cemmap
centre for microdata methods and practice

Generalized nonparametric deconvolution with an application to earnings dynamics

Stéphane Bonhomme
Jean-Marc Robin

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP3/08

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics[1]

Stéphane Bonhomme
CEMFI, Madrid

Jean-Marc Robin
Paris School of Economics,
University Paris 1 - Pantheon - Sorbonne,[2]
and University College London[3]

This version: January 2008

[1]We thank Laura Hospido for providing the data. We also thank Manuel Arellano, Jean-Pierre Florens and Eric Renault for comments and suggestions. The usual disclaimer applies.

[2]Centre d'Economie de la Sorbonne, Université Paris 1 - Panthéon - Sorbonne, 106/112 bd de l'Hôpital, 75647 Paris Cedex 13, e-mail: jmrobin@univ-paris1.fr.

[3]Jean-Marc Robin gratefully acknowledges the financial support from the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice, "Cemmap" (grant reference RES-589-28-0001).

Abstract

In this paper, we construct a nonparametric estimator of the distributions of latent factors in linear independent multi-factor models under the assumption that factor loadings are known. Our approach allows to estimate the distributions of up to $L(L+1)/2$ factors given L measurements. The estimator works through empirical characteristic functions. We show that it is consistent, and derive asymptotic convergence rates. Monte-Carlo simulations show good finite-sample performance, less so if distributions are highly skewed or leptokurtic. We finally apply the generalized deconvolution procedure to decompose individual log earnings from the PSID into permanent and transitory components.

JEL codes: C13, C14.

Keywords: Factor models, nonparametric estimation, deconvolution, Fourier transformation, earnings dynamics.

1 Introduction

In this paper, we consider linear multi-factor models of the form: $Y = AX$, where Y is a vector of L observed measurements, X is a vector of K unobserved and mutually independent latent factors, and A is a L-by-K matrix of parameters. Throughout the analysis, we assume that a root-N consistent estimator of the matrix of factor loadings A is available and that the number of factors is known. The contribution of this paper is to provide a nonparametric estimator of the distribution function of X for up to $K = L(L+1)/2$ factors.

Applications of factor models are numerous in social sciences, and economics in particular. For example, standard models of individual earnings dynamics write log earnings as the sum of a fixed effect and several independent shocks, modelled as ARMA processes with different persistence. Then, it is useful to estimate the full distribution of factor components– as opposed to their first two moments– as an input to life-cycle consumption models (e.g., Guvenen, 2007a, 2007b), or to forecast transitions into and out of poverty (Lillard and Willis, 1978).

In the absence of general nonparametric deconvolution techniques, the usual approach is to specify flexible parametric distributions.[1] We are aware of the existence of nonparametric estimators of factor densities in two special cases. Classical nonparametric deconvolution focuses on the case where $L = 1$, $K = 2$ and the distribution function of the second factor (X_2) is known. The density of X_1 can then be consistently estimated by inverse Fourier transformation with trimming, the estimator showing slow convergence rates (e.g., Carroll and Hall, 1988, Fan, 1991, and Carroll *et al.*, 1995 for a survey).

The second case has $L = 2$ and $K = 3$. Repeated measurements allow the three factor distributions to be nonparametrically identified, as shown in Kotlarski (1967, see also Rao, 1992, p.21). Horowitz and Markatou (1996) provide a consistent estimator, assuming that factors are symmetrically distributed. Li and Vuong (1998) relax the symmetry assumption. Their estimator is based on first derivatives of the empirical characteristic function of the data and trimmed inverse Fourier transformation. Li and Vuong's estimator has been used in Li *et al.* (2000) in the context of a structural auction model, and in Li (2002) in a nonlinear errors-in-variables model.[2]

[1] Recent examples are Chamberlain and Hirano (1999), Hirano (2002), Geweke and Keane (2000, 2007), and Alvarez, Browning and Ejrnaes (2007).

[2] Hall and Yao's (2003) estimator is closely related to Li and Vuong's (1998). Related methods

To our knowledge, no nonparametric estimator of factor densities is available in general linear multi-factor models.[3] In this paper, we generalize Li and Vuong's (1998) estimator to allow for any number of measurements L, and up to $K = L(L+1)/2$ factors. The application to data from the Panel Study of Income Dynamics (PSID) illustrates the benefits of this generalization in a model of earnings dynamics, allowing for permanent and transitory shocks to log-earnings (as in Hall and Mishkin, 1982, and Abowd and Card, 1989).

To construct the estimator we use a result due to Székely and Rao (2000), who show the nonparametric identification of factor distributions in the general case. Partial derivatives of the logarithms of the empirical characteristic functions of measurements identify second derivatives of the characteristic functions of factors. Then, the inverse Fourier transforms of integrated derivatives allow to recover factor densities. Our estimator is the empirical analog of this theoretical solution. It requires no optimization, unlike parametric approaches. To choose the amount of trimming that is necessary for the estimator to be well-defined, we use the "plug-in" method proposed in Delaigle and Gijbels (2004).

We show that the estimator is consistent and provide asymptotic convergence rates, allowing for unbounded factors. As in previous work on nonparametric deconvolution, the rates we obtain are slow, especially if the characteristic function of the factor to be estimated has fatter tails than the characteristic functions of the other factors. Nevertheless, Monte Carlo simulations are encouraging. When the true factor distributions are normal, we find moderate biases and tight confidence bands. Interestingly, our generalized deconvolution estimator has the same finite-sample bias and variance as the deconvolution estimator that assumes that all factor densities are known except the one to be estimated. This simulation evidence mitigates the negative conclusions of the asymptotic analysis, and suggests that the estimator may work well in practice. We also find that the shape of factor distributions strongly influences the finite-sample performances of the estimator. In particular, estimating factor distributions is more difficult when these distributions are skewed or leptokurtic.

We apply our methodology to individual earnings data from the PSID. We model

have been used by Schennach (2004a, 2004b) in the context of nonlinear regression and nonparametric regression, respectively, when the regressors are measured with error; and by Hu and Ridder (2007) in order to deal with measurement error when marginal information is available.

[3]The present problem is also more general than the measurement error model with multiple regressors considered in Li (2002), where at least two proxies are available for each latent regressor.

log earnings as the sum of an individual effect, a random walk and a white noise, and estimate the distributions of innovations from first differences. This model generalizes Horowitz and Markatou's (1996) by allowing for permanent shocks and non symmetric distributions. Our results show that both shocks exhibit more kurtosis than the normal distribution. We use the model to analyze the respective roles of permanent and transitory shocks in earnings mobility, and to correlate the variance of earnings shocks to job mobility. In particular, we find that frequent job changers face more permanent and more transitory earnings shocks than job stayers.

The outline of the paper is as follows. In section 2, we derive identifying restrictions on factor characteristic functions, that we use in section 3 to construct an estimator of factor densities. In section 4, we prove the consistency of the estimator and discuss convergence speed. Sections 5 and 6 present some simulations and the application. Lastly, section 7 concludes.

2 Identifying restrictions

In this section, we derive the identifying restrictions that will be used for estimation in the next section. These restrictions can also be found in Székely and Rao (2000, remark 6 p. 200).

2.1 Model and assumptions

The main elements of the model are defined as follows:

1. $Y = (Y_1, ..., Y_L)^{\mathrm{T}}$ is a vector of $L \geq 2$ zero-mean real-valued random variables (where $^{\mathrm{T}}$ denotes the matrix transpose operator).

2. $X = (X_1, ..., X_K)^{\mathrm{T}}$ is a random vector of K real valued, mutually independent and non degenerate random variables, with zero mean and finite variances.

3. $A = [a_{ij}]$ is a known $L \times K$ matrix of scalar parameters.

In this paper, we assume that factor loadings are known to the researcher. Alternatively, one may assume that a root-N consistent estimator of A is available. The asymptotic results derived in this paper would remain unchanged, as we find convergence rates of density estimators that are slower than root-N.

We make the following assumption.

Assumption A1 *The characteristic functions of factor variables* $X_1,\ldots,X_K$ *have no real zeros, and are twice differentiable.*

The assumption that characteristic functions have no real zeros is standard in the deconvolution literature. The characteristic functions can have complex zeros if factors have bounded support. Real zeros arise in the case of symmetric, bounded distributions, such as the symmetrically truncated normal (Hu and Ridder, 2007). A class of distributions that has no real zeros is the class of asymmetrically distributed, "range-restricted" distributions introduced in Hu and Ridder (2006), who argue that most economic variables belong to this class.

Assuming characteristic functions differentiable simplifies the construction of the estimator, although it is not a necessary condition for identification, as shown by Székely and Rao (2000).

Then, define Q as the $L(L+1)/2$-by-K matrix which generic (i,j) row, $i \leq j$, is $(a_{i1}a_{j1},\ldots,a_{iK}a_{jK})$. This matrix naturally appears when one writes the system of linear restrictions on factor variances:

$$\mathrm{Var}\,(Y) = A\,\mathrm{Var}\,(X)\,A^{\mathrm{T}}, \tag{1}$$

that is equivalent to

$$\mathrm{vec}\,(\mathrm{Var}\,(Y)) = [A \otimes A]\,\mathrm{vec}\,(\mathrm{Var}\,(X))\,. \tag{2}$$

Matrix Q corresponds to the $L\,(L+1)\,/2$-by-K matrix that selects the non redundant rows of $A \otimes A$ arranged in lexicographic order, and the columns that correspond to the nonzero entries of $\mathrm{vec}\,(\mathrm{Var}\,(X))$.

Given A, factor variances are obviously identifiable only if Q has full column rank. We thus make this assumption.[4]

Assumption A2 *Matrix* Q *has rank* K*.*

With $\mathrm{rank}(A) = K$, the distribution of X is trivially identified as that of $A^{-}Y$, where A^{-} is a generalized inverse of A. In contrast, rank condition A2 allows for up to $K = L(L+1)/2$ factors.

[4] If factor variances are known, Assumption A2 may not be required for identification (see Székely and Rao, 2000).

Note that Q cannot have full rank if some columns of A are proportional. In this case, any linear combination of the corresponding factors is observationally equivalent. We then say that these factors are observationally identical.

Example 1 The classical measurement error model:

$$\left\{ \begin{array}{lllll} Y_1 & = & \alpha X_1 & + & X_2, \\ Y_2 & = & X_1 & + & X_3, \end{array} \right. \tag{3}$$

has

$$A_1 = \begin{pmatrix} \alpha & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Q_1 = \begin{pmatrix} \alpha^2 & 1 & 0 \\ \alpha & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

So Q_1 has rank 3 unless $\alpha = 0$. Clearly, if $\alpha = 0$, X_1 and X_3 are observationally identical.

Example 2 The following simple spatial model:

$$\left\{ \begin{array}{lllllllll} Y_1 & = & X_1 & + & \rho X_2 & + & \rho X_3 & + & X_4, \\ Y_2 & = & \rho X_1 & + & X_2 & + & \rho X_3 & + & X_5, \\ Y_3 & = & \rho X_1 & + & \rho X_2 & + & X_3 & + & X_6, \end{array} \right. \tag{4}$$

has

$$A_2 = \begin{pmatrix} 1 & \rho & \rho & 1 & 0 & 0 \\ \rho & 1 & \rho & 0 & 1 & 0 \\ \rho & \rho & 1 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 1 & \rho^2 & \rho^2 & 1 & 0 & 0 \\ \rho & \rho & \rho^2 & 0 & 0 & 0 \\ \rho & \rho^2 & \rho & 0 & 0 & 0 \\ \rho^2 & 1 & \rho^2 & 0 & 1 & 0 \\ \rho^2 & \rho & \rho & 0 & 0 & 0 \\ \rho^2 & \rho^2 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

One verifies that Q_2 has rank 6 unless $\rho \in \{-2, 0, 1\}$. Model (4) is clearly underidentified if $\rho = 0$ or $\rho = 1$. If $\rho = 0$, X_1 and X_4 are observationally identical, and if $\rho = 1$ the three factors X_1, X_2, X_3 are observationally identical. To interpret the case $\rho = -2$, assume that the variances of X_1, X_2, X_3 are equal (to σ^2, say). Then covariances between measurements are zero for all values of σ^2, so factor variances are not identified.

2.2 Restrictions on cumulant generating functions

Under Assumption A1, cumulant generating functions (c.g.f.), i.e. the logarithm of the characteristic functions (c.f.), are well defined and everywhere two times differentiable. Let us denote the characteristic functions of Y and X_k as φ_Y and φ_{X_k} and their c.g.f.'s as $\kappa_Y = \ln \varphi_Y$ and $\kappa_{X_k} = \ln \varphi_{X_k}$. Let also $A_{[.,k]}$ denote the kth column of matrix A, for $k \in \{1, ..., K\}$. The independence assumptions and the linear factor structure imply that, for all $t = (t_1, ..., t_L) \in \mathbb{R}^L$,

$$\kappa_Y(t) \equiv \ln \left[\mathbb{E} \exp \left(it^{\mathrm{T}} Y\right)\right] = \sum_{k=1}^{K} \kappa_{X_k} \left(t^{\mathrm{T}} A_{[\cdot,k]}\right). \tag{5}$$

Equation (5) expresses the c.g.f. of the data as a linear function of the c.g.f's of X_k.

We then take second derivatives in (5), to obtain an explicit expression for κ_{X_k}. To proceed, let us denote as $\partial_\ell \kappa_Y(t)$ the ℓth partial derivative of $\kappa_Y(t)$ and as $\partial^2_{\ell m} \kappa_Y(t)$ the second-order partial derivative of $\kappa_Y(t)$ with respect to t_ℓ and t_m.We also denote as $\nabla \kappa_Y(t)$ the L-dimensional gradient vector and as $\nabla^2 \kappa_Y(t)$ the vector of all $L(L+1)/2$ non redundant second-order partial derivatives arranged in lexicographic order of $(\ell, m) \in \{1, ..., L\}$, $\ell \leq m$. Lastly, for any $\tau = (\tau_1, ..., \tau_K) \in \mathbb{R}^K$, let

$$\begin{aligned}
\kappa_X(\tau) &= \left(\kappa_{X_1}(\tau_1), ..., \kappa_{X_K}(\tau_K)\right)^{\mathrm{T}}, \\
\kappa'_X(\tau) &= \left(\kappa'_{X_1}(\tau_1), ..., \kappa'_{X_K}(\tau_K)\right)^{\mathrm{T}}, \\
\kappa''_X(\tau) &= \left(\kappa''_{X_1}(\tau_1), ..., \kappa''_{X_K}(\tau_K)\right)^{\mathrm{T}}.
\end{aligned}$$

First-differentiating equation (5) yields

$$\nabla \kappa_Y(t) = A \kappa'_X \left(A^{\mathrm{T}} t\right) = \sum_{k=1}^{K} \kappa'_{X_k} \left(t^{\mathrm{T}} A_{[\cdot,k]}\right) A_{[\cdot,k]}.$$

In general, $K > L$ as there are L errors and at least one common factor. So there are more functions κ'_{X_k} than $\partial_\ell \kappa_Y$. To obtain an invertible system, we differentiate once more:

$$\nabla^2 \kappa_Y(t) = Q \kappa''_X \left(A^{\mathrm{T}} t\right), \tag{6}$$

where Q is the $L(L+1)/2$-by-K matrix defined at the beginning of this section.

Evaluated at $t = 0$, equation (6) yields covariance restrictions (2). Generally, the equation shows that, if Assumption A2 holds, the second derivatives of the c.g.f.'s of factor variables are identified. As factors have zero mean, factor c.g.f.'s are thus identified by integration.

First, invert (6) as

$$\kappa''_X \left(A^{\mathrm{T}} t\right) = Q^- \nabla^2 \kappa_Y(t),$$

where Q^- is a generalized inverse of Q, for instance: $Q^- = \left(Q^{\mathrm{T}} Q\right)^{-1} Q^{\mathrm{T}}$.

Second, let $\mathcal{T}_k = \left\{ t \in \mathbb{R}^L | t^{\mathrm{T}} A_{[\cdot,k]} = 1 \right\}$. $\mathcal{T}_k$ is not empty as there is at least one non zero element in $A_{[\cdot,k]}$. Let $(Q^-)_{[k,\cdot]}$ denote the kth row of Q^-. Then, for all $t \in \mathcal{T}_k$ and $\tau_k \in \mathbb{R}$,

$$\kappa''_{X_k}(\tau_k) = \left(Q^-\right)_{[k,\cdot]} \nabla^2 \kappa_Y (\tau_k t).$$

Integrating with respect to τ_k, using the constants of integration: $\kappa'_{X_k}(0) = i\mathbb{E}X_k = 0$ and $\kappa_{X_k}(0) = 0$, yields

$$\kappa_{X_k}(\tau_k) = \int_0^{\tau_k} \int_0^u \left(Q^-\right)_{[k,\cdot]} \nabla^2 \kappa_Y (vt)\, dv du. \tag{7}$$

Equation (7) can directly be used for estimation of factor characteristic functions and densities, as we shall explain in the next section.

Example 1 (continued) In the case of model (3), we have:

$$\kappa_Y(t_1, t_2) = \kappa_{X_1}(\alpha t_1 + t_2) + \kappa_{X_2}(t_1) + \kappa_{X_3}(t_2),$$

and

$$\underbrace{\begin{pmatrix} \partial^2_{11}\kappa_Y(t_1,t_2) \\ \partial^2_{12}\kappa_Y(t_1,t_2) \\ \partial^2_{22}\kappa_Y(t_1,t_2) \end{pmatrix}}_{\nabla^2\kappa_Y(t)} = \underbrace{\begin{pmatrix} \alpha^2 & 1 & 0 \\ \alpha & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{Q_1} \underbrace{\begin{pmatrix} \kappa''_{X_1}(\alpha t_1 + t_2) \\ \kappa''_{X_2}(t_1) \\ \kappa''_{X_3}(t_2) \end{pmatrix}}_{\kappa''_X\left(A_1^{\mathrm{T}} t\right)}.$$

This yields:

$$\underbrace{\begin{pmatrix} \kappa''_{X_1}(\alpha t_1 + t_2) \\ \kappa''_{X_2}(t_1) \\ \kappa''_{X_3}(t_2) \end{pmatrix}}_{\kappa''_X\left(A_1^{\mathrm{T}} t\right)} = \underbrace{\begin{pmatrix} 0 & \alpha^{-1} & 0 \\ 1 & -\alpha & 0 \\ 0 & -\alpha^{-1} & 1 \end{pmatrix}}_{Q_1^-} \underbrace{\begin{pmatrix} \partial^2_{11}\kappa_Y(t_1,t_2) \\ \partial^2_{12}\kappa_Y(t_1,t_2) \\ \partial^2_{22}\kappa_Y(t_1,t_2) \end{pmatrix}}_{\nabla^2\kappa_Y(t)}.$$

Let us focus on the first factor X_1. Let $\mathcal{T}_1 = \left\{ (t_1, t_2) \in \mathbb{R}^2 | \alpha t_1 + t_2 = 1 \right\}$ and let $(t_1, 1 - \alpha t_1) \in \mathcal{T}_1$. Then, for all $v \in \mathbb{R}$:

$$\kappa''_{X_1}(v) = \frac{1}{\alpha} \partial^2_{12} \kappa_Y (vt_1, v - \alpha v t_1),$$

so

$$\kappa_{X_1}(\tau_1) = \frac{1}{\alpha} \int_0^{\tau_1} \int_0^u \partial^2_{12} \kappa_Y (vt_1, v - \alpha v t_1)\, dv du,$$

which is equation (7) for factor X_1, in the particular case of model (3).

Note that if we set $t_1 = 0$, the double integral simplifies into a simple integral:

$$\kappa_{X_1}(\tau_1) = \frac{1}{\alpha} \int_0^{\tau_1} \partial_1 \kappa_Y (0, u)\, du. \tag{8}$$

This is the equation used in Li and Vuong (1998) and Schennach (2004a, 2004b).[5]

[5] Note that Schennach (2004a, 2004b) shows that full independence is not necessary for (8) to hold.

Example 2 (continued) We then reconsider the case of the simple spatial model. Set $\rho = 1/2$ in (4). We obtain, for the first factor:

$$\kappa''_{X_1}\left(t_1 + \frac{1}{2}t_2 + \frac{1}{2}t_3\right) = \frac{8}{5}\partial^2_{12}\kappa_Y\left(t_1, t_2, t_3\right) + \frac{8}{5}\partial^2_{13}\kappa_Y\left(t_1, t_2, t_3\right) - \frac{12}{5}\partial^2_{23}\kappa_Y\left(t_1, t_2, t_3\right).$$

Let $(t_1, t_2, (1-t_1)/\rho - t_2) \in \mathcal{T}_1$. We have, after integrating twice:

$$\kappa_{X_1}\left(\tau_1\right) = \int_0^{\tau_1}\int_0^{u}\Big\{\frac{8}{5}\partial^2_{12}\kappa_Y\left(vt_1, vt_2, \frac{v(1-t_1)}{\rho} - vt_2\right) + \frac{8}{5}\partial^2_{13}\kappa_Y\left(vt_1, vt_2, \frac{v(1-t_1)}{\rho} - vt_2\right) - \frac{12}{5}\partial^2_{23}\kappa_Y\left(vt_1, vt_2, \frac{v(1-t_1)}{\rho} - vt_2\right)\Big\}dvdu.$$

Note that, if $t_1 = t_2 = 0$, the last two terms on the right-hand side simplify into simple integrals, but the first term does not.

3 The estimator

We here introduce our estimator of factor densities. In the next section, we shall discuss its asymptotic properties.

3.1 Characteristic functions

Estimator. Given an i.i.d. sample of size N, $Y_1, ..., Y_N$, we first estimate κ_Y and its derivatives by empirical analogs, replacing mathematical expectations by arithmetic means:

$$\widehat{\kappa}_Y(t) = \ln\left(\mathbb{E}_N\left[e^{it^{\mathrm{T}}Y}\right]\right),$$

$$\widehat{\partial_\ell\kappa_Y}(t) = i\frac{\mathbb{E}_N\left[Y_\ell e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}_N\left[e^{it^{\mathrm{T}}Y}\right]} = \partial_\ell\widehat{\kappa}_Y(t),$$

and

$$\widehat{\partial^2_{\ell m}\kappa_Y}(t) = -\frac{\mathbb{E}_N\left[Y_\ell Y_m e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}_N\left[e^{it^{\mathrm{T}}Y}\right]} + \frac{\mathbb{E}_N\left[Y_\ell e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}_N\left[e^{it^{\mathrm{T}}Y}\right]}\frac{\mathbb{E}_N\left[Y_m e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}_N\left[e^{it^{\mathrm{T}}Y}\right]} = \partial^2_{\ell m}\widehat{\kappa}_Y(t),$$

where $\mathbb{E}_N$ denotes the empirical expectation operator.

Then, as the choice of t in $\mathcal{T}_k = \left\{t \in \mathbb{R}^L | t^{\mathrm{T}}A_{[\cdot,k]} = 1\right\}$, along which to perform the integration yielding $\kappa_{X_k}\left(\tau_k\right)$, is arbitrary, one can estimate κ_{X_k} by averaging (7) over a distribution of points in $\mathcal{T}_k$, that is,

$$\begin{aligned}\widehat{\kappa}_{X_k}(\tau_k) &= \int_0^{\tau_k}\int_0^u \left(Q^-\right)_{[k,\cdot]}\left(\int \nabla^2\widehat{\kappa}_Y(vt)\,dW(t)\right)dvdu \\ &= \int_0^{\tau_k}\int_0^u \left(Q^-\right)_{[k,\cdot]}\left(\sum_{j=1}^p w_j\nabla^2\widehat{\kappa}_Y(vt_j)\right)dvdu,\end{aligned}$$

where $W = \sum_{j=1}^p w_j\delta_{t_j}$ is a discrete probability distribution on $\mathcal{T}_k$.

The characteristic function of X_k is then estimated as

$$\widehat{\varphi}_{X_k}(\tau_k) = \exp\left(\int_0^{\tau_k}\int_0^u \left(Q^-\right)_{[k,\cdot]}\left(\sum_{j=1}^p w_j\nabla^2\widehat{\kappa}_Y(vt_j)\right)dvdu\right). \tag{9}$$

Choice of W. Empirical characteristic functions are typically well estimated around the origin and badly estimated in the tails (e.g., Diggle and Hall, 1993). It thus makes sense to choose t such that $\nabla^2\kappa_Y(\tau_k t)$ is well estimated on a maximal interval. A natural choice is to minimize the Euclidian norm of $\frac{t}{t^{\mathrm{T}}A_{[\cdot,k]}}$, which yields, by Cauchy-Schwarz inequality:

$$t^* = \frac{A_{[\cdot,k]}}{A^{\mathrm{T}}_{[\cdot,k]}A_{[\cdot,k]}}.$$

For instance, in example 1, this choice yields:

$$\widehat{\kappa}_{X_1}(\tau_1) = \frac{1}{\alpha}\int_0^{\tau_1}\int_0^u \partial^2_{12}\widehat{\kappa}_Y\left(v\frac{\alpha}{1+\alpha^2}, v\frac{1}{1+\alpha^2}\right)dvdu.$$

In our simulations we found that choosing $W = \delta_{t^*}$ works well in practice. It is worth noting that the many overidentifying restrictions that the model provides could be used to improve the efficiency of the estimator. This question is very interesting, but seems also very difficult to answer. We do not address efficiency issues in this paper.

3.2 Density functions

The probability distribution function (p.d.f.) of X_k, say f_{X_k}, is obtained from its characteristic function using an inverse Fourier transformation:

$$f_{X_k}(x_k) = \frac{1}{2\pi}\int e^{-i\tau_k x_k}\varphi_{X_k}(\tau_k)d\tau_k.$$

It is well-known that the integral does not converge when the characteristic function is replaced by its empirical analog (e.g., Horowitz, 1998). To ensure convergence we truncate the integral on a compact interval $[-T_N, T_N]$, where T_N tends to infinity with the sample size N at a rate that will be discussed in the next section. The p.d.f. of X_k is then estimated as

$$\widehat{f}_{X_k}(x_k) \ = \ \frac{1}{2\pi}\int \varphi_H\left(\frac{\tau_k}{T_N}\right) e^{-i\tau_k x_k}\widehat{\varphi}_{X_k}(\tau_k)d\tau_k, \tag{10}$$

where $\widehat{\varphi}_{X_k}(\tau_k)$ is given by (9). In equation (10), φ_H is a function supported on $[-1,1]$ that is the Fourier transform of a kernel H of even order: $\varphi_H(u) = \int e^{iuv}H(v)dv$.[6]

The kernel H allows to smooth the estimation of the density, especially in the tails. We shall use the second-order kernel

$$H_2(v) = \frac{48\cos(x)}{\pi x^2}\left(1-\frac{15}{x^2}\right) - \frac{144\sin(x)}{\pi x^5}\left(2-\frac{5}{x^2}\right),$$

that corresponds to:

$$\varphi_{H_2}(u) = \left(1-u^2\right)^3 \cdot \mathbf{1}\{u \in [-1,1]\}.$$

The second-order kernel H_2 has often been used in the deconvolution literature (see, e.g., Delaigle and Gijbels, 2002, and references therein). Higher-order kernels may also be used in place of H_2, such as the fourth-order kernel H_4 given by

$$\varphi_{H_4}(u) = \left(1-u^4\right) \cdot \mathbf{1}\{u \in [-1,1]\},$$

or the infinite-order kernel $H_\infty(v) = \sin(v)/v$, that yields $\varphi_{H_\infty}(u) = \mathbf{1}\{u \in [-1,1]\}$. Higher-order kernels reduce the bias of the density estimate at the cost of higher variance. For instance, Li and Vuong (1998) use the infinite-order kernel H_∞.

4 Asymptotic properties

In this section, we study the asymptotic properties of the estimator and show that $\widehat{f}_{X_k}$ is a uniformly consistent estimator of f_{X_k}, for all $k = 1, ..., K$, provided that Assumptions A1 and A2 hold. All mathematical proofs are in the appendix.

We shall prove the result for any support of factor distributions. So the analysis contains the case of bounded support, assumed e.g. in Li and Vuong (1998) and Hall and Yao (2003), as well as the "range-restricted" distributions considered in Hu and Ridder (2006, 2007). An important assumption is that factor characteristic functions have no real zeros.

[6]A kernel of order q is a function H, not necessarily nonnegative, such that $v^kH(v)$ is integrable for all $k \leq q$, $\int v^kH(v)dv = 0$ for all $k \leq q-1$, and $\int v^qH(v)dv \neq 0$. See, e.g., Rao (1983), p. 40.

4.1 Characteristic functions

To prove the consistency of the estimator in the case of not necessarily bounded support, we first prove a uniform consistency result for the derivatives of empirical characteristic functions. For any vector $t \in \mathbb{R}^S$ $(S > 0)$, we denote the sup norm: $|t| = \max_s |t_s|$.

Lemma 1 *Let X be a scalar random variable and let Y be a vector of L scalar random variables. Let $Z = \left(X, Y^{\mathrm{T}}\right)^{\mathrm{T}}$. Let F denote the c.d.f. of Z ($\mathbb{E}$ denotes the corresponding expectation operator) and let F_N (resp. $\mathbb{E}_N$) denote the empirical c.d.f. (resp. mean) corresponding to a sample $\mathbf{Z}_N \equiv (Z_1, ..., Z_N)$ of N i.i.d. draws from F. Assume that the first J moments of X^2 and $|XY|$ are finite ($J \geq 2$). Lastly, define $f_t(x, y) = x \exp(it^{\mathrm{T}}y)$, for $t \in \mathbb{R}^L$.*

For any $0 < \gamma < (1 - \frac{1}{J})/2$, let T_N tend to infinity at (at most) a polynomial rate. Then:

$$\sup_{|t| \leq T_N} |\mathbb{E}_N f_t - \mathbb{E} f_t| = O\left(N^{-\frac{1}{2}\left(1-\frac{1}{J}\right)+\gamma}\right) \quad a.s.$$

Lemma 1 shows that the rate of convergence of the empirical mean of f_t is at most $N^{-(1-1/J)/2}$ on an interval that grows with N at a polynomial rate. So, allowing for unbounded X slowers the rate of convergence. However, in the particular case where all the moments of X and $|XY|$ exist, Lemma 1 delivers a rate of $N^{-1/2+\gamma}$, for any $0 < \gamma < 1/2$. When $X = 1$ and L is either 1 or 2, the result coincides with Lemma 6 in Hu and Ridder (2007).

Applying Lemma 1 to $\mathbb{E}\left(Y_\ell \exp(it^{\mathrm{T}}Y)\right)$ and $\mathbb{E}\left(Y_\ell Y_m \exp(it^{\mathrm{T}}Y)\right)$, for $\ell, m = 1, ..., L$, then yields the following uniform consistency result for the characteristic functions of factors.

Theorem 2 *Suppose that there exists an integrable, decreasing function $g_Y : \mathbb{R}^+ \to [0, 1]$, such that $|\varphi_Y(t)| \geq g_Y(|t|)$ as $|t| \to \infty$. Suppose also that the first $2J$ moments of $|Y|$ are finite.*

For any $0 < \gamma < \left(1 - \frac{1}{J}\right)/2$, let T_N tend to infinity at (at most) a polynomial rate, and be such that $\frac{T_N^2}{g_Y(T_N)^3} = o\left(N^{\frac{1}{2}\left(1-\frac{1}{J}\right)-\gamma}\right)$. Then:

$$\sup_{|\tau_k| \leq T_N} \left|\widehat{\varphi}_{X_k}(\tau_k) - \varphi_{X_k}(\tau_k)\right| = \frac{T_N^2}{g_Y(T_N)^3} O\left(N^{-\frac{1}{2}\left(1-\frac{1}{J}\right)+\gamma}\right) \quad a.s. \tag{11}$$

Theorem 2 shows that the rate of convergence of the e.c.f. of factors is governed by the tails of the characteristic functions, i.e. the smoothness of factor distribution functions. In the deconvolution problem where one factor distribution is known, Fan (1991) distinguished two classes of distributions: ordinary smooth, for which the c.f. converges to zero at a polynomial rate (e.g., Laplace or Gamma), and supersmooth distributions, for which the c.f. converges to zero at an exponential rate (e.g., normal). Theorem 2 implies that the intervals on which uniform convergence holds are very different in these two cases.

Let us consider the result in Theorem 2 in the case where all factor distributions are ordinary smooth with $g_Y(|t|) = |t|^{-\beta}$, where $\beta > 0$. Then, if $T_N = N^\delta$, for any $\delta < \frac{(1-1/J)/2-\gamma}{2+3\beta}$, the e.c.f. converges uniformly on $[-T_N, T_N]$. If instead factor distributions are supersmooth, one needs to restrict T_N to a logarithmic function of N in order to ensure uniform convergence. The next paragraph shows that this weaker uniform convergence result implies a slower rate of convergence of density estimates, as in Fan (1991).

4.2 Density functions

The following theorem gives conditions under which $\widehat{f}_{X_k}$ converges uniformly to f_{X_k} when the sample size tends to infinity.

Theorem 3 *Suppose that there exists an integrable, decreasing function $g_X : \mathbb{R}^+ \to [0,1]$ such that $|\varphi_X(\tau)| \geq g_X(|\tau|)$ as $|\tau| \to \infty$. Suppose also that there exist K integrable functions $h_{X_k} : \mathbb{R}^+ \to [0,1]$ such that $h_{X_k}(|\tau_k|) \geq |\varphi_{X_k}(\tau_k)|$ as $|\tau_k| \to \infty$. Lastly, suppose that the first $2J$ moments of $|Y|$ are finite.*

For any $0 < \gamma < \left(1 - \frac{1}{J}\right)/2$, let T_N tend to infinity at (at most) a polynomial rate, and be such that $\frac{T_N^3}{g_Y(T_N)^3} = o\left(N^{\frac{1}{2}\left(1-\frac{1}{J}\right)-\gamma}\right)$. Then:

$$\begin{aligned}\sup_{x_k}\left|\widehat{f}_{X_k}(x_k) - f_{X_k}(x_k)\right| &= \frac{T_N^3}{g_X(T_N)^3} O\left(N^{-\frac{1}{2}\left(1-\frac{1}{J}\right)+\gamma}\right) \\ &\quad + O\left(\int \left|1 - \varphi_H\left(\frac{v}{T_N}\right)\right| h_{X_k}(|v|)dv\right) \quad a.s. \quad (12)\end{aligned}$$

Assume in addition that $\frac{1}{T_N^2}\int_{-T_N}^{T_N} v^2 h_{X_k}(|v|)dv = o(1)$, a.s, and that H is a kernel of order $q \geq 2$. Then we have:

$$\sup_{x_k}\left|\widehat{f}_{X_k}(x_k) - f_{X_k}(x_k)\right| = o(1) \quad a.s.$$

The two terms on the right-hand side of (12) are the variance and the bias of $\widehat{f}_{X_k}$, respectively. Both terms depend on the smoothness of factor distributions, though in an opposite way. The variance term is larger when factor distributions are smoother. In the deconvolution problem $Y_1 = X_1 + X_2$, with f_{X_2} known, it is well-known that estimation is more difficult when the distribution of X_2 is supersmooth. This is because the c.f. of X_2 appears in the denominator in the Fourier inversion. A similar argument applies here. In contrast, the bias term in (12) decreases when the smoothness of factor X_k increases.

To better understand Theorem 3, let us consider the polar case when all factors are ordinary smooth. Let us take $g_X(|t|) = |t|^{-\beta}$ and $h_{X_k}(|t|) = |t|^{-\alpha}$, $\beta \geq \alpha > 1$. We can take

$$T_N = N^{\delta}, \text{ with } 0 < \delta < \frac{\frac{1}{2}\left(1 - \frac{1}{J}\right) - \gamma}{2 + 3\beta + \alpha}, \tag{13}$$

so that the rate of convergence of the density estimator is at most $N^{-\delta(\alpha-1)}$. In the case where we characterize exactly the degree of smoothness of factor variables ($\alpha = \beta$) the rate becomes:

$$\sup_{x_k} \left| \widehat{f}_{X_k}(x_k) - f_{X_k}(x_k) \right| = O\left(N^{\frac{\left(-\frac{1}{2}\left(1-\frac{1}{J}\right)+\gamma\right)(\beta-1)}{2+4\beta}} \right). \tag{14}$$

Another case of interest is when X_k is ordinary smooth, so $h_{X_k}(|t|) = |t|^{-\alpha}$, and other factors are supersmooth, so one has to take $g_X(|t|) = \exp\left(-|t|^{\beta}\right)$. In that case one obtains a logarithmic convergence rate.

Equation (14) shows that, even if we restrict our attention to ordinary smooth distributions, the convergence rate is never faster than $N^{-\frac{1-1/J}{8}}$. So, even in the case where all moments of $|Y|$ exist, the rate is never faster than $N^{-\frac{1}{8}}$. This convergence rate is slower than the one in Li and Vuong (1998), $N^{-\frac{1}{6}+\gamma}$ (for $\gamma > 0$) in the ordinary smooth case. This difference is due to the fact that our estimator relies on a double integral, instead of a single integral. This is the price to pay for dealing with more general models. The convergence rate we obtain is also slower than in the classical deconvolution problem $Y_1 = X_1 + X_2$, with f_{X_2} known. In this case, Hu and Ridder (2007, page 10) obtain a rate of $N^{-\frac{1}{2}+\gamma}$, when the smoothness of X_1 increases while that of X_2 stays constant. In the case where X_1 and X_2 have the same degree of smoothness β, their results imply a rate of $N^{-\frac{1}{4}+\gamma}$ when β gets large.

Convergence speed is an important issue when dealing with two-stage estimation problems in which the estimation of factor densities is the first stage. For instance, the distributions of permanent/transitory components can be inputs of an intertemporal consumption problem. The convergence rate of the first stage factor density estimators is likely to be too slow to obtain root-N consistent estimates in the second stage. For this same reason Li (2002), using Li and Vuong's (1998) nonparametric estimator in the first stage of a nonlinear error-in-variables models, does not get root-N consistency.[7]

Lastly, it is worth noting that in spite of the different theoretical convergence rates, these various deconvolution estimators generated very similar biases and variances in our finite-sample experiments (see the simulation section below).

4.3 Practical choice of the trimming parameter T_N

It is tempting to use (13) as a guideline to choose T_N in practice, at least in the case where all factor distributions are smooth. However, our experiments suggest that by doing so one underestimates T_N. The reason could be that T_N maximizes an upper bound for the convergence rate, which can be very conservative, especially so in finite samples.

Instead, we use a method recently developed in deconvolution kernel density estimation to choose the trimming parameter T_N. In the context of the deconvolution problem with known error distribution, Delaigle and Gijbels (2002, 2004) propose to base the choice of the bandwidth on an approximation to the Mean Integrated Squared Error of the kernel density estimator. Comparing different approaches they find that a "plug-in" method works well in many simulation designs. We use the "plug-in" method, and provide a presentation of the method in Appendix.

To adapt Delaigle and Gijbels' method to the case of a multi-factor model $Y = AX$, we proceed as follows. For $k \in \{1, ..., K\}$, let $t^* = \frac{A_{[.,k]}}{A_{[.,k]}^{\mathrm{T}} A_{[.,k]}}$. Then

$$t^{*\mathrm{T}} Y = t^{*\mathrm{T}} AX = X_k + \sum_{m \neq k} t^{*\mathrm{T}} A_{[.,m]} X_m. \tag{15}$$

We treat the distribution of $\sum_{m \neq k} t^{*\mathrm{T}} A_{[.,m]} X_m$ in (15) as if it were known. In this case, the problem of estimating the density of factor X_k boils down to a deconvolution problem with known error distribution, and the approach in Delaigle and Gijbels (2004) can be applied.

---

[7] An alternative has been considered by Schennach (2004a), expressing the moments necessary to perform the second-stage nonlinear regression directly in terms of characteristic functions, without trimming. This method requires rather strong regularity conditions (e.g., existence of moments), but yields root-N consistency and asymptotic normality.

5 Monte-Carlo simulations

In this section, we study the finite-sample behavior of our density estimator.

5.1 Measurement error model

We start with the estimation of the density of X_1 in the measurement error model (3) with $\alpha = 1$, namely:

$$\begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = X_1 + X_3, \end{cases} \tag{16}$$

where X_1, X_2 and X_3 are mutually independent, and have mean zero and variance one.

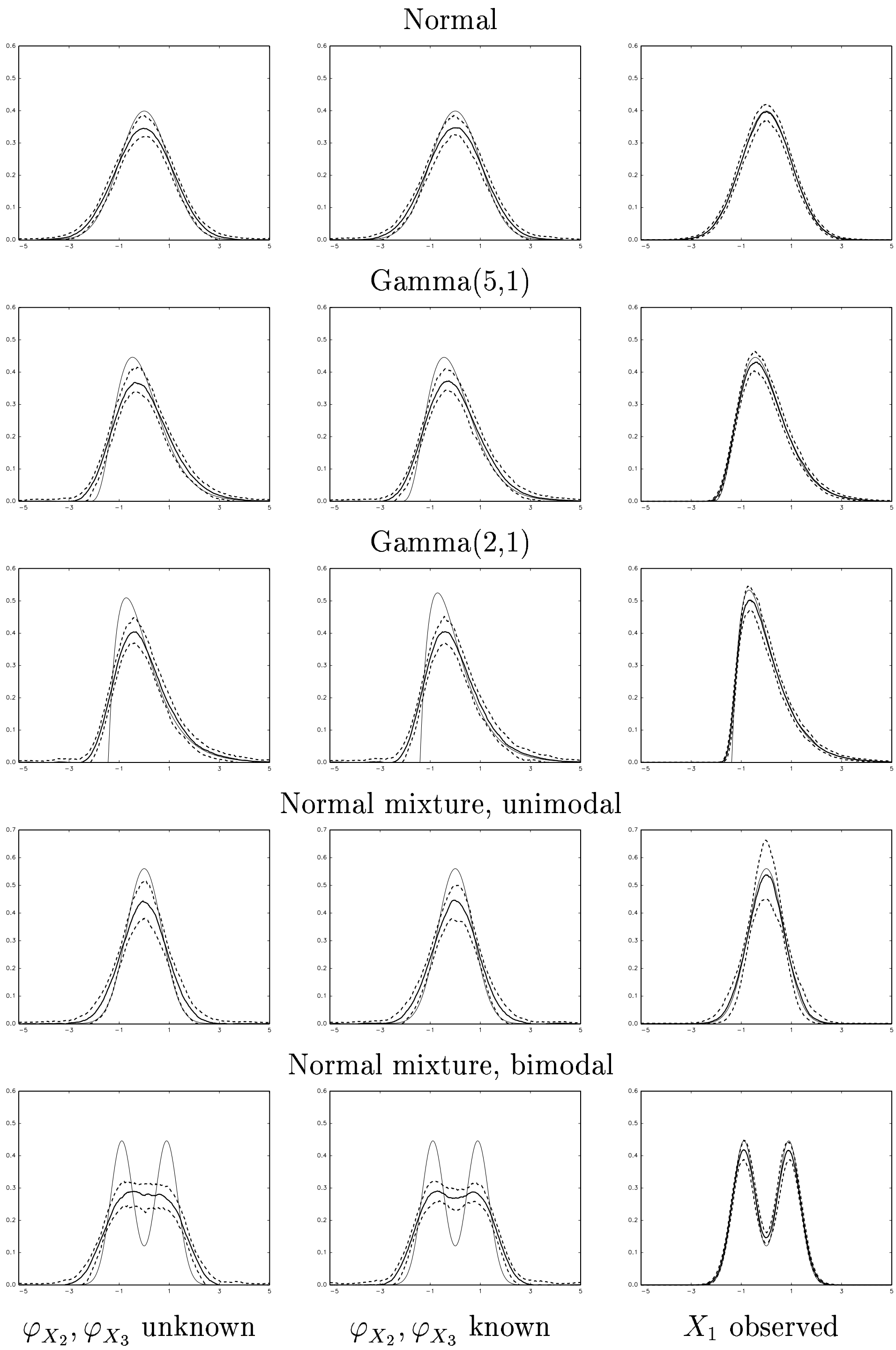
We first consider the case of normal errors X_2 and X_3, and various choices of distribution for X_1. In Figure 1 we report the outcomes of 100 simulations of samples of size $N = 1000$. In the first column we estimate the density of X_1 using the method of this paper, assuming that all three distributions are unknown. In the second column we estimate the density of X_1 from:

$$\frac{Y_1 + Y_2}{2} = X_1 + \frac{X_2 + X_3}{2},$$

assuming that $\frac{X_2+X_3}{2}$ has known c.f. $\varphi_{\frac{X_2+X_3}{2}}(u) = \exp\left(-\frac{1}{4}u^2\right)$. We use kernel deconvolution for estimation, with the second-order kernel H_2 for smoothing, and choose the trimming parameter T_N using the "plug-in" method of Delaigle and Gijbels (2004). Lastly, in the third column we show the Gaussian kernel density estimator of X_1 for comparison, using Silverman's rule of thumb for choosing the bandwidth. On each graph, the thin solid line represents the population density of X_1, and the thick solid line is the pointwise median of simulations. The dashed lines delimit the 10%-90% pointwise confidence bands.

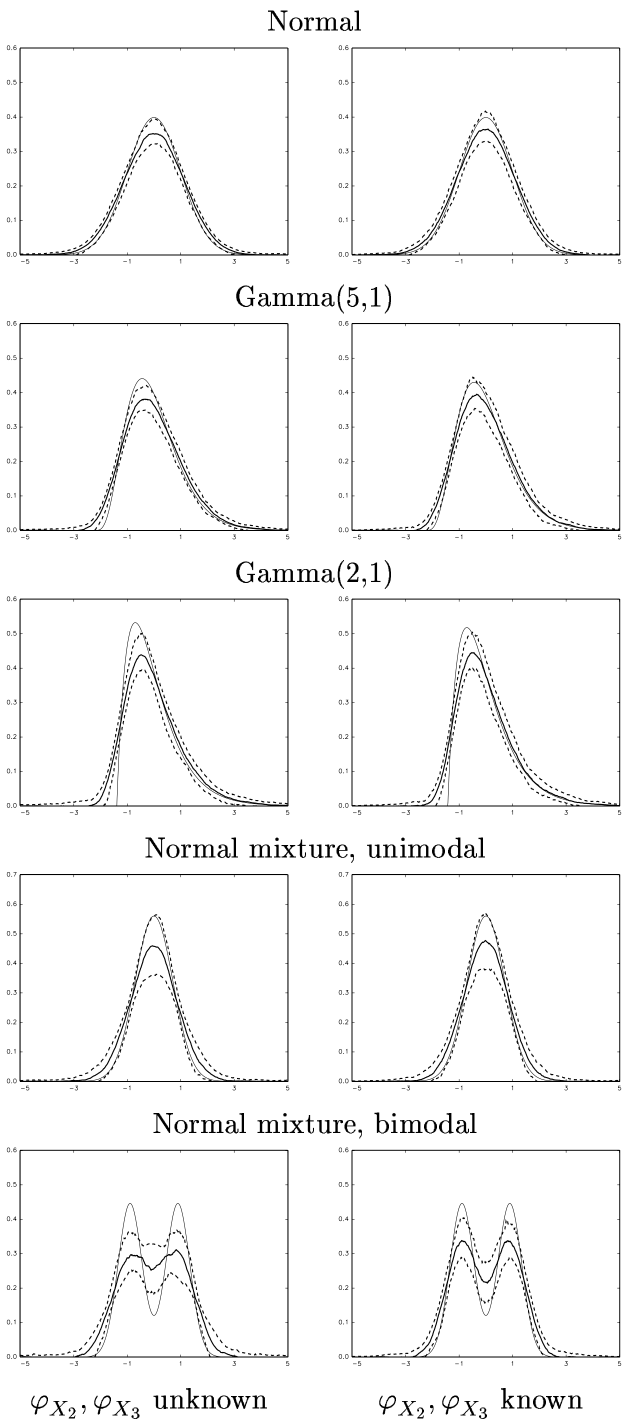
Both nonparametric deconvolution methods estimate normal factor distributions well. However, the density at the mode is biased–the true value being outside the confidence band. They both display very similar biases, and the same confidence bands, only moderately wider than when X_1 is observed without error. This suggests that repeated measurements can be very effective at providing information on the distributions of unknown latent variables. Also, the informal choice of bandwidth that we use appears to give very good results, as good as for the deconvolution problem with known error distribution for which it was initially devised.

Figure 1: Monte Carlo estimates of f_{X_1} in the measurement error model (16) with normal errors



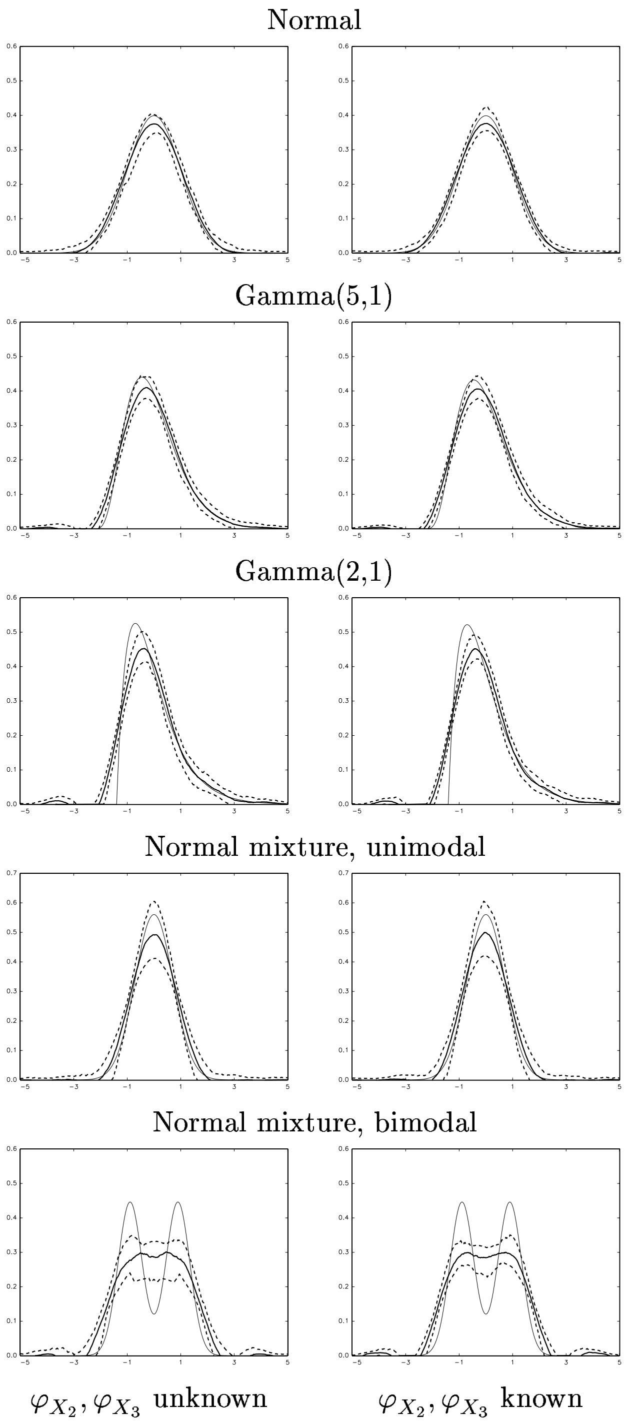
Note: Thin line=true; thick=median of 100 simulations; dashed=10%-90% confidence bands. "Normal mixture, unimodal" is $\frac{400}{403}\mathcal{N}(0, \frac{1}{2}) + \frac{3}{403}\mathcal{N}(0, \frac{406}{6})$*, "Normal mixture, bimodal" is* $\frac{1}{2}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(2, 1)$*.* $N = 1000$*.*

Figure 2: Monte Carlo estimates of f_{X_1} in the measurement error model (16) with Laplace errors



Note: Thin line=true; thick=median of 100 simulations; dashed=10%-90% confidence bands. "Normal mixture, unimodal" is $\frac{400}{403}\mathcal{N}(0, \frac{1}{2}) + \frac{3}{403}\mathcal{N}(0, \frac{406}{6})$*, "Normal mixture, bimodal" is* $\frac{1}{2}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(2, 1)$*.* $N = 1000$*.*

Figure 3: Monte Carlo estimates of f_{X_1} in the measurement error model (16) with normal errors, using a fourth-order kernel



Note: Thin line=true; thick=median of 100 simulations; dashed=10%-90% confidence bands. "Normal mixture, unimodal" is $\frac{400}{403}\mathcal{N}(0,\frac{1}{2}) + \frac{3}{403}\mathcal{N}(0,\frac{406}{6})$*, "Normal mixture, bimodal" is* $\frac{1}{2}\mathcal{N}(-2,1) + \frac{1}{2}\mathcal{N}(2,1)$*.* $N = 1000$.

For non Gaussian factor distributions, we observe that the deconvolution estimators have some difficulty to capture skewness and kurtosis. The Gamma$(5,1)$ and Gamma$(2,1)$ distributions have skewness .9 and 1.4, and kurtosis 4.2 and 6, respectively. We see that the bias is larger in the second case. Note that the Gamma distribution is smooth, while X_2 and X_3 follow supersmooth normal distributions. In this case, theory suggests that the deconvolution problem is especially difficult.

To further study the impact of factor kurtosis on estimation we consider for X_1 a two-components normal mixture that has excess kurtosis equal to 100, that is: $X_1 \sim \frac{400}{403}\mathcal{N}(0,\frac{1}{2}) + \frac{3}{403}\mathcal{N}(0,\frac{406}{6})$. The bias is also larger than in the case where X_1 is normal, although the estimator does a good job at capturing the peak of the density.

Lastly, we generate a bimodal distribution as a two-component mixture of normals with different means: $X_1 \sim \frac{1}{2}\mathcal{N}(-2,1) + \frac{1}{2}\mathcal{N}(2,1)$. The estimator fails to capture the bimodality.

It is worth noting that in these various designs, we experimented increasing the sample size to $N = 10000$, and still obtained a sizeable bias (although reduced compared to the case $N = 1000$).

Figure 2 presents simulation results for Laplace-distributed errors. Theory suggests that the deconvolution estimator should behave better, and the bias is indeed slightly lower than in the case of normal errors, especially in the most difficult case where the density of X_1 is bimodal. Still, the differences between the cases of smooth and supersmooth errors do not seem as large as theory suggests.

In a last experiment, we let again X_2 and X_3 follow standard normal distributions, but use the fourth-order kernel H_4 for estimation rather than the second-order kernel H_2 (see section 3). We also use Delaigle and Gijbels' (2004) plug-in method to choose T_N. The results displayed in Figure 3 show that the bias is reduced compared to the estimator using a second-order kernel. We simultaneously observe a slight widening of the confidence bands, especially in the tails.[8]

[8] We also tried a sixth-order kernel, and obtained a sharp reduction in the bias but a large increase in the variance.

5.2 Spatial model

We then consider the spatial model with $L = 3$ and $K = 6$:

$$\begin{cases} Y_1 &= 2X_1 &+ X_2 &+ X_3 &+ X_4 \\ Y_2 &= X_1 &+ 2X_2 &+ X_3 &+ X_5 \\ Y_3 &= X_1 &+ X_2 &+ 2X_3 &+ X_6, \end{cases} \tag{17}$$

where X_k, $k = 1, ..., 6$, are mutually independent. This corresponds to model (4) with $\rho = 1/2$.

All factor densities belong to the same parametric family. We only let their variances differ: the variances of X_1, X_2 and X_3 are equal to 1, while X_4, X_5 and X_6 have either variance 1 (first column in the figure), 4 (second column), or 16 (third column). The sample size is $N = 1000$, and the number of simulations and the conventions used in the graphs are the same as for the measurement error model.

Figure 4 presents the results. We see that when errors X_4, X_5 and X_6 have moderate variance (1 or 4) the density of X_1 is well estimated. The results are comparable to the ones obtained in Figure 1, with a slightly larger bias. When error variances increase to 16, the density of X_1 becomes badly estimated.

For other distributions, namely Gamma or mixture of normals, we generally obtain worse results than for the measurement error model (16). This is consistent with the fact that trying to estimate 6 factor densities using 3 measurements is more difficult than estimating 3 factor densities using 2 measurements. Yet, in the case of moderate error variances the shapes of the densities are reasonably well reproduced. This suggests that nonparametric deconvolution techniques can be successfully applied to difficult problems, where the number of factors one is trying to extract is large relative to the number of available measurements.

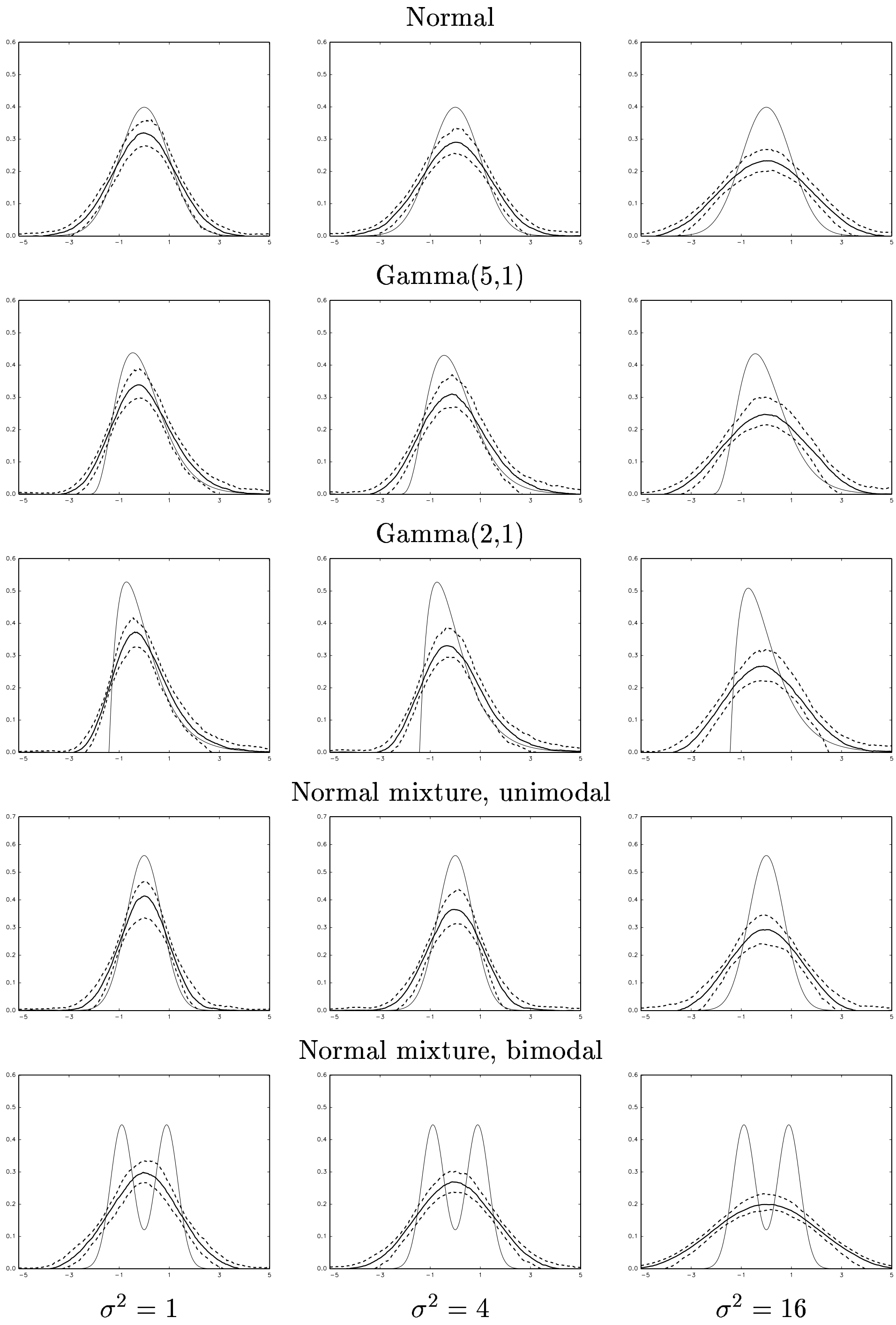
6 Application to earnings dynamics

In this section, we apply our methodology to estimate the distributions of permanent and transitory shocks in a simple model of earnings dynamics.

6.1 The data

We use PSID data, between 1978 to 1987. Let w_{it} denote the logarithm of annual earnings, and let x_{it} be a vector of regressors, namely: education dummies, a quadratic

Figure 4: Monte Carlo estimates of f_{X_1} in model (17)



Note: Density of X_1 and X_4 in model (17). X_k, $k = 1, 2, 3$, are drawn from the same distribution with mean zero and variance 1. X_k, $k = 4, 5, 6$, are drawn from the same distribution as X_1, X_2, X_3 with mean zero and variance σ^2. "Normal mixture, unimodal" is $\frac{400}{403}\mathcal{N}(0, \frac{1}{2}) + \frac{3}{403}\mathcal{N}(0, \frac{406}{6})$, "Normal mixture, bimodal" is $\frac{1}{2}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(2, 1)$. $N = 1000$.

polynomial in age, a race dummy, geographic indicators and year dummies. We compute the residuals of the OLS regression of $\Delta w_{it} = w_{it} - w_{it-1}$ on $\Delta x_{it} = x_{it} - x_{it-1}$, and denote them as Δy_{it}. In the sequel we shall refer to Δy_{it} as wage growth residuals, while keeping in mind that they reflect changes in wage rates and hours worked. We select employed male workers who have non missing observations of Δy_{it} for the whole period, and for whom wage growth does not exceed 150% in absolute value. We obtain a balanced panel of 624 individuals, for whom we have 9 observations of wage growth. Descriptive statistics are presented in the first column of Table 1.

Wage growth residuals Δy_{it} are the measurements that we use in this application. We shall also consider moving sums of wage growth residuals, defined as

$$\Delta_s y_{it} = y_{it} - y_{i,t-s} = \sum_{k=1}^{s} \Delta y_{i,t-k+1}, \text{ for } s = 1, 2, ...$$

Table 2 shows the marginal moments of these variables, as well as their first three autocorrelation coefficients. Focusing on the first row, we see that the variance of $\Delta_s y_{it}$ increases with s. This indicates that wage differences between two points in time are more dispersed the longer the lag.

Another feature of Table 2 is the high kurtosis of wage growth residuals. Figure 5 confirms that the distribution of Δy_{it} is very different from the normal. On panel a), the solid line is a kernel density estimate, and the dashed line is the density of a normal distribution with the same mean and variance. An alternative way of presenting the evidence of non-normality is to draw the normal probability plot of Δy_{it}. If the data are normally distributed, then $\Phi^{-1}(F_N(\Delta y_{it}))$, where $F_N(\Delta y_{it})$ denotes the empirical c.d.f. of Δy_{it}, is a straight line up to sampling error. Panel b) in Figure 5 shows that this is not the case, as the c.d.f. of Δy_{it} has fatter tails than the normal. This evidence on the non-normality of wage growth residuals is consistent with previous findings on U.S. data, e.g. Horowitz and Markatou (1996) who use data from the Current Population Survey.

6.2 The model

We consider the following model:

$$\begin{aligned} \Delta y_{it} &= \Delta p_{it} + \Delta r_{it}, \\ &= \varepsilon_{it} + r_{it} - r_{it-1}, \quad i = 1, ..., N, \quad t = 2, ..., T, \end{aligned} \tag{18}$$

Table 1: Means of variables

Job changes	All	None	One/two	Three/more
Annual earnings (/1000)	36.4	35.3	36.7	37.0
Age	37.4	39.6	37.3	36.1
High school dropout	.21	.22	.23	.16
High school graduate	.54	.59	.51	.54
Hours	2194	2191	2199	2191
Married	.85	.84	.84	.85
White	.70	.63	.69	.75
North east	.15	.15	.13	.17
North central	.26	.30	.24	.27
South	.43	.44	.51	.36
SMSA	.59	.60	.55	.61
Number	624	150	234	240

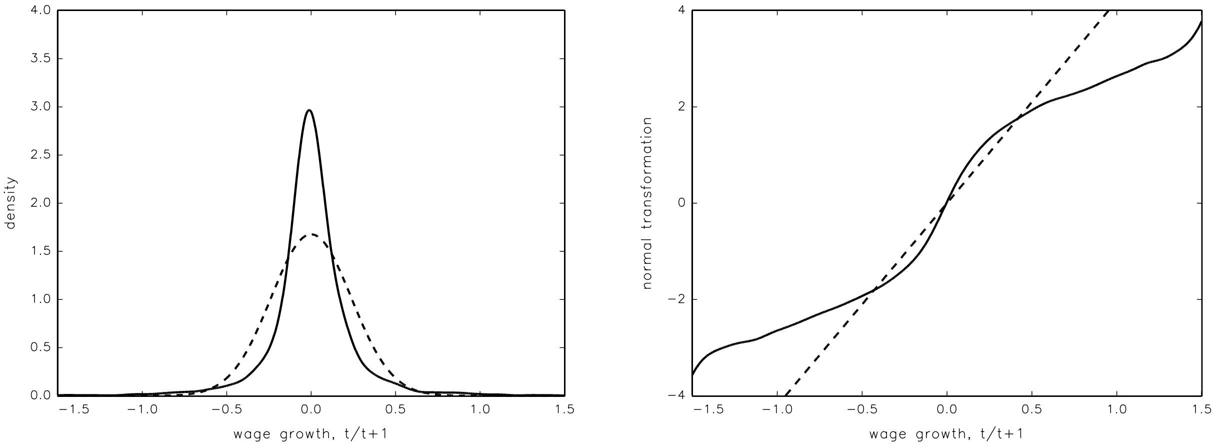
Note: Balanced subsample of 624 individuals extracted from the PSID, 1978-1987. "None"=no job change; "One/two"= one or two job changes; "Three/more"= more than three job changes.

Table 2: Moments of wage growth residuals

Wage growth	$t/t+1$	$t/t+2$	$t/t+3$	$t/t+T$
Variance	.055	.073	.086	.137
Skewness	-.077	.062	-.073	.457
Kurtosis	10.3	11.2	8.0	4.8
Autocorrelation 1	-.33	.21	.35	-
Autocorrelation 2	-.06	-.34	.08	-
Autocorrelation 3	-.02	-.06	-.34	-

Note: Balanced subsample of 624 individuals extracted from the PSID, 1978-1987. Wage growth residuals are the OLS residuals of first-differenced log earnings on regressors. Wage growth between t and $t+s$ is obtained as the sum of s consecutive wage growth residuals.

Figure 5: Non normality of wage growth residuals



a) Density b) Normal probability plot

Note: See the note to Table 2. a): density estimate of wage growth residuals (solid), and density of normal with same mean and variance (dashed); b): normal probability plot of wage growth residuals (solid), and for normal with same mean and variance (dashed).

where p_{it} follows a random walk: $p_{it} = p_{it-1} + \varepsilon_{it}$, where ε_{it} and r_{it} are white noise innovations with variances σ_ε^2 and σ_r^2. We shall refer to p_{it} as the permanent component and to r_{it} as the transitory component.

Permanent-transitory decompositions are very popular in the earnings dynamics literature, see among others Hall and Mishkin (1982) and Abowd and Card (1989). There is a growing concern that the distributions of wage shocks might be non normal (e.g., Geweke and Keane, 2000). To assess this issue, Horowitz and Markatou (1996) estimate a model of earnings levels with an individual fixed effect and a transitory i.i.d. shock. There is no permanent shock in their model. Their estimation procedure is fully nonparametric. However, one particular implication of their model is that Δy_{it}, $\Delta_2 y_{it}$, ... are identically distributed. This is clearly at odds with the evidence presented in Table 2. The introduction of a permanent component easily permits to capture the increase in $\text{Var}(\Delta_s y_{it})$ when s increases.[9] The generalized deconvolution technique of this paper allows to conduct the same fully nonparametric analysis as in Horowitz and Markatou (1996) while allowing for a permanent component in wages.

[9]Notice that model (18) implies that: $\text{Var}(\Delta_s y_{it}) - \text{Var}(\Delta y_{it}) = (s-1)\,\sigma_\varepsilon^2$. The marginal distributions of Δy_{it} and $\Delta_2 y_{it}$ thus contain all the necessary information to identify σ_ε^2 and σ_r^2.

For estimation we proceed as follows. As the first and last permanent/transitory shocks are not separately identified, we treat $\varepsilon_{i2} - r_{i1}$ and $\varepsilon_{iT} + r_{iT}$ as additional factors. We end up with $K = 2T - 3$ factors. Then we estimate different variances for all shocks, using Equally Weighted Minimum Distance. We also estimate the density of each shock. For this purpose, we use the second-order kernel H_2, and Delaigle and Gijbels' (2004) method to pick up the trimming parameter T_N. Lastly, we obtain $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\sigma}_r^2$ in (18) as the means of the estimated permanent and transitory variances, respectively, and similarly average the estimated densities to obtain the final density estimates $\widehat{f_\varepsilon}$ and $\widehat{f_r}$.

6.3 Estimation results

The estimated variance of permanent shocks is $\widehat{\sigma}_\varepsilon^2 = .0208$, and the estimated transitory variance is $\widehat{\sigma}_r^2 = .0185$, with standard errors of .0029 and .0017, respectively.[10] According to these estimates, permanent shocks account for 36% of the total variance of wage growth residuals.

Figure 6 presents the density estimates. The permanent and transitory components are shown in panels a) and b), respectively. In each panel, the thick solid line represents the density of the shock, standardized to have unit variance, and the thin solid line represents the standard normal density, that we draw for comparison. The dashed lines delimit the bootstrapped 10%-90% confidence band.[11]

Figure 6 shows that none of the two distributions is Gaussian. Both permanent and transitory shocks appear strongly leptokurtic. In particular, they have high modes and fatter tails than the normal. Moreover, the transitory part seems to have higher kurtosis than the permanent component.[12] Lastly, both densities are approximately symmetric.

6.4 Fit

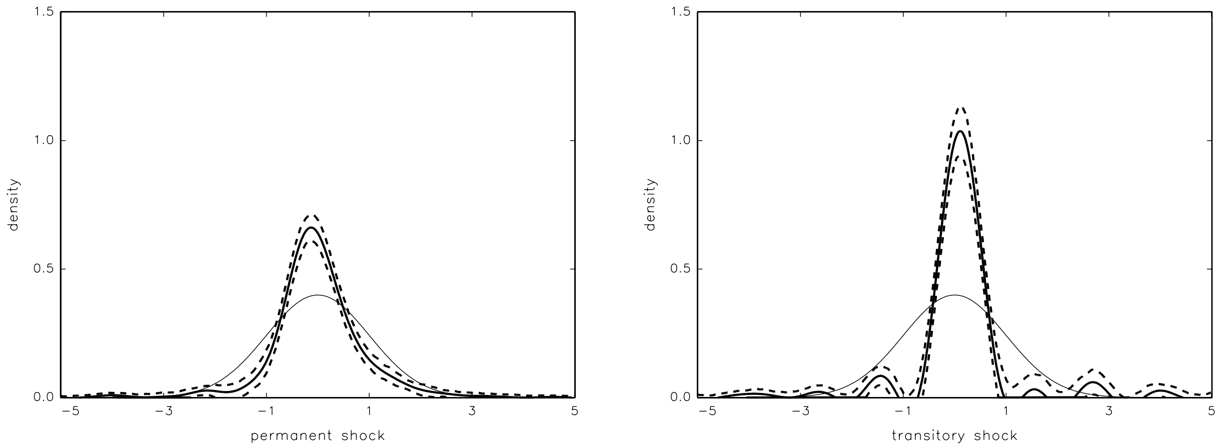
Figure 7 compares the predicted densities of $\Delta_s y_{it}$, $s = 1, 2, 3$, using the model and the estimated densities of permanent and transitory shocks, to kernel density estimates. In panels a1) to c1), the thin line is a kernel estimator of the actual distribution's density.

[10] Standard errors were computed by 1000 iterations of individual block bootstrap.

[11] Remark that, as we do not derive the asymptotic distribution of the nonparametric estimator, the validity of the bootstrap in our context is difficult to verify.

[12] We checked that varying the trimming parameter T_N around the valuethat we obtained using Delaigle and Gijbels' (2004) method had little effect on the estimate $\widehat{f_\varepsilon}$, but a stronger effect on $\widehat{f_r}$, tail oscillations increasing with T_N.

Figure 6: Nonparametric estimates of the densities of standardized permanent and transitory shocks.



a) Permanent shock   b) Transitory shock

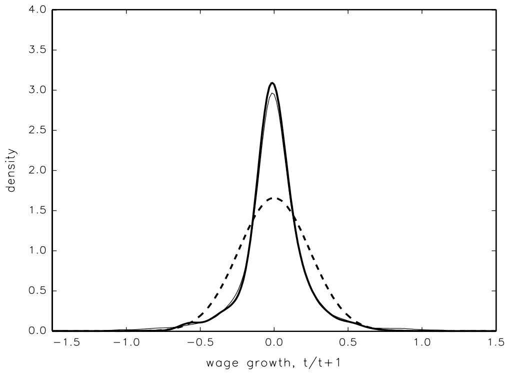
Note: Density estimates of ε_{it} *and* r_{it}*, both standardized to have unit variance. Density estimate (thick); 10%-90% confidence bands of 100 bootstrap simulations (dashed); standard normal density (thin).*

The thick line is the predicted density. The dashed line shows the density that is predicted under the assumption that shocks are normally distributed. The predicted densities of $\Delta_s y_{it}$, $s = 1, 2, 3$, where calculated analytically by convolution of the estimated densities of ε_{it} and r_{it}.

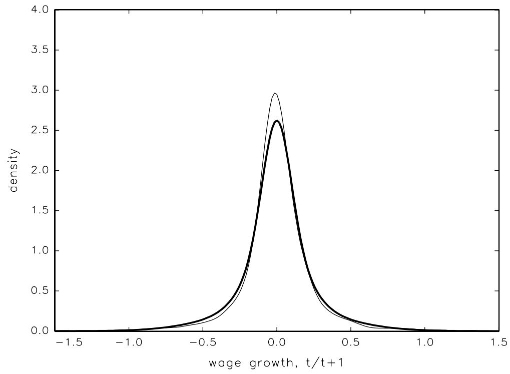
Figure 7 shows that our specification reproduces two features apparent in Table 2: the high kurtosis of wage growth residuals, and the decreasing kurtosis when the time lag increases. Note that the high mode of the density is remarkably well captured by our nonparametric method, even in the case of $\Delta_3 y_{it}$. In contrast, the normal specification gives a rather poor fit.

We then present in Table 3 the moments of wage growth residuals, as in the data and as predicted under normality and nonparametrically. We see that variances are severely underestimated, reflecting a rather bad estimation of the density in the tails. Moreover, the estimated kurtosis is 5.6, that is significantly non-normal but very different from the kurtosis of the distribution to be fitted (10.3). Overall, our method captures the shapes of the densities of wage growth variables very well, but fails at fitting the tails, which

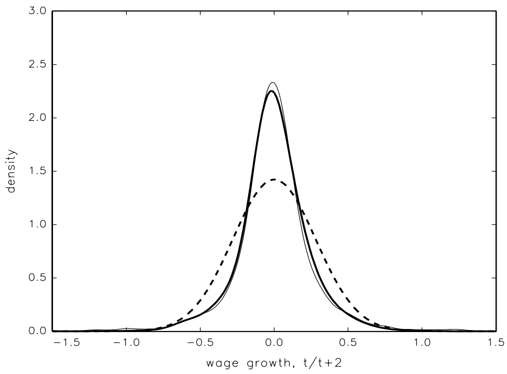
Figure 7: Fit of the model, densities of wage growth residuals.



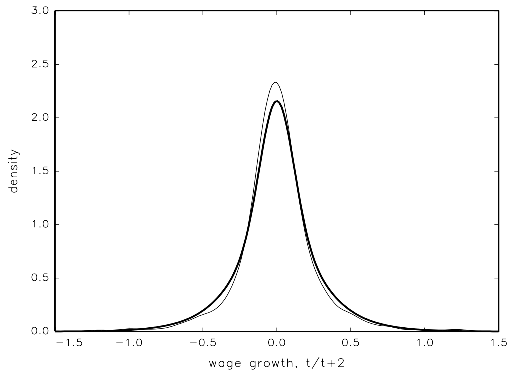
a1) wage growth $t/t+1$



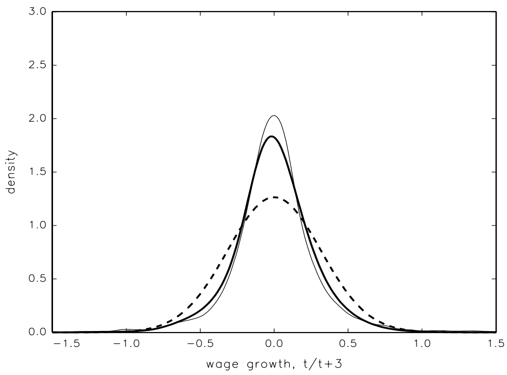
a2) wage growth $t/t+1$, normal mixture



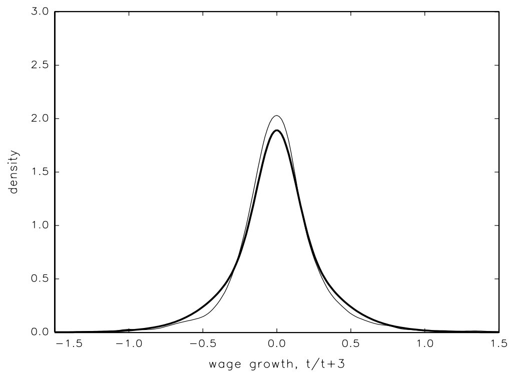
b1) wage growth $t/t+2$



b2) wage growth $t/t+2$, normal mixture



c1) wage growth $t/t+3$



c2) wage growth $t/t+3$, normal mixture

Note: Graphs a1), b1) and c1) show the fit of wage growth residuals calculated over one, two and three years, respectively, using the generalized deconvolution estimator. Graphs a2), b2) and c2): densities are estimated by Maximum Likelihood, where shocks follow two-component mixtures of zero mean normals. Predicted density (thick); kernel density estimate (thin); normal (dashed).

Table 3: Fit of the model, moments of wage growth residuals

Wage growth	$t/t+1$	$t/t+2$	$t/t+3$
	Data		
Variance	.055	.073	.086
Skewness	-.08	.06	-.07
Kurtosis	10.3	11.2	8.0
	Predicted, nonparametric		
Variance	.037	.053	.069
Skewness	-.02	-.02	-.02
Kurtosis	5.6	4.6	4.2
	Predicted, normal		
Variance	.057	.076	.096
Skewness	0	0	0
Kurtosis	3	3	3
	Predicted, normal mixture		
Variance	.058	.072	.086
Skewness	0	0	0
Kurtosis	6.3	5.3	4.8

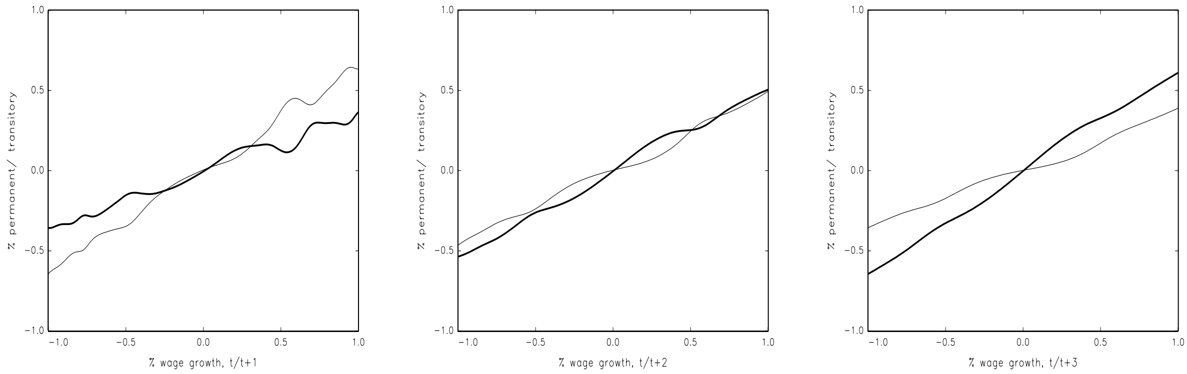
Note: See the note to Figure 7. Moments are predicted using the predicted densities shown in Figure 7, by computing the integrals numerically.

leads to underestimating higher moments.

To fit the moments better, we use the nonparametric estimates $\widehat{f}_\varepsilon$ and $\widehat{f}_r$ as a guide to find a convenient parametric form for factor densities. Figure 6 suggests that a mixture of two normals centered at zero may work well in practice. We thus estimate model (18) under this parametric specification for both ε_{it} and r_{it}. Parameters are estimated by Maximum Likelihood, using the EM algorithm of Dempster, Laird and Rubin (1977). Panels a2) to c2) in Figure 7 show the fit of the model. The shape of the densities is very well reproduced. Moreover, the last three rows of Table 3 show that the normal mixture specification yields much better estimates of the variance and kurtosis of wage growth residuals.

Notice that the normal mixture model was already used by Geweke and Keane (2000) to model earnings dynamics. Our results strongly support this modelling choice.

Figure 8: Conditional expectations of shocks given wage growth residuals



a) Wage growth, $t/t+1$ b) Wage growth, $t/t+2$ c) Wage growth, $t/t+3$

Note: See the note to Figure 7. a): conditional expectation of ε_{it} (thick) and $r_{it} - r_{i,t-1}$ (thin) given Δy_{it}; b): $\varepsilon_{it} + \varepsilon_{i,t-1}$ (thick) and $r_{it} - r_{i,t-2}$ (thin) given $\Delta_2 y_{it}$; c): $\varepsilon_{it} + \varepsilon_{i,t-1} + \varepsilon_{i,t-2}$ (thick) and $r_{it} - r_{i,t-3}$ (thin) given $\Delta_3 y_{it}$.

6.5 Wage mobility

We then use the model to weight the respective influence of permanent and transitory shocks in wage mobility. To this end, we compute the conditional expectations of the permanent and transitory components of $\Delta_s y_{it}$, $s = 1, 2, 3$: $\mathbb{E}\left(\sum_{r=0}^{s-1} \varepsilon_{it-r} | \Delta_s y_{it}\right)$ and $\mathbb{E}\left(r_{it} - r_{it-s} | \Delta_s y_{it}\right)$.

To do so, we first compute the conditional distribution of permanent and transitory shocks using Bayes rule. For instance the conditional density of the permanent shock given wage observations is given by:

$$f(\varepsilon|\Delta y) = \frac{f_\varepsilon(\varepsilon) f(\Delta y|\varepsilon)}{\int f_\varepsilon(\widetilde{\varepsilon}) f(\Delta y|\widetilde{\varepsilon}) d\widetilde{\varepsilon}} = \frac{f_\varepsilon(\varepsilon) \int f_r(r) f_r(\Delta y - \varepsilon + r) dr}{\int f_\varepsilon(\widetilde{\varepsilon}) \int f_r(r) f_r(\Delta y - \widetilde{\varepsilon} + r) dr d\widetilde{\varepsilon}},$$

where f_ε is the p.d.f. of ε and f_r is the p.d.f. of r. We proceed similarly for transitory shocks $r_{it} - r_{it-1}$.

Figure 8 plots these conditional expectations. We verify that the volatility of earnings is more likely to have a permanent origin if s is large. In panel a), we see for example that a log wage growth of $\pm 100\%$ has a transitory origin for more than $\pm 60\%$ and a permanent origin for less that $\pm 30\%$. In panel c), we see that a change $\Delta_3 y_{it}$ of $\pm 100\%$ is almost twice more likely to be permanent than transitory.

Table 4: Variances of the shocks by categories of job changers

Job changes	None	One/two	Three/more
	wage growth, $t/t+1$		
total	.034	.039	.068
permanent	.014	.016	.022
transitory	.020	.023	.046
	wage growth, $t/t+2$		
total	.041	.053	.089
permanent	.025	.032	.053
transitory	.016	.021	.036
	wage growth, $t/t+3$		
total	.054	.063	.108
permanent	.037	.044	.076
transitory	.017	.019	.032

Note: See the note to Figure 7. "None"=no job change in the observation period; "One/two"= one or two job changes; "Three/more"= more than three job changes. Variances of wage growth residuals ("Total") and the variances of the permanent and transitory parts, conditional on having experienced a given number of job changes.

6.6 Job changes

Finally, we address the issue of the link between the degree of permanence of wage shocks and job-to-job mobility. It is notoriously difficult to identify job changes precisely in the PSID (see Brown and Light, 1992), so we tend to think of this exercise as tentative. We adopt the simplest criteria to identify job changes, setting the job change dummy equal to one if tenure is less than 12 months.[13] We then classify individuals into job stayers (no job change during the period), infrequent job changers (one or two job changes) and frequent job changers (more than three job changes). The last three columns of Table 1 in Appendix give descriptive statistics for these three groups of individuals.

Then we compute the densities of permanent and transitory shocks given wage growth residuals, separately for each category of job changers by averaging within each group the conditional densities that we have already calculated. Table 4 presents the variances of permanent and transitory shocks for each mobility group. Focusing on the first three

[13]Note that there were two "tenure" variables before 1987 in the PSID: time in position and time with employer. We take the former as our definition of tenure.

rows we see that wage volatility, as measured by the variance, is higher for frequent job changers. Moreover, these individuals are more likely to experience both permanent and transitory wage changes. The transitory variance is about 15% higher for infrequent job movers than for job stayers (.023 versus .020), and about 2.3 times higher for frequent job movers (.046). At the same time, the permanent variance is about 15% higher for infrequent job movers than for job stayers (.016 versus .014), and about 60% higher for frequent job movers (.022). As permanent shocks accumulate over time while transitory shocks do not, the difference in wage growth volatility increases with the length of time over which wage growth is computed. For example, the variance of wage growth over ten years is .16 ($= .020 + 10 * .014$) for an individual who stayed with the same employer over the whole period, while it is about .27 ($= .046 + 10 * .022$) for an individual who has changed job three times or more.

These results give some basis to the interpretation of permanent shocks to log earnings as resulting for a large part from job changes.[14] Nevertheless, identifying permanent shocks with job changes is likely to be wrong for two reasons. First, part of the shocks faced by job stayers is permanent. Indeed, the share of permanent variance in total variance is higher for job stayers (40%) than for frequent job changers (30%). This finding suggests that there might be other permanent wage movements, caused for instance by within-job promotions. Second, job changers also face more transitory shocks. Describing precisely these effects requires modelling job change decisions together with wage profiles.

7 Conclusion

This paper provides a generalization of the nonparametric estimator of Li and Vuong (1998) to linear independent factor models, allowing for any number of measurements, L, and at most $K = \frac{L(L+1)}{2}$ latent factors. On the theoretical side, the main lessons of the standard deconvolution literature carry over to the more general context that we consider in this paper. In particular, asymptotic convergence rates are slow, and it is more difficult to estimate the distribution of one factor if the characteristic functions of the other factors have thinner tails.

Our Monte Carlo results yield interesting insights. The finite-sample performance of

[14]Note that we do not identify the part of the wage growth variance that comes from differences in hours worked from the one coming from differences in wage rates. Nor are we able to tell whether job or individual-specific components are mostly responsible for the results.

our estimator seems rather good, remarkably similar to the performance of the kernel deconvolution estimator that assumes that the distributions of all factors but one are known. Moreover, the performance critically depends on the shape of the distributions to be estimated, as we find that it is easier to estimate distributions with little skewness or excess kurtosis.[15]

In any case, identifying the distributions of more factors than measurements should be viewed as considerably more difficult than the classical nonparametric deconvolution problem. Given the difficulty of the problem at hand, we view the results of our simulations and the application as a confirmation that the generalized nonparametric deconvolution approach that we propose can be successfully applied to a wide range of distributions.

The empirical application shows that the permanent and transitory components of individual earnings dynamics are clearly non normal. Predicting transitory and permanent shocks for the individuals in the sample, we see that frequent job changers face more permanent and transitory earnings shocks than job stayers. These results have important consequences for welfare analysis. For instance, savings and insurance could be very different if the risk of large deviations is much higher than is usually assumed with normal shocks. Of course, the model of earnings dynamics that we have considered is very limited. One might want to add non i.i.d. transitory shocks and yet allow for measurement error (as in Abowd and Card, 1989). We experimented with a MA(1) transitory shock without much success. It seems very difficult to nonparametrically identify the MA(1) component from the PSID data. Thus, maybe the sample is not appropriate, or a single non normal MA(0) transitory shock/measurement error is enough to describe the PSID data.

Another interesting issue is the assumption of independence between factors that we maintain throughout this analysis. Meghir and Pistaferri (2004) shows evidence of autoregressive conditional heteroskedasticity in permanent and transitory components. It is not straightforward at all to extend the study of the nonparametric identification

[15] In Bonhomme and Robin (2007), we show that skewness and peakedness are required for the matrix of factor loadings to be identified from higher-order moments. There is thus a tension between obtaining a precise estimate of factor loadings and a precise estimate of the distribution of factors in models where second-order information is not sufficient to ensure the identification of the factor loadings.

and estimation of factor densities in conditionally heteroskedastic factor models like:

$$y_{it} = A\varepsilon_{it}, \quad \varepsilon_{it}^{k} = \sigma(\varepsilon_{it-1})\eta_{it}^{k}, \quad k = 1, ..., K,$$

where $\eta_{it} = (\eta_{it}^{1}, ..., \eta_{it}^{K})^{\mathrm{T}}$ is a $K \times 1$ vector of i.i.d. random variables. But this is a very interesting problem for future research.

References

[1] ABOWD, J., and D. CARD (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica,* 57, 411-445.

[2] ALVAREZ, A., M. BROWNING, and M. EJRNAES (2007): "Modelling Income Processes with Lots of Heterogeneity," *mimeo.*

[3] ANDERSON, T. W., and H. RUBIN (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Symposium in Mathematical Statistics and Probability*, Vol. 5. University of California press.

[4] BONHOMME, S., and J.-M. ROBIN (2007): "Consistent Noisy Independent Component Analysis," *mimeo.*

[5] BROWN, J., and A. LIGHT (1992): "Interpreting Panel Data on Job Tenure," *Journal of Labor Economics,* 10, 219-257.

[6] CARROLL, R. J., D. RUPPERT, and L. A. STEFANSKI (1995): *Measurement Error in Nonlinear Models.* New York: Chapman & Hall.

[7] CARROLL, R. J., and P. HALL (1988): "Optimal rates of Convergence for Deconvoluting a Density," *Journal of the American Statistical Association*, 83, 1184-1186.

[8] CHAMBERLAIN, G. (1984): "Panel Data," in Griliches, Z., and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[9] CHAMBERLAIN, G., and K. HIRANO (1999): "Predictive Distributions Based on Longitudinal Earnings Data ," *Annales d'Économie et de Statistiques*, 55-56, 211-242.

[10] DELAIGLE, A., and I. GIJBELS (2002): "Estimation of Integrated Squared Density Derivatives From a Contaminated Sample," *Journal of the Royal Statistical Society*, B, 64, 869-886.

[11] DELAIGLE, A., and I. GIJBELS (2004): "Comparison of Data-Driven Bandwidth Selection Procedures in Deconvolution Kernel Density Estimation," *Computational Statistics and Data Analysis*, 45, 249-267.

[12] DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B 39(1), 1-38.

[13] DIGGLE, P. J., and P. HALL (1993): "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *Journal of the Royal Statistical Society Series B*, 55, 523-531.

[14] FAN, J. Q. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of statistics*, 19, 1257-1272.

[15] GEWEKE, J., and M. KEANE (2000): "An Empirical Analysis of Earnings Dynamics Among Men in the PSID: 1968-1989," *Journal of Econometrics,* 96, 293-356.

[16] GEWEKE, J., and M. KEANE (2007): "Smoothly Mixing Regressions," *Journal of Econometrics*, 138, 252-290.

[17] GUVENEN, F. (2007a): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97, 687-712.

[18] GUVENEN, F. (2007b): "An Empirical Investigation of Labor Income Processes," *mimeo.*

[19] HALL, P., and Q. YAO (2003): "Inference in Components of Variance Models with Low Replications," *Annals of Statistics*, 31, 414-441.

[20] HALL, R., and F. MISHKIN (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461-481.

[21] HIRANO, K. (2002): "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica*, 70, 780-799.

[22] HOROWITZ, J. L. (1998): *Semiparametric Methods in Econometrics.* New-York: Springer-Verlag.

[23] HOROWITZ, J. L., and M. MARKATOU (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63, 145-168.

[24] HU, Y., and G. RIDDER (2006): "On Deconvolution as a First Stage Nonparametric Estimator", *mimeo.*

[25] HU, Y., and G. RIDDER (2007): "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information", *mimeo.*

[26] KOTLARSKI, I. (1967): "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69-76.

[27] LI, T. (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models, " *Journal of Econometrics*, 110, 1-26.

[28] LI, T., I. PERRIGNE, and Q. VUONG (2000): "Conditionally Independent Private Information in OSC Wildcat Auctions," *Journal of Econometrics*, 98, 129-161.

[29] LI, T., and Q. VUONG (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139-165.

[30] LILLARD, L., and R. WILLIS (1978): "Dynamic Aspects of Earnings Mobility," *Econometrica*, 46, 985-1012.

[31] MEGHIR, C., and L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica,* 72, 1-32.

[32] POLLARD, D. (1984): *Convergence of Stochastic Processes.* Springer: New-York.

[33] RAO, P. (1983): *Nonparametric Functional Estimation.* New-York: Academic Press.

[34] RAO, P. (1992): *Identifiability in Stochastic Models.* New-York: Academic Press.

[35] SCHENNACH, S. (2004a): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33-75.

[36] SCHENNACH, S. (2004b): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046-94.

[37] SZÉKELY, G. J., and C. R. RAO (2000): "Identifiability of Distributions of Independent Random Variables by Linear Combinations and Moments," *Sankhyä*, 62, 193-202.

APPENDIX

A Proof of Lemma 1

For any T, let $\mathcal{G} = \{f_t(x,y), |t| \leq T\}$. The first step of the proof is to find the L_1 covering number of $\mathcal{G}$. We need to cover: $\mathcal{G}_1 = \{x\cos\left(t^{\mathrm{T}}y\right), |t| \leq T\}$, and $\mathcal{G}_2 = \{x\sin\left(t^{\mathrm{T}}y\right), |t| \leq T\}$. Now, for any couple (t_1, t_2),

$$\begin{aligned}
\left|x\cos(t_1^T y) - x\cos(t_2^T y)\right| &\leq \left|x(t_1^T y - t_2^T y)\right| \\
&\leq \sum_\ell |xy_\ell\,(t_{1\ell} - t_{2\ell})| \\
&\leq \sum_\ell |xy_\ell| \cdot |t_1 - t_2| \\
&\leq L\,|xy| \cdot |t_1 - t_2|\,.
\end{aligned}$$

It follows that the L_1 covering number of $\mathcal{G}_1$ satisfies

$$\mathcal{N}_1\left(\frac{\varepsilon}{2}, P_N, \mathcal{G}_1\right) \leq C\left(\frac{T\mathbb{E}_N|XY|}{\varepsilon}\right)^L,$$

where P_N is the probability measure obtained by independent sampling from F, and $C > 0$ is a constant independent of N. We obtain a similar expression for $\mathcal{G}_2$; hence:

$$\mathcal{N}_1\left(\varepsilon, P_N, \mathcal{G}\right) \leq C\left(\frac{T\mathbb{E}_N|XY|}{\varepsilon}\right)^L,$$

Equation (31) in Pollard (p. 31) then implies, for a given sample $\mathbf{Z}_N$:

$$\Pr\left\{\sup_{|t|\leq T}|\mathbb{E}_N f_t - \mathbb{E}f_t| \geq \varepsilon \,\middle|\, \mathbf{Z}_N\right\} \leq C\left(\frac{T\mathbb{E}_N|XY|}{\varepsilon}\right)^L \exp\left[-\frac{N\varepsilon^2}{128}\Big/\mathbb{E}_N X^2\right], \tag{D1}$$

provided that $N \geq \frac{8\,\mathrm{Var}\,f_t}{\varepsilon^2}$. Moreover, $\mathrm{Var}\,f_t = \mathrm{Var}\left[X\cos\left(t^T Y\right)\right] \leq \mathbb{E}X^2 + \left(\mathbb{E}|X|\right)^2 \equiv M_1$ is finite. So inequality (D1) is true for $N \geq \frac{8M_1}{\varepsilon^2}$.

Then, bounding some probabilities by one, we obtain, for all $k > 0$:

$$\begin{aligned}
\Pr\left\{\sup_{|t|\leq T}|\mathbb{E}_N f_t - \mathbb{E}f_t| \geq \varepsilon\right\} &= \Pr\left\{\sup_{|t|\leq T}|\mathbb{E}_N f_t - \mathbb{E}f_t| \geq \varepsilon \,\middle|\, \mathbb{E}_N X^2 \leq k, \mathbb{E}_N|XY| \leq k\right\} \\
&\quad \times \Pr\left\{\mathbb{E}_N X^2 \leq k, \mathbb{E}_N|XY| \leq k\right\} \\
&\quad + \Pr\left\{\sup_{|t|\leq T}|\mathbb{E}_N f_t - \mathbb{E}f_t| \geq \varepsilon \,\middle|\, \mathbb{E}_N X^2 \geq k \text{ or } \mathbb{E}_N|XY| \geq k\right\} \\
&\quad \times \Pr\left\{\mathbb{E}_N X^2 \geq k \text{ or } \mathbb{E}_N|XY| \geq k\right\} \\
&\leq \Pr\left\{\sup_{|t|\leq T}|\mathbb{E}_N f_t - \mathbb{E}f_t| \geq \varepsilon \,\middle|\, \mathbb{E}_N X^2 \leq k, \mathbb{E}_N|XY| \leq k\right\} \\
&\quad + \Pr\left\{\mathbb{E}_N X^2 \geq k \text{ or } \mathbb{E}_N|XY| \geq k\right\}.
\end{aligned}$$

To obtain a final inequality, apply Markov inequality as follows:

$$\begin{aligned}\Pr\left\{\mathbb{E}_N X^2 \geq k \text{ or } \mathbb{E}_N |XY| \geq k\right\} &\leq \Pr\left\{\mathbb{E}_N X^2 \geq k\right\} + \Pr\left\{\mathbb{E}_N |XY| \geq k\right\} \\ &= \Pr\left\{\left(\mathbb{E}_N X^2\right)^J \geq k^J\right\} + \Pr\left\{\left(\mathbb{E}_N |XY|\right)^J \geq k^J\right\} \\ &\leq \frac{\mathbb{E}\left(\left(\mathbb{E}_N X^2\right)^J\right) + \mathbb{E}\left(\left(\mathbb{E}_N |XY|\right)^J\right)}{k^J}. \qquad \text{(D2)}\end{aligned}$$

As X^2 and $|XY|$ have finite moments up to the Jth order by assumption, the numerator in (D2) is bounded, say by a constant $M_2 < \infty$.

Therefore,

$$\Pr\left\{\sup_{|t|\leq T} |\mathbb{E}_N f_t - \mathbb{E} f_t| \geq \varepsilon\right\} \leq C\left(\frac{Tk}{\varepsilon}\right)^L \exp\left[-\frac{N\varepsilon^2}{128k}\right] + \frac{M_2}{k^J},$$

for any ε such that $\varepsilon^2 \geq \frac{8M_1}{N}$.

Now, index ε, T and k on N. For any $0 < \gamma < (1-\frac{1}{J})/2$, let $k_N = N^{\frac{1}{J}+\gamma}$, and let $\varepsilon_N = N^{-\frac{1}{2}(1-\frac{1}{J})+\gamma}$. Then

$$\sum_N \frac{1}{k_N^J} < \infty,$$

and

$$\sum_N \exp\left\{L \ln\left(\frac{T_N k_N}{\varepsilon_N}\right) - \frac{N\varepsilon_N^2}{128 k_N}\right\} = \sum_N \exp\left\{L \ln\left(T_N N^{\frac{1}{2}(1+\frac{1}{J})}\right) - \frac{N^\gamma}{128}\right\} < \infty,$$

if T_N tends to infinity at (at most) a polynomial rate.

The Borel-Cantelli Lemma then implies that only a finite number of events are such that

$$\sup_{|t|\leq T_N} |\mathbb{E}_N f_t - \mathbb{E} f_t| \geq \varepsilon_N.$$

Hence,

$$\sup_{|t|\leq T_N} |\mathbb{E}_N f_t - \mathbb{E} f_t| = O(\varepsilon_N), \qquad \text{a.s.}$$

This achieves to prove Lemma 1.

B Proof of Theorem 2

In this proof and the next, all convergence statements are implicitly understood to hold almost surely.

(i) Fix any $t \in \mathbb{R}^L$, let $\varphi(t) \equiv \varphi_Y(t) = \mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]$, $\psi_\ell(t) = \mathbb{E}\left[Y_\ell e^{it^{\mathrm{T}}Y}\right]$ and $\xi_{\ell m}(t) = \mathbb{E}\left[Y_\ell Y_m e^{it^{\mathrm{T}}Y}\right]$, for any $\ell, m = 1, ..., L$. Then, for all $\gamma > 0$, Lemma 1 implies that, for all $f \in \{\varphi, \{\psi_\ell\}_\ell, \{\xi_{\ell m}\}_{\ell,m}\}$:

$$\sup_{|t|\leq T_N} \left|\widehat{f}(t) - f(t)\right| = O(\varepsilon_N),$$

where T_N tends to infinity at (at most) a polynomial rate and $\varepsilon_N = N^{-\frac{1}{2}(1-\frac{1}{J})+\gamma}$.

(ii) Removing the subscript Y from φ_Y and g_Y to simplify the notation, as $|\varphi(t)| \geq g(|t|)$ when $|t| \to \infty$, and as φ is nonvanishing everywhere, then for T_N large enough

$$\inf_{|t|\leq T_N} |\varphi(t)| \geq g(T_N),$$

and

$$\sup_{|t|\leq T_N} \left| \frac{\widehat{\varphi}(t) - \varphi(t)}{\varphi(t)} \right| = \frac{O(\varepsilon_N)}{g(T_N)} = o\,(1)\,.$$

The last equality follows from the fact that $\frac{T_N^2 \varepsilon_N}{g(T_N)^3} \geq \frac{\varepsilon_N}{g(T_N)}$ for N large enough, and that, by assumption, $\frac{T_N^2 \varepsilon_N}{g(T_N)^3} = o(1)$.

(iii) We have

$$\frac{\partial \kappa_Y(t)}{\partial t_\ell} = i\frac{\psi_\ell(t)}{\varphi(t)} = i\frac{\mathbb{E}\left[Y_\ell e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]},$$

and

$$\begin{aligned} \frac{\widehat{\psi}_\ell(t)}{\widehat{\varphi}(t)} - \frac{\psi_\ell(t)}{\varphi(t)} &= \frac{\widehat{\psi}_\ell(t)}{\widehat{\varphi}(t)} - \frac{\widehat{\psi}_\ell(t)}{\varphi(t)} + \frac{\widehat{\psi}_\ell(t)}{\varphi(t)} - \frac{\psi_\ell(t)}{\varphi(t)} \\ &= -\frac{\widehat{\psi}_\ell(t)}{\varphi(t)} \frac{\frac{\widehat{\varphi}(t)-\varphi(t)}{\varphi(t)}}{\frac{\widehat{\varphi}(t)-\varphi(t)}{\varphi(t)}+1} + \frac{1}{\varphi(t)}\left[\widehat{\psi}_\ell(t) - \psi_\ell(t)\right]. \end{aligned}$$

One can bound $\widehat{\psi}_\ell(t)$ as follows:

$$\begin{aligned} \sup_{|t|\leq T_N} \left|\widehat{\psi}_\ell(t)\right| &\leq \sup_{|t|\leq T_N} \left|\widehat{\psi}_\ell(t) - \psi_\ell(t)\right| + \sup_{t\in[-T_N,T_N]} |\psi_\ell(t)| \\ &\leq \sup_{|t|\leq T_N} \left|\widehat{\psi}_\ell(t) - \psi_\ell(t)\right| + \mathbb{E}|Y_\ell| = O(1), \end{aligned}$$

as the first moments of Y are finite by assumption.

It follows that

$$\sup_{|t|\leq T_N} \left| \frac{\widehat{\psi}_\ell(t)}{\widehat{\varphi}(t)} - \frac{\psi_\ell(t)}{\varphi(t)} \right| = \frac{O(\varepsilon_N)}{g(T_N)^2} = o\,(1)\,.$$

The same argument applies to show that

$$\sup_{|t|\leq T_N} \left| \frac{\widehat{\xi}_{\ell m}(t)}{\widehat{\varphi}(t)} - \frac{\xi_{\ell m}(t)}{\varphi(t)} \right| = \frac{O(\varepsilon_N)}{g(T_N)^2} = o\,(1)$$

for all ℓ, m.

(iv) It is easy to extend these results to second derivatives of cumulant generating functions:

$$\begin{aligned} \zeta_{\ell m}(t) &\equiv \frac{\partial^2 \kappa_Y}{\partial t_\ell \partial t_m}(t) \\ &= -\frac{\mathbb{E}\left[Y_\ell Y_m e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]} + \frac{\mathbb{E}\left[Y_\ell e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]} \frac{\mathbb{E}\left[Y_m e^{it^{\mathrm{T}}Y}\right]}{\mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]} \\ &= -\frac{\xi_{\ell m}(t)}{\varphi(t)} + \frac{\psi_\ell(t)}{\varphi(t)}\frac{\psi_m(t)}{\varphi(t)}. \end{aligned}$$

Let $\widehat{\zeta}_{\ell m}(t) = -\frac{\widehat{\xi}_{\ell m}(t)}{\widehat{\varphi}(t)} + \frac{\widehat{\psi}_{\ell}(t)}{\widehat{\varphi}(t)}\frac{\widehat{\psi}_{m}(t)}{\widehat{\varphi}(t)}$. Then,

$$\begin{aligned}
\widehat{\zeta}_{\ell m}(t) - \zeta_{\ell m}(t) &= -\left[\frac{\widehat{\xi}_{\ell m}(t)}{\widehat{\varphi}(t)} - \frac{\xi_{\ell m}(t)}{\varphi(t)}\right] \\
&+ \left[\frac{\widehat{\psi}_{\ell}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{\ell}(t)}{\varphi(t)}\right]\frac{\psi_{m}(t)}{\varphi(t)} + \left[\frac{\widehat{\psi}_{m}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{m}(t)}{\varphi(t)}\right]\frac{\psi_{\ell}(t)}{\varphi(t)} \\
&+ \left[\frac{\widehat{\psi}_{\ell}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{\ell}(t)}{\varphi(t)}\right]\left[\frac{\widehat{\psi}_{m}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{m}(t)}{\varphi(t)}\right].
\end{aligned}$$

Since

$$\sup_{|t|\leq T_N}\left|\frac{\psi_{\ell}(t)}{\varphi(t)}\right| \leq \frac{\mathbb{E}|Y_{\ell}|}{g(T_N)}$$

for all ℓ, it follows that

$$\sup_{|t|\leq T_N}\left|\widehat{\zeta}_{\ell m}(t) - \zeta_{\ell m}(t)\right| = \frac{O(\varepsilon_N)}{g(T_N)^2} + \frac{O(\varepsilon_N)}{g(T_N)^3} + \left(\frac{O(\varepsilon_N)}{g(T_N)^2}\right)^2 = \frac{O(\varepsilon_N)}{g(T_N)^3}$$

because

$$\frac{\varepsilon_N}{g(T_N)^3} > \frac{\varepsilon_N^2}{g(T_N)^4} \Leftrightarrow 1 > \frac{\varepsilon_N}{g(T_N)}$$

for N large enough.

(v) For any vector $t = (t_1, ..., t_L)^{\mathrm{T}} \in \mathbb{R}^L$ and $\tau \in \mathbb{R}$, then

$$\begin{aligned}
B_{\ell}(t) &= \sup_{\tau\in[-T_N,T_N]}\left|\int_0^{\tau}\frac{\widehat{\psi}_{\ell}(ut)}{\widehat{\varphi}(ut)}du - \int_0^{t_\ell}\frac{\psi_{\ell}(ut)}{\varphi(ut)}du\right| \\
&\leq \sup_{\tau\in[-T_N,T_N]}\left(|\tau|\sup_{|t|\leq T_N}\left|\frac{\widehat{\psi}_{\ell}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{\ell}(t)}{\varphi(t)}\right|\right) \\
&\leq T_N\sup_{|t|\leq T_N}\left|\frac{\widehat{\psi}_{\ell}(t)}{\widehat{\varphi}(t)} - \frac{\psi_{\ell}(t)}{\varphi(t)}\right| \\
&= \frac{T_N}{g(T_N)^2}O(\varepsilon_N).
\end{aligned}$$

Similarly,

$$\begin{aligned}
C_{\ell m}(t) &= \sup_{\tau\in[-T_N,T_N]}\left|\int_0^{\tau}\int_0^{u}\widehat{\zeta}_{\ell m}(vt)dvdu - \int_0^{\tau}\int_0^{u}\zeta_{\ell m}(vt)dvdu\right| \\
&\leq \sup_{\tau\in[-T_N,T_N]}\left(\frac{\tau^2}{2}\sup_{|t|\leq T_N}\left|\widehat{\zeta}_{\ell m}(t) - \zeta_{\ell m}(t)\right|\right) \\
&\leq T_N^2\sup_{|t|\leq T_N}\left|\widehat{\zeta}_{\ell m}(t) - \zeta_{\ell m}(t)\right| \\
&= \frac{T_N^2}{g(T_N)^3}O(\varepsilon_N).
\end{aligned}$$

Moreover, for any distribution W on $\mathcal{T}_k$,

$$\int B_\ell(t)\,dW(t) \le \sup_{|t|\le T_N} B_\ell(t)\cdot\int dW(t) = \frac{T_N}{g(T_N)^2}O(\varepsilon_N)$$

and

$$\int C_{\ell m}(t)\,dW(t) \le \sup_{|t|\le T_N} C_{\ell m}(t)\cdot\int dW(t) = \frac{T_N^2}{g(T_N)^3}O(\varepsilon_N).$$

(vi) It easily follows from the previous step that:

$$\sup_{\tau\in[-T_N,T_N]} \left|\widehat{\kappa}_{X_k}(\tau) - \kappa_{X_k}(\tau)\right| = \frac{T_N^2}{g(T_N)^3}O(\varepsilon_N) = o(1).$$

In particular, $\sup_{\tau\in[-T_N,T_N]}\left|\widehat{\kappa}_{X_k}(\tau)-\kappa_{X_k}(\tau)\right| < 1$ for N large enough. Therefore, for N large enough

$$\begin{aligned}\sup_{\tau\in[-T_N,T_N]} \left|\widehat{\varphi}_{X_k}(\tau) - \varphi_{X_k}(\tau)\right| &= \sup_{\tau\in[-T_N,T_N]} \left|\exp\left(\widehat{\kappa}_{X_k}(\tau)\right) - \exp\left(\widehat{\kappa}_{X_k}(\tau)\right)\right|,\\ &\le \sup_{\tau\in[-T_N,T_N]} \left|\widehat{\kappa}_{X_k}(\tau) - \kappa_{X_k}(\tau)\right|,\end{aligned}$$

from which it follows that

$$\sup_{\tau\in[-T_N,T_N]} \left|\widehat{\varphi}_{X_k}(\tau) - \varphi_{X_k}(\tau)\right| = \frac{T_N^2}{g(T_N)^3}O(\varepsilon_N).$$

This ends the proof of Theorem 2.

C Proof of Theorem 3

For all x_k in the support of X_k:

$$\begin{aligned}\widehat{f}_{X_k}(x_k) - f_{X_k}(x_k) &= \frac{1}{2\pi}\int \varphi_H\left(\frac{v}{T_N}\right) e^{-ivx_k}\left(\widehat{\varphi}_{X_k}(v) - \varphi_{X_k}(v)\right)dv\\ &\quad + \frac{1}{2\pi}\int\left(\varphi_H\left(\frac{v}{T_N}\right) - 1\right)e^{-ivx_k}\varphi_{X_k}(v)dv.\end{aligned}$$

So:

$$\begin{aligned}\left|\widehat{f}_{X_k}(x_k) - f_{X_k}(x_k)\right| &\le \frac{1}{2\pi}\left(\int_{-T_N}^{T_N}\left|\varphi_H\left(\frac{v}{T_N}\right)\right|\left|\widehat{\varphi}_{X_k}(v) - \varphi_{X_k}(v)\right|dv + \int\left|\varphi_H\left(\frac{v}{T_N}\right) - 1\right| h_{X_k}(|v|)dv\right)\\ &\le \frac{T_N}{\pi}\sup_{|\tau|\le T_N}\left|\widehat{\varphi}_{X_k}(\tau) - \varphi_{X_k}(\tau)\right| + \frac{1}{2\pi}\int\left|\varphi_H\left(\frac{v}{T_N}\right) - 1\right| h_{X_k}(|v|)dv.\end{aligned}$$

Note that

$$|\varphi_Y(t)| = \left|\mathbb{E}\left[e^{it^{\mathrm{T}}Y}\right]\right| = \left|\mathbb{E}\left[e^{it^{\mathrm{T}}AX}\right]\right| = |\varphi_X(A^{\mathrm{T}}t)| \ge g_X(\left|A^{\mathrm{T}}t\right|) \ge g_X(L\,|A|\,|t|),$$

where $|A| = \max_{i,j}(|a_{ij}|)$. Moreover, function g_Y inherits g_X's properties: it maps $\mathbb{R}^+$ onto $[0,1]$, it is decreasing and it is integrable, so that in particular $g_Y(|t|) \to 0$ when $|t| \to \infty$. We can thus apply Theorem 2 and obtain:

$$\sup_{x_k} \left| \widehat{f}_{X_k}(x_k) - f_{X_k}(x_k) \right| = \frac{T_N^3}{g(T_N)^3} O(\varepsilon_N) + O\left(\int \left| \varphi_H\left(\frac{v}{T_N}\right) - 1 \right| h_{X_k}(|v|) dv \right).$$

where $g(|t|) = g_X(L|A||t|)$.

If H is a higher-order kernel of order $q \geq 2$, then there exists a function m such that $\varphi_H(v) = 1 + m(v)v^2$ for all $v \in [-1,1]$, and $\varphi_H(v) = 0$ for $v \notin [-1,1]$, where m is continuous on $[-1,1]$. So the bias term is:

$$\begin{aligned} O\left(\int \left| \varphi_H\left(\frac{v}{T_N}\right) - 1 \right| h_{X_k}(|v|) dv \right) &= O\left(\int_{-T_N}^{T_N} \left| m\left(\frac{v}{T_N}\right) \right| \left(\frac{v}{T_N}\right)^2 h_{X_k}(|v|) dv \right) \\ &\quad + O\left(\int_{T_N}^{+\infty} h_{X_k}(|v|) dv \right) \\ &= O\left(\frac{1}{T_N^2} \int_{-T_N}^{T_N} v^2 h_{X_k}(|v|) dv \right) + O\left(\int_{T_N}^{+\infty} h_{X_k}(|v|) dv \right) \\ &= o(1), \end{aligned}$$

as $\sup_{v \in [-1,1]} m(v) = O(1)$. This ends the proof of Theorem 3.

D "Plug-in" bandwidth selection

We here present the "plug-in" method of Delaigle and Gijbels (2004) to choose the bandwidth in deconvolution kernel density estimation. We focus on second-order kernels in the presentation. Extension to higher-order kernels of even order is direct.

To present the method, let us consider the deconvolution problem with known error distribution $Y = X_1 + X_2$, where f_{X_2}, or equivalently φ_{X_2}, is known. Based on a random sample $Y_1, ..., Y_N$, the deconvolution kernel density estimator of f_{X_1} is given by:

$$\widehat{f}_{X_1}(x_1) = \frac{1}{2\pi} \int \varphi_H\left(\frac{v}{T_N}\right) e^{-ivx_1} \frac{\widehat{\varphi}_Y(v)}{\varphi_{X_2}(v)} dv,$$

where $\widehat{\varphi}_Y(v) = \mathbb{E}_N e^{ivY}$ is the empirical characteristic function of Y.

Let the Mean Integrated Squared Error (MISE) of $\widehat{f}_{X_1}$ be:

$$\text{MISE}(T_N) = \mathbb{E}\left(\int \left(\widehat{f}_{X_1}(x_1) - f_{X_1}(x_1) \right)^2 dx_1 \right).$$

The choice of T_N relies on the following approximation of the MISE:

$$\text{MISE}(T_N) \approx \frac{1}{2\pi N} \int \left| \varphi_H\left(\frac{v}{T_N}\right) \right|^2 \left| \varphi_{X_2}(v) \right|^{-2} dv + \frac{\mu_{H,2}^2 R\left(f_{X_1}''\right)}{4T_N^4}. \tag{D3}$$

In this expression: $\mu_{H,2} = \int v^2 H(v) dv$, so for instance $\mu_{H_2,2} = 6$. Moreover:

$$R\left(f_{X_1}''\right) = \int \left[f_{X_1}''(x_1) \right]^2 dx_1.$$

This is an unknown quantity. The plug-in method estimates $R\left(f''_{X_1}\right)$ by the following algorithm.

1. Estimate $R\left(f''''_{X_1}\right)$ as if X_1 was normally distributed:

$$\widehat{R}\left(f''''_{X_1}\right) = \frac{8!}{2^9 4! \sqrt{\pi}\left[\widehat{\mathrm{Var}}(X_1)\right]^{\frac{9}{2}}}.$$

2. Minimize the following quantity with respect to T:

$$-\frac{\mu_{H,2}\widehat{R}\left(f''''_{X_1}\right)}{T^2} + \frac{1}{2\pi N}\int v^6 \left|\varphi_H\left(\frac{v}{T}\right)\right|^2 \left|\varphi_{X_2}(v)\right|^{-2} dv.$$

   This quantity can be interpreted as the squared asymptotic bias of $\widehat{R}\left(f'''_{X_1}\right)$. This step yields $\widehat{T}$.

3. Compute:

$$\widehat{R}\left(f'''_{X_1}\right) = \frac{1}{2\pi}\int v^6 \left|\varphi_H\left(\frac{v}{\widehat{T}}\right)\right|^2 \left|\frac{\widehat{\varphi}_Y(v)}{\varphi_{X_2}(v)}\right|^2 dv.$$

4. Iterate one more time steps 2 and 3. This yields $\widehat{R}\left(f''_{X_1}\right)$.

Finally, once $R\left(f''_{X_1}\right)$ has been estimated, $\widehat{T}_N$ is obtained as the minimizer of the approximated MISE given by the right-hand side of (D3).