Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics

Stéphane Bonhomme
CEMFI, Madrid

Jean-Marc Robin
Université de Paris I,
Paris School of Economics,
CeMMAP, UCL and IFS, London

AIM

- In this paper, we study nonparametric methods for identifying and estimating the distributions of factor and error variables in linear multi-factor models.
- Data: a vector of L measurements Y, modelled as $Y = \Lambda X + U$, where X is a vector of unobservable common factors, U is a vector of errors and Λ is a matrix of parameters (factor loadings).
- The critical assumption is that all components of X and U are mutually independent.
- We assume Λ known or that a root-N consistent estimator is available and that the number of factors is known.
- Problem: identification and estimation of the distributions of X and U.

Examples of application

- Numerous examples of econometric applications in finance, macroeconomics (structural VAR models).
- Examples of microeconometric applications:
  - Models of individual earnings dynamics:

$$\begin{aligned} y_{it} &= y_{it}^P + y_{it}^T \\ &= \left(y_{i0}^P + u_{i1} + ... + u_{it}\right) + \left(\phi v_{i,t-1} + v_{it}\right), \quad t = 1, ..., L, \end{aligned}$$

    where u_{it} is a permanent shock and v_{it} a transitory shock.
  - Linear multifactor heterogeneity structure for latent variables of selection models. See series of recent papers by Heckman and various coauthors, starting with Carneiro, Hansen, Heckman (IER, 2003) and most recent application in Cunha, Heckman, Schennach (2006).
- In all these applications, it is important to estimate both factor loadings and/or a prediction of underlying factors given observables.

Available techniques

- PCA delivers a point estimate of both Λ and X. But factor structure identified only up to a rotation.
- ICA assumes no noise and less factors than measurements and recovers a prediction of X as $\Lambda^{-}Y$.
- Flexible parametric factor distributions like normal mixtures. Difficult to identify the number of mixture components. Computationally intensive (Gibbs sampling, MCMC).

The alternative approach that we adopt in this paper extends nonparametric deconvolution methods.

Nonparametric deconvolution

- Abundant literature on deconvolution:
  - Classical deconvolution problem assumes one measurement, one factor and one error, with known error distribution; see Caroll and Hall (1988), Stefanski and Caroll (1990).
  - Horowitz and Markatou (1996), Li and Vuong (1998), Linton and Whang (2002) and Hall and Yao (2003) have proposed estimators for one-factor models with unknown error distributions.
  - Related methods have been developed by Li (2002), Schennach (2004) and Hu and Ridder (2005) for the problem of estimating nonlinear models with measurement error.
- No available result for multifactor models.

WHAT THE PAPER DOES

- **Identification:** use a powerful identification result due to G.J. Székely and C.R. Rao (Sankhyā, 2000) who show that $\frac{L(L+1)}{2}$ factor and error distributions are nonparametrically identified under a simple condition on factor loadings.

- **Estimation:** We propose a nonparametric estimation procedure for factor and error distributions based on these second-order functional restrictions. Factor and error densities follow by integration and inverse Fourier transformation. This procedure generalizes Li and Vuong's (1998) estimator to the multi-factor case.

- **Asymptotic theory:** We prove that our estimator converges uniformly to the true density when the sample size tends to infinity. Contrary to Li and Vuong (1998), but following Hu and Ridder (2005), we do not assume bounded support for factor and error variables.

- **Monte Carlo simulations**.

- **Empirical application:** wage dynamics.

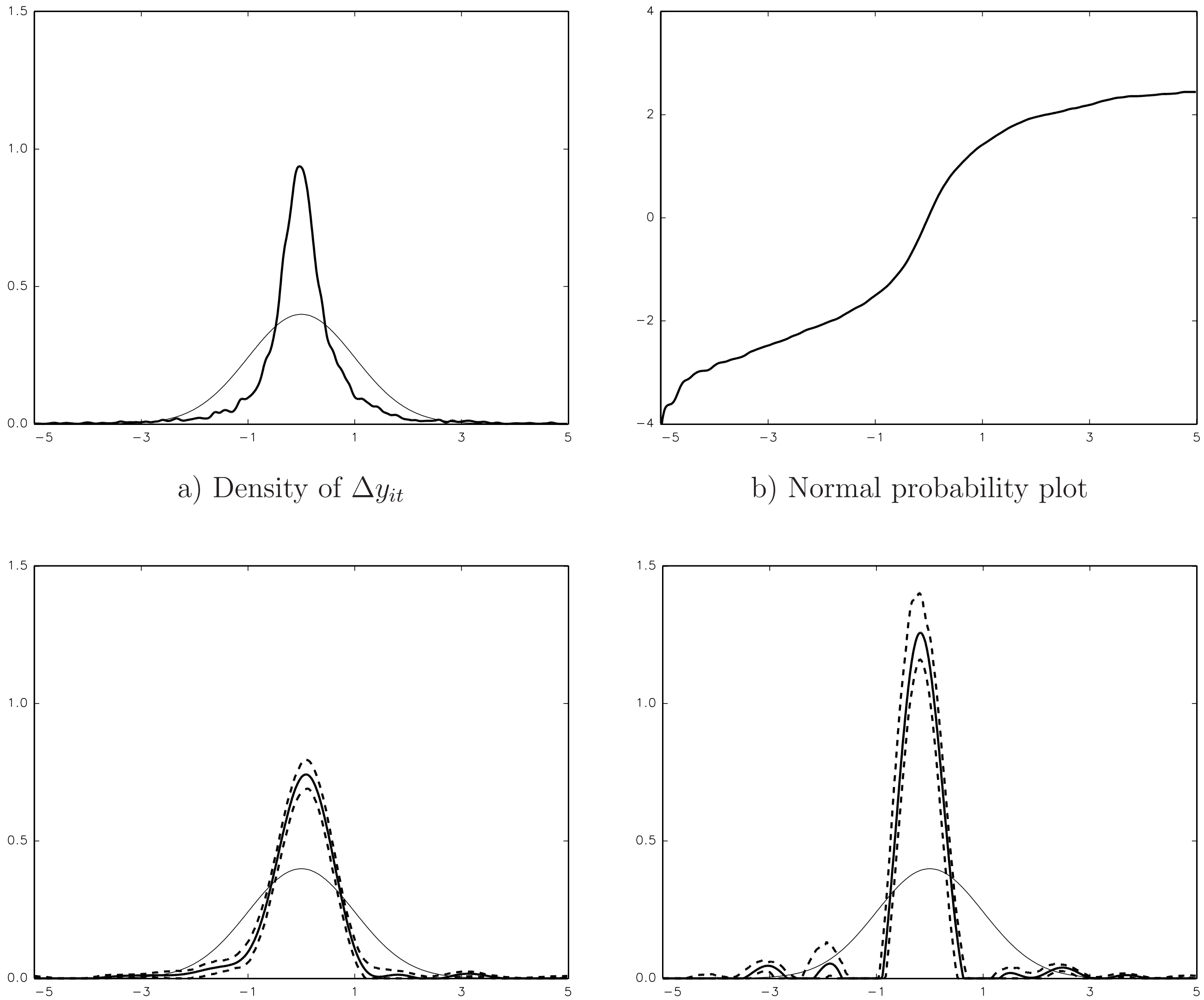
APPLICATION: WAGE DYNAMICS

- PSID, 1978 to 1987
- We consider the following model:

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

  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.

- Estimation results (GMM): $\sigma_\varepsilon^2 = .0256\ (.0050)$, $\sigma_r^2 = .0361\ (.0045)$. Permanent shocks account for 26% of the total variance of wage growth residuals.



a) Density of Δy_{it}

b) Normal probability plot

a) Permanent shock

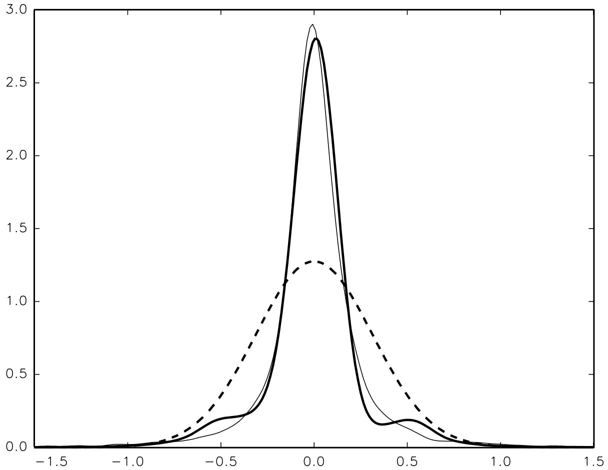
b) Transitory shock

Figure 1: Nonparametric density estimates for standardized permanent and transitory shocks

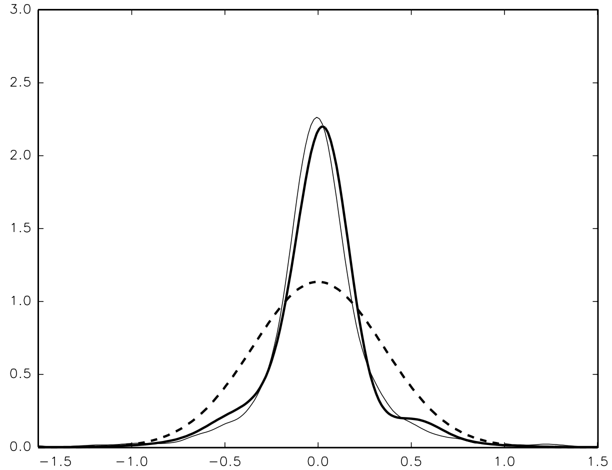
Model fit

	$\Delta y_{it} = y_{it} - y_{i,t-1}$	$\Delta^2 y_{it} = y_{it} - y_{i,t-2}$	$\Delta^3 y_{it} = y_{it} - y_{i,t-3}$
		Data	
Variance	.0915	.1206	.1374
Skewness	-.244	-.440	-.377
Kurtosis	24.2	22.9	19.7
		Predicted, nonparametric	
Variance	.0579	.0734	.0892
Skewness	-.031	-.0034	-.0093
Kurtosis	6.46	5.46	4.87
		Predicted, normal	
Variance	.0978	.1235	.1492
Skewness	0	0	0
Kurtosis	3	3	3
		Predicted, normal mixture	
Variance	.1007	.1275	.1543
Skewness	0	0	0
Kurtosis	11.5	8.66	7.28

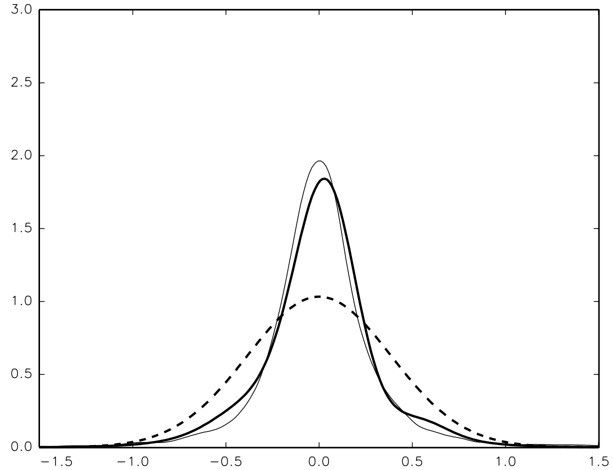
Table 1: Fit of the model, moments of wage growth residuals



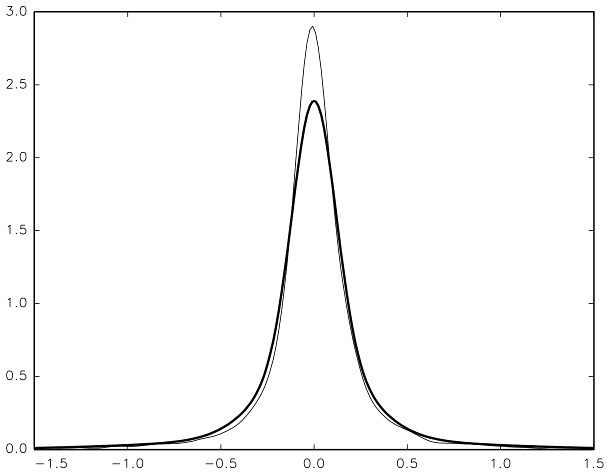
a1) $\Delta y_{it} = y_{it} - y_{i,t-1}$



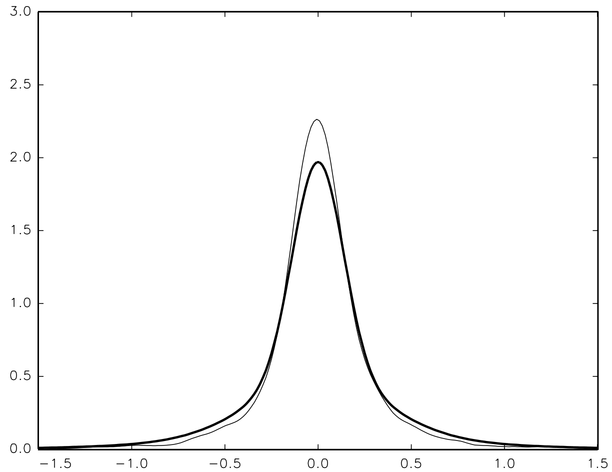
b1) $\Delta_2 y_{it} = y_{it} - y_{i,t-2}$



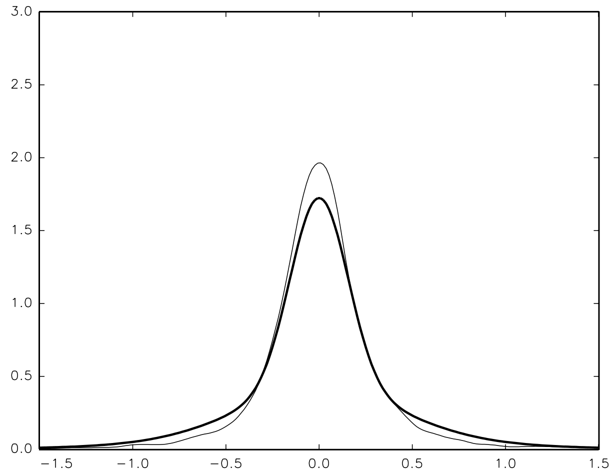
c1) $\Delta_3 y_{it} = y_{it} - y_{i,t-3}$



a2) Δy_{it}, normal mixture



b2) $\Delta_2 y_{it}$, normal mixture



c2) $\Delta_3 y_{it}$, normal mixture

Figure 2: Fit of the model, densities of wage growth residuals (data: thin line; nonparametric/normal mixture estimate: thick line; normal estimate: dashed line).

Components of wage mobility

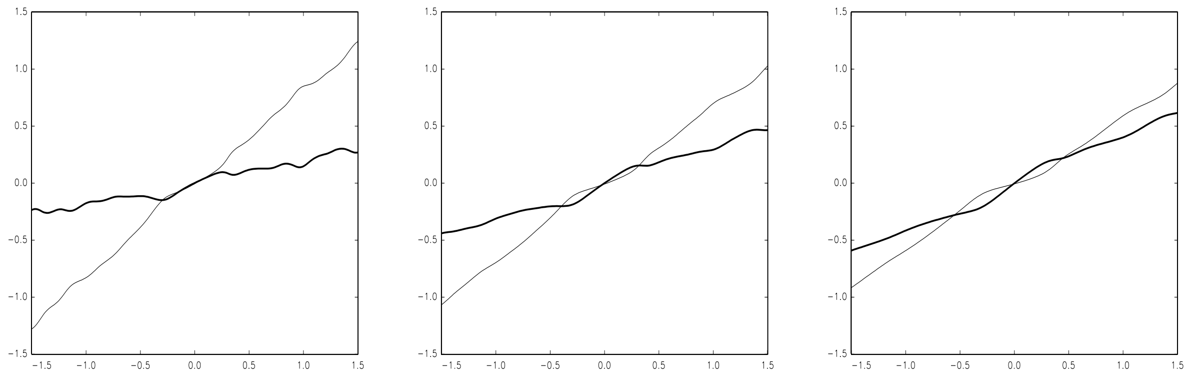
- We compute the conditional expectations of permanent and transitory components given observations of $\Delta_s y_{it}$, $s = 1, 2, 3$, i.e.

$$\mathbb{E}(\sum_{r=0}^{s-1} \varepsilon_{it-r} | \Delta_s y_{it}) \text{ and } \mathbb{E}(r_{it} - r_{it-s} | \Delta_s y_{it}),$$

  using Bayes formula.

- For example

$$f_{\varepsilon_{it}}(\varepsilon | \Delta y_{it} = \Delta y) \propto f_{\varepsilon_{it}}(\varepsilon) f_{\Delta y_{it}}(\Delta y | \varepsilon_{it} = \varepsilon) = f_{\varepsilon_{it}}(\varepsilon) \int f_{\eta_{it}}(r) \, f_{\eta_{i,t-1}}(-\Delta y + \varepsilon + r) dr.$$



a) Δy_{it} b) $\Delta_2 y_{it}$ c) $\Delta_3 y_{it}$

Figure 3: Conditional expectation of shocks given wage growth residuals (permanent: thick line; transitory: thin line)

Job changes

- We compute permanent and transitory shocks given number of job changes.
- We find that job movers have bigger variance components and wage innovations are much more transitory.

Job changes	0	1/2	3+
		Δy_{it}	
total	.04048	.05334	.10929
permanent	.01244	.01400	.02195
transitory	.02804	.03934	.08734
		$\Delta^2 y_{it}$	
total	.04793	.07045	.13705
permanent	.02239	.02858	.04702
transitory	.02554	.04187	.09003
		$\Delta^3 y_{it}$	
total	.06140	.08155	.15740
permanent	.03310	.04170	.07253
transitory	.02830	.03985	.08487

Table 2: Variances of shocks by categories of job changers

IDENTIFICATION THEORY

- Drop the distinction between factor and error
- and consider a DGP: $Y = AX$, where

1. $Y = (Y_1, ..., Y_L)^T$ is a vector of $L \geq 2$ zero-mean real-valued random variables,
2. $X = (X_1, ..., X_K)^T$ is a random vector of K real valued, mutually independent and non degenerate random variables with zero means and finite variances,
3. $A = [a_{\ell k}]$ is a known $L \times K$ matrix of scalar parameters such that any two columns are linearly independent,

   $\rightarrow$ write $A_{[\cdot,k]}$ for column k and $A_{[\ell,\cdot]}$ for row ℓ.
4. Assume that the characteristic functions of factor variables are everywhere non vanishing.

G.J. Székely and C.R. Rao (Sankhyā, 2000)

They show the following general identification theorem.

- Let $\mathbf{B}_p = \left[b_{ik}^{(p)}\right]$ where
$$b_{ik}^{(p)} = a_{\ell_1 k} a_{\ell_2 k} \ldots a_{\ell_{p+1},k}$$
for $k = 1, \ldots, K$ and $i = (\ell_1, \ldots, \ell_{p+1})$ represents one of the $\binom{L+p}{p+1}$ unordered samples with replacement from $\{1, 2, \ldots, L\}$.
- Assume that $\mathbb{E}X_k^s$, $s = 1, 2, \ldots, p$, $k = 1, \ldots, K$, exist and are known.
- Then the distribution of $X_1, \ldots, X_K$ is uniquely determined if and only if $\operatorname{rank}(\mathbf{B}_p) \geq K$.
- When the characteristic function (cf) of Y is differentiable $p+1$ times (and not only p times as implied by the existence of the first p moments), it is easy to construct the cf's of factors $X_1, .., X_K$ from the set of all partial derivatives of $\kappa_Y(t) = \ln\left[\mathbb{E}\exp\left(it^T Y\right)\right]$.

Example for $p = 1$: the measurement error model

- Consider the measurement error model:

$$\begin{cases} Y_1 = X + U_1, \\ Y_2 = aX + U_2, \end{cases} \qquad a \neq 0, \quad X \in \mathbb{R}.$$

- If $\mathbb{E}X^3 \equiv 0$, a can be estimated by regressing Y_2 on Y_1 using $Y_1 Y_2$ as instrument (Geary, 1942).

- The cumulant generating function (cgf) of $Y = (Y_1, Y_2)$ is the log of the characteristic function (cf):

$$\begin{aligned} \kappa_Y(t_1, t_2) &= \ln \mathbb{E} \exp i \left(t_1 Y_1 + t_2 Y_2\right) \\ &= \ln \mathbb{E} \exp i \left(\left(t_1 + a t_2\right) X + t_1 U_1 + t_2 U_2\right) \\ &= \ln \mathbb{E} e^{i(t_1 + a t_2) X} + \ln \mathbb{E} e^{i t_1 U_1} + \ln \mathbb{E} e^{i t_2 U_2} \\ &\equiv \kappa_X\left(t_1 + a t_2\right) + \kappa_{U_1}\left(t_1\right) + \kappa_{U_2}\left(t_2\right) \end{aligned}$$

- Differentiate two times:

$$\begin{pmatrix} \partial_{11}^2 \kappa_Y(t_1,t_2) \\ \partial_{12}^2 \kappa_Y(t_1,t_2) \\ \partial_{22}^2 \kappa_Y(t_1,t_2) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ a & 0 & 0 \\ a^2 & 0 & 1 \end{pmatrix}}_{=\mathbf{B}_1} \begin{pmatrix} \kappa_X''(t_1+at_2) \\ \kappa_{U_1}''(t_1) \\ \kappa_{U_2}''(t_2) \end{pmatrix}$$

- This implies that

$$\kappa_X''(t_1+at_2) = \frac{1}{a}\partial_{12}^2 \kappa_Y(t_1,t_2)$$

or

$$\kappa_X''(\tau) = \frac{1}{a}\partial_{12}^2 \kappa_Y\left(t_1, \frac{1}{a}\tau - \frac{1}{a}t_1\right), \ \forall \tau, t_1.$$

- Integrate using $\kappa_X(0) = \kappa_X'(0) = 0$ as integration constants:

$$\begin{aligned} \kappa_X(\tau) &= \frac{1}{a}\int_0^\tau \int_0^u \partial_{12}^2 \kappa_Y\left(t_1, \frac{1}{a}v - \frac{1}{a}t_1\right) dv du \\ &= \int_0^\tau \left[\partial_1 \kappa_Y\left(t_1, \frac{1}{a}u - \frac{1}{a}t_1\right) - \partial_1 \kappa_Y\left(t_1, -\frac{1}{a}t_1\right)\right] du. \end{aligned}$$

- Setting $t_1 = 0$ yields Li and Vuong's (1998) solution:

$$\kappa_X(\tau) = \int_0^\tau \partial_1 \kappa_Y \left(0, \frac{1}{a}u\right) du = \int_0^\tau i \frac{\mathbb{E}\left[Y_1 e^{iuY_2/a}\right]}{\mathbb{E}\left[e^{iuY_2/a}\right]} du.$$

(Li &Vuong assume $a = 1$.)

- Remarks:

1. As pointed out by Schennach, Li & Vuong's particular form of the general estimator only requires $\mathbb{E}(U_1|X, U_2) = 0$ and $U_2 \perp\!\!\!\perp X$.
2. The double integrals of second-order derivatives of κ_Y will not simplify into a simple integral of first derivatives for all different possible choices of t_1.
3. Apart from that the choice of t_1 is arbitrary. Our preferred choice takes t_1 so as to minimize

$$\left\| \left(t_1, \frac{1}{a}\tau - \frac{1}{a}t_1\right) \right\|^2 = t_1^2 + \left(\frac{1}{a}\tau - \frac{1}{a}t_1\right)^2,$$

as cf's are more difficult to estimate at large frequencies, i.e. $t_1 = \tau / \left(1 + \frac{1}{a}\right)$.

General constructive identification proof for $p = 1$

(maybe skipped)

- Consider cumulant generating functions (cgf; = log of characteristic function) of measurements and factors:

$$\kappa_Y(t) = \ln\left[\mathbb{E}\exp\left(it^T Y\right)\right] \quad \text{and} \quad \kappa_{X_k}(\tau) = \ln\left[\mathbb{E}\exp\left(i\tau X_k\right)\right].$$

- Independence and linearity imply that, for all $t = (t_1, ..., t_L) \in \mathbb{R}^L$,

$$\boxed{\kappa_Y(t) = \textstyle\sum_{k=1}^K \kappa_{X_k}\left(t^T A_{[\cdot,k]}\right).}$$

- Differentiate the cgf of Y:

$$\begin{aligned} \kappa_Y(t) &= \sum_{k=1}^K \kappa_{X_k}\left(t^T A_{[\cdot,k]}\right) \\ \Longrightarrow \partial_\ell \kappa_Y(t) &= i\frac{\mathbb{E}\left[Y_\ell e^{it^T Y}\right]}{\mathbb{E}\left[e^{it^T Y}\right]} = \sum_{k=1}^K a_{\ell k}\kappa'_{X_k}\left(t^T A_{[\cdot,k]}\right). \end{aligned}$$

- In general, $K > L$ as there are L errors and at least one common factor. So there are more function κ'_{X_k} than partial derivatrives $\partial_\ell \kappa_Y$.

- To obtain an invertible system, differentiate one more time:

$$\begin{aligned}\partial^2_{\ell m}\kappa_Y(t) &= -\frac{\mathbb{E}\left[Y_\ell Y_m e^{it^TY}\right]}{\mathbb{E}\left[e^{it^TY}\right]} + \frac{\mathbb{E}\left[Y_\ell e^{it^TY}\right]}{\mathbb{E}\left[e^{it^TY}\right]}\frac{\mathbb{E}\left[Y_m e^{it^TY}\right]}{\mathbb{E}\left[e^{it^TY}\right]} \\ &= \sum_{k=1}^{K} a_{\ell k}a_{mk}\kappa''_{X_k}\left(t^TA_{[\cdot,k]}\right), \qquad \ell \leq m.\end{aligned}$$

- This is a system of $\frac{L(L+1)}{2}$ equations and K variables $\kappa''_{X_1}\left(t^TA_{[\cdot,1]}\right), ..., \kappa''_{X_K}\left(t^TA_{[\cdot,K]}\right)$.
- Assume that $\mathbf{B}_1 = [a_{\ell 1}a_{m1}, ..., a_{\ell K}a_{mK}]_{\ell\leq m} \in \mathbb{R}^{\frac{L(L+1)}{2}\times K}$ is full column rank (so $K \leq \frac{L(L+1)}{2}$).
- One can invert the identifying system of second-order restrictions as

$$\kappa''_{X_k}\left(t^TA_{[\cdot,k]}\right) = \left(\mathbf{B}_1^-\right)_{[k,\cdot]}\nabla^2\kappa_Y(t) \quad (k = 1, ..., K)$$

where

  - $\mathbf{B}_1^-$ is a generalized inverse of $\mathbf{B}_1$,
  - $\left(\mathbf{B}_1^-\right)_{[k,\cdot]}$ denote the kth row of $\mathbf{B}_1^-$
  - and $\nabla^2\kappa_Y(t) = \left[\partial^2_{\ell m}\kappa_Y(t); \ell \leq m\right]$ is the $\frac{L(L+1)}{2}$-vector of non redundant second-order partial derivatives.

- Let $\mathcal{T}_k = \left\{ t \in \mathbb{R}^L | t^T A_{[\cdot,k]} = 1 \right\}$. Then, for all $t \in \mathcal{T}_k$ and $\tau \in \mathbb{R}$,

$$\kappa''_{X_k}(\tau) = \left(\mathbf{B}_1^-\right)_{[k,\cdot]} \nabla^2 \kappa_Y(\tau t).$$

- Integrating with respect to τ yields

$$\kappa_{X_k}(\tau) = \int_0^\tau \int_0^u \left(\mathbf{B}_1^-\right)_{[k,\cdot]} \nabla^2 \kappa_Y(vt)\, dv du.$$

Example for $p = 2$: Two measurements, two factors

- Consider model
$$\begin{cases} Y_1 = X_1 + X_2 + U_1, \\ Y_2 = X_1 + aX_2 + U_2, \end{cases}$$
with $a \neq 1$ and $a \neq 0$, is identified provided that the variances of X_1 and X_2 are known (preestimated) and normalized to one.

- $\mathbf{B}_2$ full column rank is a necessary condition for identification as $K = 4 = \binom{L+p}{p+1}$.

- One can show that
$$\begin{cases} \kappa_{X_1}(\tau_1) = \int_0^{\tau_1} \int_0^u \partial_{12}^2 \kappa_Y(2v, -v)\, dv du - a\frac{\tau_1^2}{2}, \\ \kappa_{X_2}(\tau_2) = \int_0^{\tau_2} \int_0^u \partial_{12}^2 \kappa_Y(-v, v)\, dv du - \frac{1}{a}\frac{\tau_2^2}{2}. \end{cases}$$

ESTIMATION

- **First step – Empirical characteristic functions**:

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

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

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

- **Second step – Integration**:

$$\widehat{\kappa}_{X_k}(\tau) = \int_0^\tau \int_0^u \left(\mathbf{B}_1^-\right)_{[k,\cdot]}\left(\int \widehat{\nabla^2\kappa_Y}(vt)\, dW(t)\right) dv du,$$

where W is some probability distribution on $\mathcal{T}_k = \left\{t \in \mathbb{R}^L | t^T A_{[\cdot,k]} = 1\right\}$.

- **Third step – Inverse Fourier transformation**:

$$\widehat{f}_{X_k}(x) = \frac{1}{2\pi}\int_{-T_N}^{T_N} \widehat{\varphi}_{X_k}(\tau)e^{-i\tau x} d\tau = \frac{1}{2\pi}\int_{-T_N}^{T_N} \exp\left[-i\tau x + \widehat{\kappa}_{X_k}(\tau)\right] d\tau,$$

where the smoothing parameter T_N tends to infinity at a rate to be specified.

Choice of the weighting distribution W

- Empirical characteristic functions are well estimated around the origin and badly estimated in the tails.
- It 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 $\|t\|$, for $t \in \mathcal{T}_k = \left\{ t \in \mathbb{R}^L | t^T A_{[\cdot,k]} = 1 \right\}$:

$$\arg\min_t \left\| \frac{t}{t^T A_{[.,k]}} \right\| \equiv t^* = \left(A_{[\cdot,k]} \right)^{-T} = \frac{A_{[\cdot,k]}}{A_{[\cdot,k]}^T A_{[\cdot,k]}}.$$

- The simulation section will provide evidence that choosing $W = \delta_{t^*}$ works well in practice.

Li and Vuong (1998) and Hall and Yao (2003) assume bounded supports. Clashes with assumption that cf $\neq 0$ (same remark in Hu and Ridder, 2005). We relax the boundedness assumption.

Lemma 1 *Let X be a scalar random variable and let Y be a vector of L scalar random variables. Let there be an iid sample of N observations of (X, Y).*
Assume that $EX^2 \leq M_1 < \infty$ and that $E\left|Y\right|^i < \infty$ for all $i \in \{1, ..., L\}$.
Define $f_t(x, y) = x \exp(it^T y)$ for $t \in \mathbb{R}^L$.
Let $K_{|X|}(\varepsilon)$ be such that

$$\mathbb{E}\left[|X| \mathbf{1}\left\{|X| > K_{|X|}(\varepsilon)\right\}\right] = \int_{K_{|X|}(\varepsilon)}^{\infty} u f_{|X|}(u)\, du = \varepsilon.$$

Then,

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

(where $|t| = \max_\ell |t_\ell|$) for all ε_N, T_N such that

$$\ln T_N = O\left(\ln N\right) \quad \text{and} \quad \frac{K_{|X|}(\varepsilon_N)}{\varepsilon_N} = o\left[\left(\frac{N}{\ln N}\right)^{\frac{1}{2}}\right].$$

Consistency of factor cf's

Theorem 2 *Suppose that there exists an integrable, decreasing function $g_Y : R^+ \to [0,1]$, such that $|\varphi_Y(t)| \geq g_Y(|t|)$ as $|t| \to \infty$. Then, there exists $\varepsilon_N \downarrow 0$ and $T_N \to \infty$ such that*

$$\sup_{|\tau| \leq T_N} \left| \widehat{\varphi}_{X_k}(\tau) - \varphi_{X_k}(\tau) \right| = \frac{T_N^2}{g_Y(T_N)^3} O(\varepsilon_N) \quad a.s., \tag{1}$$

where ε_N is the minimum convergence rate satisfying the conditions of Lemma 1 for all functions f_t of the form $\exp\left(it^T Y\right)$, $Y_\ell \exp\left(it^T Y\right)$ and $Y_\ell Y_m \exp\left(it^T Y\right)$, $\ell, m \in \{1, ..., L\}$, and T_N satisfies two constraints: $\ln T_N = O\left(\ln N\right)$, and $\frac{T_N^2}{g_Y(T_N)^3} \varepsilon_N = o(1)$.

- More difficult to identify factor characteristic function if measurement distribution is smooth ($g_Y(|t|)$ decays rapidly to zero).

Consistency of factor pdf's

Theorem 3 *Suppose that there exists an integrable, decreasing function* $g_X : R^+ \to [0,1]$ *such that* $|\varphi_X(\tau)| \geq g_X(|\tau|)$ *as* $|\tau| \to \infty$. *Suppose also that there exist* K *integrable functions* $h_{X_k} : R^+ \to [0,1]$ *such that* $h_{X_k}(|\tau|) \geq \left|\varphi_{X_k}(\tau)\right|$ *as* $|\tau| \to \infty$. *Then,* $\widehat{f}_{X_k}$ *is a uniformly convergent estimator of the pdf* f_{X_k} *of* X_k*, i.e.*

$$\sup_x \left|\widehat{f}_{X_k}(x) - f_{X_k}(x)\right| = \frac{T_N^3}{g_X(T_N)^3} O(\varepsilon_N) + O\left(\int_{T_N}^{+\infty} h_{X_k}(v)dv\right) = o(1) \quad a.s., \tag{2}$$

where ε_N *and* T_N *are given by Theorem 2 applied to* $g_Y(|t|) = g_X(L|A||t|)$.

Interestingly, smoothness is both good and bad.

- On one hand, smooth distributions require less trimming (less regularization).
- On the other hand, it is more difficult to separate the different sources of information if factors are smooth.
- A more precise assessment of this trade-off requires a more precise, parametric specification of factor distributions' tails and smoothness. In which case, one can optimize the choice of T_N wrt ε_N in (2).

Particular case: Pareto tailed, polynomially-smooth factor distributions

Corollary 4 *Assume that there exists* $1 < \beta_k \leq \alpha_k$ *such that*

$$|\tau|^{-\alpha_k} \leq \left|\varphi_{X_k}(\tau)\right| \leq |\tau|^{-\beta_k}, \quad |\tau| \to \infty,$$

and $a > 1$ *such that* $K_{|X_k|}(\varepsilon) \leq (1/\varepsilon)^{a-1}$. *Then*

$$\sup_x \left|\widehat{f}_{X_k}(x) - f_{X_k}(x)\right| = O\left(\left(\frac{\ln N}{N}\right)^{\frac{\beta_k - 1}{2+3\alpha+\beta_k} a(1/2-\gamma)}\right) \quad a.s.,$$

for $\alpha = \sum_{k=1}^{K} \alpha_k$, *and for a trimming parameter* T_N *in* $\widehat{f}_{X_k}$ *chosen such as*

$$T_N = O\left(\left(\frac{N}{\ln N}\right)^{\frac{a(1/2-\gamma)}{2+3\alpha+\beta_k}}\right).$$

- The convergence rate is polynomial in $\frac{\ln N}{N}$ instead of $\frac{\ln \ln N}{N}$ as in Li and Vuong (1998). This is because factor distributions do not necessarily have bounded support.
- Thick tails (small a) require more trimming and yield lower convergence rates.

Is smoothness good or bad?

- The rate of convergence increases with β_k (smoothness is good) and decreases with α_k (smoothness is bad), the other α_m, $m \neq k$, remaining constant.

- Makes little sense to vary α_k independently from β_k. Let $\alpha_k = \beta_k$. Then, the optimal convergence rate unambiguously increases with β_k (smoothness is good) and less trimming is necessary to achieve a given rate.

- Furthermore, the optimal uniform rate of convergence of $\widehat{f}_{X_k}(x)$ unambiguously decreases with $\alpha - \alpha_k = \sum_{m \neq k} \alpha_m$. That is to say that it is more difficult to identify and estimate the distribution of one factor if the other factors and errors are smooth.

Practical choice of the smoothing parameter T_N

- Asymptotic bounds do not yield practical ways of choosing T_N.
- We use method in Diggle and Hall (1988) who consider the deconvolution problem, $Y = X + U$, with independent random samples for Y and U (!). Our simulations proved that their method also works well in our setup.

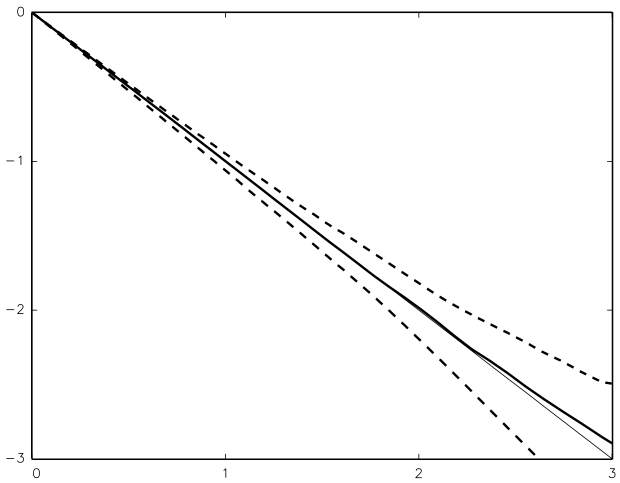
MONTE-CARLO SIMULATIONS

A measurement error model ($L = 2$, $K = 3$)

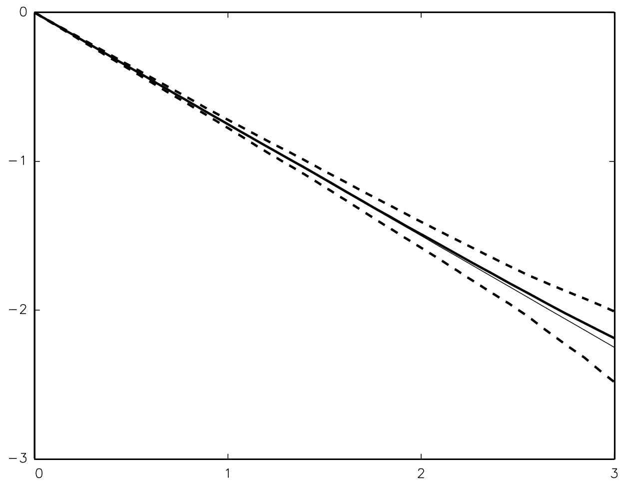
- We simulate

$$\begin{cases} Y_1 = X_1 + U_1, \\ Y_2 = X_1 + U_2, \end{cases}$$

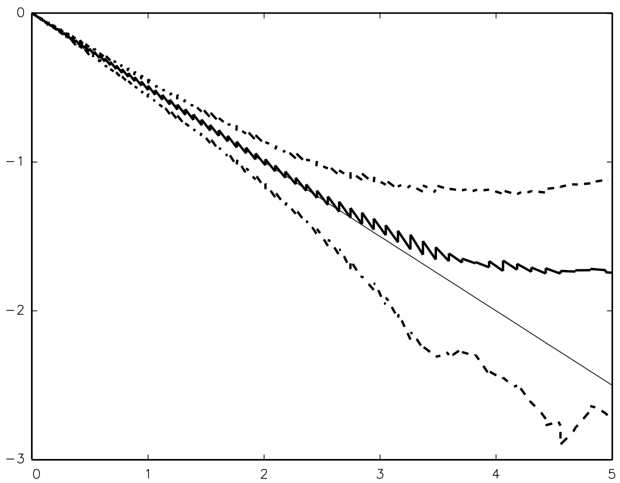
where $(X_1, U_1, U_2) \sim N(0, I_3)$.



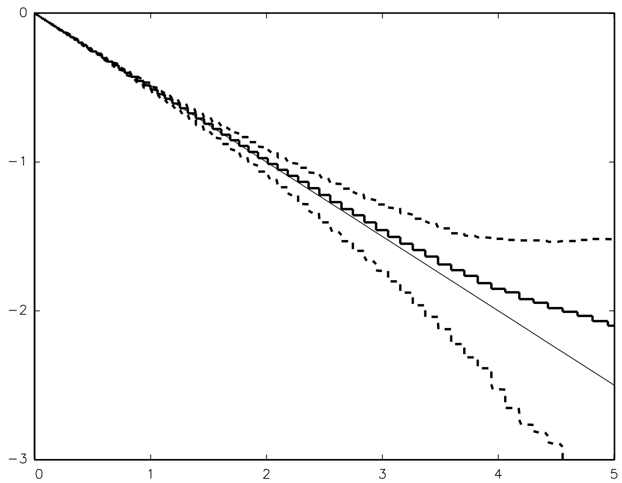
a) $\ln|\varphi_Y(t)|$, $t = (0, \tau)$



b) $\ln|\varphi_Y(t)|$, $t = (\tau/2, \tau/2)$



c) $\ln\left|\varphi_{X_1}(\tau)\right|$, $t = (0, \tau)$



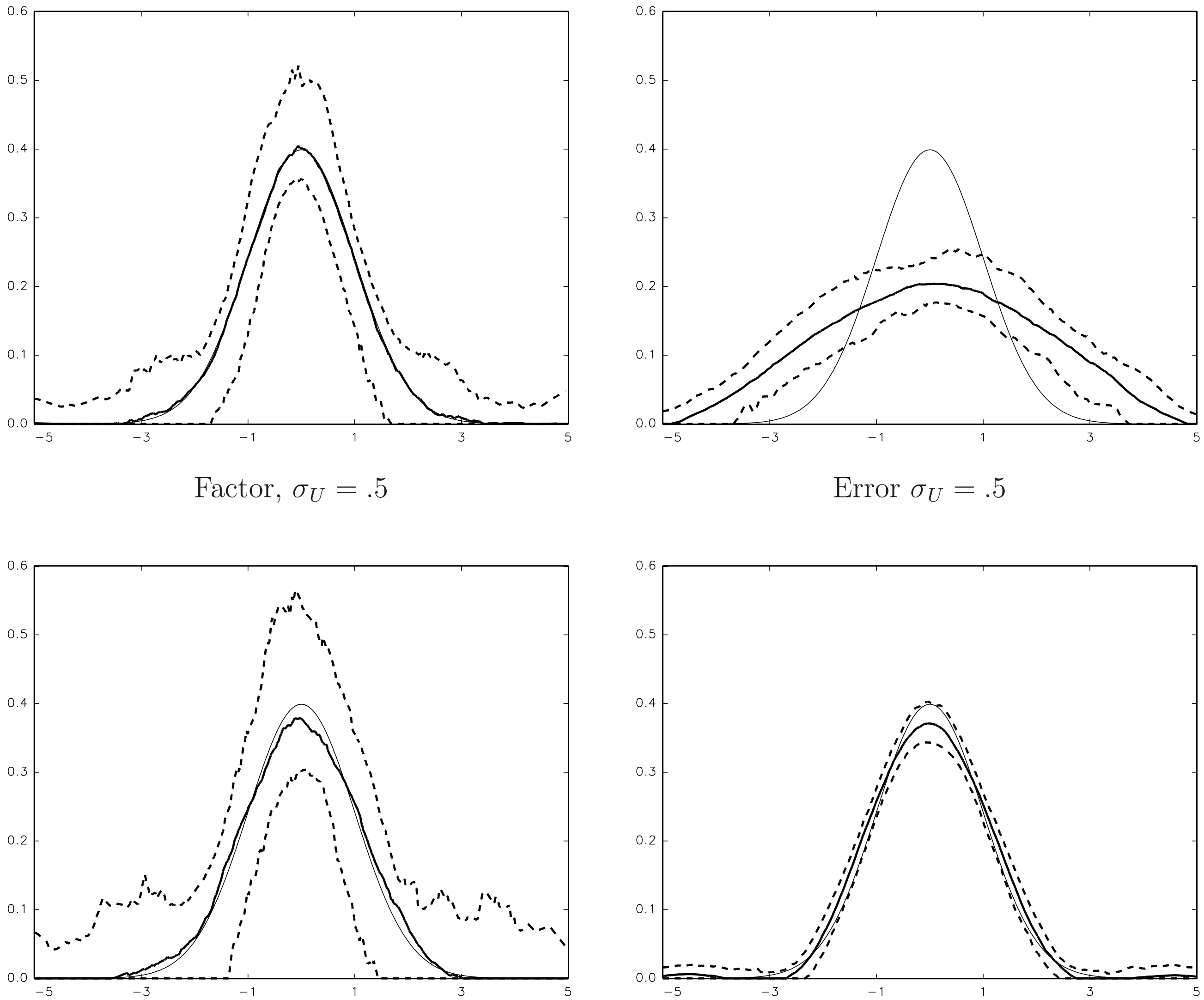
d) $\ln\left|\varphi_{X_1}(\tau)\right|$, $t = (\tau/2, \tau/2)$

Figure 4: Monte Carlo simulations for the estimated characteristic functions in the measurement error model (left: Li and Vuong; right: our preferred direction of integration)

A 3-measurement, 3-factor, 2-error case ($L = 3$ and $K = 6$!)

- We simulate $Y = \Lambda X + U$, with $Y \in \mathbb{R}^3$, $X \in \mathbb{R}^3$ and $U \in \mathbb{R}^3$.

- We set
$$\Lambda = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},$$
and assume that all factors follow the same distribution. Idem for all errors.



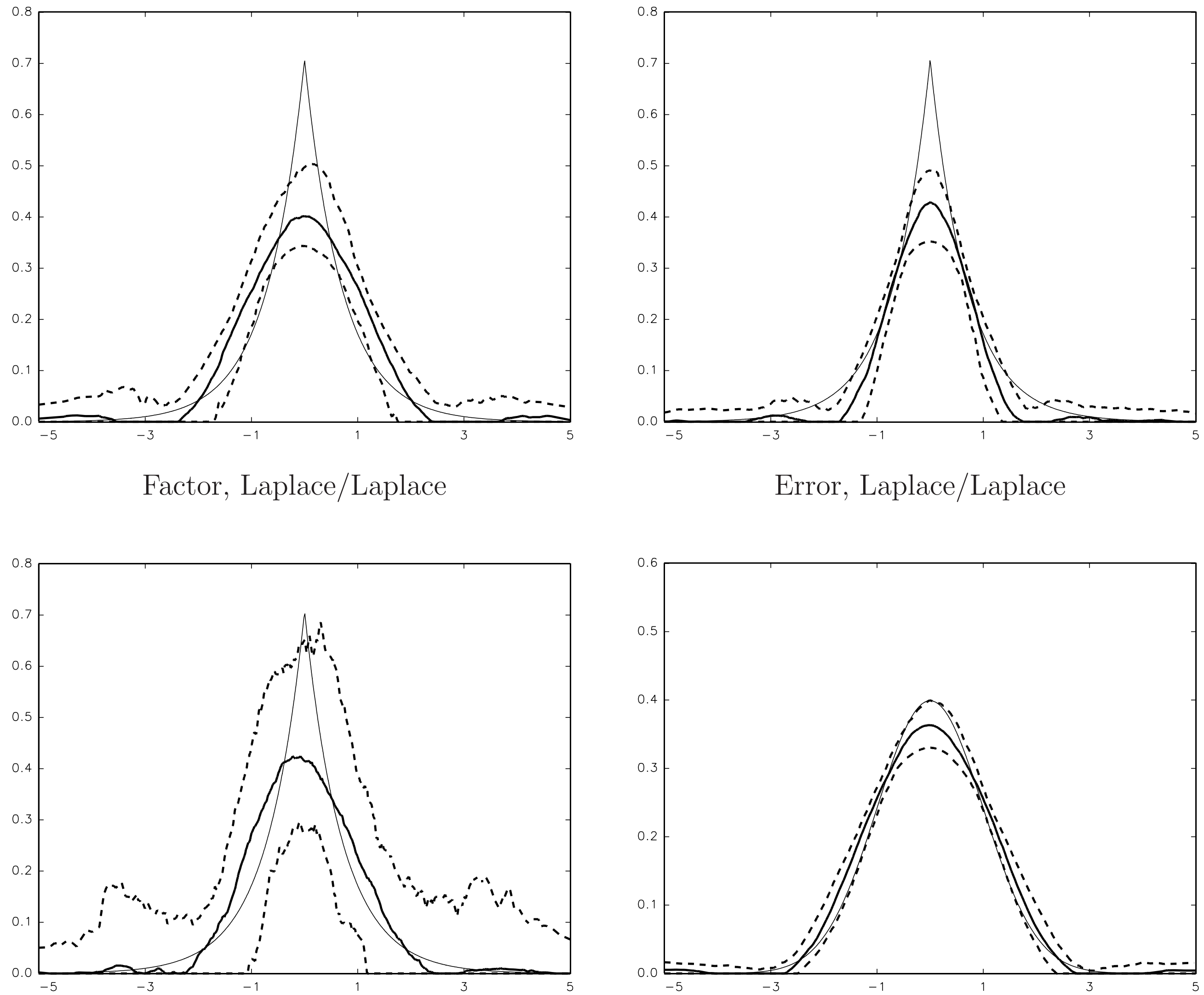
Factor, $\sigma_U = .5$

Error $\sigma_U = .5$

Factor, $\sigma_U = 2$

Error, $\sigma_U = 2$

Figure 5: Monte Carlo simulations for density estimates in the linear factor model with 3 measurements, 3 factors and 3 errors — normal distributions



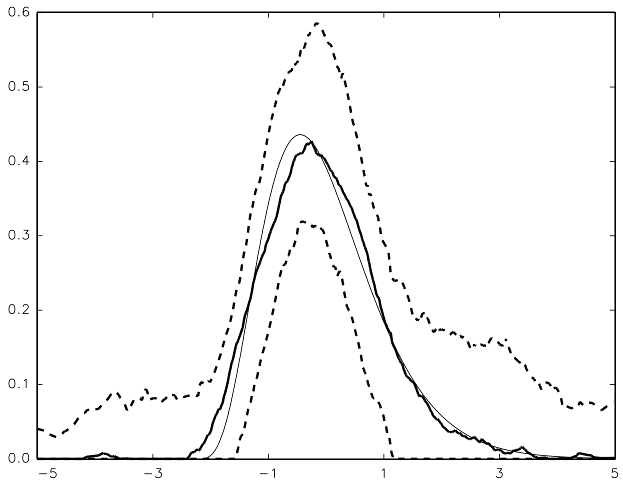
Factor, Laplace/Laplace

Error, Laplace/Laplace

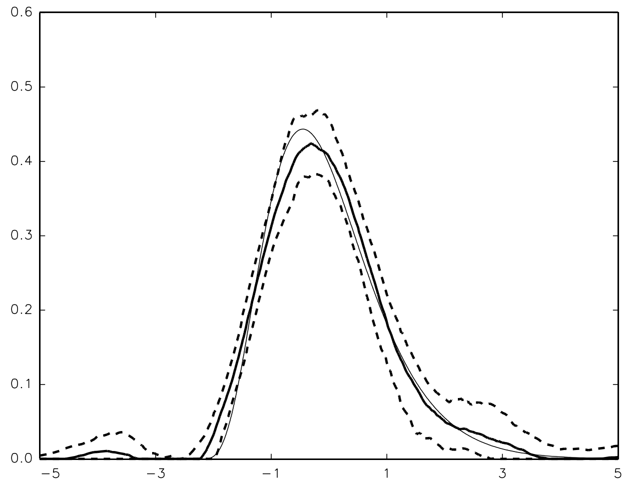
Factor, Laplace/normal

Error, Laplace/normal

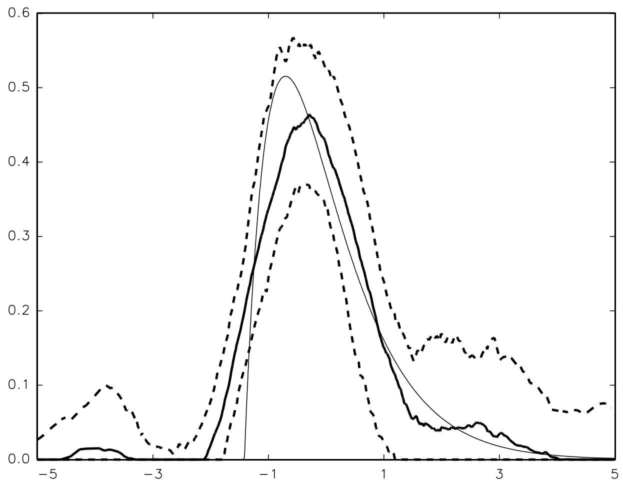
Figure 6: Monte Carlo simulations for density estimates in the linear factor model with 3 measurements, 3 factors and 3 errors — Laplace and normal distributions



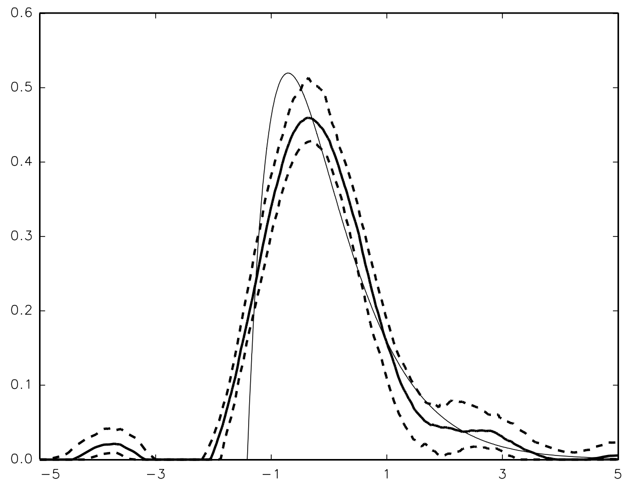
Factor, gamma $(5, 1)$



Error, gamma $(5, 1)$

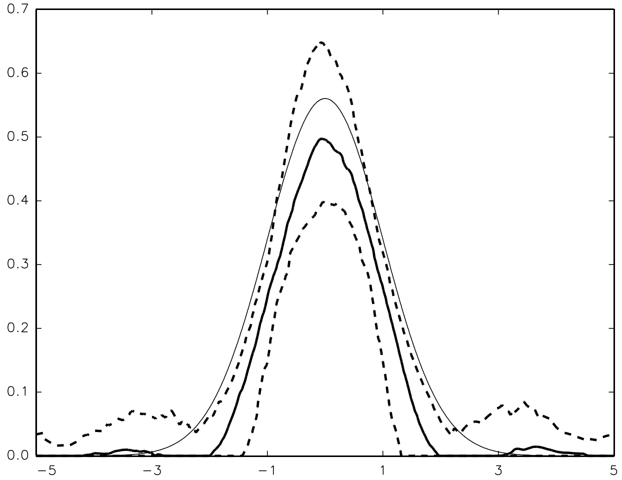


Factor, gamma $(2, 1)$

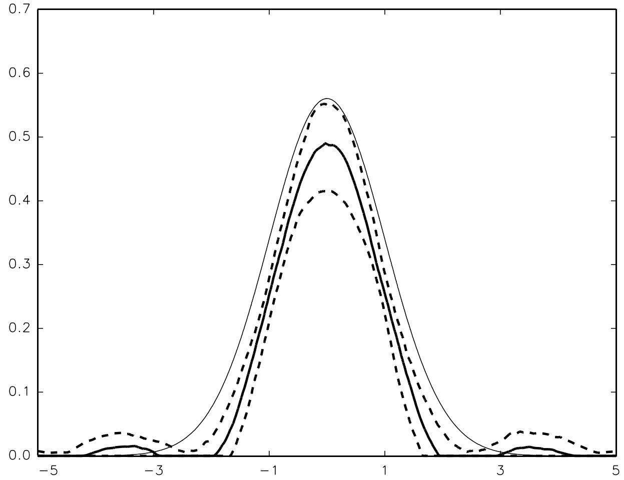


Error, gamma $(2, 1)$

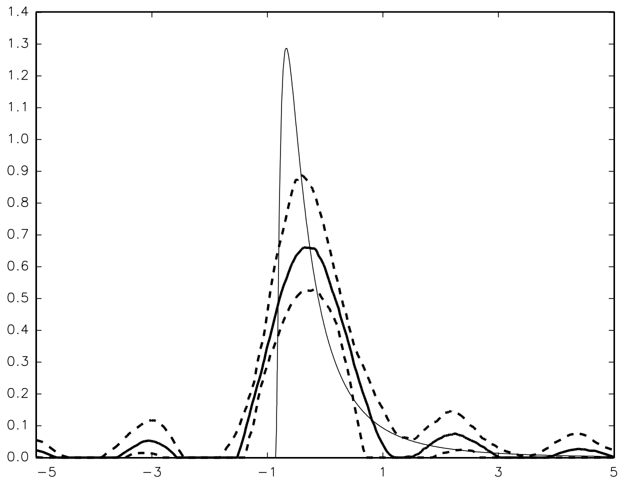
Figure 7: Monte Carlo simulations for density estimates in the linear factor model with 3 measurements, 3 factors and 3 errors — Gamma distributions



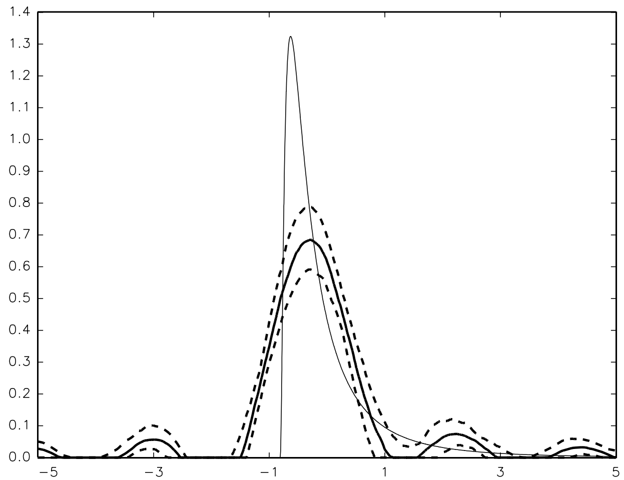
Factor, normal mixture ($\kappa_4 = 100$)



Error, normal mixture ($\kappa_4 = 100$)



Factor, log-normal



Error, log-normal

Figure 8: Monte Carlo simulations for density estimates in the linear factor model with 3 measurements, 3 factors and 3 errors — normal mixtures and log-normal distributions

CONCLUSION

- This paper provides a generalization of the nonparametric estimator of Li and Vuong (1998) to the case of a general linear independent factor structure, allowing for any number of measurements, L, and at most $\frac{L(L+1)}{2}$ factors (including errors).
- The main lessons of the standard deconvolution literature carry over:
  - Convergence rates are slow; it is easier to identify the distribution of a smooth factor;
  - and it is easier to identify the distribution of one factor if the other factor distributions are not smooth.
- Monte Carlo simulations show that the method works reasonably well in practice.
- However, it is easier to identify smooth distributions with little kurtosis excess.
- In Bonhomme and Robin (2005), 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 precise estimates of factor loadings and precise estimates of factor distributions.
- Future work: nonlinear factor models, ARCH factors, ...