2008

Semiparametric Deconvolution with Unknown Error Variance

William C. Horrace
Syracuse University. Center for Policy Research, whorrace@maxwell.syr.edu

Christopher F. Parmeter
Virginia Polytechnic Institute and State University, Department of Agriculural and Applied Economics, parms@vt.edu

ISSN: 1525-3066

Center for Policy Research
Working Paper No. 104

SEMIPARAMETRIC DECONVOLUTION WITH UNKNOWN ERROR VARIANCE

William C. Horrace and Christopher F. Parmeter

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

April 2008

$5.00

CENTER FOR POLICY RESEARCH – Spring 2008

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi	Economics
Kalena Cortes	Education
Thomas Dennison	Public Administration
William Duncombe	Public Administration
Gary Engelhardt	Economics
Deborah Freund	Public Administration
Madonna Harrington Meyer	Sociology
Christine Himes	Sociology
William C. Horrace	Economics
Duke Kao	Economics
Eric Kingson	Social Work
Thomas Kniesner	Economics
Jeffrey Kubik	Economics
Andrew London	Sociology
Len Lopoo	Public Administration
Amy Lutz	Sociology
Jerry Miner	Economics
Jan Ondrich	Economics
John Palmer	Public Administration
Lori Ploutz-Snyder	Exercise Science
David Popp	Public Administration
Christopher Rohlfs	Economics
Stuart Rosenthal	Economics
Ross Rubenstein	Public Administration
Perry Singleton	Economics
Margaret Usdansky	Sociology
Michael Wasylenko	Economics
Janet Wilmoth	Sociology

GRADUATE ASSOCIATES

Amy Agulay	Public Administration
Javier Baez	Economics
Sonali Ballal	Public Administration
Jesse Bricker	Economics
Maria Brown	Social Science
Il Hwan Chung	Public Administration
Mike Eriksen	Economics
Qu Feng	Economics
Katie Fitzpatrick	Economics
Chantell Frazier	Sociology
Alexandre Genest	Public Administration
Julie Anna Golebiewski	Economics
Nadia Greenhalgh-Stanley	Economics
Sung Hyo Hong	Economics
Neelakshi Medhi	Social Science
Larry Miller	Public Administration
Phuong Nguyen	Public Administration
Wendy Parker	Sociology
Shawn Rohlin	Economics
Carrie Roseamelia	Sociology
Jeff Thompson	Economics
Coady Wing	Public Administration
Ryan Yeung	Public Administration
Can Zhou	Economics

STAFF

Kelly Bogart	Administrative Secretary
Martha Bonney	Publications/Events Coordinator
Karen Cimilluca	Administrative Secretary
Roseann DiMarzo	Receptionist/Office Coordinator
Kitty Nasto	Administrative Secretary
Candi Patterson	Computer Consultant
Mary Santy	Administrative Secretary

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 Monte Carlo evidence that demonstrates the merits of the method.

William C. Horrace: Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020, US. Email: whorrace@maxwell.syr.edu
Christopher F. Parmeter: Department of Agricultural and Applied Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0401. Email: parms@vt.edu

JEL Classification: C14; C21.

Keywords: Error Component; Ordinary Smooth; Semi-Uniform Consistency

Semiparametric Deconvolution with Unknown Error Variance

William C. Horrace*
Center for Policy Research
Syracuse University
Syracuse, NY 13244-1020
whorrace@maxwell.syr.edu

Christopher F. Parmeter
Department of Agricultural and Applied Economics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0401
parms@vt.edu

April 10, 2008

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 Monte Carlo evidence that demonstrates the merits of the method.

Keywords: Error Component, Ordinary Smooth, Semi-Uniform Consistency.

JEL Classification No.: C14, C21.

---

*Corresponding author. Tel: (315) 443-9061; Fax: (315) 443-1081. The authors thank Joel Horowitz, Nicolai Kuminoff, Hailong Qian, Peter Schmidt and Leonard Stefanski. Comments from seminar participants at Binghamton University, Syracuse University, the University of Richmond and Virginia Tech as well as EWEPA X are gratefully acknowledged. All code used in this paper is available upon request. The usual disclaimer applies.

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. 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 error 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 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.[1] Essentially, the information contained in the time-dimension of the panel replaces the Normal error assumption to achieve uniformly consistent estimates 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

[1] 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.

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 the density of the error is from a zero-mean, Normal family with *unknown* variance. In this paper we show that if the 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, Robin and van den Berg (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, Perrigne, and Vuong (2000) and Krusnatskaya (2008). We also note that research geared towards recovering the distribution of unobservable heterogeneity (e.g., hedonic models) may benefit from deconvolution methods, see Bajari and Benkard (2005).

The paper is organized as follows. In section 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 semi-uniform consistency of the estimator. Section 4 contains a Monte Carlo study of the finite sample performance of the estimator. Directions for future research and conclusions are in section 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-c, 1992, 1993). Bandwidth selection issues are considered in Barry and Diggle (1995), Hesse (1999) and Delaigle and Gijbels (2002, 2004a). Other practical issues relating to kernel deconvolution are in

Zhang and Karunramuni (2000) for boundary corrections, Neumann (1997) dealing with the estimation of the unknown measurement error as opposed to assuming its family, Hesse (1995) for the case when only some of the data are measured with error, and Hesse (1996) for the case when the data of interest are dependent.

A recent explosion of papers on theoretical and computational aspects of deconvolution estimators has rekindled interest in the area. Meister (2004a,b) develops procedures for testing whether the assumed measurement error density is correct while Meister (2006) proposes the first estimator for the variance of the known noise distribution. Delaigle and Meister (2007a,b, 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 low order approximation for deconvolution, and Hall and Qui (2005) consider a trigonometric expansion for deconvolution that is simpler than kernel methods. Delaigle and Gijbels (2007) discuss key issues with calculating the integrals arising in deconvolution settings. Delaigle and Hall (2007) and Delaigle (2007) discuss issues associated with optimal kernel choice and the appropriateness of assuming a diminishing error variance as the sample size grows. Also, Delaigle, Hall and Meister (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.

In economics, the usefulness of deconvolution estimators has not been fully realized. The only mainstream papers that have employed these techniques are Horowitz and Markatou (1996) and Schennach (2004). Here, our goal is to develop a similar estimator for the cross sectional setting where the noise distribution is assumed Normal with unknown variance.

3 The convolution problem

Consider the specification:

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

$$y_j = m(x_j; \beta) + \varepsilon_j, \ \ j = 1, ..., n, \tag{2}$$

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 $m(\cdot)$ is a production function and $u_j < 0$ then this would be a standard stochastic frontier model. If one were in a panel setting and assumed that x_j was uncorrelated with u_j then this would be a classic random effects scenario. 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 $j = 1, ..., n$.*

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. $\mathcal{F}_u = \{u \text{ density}; C_1|\tau|^{-\delta} \leq |h_u(\tau)| \leq C_2|\tau|^{-\delta}, \forall|\tau| \geq T > 0\}$. 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 of densities are the Laplace and Gamma. 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, p14). 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, as in previous work on deconvolution. See Meister (2003) for more on concepts related to semi-uniform consistency.

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

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

However, since we are estimating the errors from equation (2) this estimate is not useful empirically. We will use the empirical characteristic function of the residuals which is defined as

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

Unfortunately, replacing h_ε with $\widehat{h}_\varepsilon$ or $\widehat{h}_{\hat{\varepsilon}}$ in equation (4), does not ensure that the integration will exist, so we convolute the integrand with a smoothing kernel.[2] 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 in equation (7) 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,

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

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

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

where $\tilde{\sigma}_n^2 = -2k_n^{-2}\ln\left(\frac{\hat{h}_{\hat{\varepsilon}}(k_n)}{C_1 k_n^\delta}\right)$ and $\sigma_n^2 = \ln(\ln(n))/4$.[3] 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 $m(x;\beta)$ 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), is used for simplification of the proofs and can be easily satisfied by dropping extreme values of X. Assumption 3.5 guaran-

[2]See Stefanski and Carroll (1990).

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

tees that the random sampling error between the composed errors and the estimated ones is asymptotically negligible.

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

Lemma 3.1 *Under Assumptions 3.4 and 3.5* $\tilde{\sigma}_n^2 = \tilde{\sigma}_{1,n}^2 + \tilde{\sigma}_{2,n}^2$, *where*

$$\tilde{\sigma}_{1,n}^2 = -2k_n^{-2} \ln \left(\frac{\hat{h}_\varepsilon(k_n)}{C_1 k_n^\delta} \right) \tag{9}$$

and

$$\tilde{\sigma}_{2,n}^2 = -2k_n^{-2} \ln \left(1 + O_p(k_n n^{-1/2}) \right), \tag{10}$$

where $\tilde{\sigma}_{2,n}^2$ *is an additional component of the variance due to using* $\hat{\varepsilon}$ *instead of* ε.

Lemma 3.1 shows that our variance estimator is equivalent to that of Meister (2006), plus an additional term that arises from estimation of the errors as opposed to direct observation.

Lemma 3.2 *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*

$$\sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} E_{f_v,f_u} \|\hat{f}_u - f_u\|_{L_2}^2 \leq B + V_1 + V_2 + E, \tag{11}$$

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

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

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

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.[4] 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 a 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

[4]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.

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

$$\sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} P_{f_v,f_u}\left(|\hat{\sigma}_n^2 - \sigma^2| > d_n\right) \leq const. k_n^{2\delta} \exp(\sigma_n^2 k_n^2)(1 + k_n^2)/n. \tag{12}$$

Given that we have to replace ε with an estimate our lemma differs from Meister (2006) through the addition of k_n^2. 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.

Lemmas 3.1 through 3.3 can be used to show

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

$$\sup_{f_u \in \mathcal{F}_u} E_{f_v,f_u} \|\hat{f}_u - f_u\|_{L_2}^2 \leq \begin{cases} (\ln(\ln(n)))^{3\delta-1.5}(\ln(n))^{0.5-\delta}, & \text{if } \delta < 2.5 \\ (\ln(\ln(n)))^{6}(\ln(n))^{-2}, & \text{if } \delta = 2.5 \\ (\ln(\ln(n)))^{7}(\ln(n))^{-2}, & \text{if } \delta > 2.5, \end{cases} \tag{13}$$

The rates here are identical to those found in Meister (2006) and show the difficulty of deconvolution with Normal error.[5] 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 3.2. However, the residuals only

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

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 still guarantees a semi-uniformly consistent estimator. This is essentially what occurs in the deconvolution estimator of Horowitz and Markatou (1996) who also do 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. For this case the parameters of the distribution would still have to be assumed known as no estimation strategy has been proposed for this deconvolution setting. 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.[6] 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 Finite sample properties of the estimator

We draw from several other deconvolution simulation studies to examine the small sample properties of our estimator.[7] We consider sample sizes of 200 and 1000. Our model is

$$y_i = 4 + 3x_i + v_i + u_i \tag{14}$$

The x_is are generated from a standard Normal. The u_is are generated from the two times self-convoluted, zero-mean Laplace density.[8] To determine the impact of the noise variance we generate v_i 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 through 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 the paper using $\hat{\varepsilon}$ (labelled ME in the legends). We can see that both our estimator and the Meister estimator provide a similar depiction 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

[6]Here loss is taken to be $MISE$.

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

[8]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 $\tilde{L}(x) = (4b^2)^{-1}e^{-|x|/b}(|x| + b)$. We choose b so that this density has variance 1, which corresponds to $b = 1$. In this setting $C_1 = 1/4$ and $\delta = 2$. We are not concerned with C_2 as it has no bearing on any calculations for the estimator.

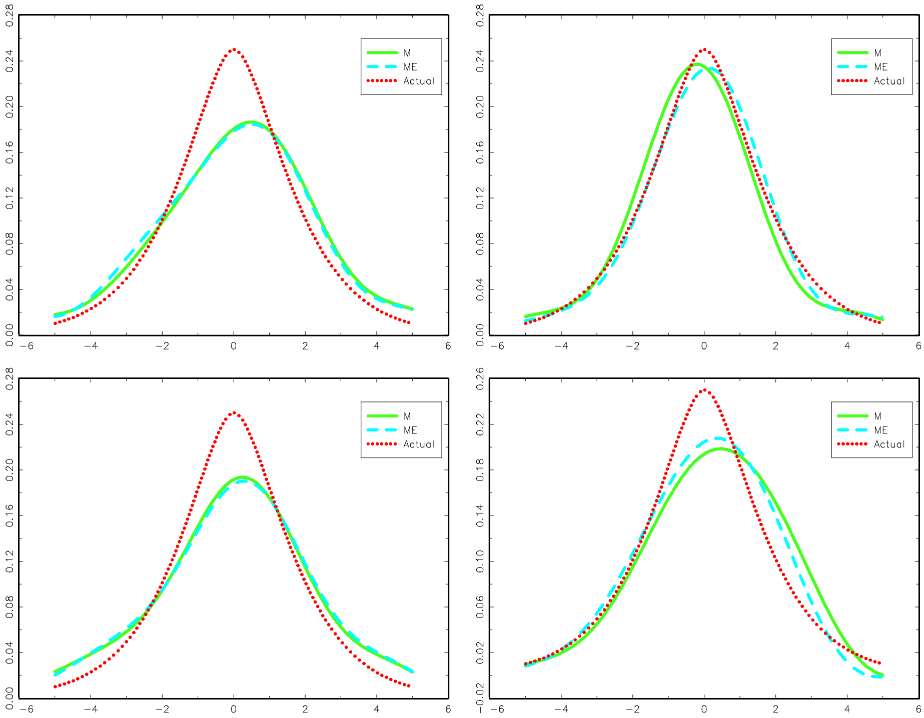

Figure 1: Single Runs for $n = 200$ and $\sigma^2 = 1$.

residuals. In Figure 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 Figure 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 of 1000 further illustrates the superiority of the $\sqrt{n}$ rate of convergence of the residuals as opposed to the logarithmic rates for the deconvolution density estimator. In Figure 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. Both estimators are indistinguishable from one another. 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 with these types of 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

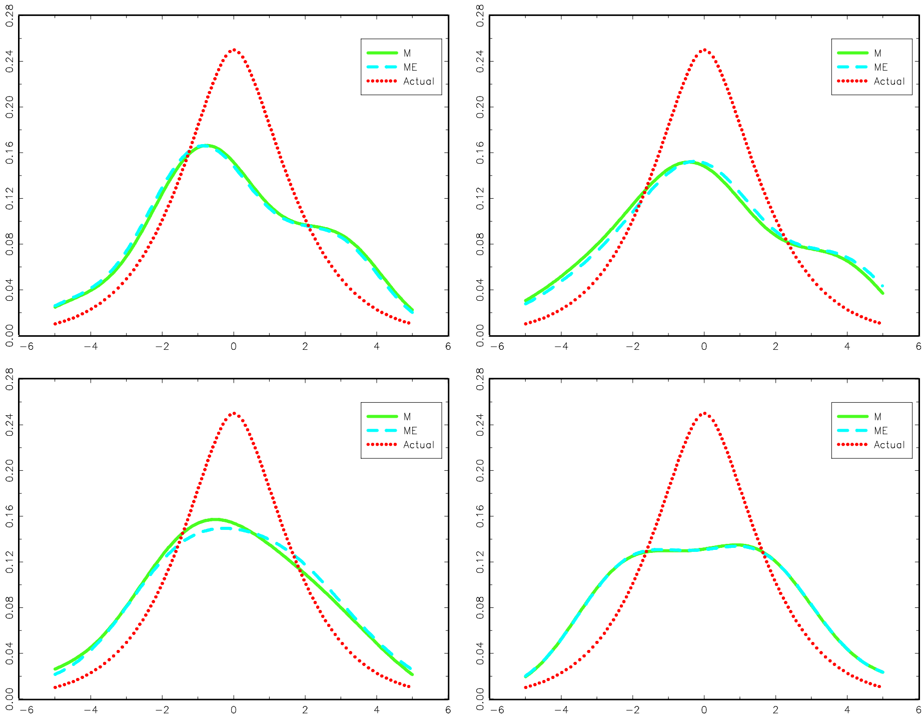

Figure 2: Single Runs for $n = 200$ and $\sigma^2 = 4$.

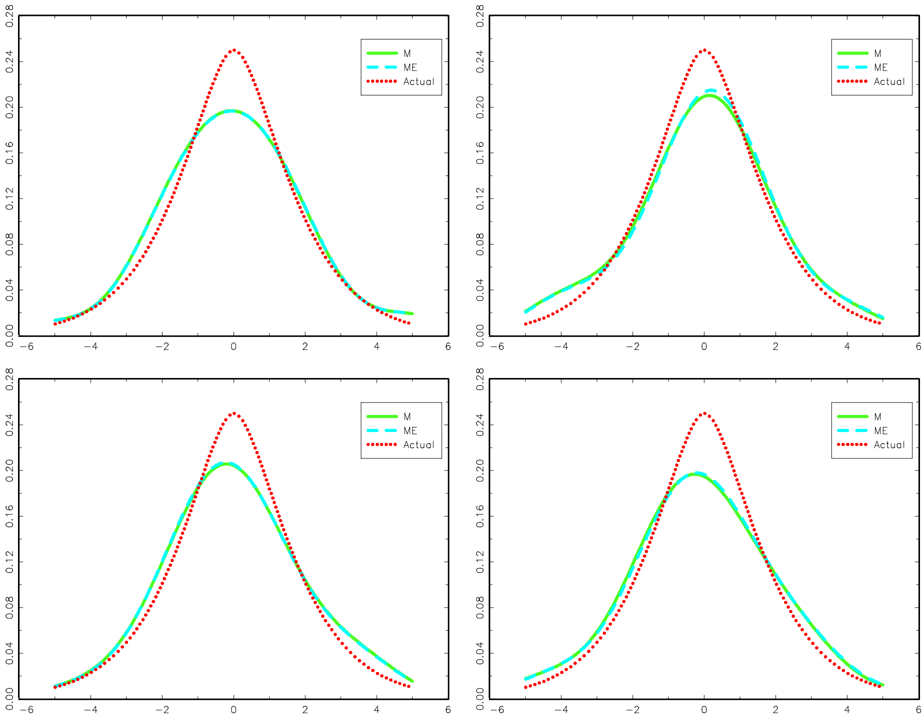

Figure 3: Single Runs for $n = 1000$ and $\sigma^2 = 1$.

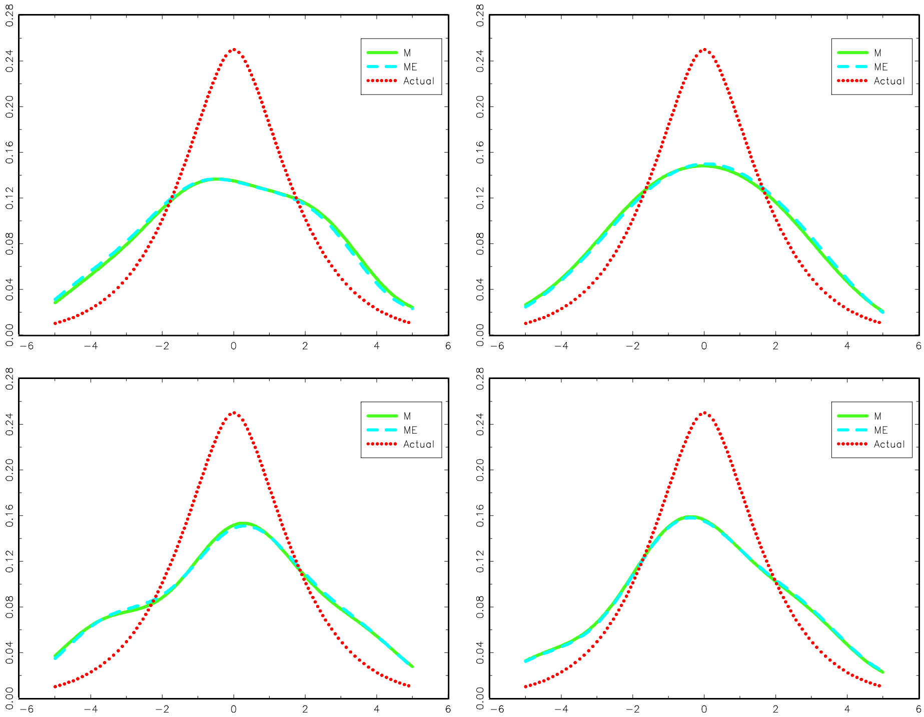

Figure 4: Single Runs for $n = 1000$ and $\sigma^2 = 4$.

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.

5 Conclusions

This paper proposes a semiparametric estimator for cross-sectional 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 rely on a logarithmic rate.

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 the results here. As previously stated, developing a positive variance estimate is also an interesting extension of this work. These additional pieces should enhance the appeal of deconvolution methods in cross-sectional error component settings.

References

[1] Aigner, D., C. K. Lovell, and P. Schmidt (1977) Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6, 21-37.

[2] Benkard, C. L. and P. Bajari (2005) Hedonic Price indices with unobserved product characteristics, and application to personal computers. *Journal of Business & Economic Statistics* 23, 61-75.

[3] Barry, J. and P. Diggle (1995) Choosing the smoothing parameter in a Fourier approach to nonparametric deconvolution of a density estimate. *Journal of Nonparametric Statistics* 4, 223-232.

[4] Carroll, R. J. and P. Hall (1988) Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83, 1184-1186.

[5] Delaigle A. and I. Gijbels (2002) Estimation of integrated squared density derivatives from a contaminated sample. *Journal of the Royal Statistical Society, Series B* 64, 869-886.

[6] Delaigle A. and I. Gijbels (2004a) Practical bandwidth selection in deconvolution density estimation. *Computational Statistics & Data Analysis* 45, 249-267.

[7] Delaigle A. and I. Gijbels (2004b) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *The Annals of the Institute of Statistical Mathematics* 56, 19-47.

[8] Delaigle A. and I. Gijbels (2006a) Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica* 16, 773-788.

[9] Delaigle A. and I. Gijbels (2006b) Data-driven boundary estimation in deconvolution problems. *Computational Statistics & Data Analysis* 50, 1965-1994.

[10] Delaigle A. and A. Meister (2007) Nonparametric Regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association* 102, 1416-1426.

[11] Devroye, P. L. (1989) Consistent deconvolution in denstiy estimation. *Canadian Journal of Statistics* 17, 235-239.

[12] Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* 19, 1257-1272.

[13] Fan, J. (1991b) Global behavior of deconvolution kernel estimates. *Statistica Sinica* 1, 541-551.

[14] Fan, J. (1991c) Asymptotic normality for deconvolution kernel estimators. *Sankhȳa, Series A* 53, 97-100.

[15] Fan, J. (1992) Deconvolution with supersmooth distributions. *Canadian Journal of Statistics* 20, 155-169.

[16] Fan, J. (1993) Adaptively local one-dimensional subproblems with application to a deconvolution problem. *The Annals of Statistics* 21, 600-610.

[17] Hall, P. and L. Simar (2002) Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association* 97, 523-534.

[18] Hesse, C. H. (1995) Deconvolving a density from contaminated dependent observations. *The Annals of the Institute of Statistical Mathematics* 47, 645-663.

[19] Hesse, C. H. (1996) How many "good" observations do you need for "fast" density deconvolution from supersmooth errors. *Sankhȳa, Series A* 58, 491-506.

[20] Hesse, C. H. (1999) Data-driven deconvolution. *Journal of Nonparametric Statistics* 10, 343-373.

[21] Horowitz, J.L. and M. Markatou (1996) Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63, 145-168.

[22] Krasnokutskaya, E., 2004. Identification and estimation in highway procurement auctions under unobserved auction heterogeneity. Working Paper, University of Pennsylvania.

[23] Li, T., Perrigne, I., Vuong, Q., (2000) Conditionally independent private information in OCS wildcat auctions. *Journal of Econometrics* 98, 129-161.

[24] Liu, M. C. and R. L. Taylor (1989) A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics* 17, 427-438.

[25] Meister, A. (2003) Robustheitseigenschaften von Dekonvolutionsdichteschätzern bezüglich Missspezifikation der Fehlerdichte. Dissertation, University of Stuttgart, Germany.

[26] Meister, A. (2004a) Deconvolution density estimation with a testing procedure for the error distribution. Technical report, University of Stuttgart, Germany.

[27] Meister, A. (2004b) On the effect of misspecifying the error density in a nonparametric deconvolution problem. *Canadian Journal of Statistics* 32, 439-449.

[28] Meister, A. (2006) Density Estimation with Normal Measurement Error with Unknown Variance. *Statistica Sinica* 16, pp. 195-211.

[29] Mendelsohn, J. and J. Rice (1982) Deconvolution of microfluorometric histograms with *B* splines. *Journal of the American Statistical Association* 77, 748-753.

[30] Neumann, M. H. (1997) On the effect of estimating the error density in nonparametric deconvolution. *Nonparametric Statistics* 7, 307-330.

[31] Lukacs, E. (1968) *Stochastic Convergence*. Ratheon Education Company.

[32] Rebitzer (1987) Unemployment, long-term employment relations, and productivity growth. *The Review of Economic and Statistics* 69, 627-635.

[33] Postel-Vinay, F. and J.-M. Robin (2002) Equilibrium wage dispersion with heterogeneous workers and firms. *Econometrica* 70, 2295-2350.

[34] Postel-Vinay, F. and J.-M. Robin (2003) The distribution of earnings in an equilibrium search model with state-dependent offers and counter-offers. *International Economic Review* 43, 989-1016.

[35] Parzen, E. (1962) On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics* 35, 1065-1076.

[36] Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman Hall, London.

[37] Stefanski, L and R.J. Carroll (1990) Deconvoluting kernel density estimators. *Statistics* 21, 169-184.

[38] Zhang S. and R. J. Karunramuni (2000) Boundary bias correction for nonparametric deconvolution. *Annals of the Institute of Statistical Mathematics* 52, 612-629.

Appendix: Proofs of Lemmas and Theorems

Proof of Lemma 3.1:

Rewrite the empirical characteristic function of the estimated residuals as

$$\hat{h}_{\hat{\varepsilon}}(\tau) = \frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j} e^{i\tau(\hat{\varepsilon}_j - \varepsilon_j)}, \tag{15}$$

and note that by the Mean Value Theorem of Calculus

$$\begin{aligned}\hat{h}_{\hat{\varepsilon}}(\tau) =& \frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j}\left(1 + O_p(\tau n^{-1/2})\right)\\ =& \left(1 + O_p(\tau n^{-1/2})\right)\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j} = \left(1 + O_p(\tau n^{-1/2})\right)\hat{h}_{\varepsilon}(\tau).\end{aligned} \tag{16}$$

We can use this derivation to complete the Lemma. The variance estimator is defined as $\tilde{\sigma}_n^2 = -2k_n^{-2}\ln\left(\frac{\hat{h}_{\hat{\varepsilon}}(k_n)}{C_1 k_n^{\beta}}\right)$ which by the above argument can be rewritten as

$$\tilde{\sigma}_n^2 = -2k_n^{-2}\ln\left(\frac{\hat{h}_{\varepsilon}(k_n)\left(1 + O_p(k_n n^{-1/2})\right)}{C_1 k_n^{\delta}}\right) = -2k_n^{-2}\ln\left(\frac{\hat{h}_{\varepsilon}(k_n)}{C_1 k_n^{\delta}}\right) - 2k_n^{-2}\ln\left(1 + O_p(k_n n^{-1/2})\right). \tag{17}$$

which completes the proof.

Proof of Lemma 3.2:

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

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

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, V_1 and V_2, where V_1 is identical to V in Lemma 1 of Meister (2006) while V_2 is the additional component of variance due to estimating the 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 (V_1 and V_2) using the results in Lemma 3.1 and let $A(\hat{h}_\varepsilon) = \int\limits_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| \left[\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau) \right] \right|^2 d\tau$:

$$\sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} 2 \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| \exp\left(\widehat{\sigma}_n^2 \tau^2/2\right) \left[\hat{h}_{\hat{\varepsilon}}(\tau) - h_\varepsilon(\tau) \right] \right|^2 d\tau \tag{18}$$

$$\leq 2e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| \left[n^{-1} \sum_{j=1}^n \exp\left(i\tau\varepsilon_j\right) \exp\left(i\tau(\hat{\varepsilon}_j - \varepsilon_j)\right) - h_\varepsilon(\tau) \right] \right|^2 d\tau \tag{19}$$

$$= 2e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| n^{-1} \sum_{j=1}^n \exp\left(i\tau\varepsilon_j\right) \left(1 + O_p(\tau n^{-1/2})\right) - h_\varepsilon(\tau) \right|^2 d\tau \tag{20}$$

$$= 2e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| \hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau) + \left(O_p(\tau n^{-1/2}) \hat{h}_\varepsilon(\tau)\right) \right|^2 d\tau \tag{21}$$

$$\leq 2e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| \left[\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau) \right] + \left(O_p(n^{-1/2}\lambda_n^{-1}) \hat{h}_\varepsilon(\lambda_n^{-1})\right) \right|^2 d\tau \tag{22}$$

$$\leq 4e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \left(\int_{-1/\lambda_n}^{1/\lambda_n} E_{f_v,f_u} \left| O_p(n^{-1/2}\lambda_n^{-1}) \hat{h}_\varepsilon(\lambda_n^{-1}) \right|^2 d\tau + A(\hat{h}_\varepsilon) \right) \tag{23}$$

$$= 4e^{\sigma_n^2/\lambda_n^2} \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \left(O_p(n^{-1}\lambda_n^{-2}) E_{f_v,f_u} \left| \hat{h}_\varepsilon(\lambda_n^{-1}) \right|^2 \int_{-1/\lambda_n}^{1/\lambda_n} d\tau + A(\hat{h}_\varepsilon) \right) \tag{24}$$

$$= \sup_{f_v \in \mathcal{F}_v} \sup_{f_u \in \mathcal{F}_u} \left(8e^{\sigma_n^2/\lambda_n^2} O_p(n^{-1}\lambda_n^{-3}) E_{f_v,f_u} \left| \hat{h}_\varepsilon(\lambda_n^{-1}) \right|^2 + 4e^{\sigma_n^2/\lambda_n^2} A(\hat{h}_\varepsilon) \right). \tag{25}$$

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

$$V_1 \leq const.(n\lambda_n)^{-1} \exp\left(\sigma_n^2/\lambda_n^2\right). \tag{26}$$

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 with bound:

$$V_2 \leq const.(n\lambda_n^3)^{-1} \exp(\sigma_n^2/\lambda_n^2). \tag{27}$$

Proof of Lemma 3.3:

We can bound the term $P_{f_v,f_u}\left(|\hat{\sigma}_n^2-\sigma^2|>d_n\right)$ 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

$$
\begin{aligned}
P_{f_v,f_u}\left(\hat{\sigma}_n^2-\sigma^2>d_n\right) =& P_{f_v,f_u}\left(-2k_n^{-2}\ln(\hat{h}_{\hat{\varepsilon}}(k_n)/C_1k_n^{-\delta})>d_n+\sigma^2\right) && (28)\\
=& P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)/C_1k_n^{-\delta}<\exp(-k_n^2d_n/2-\sigma^2k_n/2))\right) && (29)\\
=& P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<C_1k_n^{-\delta}\exp(-k_n^2d_n/2)\exp(-\sigma^2k_n/2))\right) && (30)\\
=& P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<\frac{2C_1}{C_1}k_n^{-\delta}\exp(-k_n^2d_n/2)\exp(-\sigma^2k_n/2)\right) && (31)\\
=& P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<\alpha_nC_1k_n^{-\delta}\exp(-k_n^2d_n/2)\exp(-\sigma^2k_n/2)\right) && (32)\\
\leq& P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<\alpha_n|h_\varepsilon(k_n)|\right), && (33)
\end{aligned}
$$

where $\alpha_n=(2C_1)\exp(-d_nk_n^2/2)$. 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

$$
P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<\alpha_n|h_\varepsilon(k_n)|\right)\leq P_{f_v,f_u}\left(\hat{h}_{\hat{\varepsilon}}(k_n)<c|h_\varepsilon(k_n)|\right) \tag{34}
$$

which by Chebyshev's inequality yields

$$
\begin{aligned}
&\leq (1-c)^{-2}\sup_{f_v\in\mathcal{F}_n}\sup_{f_u\in\mathcal{F}_u}|h_\varepsilon(k_n)|^{-2}\,E_\varepsilon\left|\frac{(1+O_p(k_nn^{-1/2}))}{n}\sum_{j=1}^{n}\exp(ik_n\varepsilon_j)-h_\varepsilon(k_n)\right|^2 && (35)\\
&= (1-c)^{-2}\sup_{f_v\in\mathcal{F}_n}\sup_{f_u\in\mathcal{F}_u}|h_\varepsilon(k_n)|^{-2}\,E_\varepsilon\left|\hat{h}_\varepsilon(k_n)-h_\varepsilon(k_n)+O_p(k_nn^{-3/2}))\sum_{j=1}^{n}\exp(ik_n\varepsilon_j)\right|^2 && (36)\\
&\leq 2(1-c)^{-2}\sup_{f_v\in\mathcal{F}_n}\sup_{f_u\in\mathcal{F}_u}|h_\varepsilon(k_n)|^{-2}\left(E_\varepsilon\left|O_p(k_nn^{-3/2}))\sum_{j=1}^{n}\exp(ik_n\varepsilon_j)\right|^2+E_\varepsilon\left|\hat{h}_\varepsilon(k_n)-h_\varepsilon(k_n)\right|^2\right) && (37)\\
&= const.(E_1+E_2). && (38)
\end{aligned}
$$

The second addend is bound by

$$
E_2\leq const.k_n^{2\delta}\exp(\sigma_n^2k_n^2)n^{-1}, \tag{39}
$$

as in Lemma 2 of Meister (2006), and the first addend is bound as in our Lemma 3.2 by

$$E_1 \leq const.k_n^{2(1+\delta)} \exp(\sigma_n^2 k_n^2) n^{-1}. \tag{40}$$

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.