\left(\sigma_n^2/\lambda_n^2\right). \tag{17}$$

Proof of Lemma 3.2:

We can bound the term $P_{f_v f_u}\left(\left|\hat{\sigma}_n^2 - \sigma^2\right| > d_n\right)$ from above by two addends; we derive upper bounds for both of these addends. Since we are selecting an $f_v \in \mathcal{F}_v$, we have that $\sigma^2 \in (0, \sigma_n^2]$. Now

$$\begin{aligned}
P_{f_v f_u}\left(\hat{\sigma}_n^2 - \sigma^2 > d_n\right) &= P_{f_v f_u}\left(-2k_n^{-2} \ln\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| / C_1 k_n^{-\delta}\right) > d_n + \sigma^2 \right) \\
&= P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| / C_1 k_n^{-\delta} \quad < \exp\left(-k_n^2 d_n/2 - \sigma^2 k_n/2\right)\right) \\
&= P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < C_1 k_n^{-\delta} \exp\left(-k_n^2 d_n/2\right) \exp\left(-\sigma^2 k_n/2\right)\right) \\
&= P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < \frac{2C_1}{C_1} k_n^{-\delta} \exp\left(-k_n^2 d_n/2\right) \exp\left(-\sigma^2 k_n/2\right)\right) \\
&= P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < \alpha_n C_1 k_n^{-\delta} \exp\left(-k_n^2 d_n/2\right) \exp(-\sigma^2 k_n/2)\right) \leq P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < \alpha_n \left|h_\varepsilon(k_n)\right|\right),
\end{aligned} \tag{18}$$

where $\alpha_n = (2C_1)\exp\left(-d_n k_n^2/2\right)$. The last inequality follows from the bounds on a characteristic function for a normally distributed random variable and the ordinary smooth characteristic function. Given the description of k_n and d_n above we know that $\alpha_n \to 0$ as $n \to \infty$. At this point we know a constant $c \in (0,1)$ exists such that

$$P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < \alpha_n \left|h_\varepsilon(k_n)\right|\right) \leq P_{f_v f_u}\left(\left|\hat{h}_{\hat{\varepsilon}}(k_n)\right| < c\left|h_\varepsilon(k_n)\right|\right) \tag{19}$$

which by Chebyshev's inequality yields

$$\leq (1-c)^{-2} \sup_{f_v \in \mathcal{F}_n} \sup_{f_u \in \mathcal{F}_u} |h_\varepsilon(k_n)|^{-2} E\left|\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)\right|^2$$

$$\leq 2(1-c)^{-2} \sup_{f_v \in \mathcal{F}_n} \sup_{f_u \in \mathcal{F}_u} |h_\varepsilon(k_n)|^{-2}$$

$$\times \left[E\left|\hat{h}_\varepsilon(k_n) - \hat{h}_\varepsilon(k_n)\right|^2 + E\left|\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)\right|^2\right]$$

$$= const.(E_1 + E_2).$$

The second addend is bound by

$$E_2 \leq const.k_n^{2\delta} \exp\left(\sigma_n^2 k_n^2\right) n^{-1}, \qquad (20)$$

as in Lemma 2 of Meister (2006), and the first addend is bound as in our Lemma 3.1 by

$$E_1 \leq const.k_n^{2(1+\delta)} \exp\left(\sigma_n^2 k_n^2\right) n^{-1}. \qquad (21)$$

The term $P_{f_v f_u}\left(\hat{\sigma}_n^2 - \sigma^2 < -d_n\right)$ can be bound in identical fashion.

Proof of Theorem 3.1:
Our proof follows Meister (2006) Theorem 2 except that in the corresponding max operators we have $(\ln(n))^{2\delta+1/2} n^{-1/2}$ instead of $(\ln(n))^{\delta+1/2} n^{-1/2}$ in the second argument. However, the presence of $n^{-1/2}$ makes these terms asymptotically irrelevant to the other arguments. We still have the three cases that Meister (2006) considers: (a) $1 < \delta < 5/2$, (b) $\delta = 5/2$, and (c) $\delta > 5/2$, so the theorem follows by the same arguments in Meister (2006). It turns out that for the case $1 < \delta < 5/2$, the bias (B) dictates the uniform rate of convergence, while for the case $\delta \geq 5/2$ the first integral of the bound on E dictates the rates.

References

Aigner D, Lovell CK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econometr 6:21–37

Benkard CL, Bajari P (2005) Hedonic Price indices with unobserved product characteristics, and application to personal computers. J Bus Econ Stat 23:61–75

Barry J, Diggle P (1995) Choosing the smoothing parameter in a Fourier approach to nonparametric deconvolution of a density estimate. J Nonparametr Stat 4:223–232

Butucea C, Matias C (2005) Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolutionm model. Bernoulli 11(2):309–340

Carroll RJ, Hall P (1988) Optimal rates of convergence for deconvolving a density. J Am Stat Assoc 83:1184–1186

Carroll RJ, Hall P (2004) Low order approximations in deconvolution and regression with errors in variables. J R Stat Soc Ser B 66(1):31–46

Delaigle A (2008) An alternative view of the deconvolution problem. Stat Sin 18:1025–1045

Delaigle A, Gijbels I (2002) Estimation of integrated squared density derivatives from a contaminated sample. J R Stat Soc Ser B 64: 869–886

Delaigle A, Gijbels I (2004a) Practical bandwidth selection in deconvolution density estimation. Comput Stat Data Anal 45:249–267

Delaigle A, Gijbels I (2004b) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. Ann Inst Stat Math 56:19–47

Delaigle A, Gijbels I (2006a) Estimation of boundary and discontinuity points in deconvolution problems. Stat Sin 16:773–788

Delaigle A, Gijbels I (2006b) Data-driven boundary estimation in deconvolution problems. Comput Stat Data Anal 50:1965–1994

Delaigle A, Gijbels I (2007) Frequent problems in calculating integrals and optimizing objective functions: a case study in density deconvolution. Stat Comput 17:349–355

Delaigle A, Hall P (2007) On optimal kernel choice for deconvolution. Stat Probab Lett 76:1594–1602
Delaigle A, Hall P, Meister A (2008) On deconvolution with repeated measurements. Ann Stat 36(2):665–685
Delaigle A, Meister A (2007) Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. J Am Stat Assoc 102:1416–1426
Delaigle A, Meister A (2008) Density estimation with heteroscedastic error. Bernoulli 14(2):562–579
Devroye PL (1989) Consistent deconvolution in denstiy estimation.Can J Stat 17:235–239
Fan J (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. Ann Stat 19:1257–1272
Fan J (1991b) Global behavior of deconvolution kernel estimates. Stat Sin 1:541–551
Fan J (1991c) Asymptotic normality for deconvolution kernel estimators. Sankhy¯a Ser A 53:97–100
Fan J (1992) Deconvolution with supersmooth distributions. Can J Stat 20:155–169
Fan J (1993) Adaptively local one-dimensional subproblems with application to a deconvolution problem. Ann Stat 21:600–610
Hall P, Qiu P (2005) Discrete-transform approach to deconvolution problems. Biometrika 92(1):135–148
Greene WH (1990) A gamma-distributed stochastic frontier model. J Econometr 46(1/2):141–164
Hall P, Simar L (2002) Estimating a changepoint, boundary, or frontier in the presence of observation error. J Am Stat Assoc 97:523–534
Hesse CH (1995) Deconvolving a density from contaminated dependent observations. Ann Inst Stat Math 47:645–663
Hesse CH (1996) How many ‘‘good’’ observations do you need for ‘‘fast’’ density deconvolution from supersmooth errors. Sankhy¯a Ser A 58:491–506
Hesse CH (1999) Data-driven deconvolution. J Nonparametr Stat 10:343–373
Horowitz JL, Markatou M (1996) Semiparametric estimation of regression models for panel data. Rev Econ Stud 63:145–168
Horrace WC (2009) Moments of the singly truncated normal distribution. Working paper, Center for Policy Research, Syracuse University
Jondrow J, Lovell CAK, Matertov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier
model. J Econometr 19:233–238
Krasnokutskaya E (2009) Identification and estimation in highway procurement auctions under unobserved auction heterogeneity. Rev Econ Stud (forthcoming)
Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, New York
Kumbhakar SC, Tsionas E (2005) Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. J Econometr 126:355–384
Li T, Perrigne I, Vuong Q (2000) Conditionally independent private information in OCS wildcat auctions. J Econometr 98:129–161
Liu MC, Taylor RL (1989) A consistent nonparametric density estimator for the deconvolution problem. Can J Stat 17:427–438
Lukacs E (1968) Stochastic convergence. Ratheon Education Company, USA
Matias C (2002) Semiparametric deconvolution with unknown noise variance. ESAIM Stat Probab 6:271–292
Meister A (2003) Robustheitseigenschaften von Dekonvolutionsdichtescha ¨tzern bezu¨glich Missspezifikation der Fehlerdichte. Dissertation, University of Stuttgart, Germany
Meister A (2004a) Deconvolution density estimation with a testing procedure for the error distribution. Technical report, University of Stuttgart, Germany
Meister A (2004b) On the effect of misspecifying the error density in a nonparametric deconvolution problem. Can J Stat 32:439–449
Meister A (2006) Density estimation with normal measurement error with unknown variance. Stat Sin 16:195–211
Mendelsohn J, Rice J (1982) Deconvolution of microfluorometric histograms with B splines. J Am Stat Assoc 77:748–753
Neumann MH (1997) On the effect of estimating the error density in nonparametric deconvolution. Nonparametr Stat 7:307–330
Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 35:1065–1076
Postel-Vinay F, Robin J-M (2002) Equilibrium wage dispersion with heterogeneous workers and firms. Econometrica 70:2295–2350
Postel-Vinay F, Robin J-M (2003) The distribution of earnings in an equilibrium search model with state-dependent offers and counter-offers. Int Econ Rev 43:989–1016

Stefanski L, Carroll RJ (1990) Deconvoluting kernel density estimators. Statistics 21:169–184
Wang WS, Amsler C, Schmidt P (2008) Goodness of fit tests in stochastic frontier models. Unpublished manuscript, Michigan State University
Wang WS, Schmidt P (2009) On the distribution of estimated technical efficiency in stochastic frontier models. J Econometr 148(1):36–45
Zhang S, Karunramuni RJ (2000) Boundary bias correction for nonparametric deconvolution. Ann Inst Stat Math 52:612–629