Econometrica Supplementary Material

SUPPLEMENT TO "FUNCTIONAL DIFFERENCING"

(*Econometrica*, Vol. 80, No. 4, July 2012, 1337–1385)

BY STÉPHANE BONHOMME

THIS SUPPLEMENT CONTAINS analytical and numerical results on various models. It also presents a method to numerically compute information bounds and check the non-surjectivity condition. Last, it outlines a specification test of parametric random-effects models.

S1. EXAMPLES: ANALYTICAL RESULTS

S1.1. *Details About Derivations in the Text*

EXAMPLE 1A: Note the identity

$$\begin{aligned}
&(y-a-B\alpha)'\Sigma^{-1}(y-a-B\alpha) \\
&\quad = (y-a-B\alpha)'\Sigma^{-1/2}Q\Sigma^{-1/2}(y-a-B\alpha) \\
&\qquad + (y-a)'\Sigma^{-1/2}W\Sigma^{-1/2}(y-a).
\end{aligned}$$

We have, for every function $g(\alpha)$, and denoting $q=\dim\alpha$,

$$\begin{aligned}
[L_{\theta,x}g](y) &= \int_{\mathbb{R}^q} f_{y|x,\alpha}(y|x,\alpha;\theta)g(\alpha)\,d\alpha \\
&= (2\pi)^{-T/2}|\Sigma|^{-1/2}\left\{\int_{\mathbb{R}^q}\exp\left[-\frac{1}{2}(y-a-B\alpha)'\right.\right. \\
&\qquad \left.\left.\times\,\Sigma^{-1/2}Q\Sigma^{-1/2}(y-a-B\alpha)\right]g(\alpha)\,d\alpha\right\} \\
&\qquad \times\left\{\exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}W\Sigma^{-1/2}(y-a)\right]\right\}.
\end{aligned} \tag{S1}$$

This shows (11).

EXAMPLE 1B: Here we extend the analysis of Example 1B to censored regression models with random coefficients.

Let us assume for simplicity that $B(x,\theta)$ has full-column rank q, for all θ, almost surely in x. Let V be a $T\times q$ matrix such that $Q=VV'$ and $V'V=I_q$. Let also U be a $T\times(T-q)$ matrix such that $W=UU'$, and $U'U=I_{T-q}$. Last, let $(\mu,\nu)=(V'\Sigma^{-1/2}y, U'\Sigma^{-1/2}y)$.

Then let us consider a region in $\mathbb{R}^T$ of the form

$$\{y\in\mathbb{R}^T, (\mu,\nu)\in R_1\times R_2\}\subset\{y\in\mathbb{R}^T, y_1>c_1,\dots,y_T>c_T\},$$

 DOI: 10.3982/ECTA9311

where R_1 and R_2 are subsets of $\mathbb{R}^q$ and $\mathbb{R}^{T-q}$, respectively. Finally, let us define the function supported on that Cartesian product:

$$\varphi(y) = \varphi_2(\nu)\mathbf{1}\{\mu \in R_1\}\mathbf{1}\{\nu \in R_2\}.$$

Then (9) will hold if φ_2 and R_2 are chosen such that

$$\int_{R_2} \varphi_2(\nu) \exp\left[-\frac{1}{2}\left(\nu - U'\Sigma^{-1/2}a\right)'\left(\nu - U'\Sigma^{-1/2}a\right)\right] d\nu = 0. \tag{S2}$$

In particular, if R_2 is chosen such that $\{\nu - U'\Sigma^{-1/2}a, \nu \in R_2\}$ is symmetric around zero, then

$$\mathbb{E}\left[U'\Sigma^{-1/2}(y_i - a)\mathbf{1}\left\{V'\Sigma^{-1/2}y_i \in R_1\right\}\mathbf{1}\left\{U'\Sigma^{-1/2}y_i \in R_2\right\}|x_i\right] = 0. \tag{S3}$$

Restrictions (S3) are valid under nonnormality if the distribution of $U'\Sigma^{-1/2}v_i$ is symmetric around the origin.

EXAMPLE 2: To see why finding a nonzero $\{\varphi_y\}$ that satisfies (16) is equivalent to all 2^T products of distinct F's being linearly dependent, $F_1^{k_1} \times \cdots \times F_T^{k_T}$, $(k_1, \ldots, k_T) \in \{0, 1\}^T$, consider the case $T = 2$. Then (16) can be written as

$$\begin{aligned}&\varphi_{00} + (\varphi_{10} - \varphi_{00})F_1 + (\varphi_{01} - \varphi_{00})F_2 \\ &\quad + (\varphi_{11} - \varphi_{10} - \varphi_{01} + \varphi_{00})F_1F_2 = 0,\end{aligned}$$

and we have

$$\begin{pmatrix} \varphi_{00} \\ \varphi_{10} - \varphi_{00} \\ \varphi_{01} - \varphi_{00} \\ \varphi_{11} - \varphi_{10} - \varphi_{01} + \varphi_{00} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_{00} \\ \varphi_{10} \\ \varphi_{01} \\ \varphi_{11} \end{pmatrix}.$$

This triangular structure holds for all $T \geq 2$.

Last, we prove (18). We have

$$\begin{aligned}
(17) \quad &\Leftrightarrow \quad \sum_{y \in \{0,1\}^T} \varphi_y(x, \theta) \prod_{t=1}^{T} \Lambda(x_t'\theta + \alpha)^{y_t}(1 - \Lambda(x_t'\theta + \alpha))^{1-y_t} = 0 \\
&\Leftrightarrow \quad \sum_{y \in \{0,1\}^T} \varphi_y(x, \theta) \prod_{t=1}^{T} \left[\frac{e^{x_t'\theta+\alpha}}{1 + e^{x_t'\theta+\alpha}}\right]^{y_t} \left[\frac{1}{1 + e^{x_t'\theta+\alpha}}\right]^{1-y_t} = 0 \\
&\Leftrightarrow \quad \sum_{y \in \{0,1\}^T} \varphi_y(x, \theta) e^{\sum_{t=1}^T y_t(x_t'\theta+\alpha)} = 0 \\
&\Leftrightarrow \quad \sum_{y \in \{0,1\}^T} \varphi_y(x, \theta) e^{\sum_{t=1}^T y_t x_t'\theta} e^{\alpha \sum_{t=1}^T y_t} = 0.
\end{aligned}$$

So, as $e^{s\alpha}$, $s=0,\dots,T$, are linearly independent, (18) follows.

S1.2. *Nonlinear Regression Model*

Let us consider the model

$$y_i = m(x_i, \alpha_i, \theta_0) + v_i, \quad i = 1, \dots, N, \tag{S4}$$

where $m(\cdot,\cdot,\cdot)$ is a known $T \times 1$ function. The distribution of v_i given x_i and α_i is known given θ_0, and is independent of α_i. The non-Gaussian random coefficients model is covered as a special case, with $m(x,\alpha,\theta) = a(x,\theta) + B(x,\theta)\alpha$. We take $\mathcal{Y} = \mathbb{R}^T$, $\mathcal{A} \subset \mathbb{R}^q$ (where $q = \dim \alpha$), and $\pi_\alpha = 1$ and $\pi_y = 1$. Last, we let $x \in \mathcal{X}$.

Let $g \in L^1(\mathcal{A}) \cap L^2(\mathcal{A})$. We have

$$[L_{\theta,x} g](y) = \int_{\mathcal{A}} f_{v|x}(y - m(x,\alpha,\theta); \theta) g(\alpha)\, d\alpha. \tag{S5}$$

Let us define the operator

$$\mathcal{F} \circ L_{\theta,x} : L^2(\mathcal{A}) \to L^2(\mathbb{R}^T),$$

where $\mathcal{F}$ is the L^2-Fourier transform (e.g., Yoshida (1971, p. 154)). Taking Fourier transforms in (S5), we obtain, for all $g \in L^1(\mathcal{A}) \cap L^2(\mathcal{A})$,

$$[\mathcal{F}[L_{\theta,x} g]](\xi) = \left(\int_{\mathcal{A}} e^{\sqrt{-1}\xi' m(x,\alpha,\theta)} g(\alpha)\, d\alpha \right) \cdot \Psi_{v|x}(\xi | x; \theta), \quad \xi \in \mathbb{R}^T, \tag{S6}$$

where $\Psi_{v|x} = \mathcal{F} f_{v|x}$ is the conditional characteristic function of v_i given x_i. Here we assume that $\Psi_{v|x}$ is nonvanishing (as in Carrasco and Florens (2009), among other references).

Note that $L^1(\mathcal{A}) \cap L^2(\mathcal{A})$ is dense in $L^2(\mathcal{A})$, and that $\mathcal{F}$ is one-to-one. Hence $L_{\theta,x}$ is surjective if and only if $\mathcal{F} \circ L_{\theta,x}$ is surjective. It thus follows from (S6) that $L_{\theta,x}$ is surjective if and only if

$$\left\{ \xi \mapsto e^{\sqrt{-1}\xi' m(x,\alpha,\theta)}, \alpha \in \mathcal{A} \right\} \tag{S7}$$

is dense in the Hilbert space $L^2(|\Psi_{v|x}(\cdot | x; \theta)|^2)$.

In particular, if $\{m(x,\alpha,\theta), \alpha \in \mathcal{A}\}$ has nonempty interior in $\mathbb{R}^T$, then $L_{\theta,x}$ is generally surjective. As an example, $L_{\theta,x}$ is surjective when $T = q$ and $m(x,\cdot,\theta)$ is one-to one. When $T > q$, surjectivity will hold when $m(x,\cdot,\theta)$ is a space-filling mapping (such as a Peano curve) that maps surjectively $\mathcal{A}$ onto $\mathbb{R}^T$ (or an open ball in $\mathbb{R}^T$).

S1.3. *The Linear Dynamic Panel Data Model*

In this subsection we derive the efficient functional differencing restrictions for Example 1A. We will distinguish two cases.

- Case I: $B(x,\theta)$ does not depend on θ. As an example, the static linear model falls in that category:

$$y_{it} = x_{it}'\beta_0 + \alpha_i + v_{it},$$

where $v_{it}|x_i, \alpha_i \sim N[0, \sigma_0^2]$, with $\theta_0 = (\beta_0', \sigma_0^2)'$, and $B(x,\theta) = (1, \dots, 1)'$ is independent of θ.

- Case II: $B(x,\theta)$ depends on θ. As an example, the dynamic AR(1) model

$$y_{it} = \rho_0 y_{i,t-1} + \alpha_i + v_{it}, \quad t = 1, \dots, T,$$

where $v_{it}|y_{i0}, \alpha_i \sim N[0, \sigma_0^2]$, falls in that category. For simplicity, we have assumed that the initial condition y_{i0} is observed (and is thus the only covariate in the model). To see the correspondence with the general formulation, notice that

$$y_{it} = \rho_0^t y_{i0} + (1 + \rho_0 + \dots + \rho_0^{t-1})\alpha_i + v_{it} + \rho_0 v_{i,t-1} + \dots + \rho_0^{t-1} v_{i1},$$
$$t = 1, \dots, T.$$

So, $\theta_0 = (\rho_0, \sigma_0^2)'$, and

$$B(x,\theta) = (1, 1+\rho, \dots, 1 + \rho + \dots + \rho^{T-1})'.$$

We have the next result, where for conciseness we omit the reference to x and θ_0 throughout.

PROPOSITION S1: *The efficient moment function for θ_k is*

$$\begin{aligned}
\mathbb{S}_k^*(y, x; \theta_0) = {} & \left(\frac{\partial a}{\partial \theta_k} + \frac{\partial B}{\partial \theta_k}\mathbb{E}\big(\alpha_i | Q\Sigma^{-1/2}y\big)\right)' \Sigma^{-1/2} W \Sigma^{-1/2}(y - a) \\
& - \frac{1}{2}\Bigg[(y-a)'\Sigma^{-1/2} W \Sigma^{1/2}\frac{\partial \Sigma^{-1}}{\partial \theta_k}\Sigma^{1/2} W \Sigma^{-1/2}(y-a) \\
& - \operatorname{Tr}\left(\Sigma^{1/2} W \Sigma^{1/2}\frac{\partial \Sigma^{-1}}{\partial \theta_k}\right)\Bigg] \\
& - (y-a)'\Sigma^{-1/2} W \Sigma^{1/2}\frac{\partial \Sigma^{-1}}{\partial \theta_k}\Sigma^{1/2} Q \Sigma^{-1/2} \\
& \times \big(y - a - B\mathbb{E}\big(\alpha_i | Q\Sigma^{-1/2}y\big)\big).
\end{aligned}$$

PROOF: We will need the following result.

LEMMA S1: *$\frac{1}{f_{y|x}} \cdot \overline{\mathcal{R}(L)}$ coincides with the set of all zero-mean functions of $Q\Sigma^{-1/2}y$.*

PROOF: It is easy to show that (S1) implies, for all $g \in \mathcal{G}_\alpha$,

$$\frac{1}{[Lf_\alpha](y)}[Lg](y) = \mathbb{E}\big[g(\alpha_i)|Q\Sigma^{-1/2}y, x\big].$$

The result follows. *Q.E.D.*

By Lemma S1, the efficient moment function with respect to θ_k ($k \in \{1, \dots, \dim\theta\}$) is given by

$$\mathbb{S}_k^*(y, x; \theta_0) = \frac{\partial}{\partial\theta_k}\ln([Lf_\alpha](y)) - \mathbb{E}\left(\frac{\partial}{\partial\theta_k}\ln([Lf_\alpha](y_i))\Big|Q\Sigma^{-1/2}y, x\right). \tag{S8}$$

Now, from (4),

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k}\ln([Lf_\alpha](y)) \\
&\quad = -\frac{1}{2}\frac{\partial}{\partial\theta_k}\ln|\Sigma| + \frac{1}{[Lf_\alpha](y)} \\
&\qquad \times \int_{\mathbb{R}^q}\frac{\partial}{\partial\theta_k}\left[-\frac{1}{2}(y-a-B\alpha)'\Sigma^{-1}(y-a-B\alpha)\right] \\
&\qquad \times \exp\left[-\frac{1}{2}(y-a-B\alpha)'\Sigma^{-1}(y-a-B\alpha)\right]f_{\alpha|x}(\alpha|x)\,d\alpha \\
&\quad = -\frac{1}{2}\frac{\partial}{\partial\theta_k}\ln|\Sigma| + \int_{\mathbb{R}^q}\frac{\partial}{\partial\theta_k}\left[-\frac{1}{2}(y-a-B\alpha)'\Sigma^{-1}(y-a-B\alpha)\right] \\
&\qquad \times \widetilde{f}\big(\alpha|Q\Sigma^{-1/2}y, x\big)\,d\alpha,
\end{aligned}$$

where $\widetilde{f}$ denotes the distribution function of α_i given $Q\Sigma^{-1/2}y_i$ and x_i, and where we have used the factorization (S1).

Moreover, we have

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k}\left[-\frac{1}{2}(y-a-B\alpha)'\Sigma^{-1}(y-a-B\alpha)\right] \\
&\quad = \left(\frac{\partial a}{\partial\theta_k} + \frac{\partial B}{\partial\theta_k}\alpha\right)'\Sigma^{-1}(y-a-B\alpha) \\
&\qquad - \frac{1}{2}(y-a-B\alpha)'\frac{\partial\Sigma^{-1}}{\partial\theta_k}(y-a-B\alpha).
\end{aligned}$$

So, by (S8),

$$\mathbb{S}_k^* = \underbrace{\mathbb{E}\left[\left(\frac{\partial a}{\partial \theta_k} + \frac{\partial B}{\partial \theta_k}\alpha_i\right)' \Sigma^{-1}(y - a - B\alpha_i) \Big| Q\Sigma^{-1/2}y\right]}_{=A(y)}$$

$$+ \underbrace{\mathbb{E}\left[-\frac{1}{2}(y - a - B\alpha_i)' \frac{\partial \Sigma^{-1}}{\partial \theta_k}(y - a - B\alpha_i) \Big| Q\Sigma^{-1/2}y\right]}_{=B(y)}$$

$$- \mathbb{E}\big(A(y_i) + B(y_i) | Q\Sigma^{-1/2}y\big).$$

To simplify this expression, note that

$$y - a - B\alpha = \Sigma^{1/2}Q\Sigma^{-1/2}(y - a - B\alpha) + \Sigma^{1/2}W\Sigma^{-1/2}(y - a).$$

So

$$A(y) = \mathbb{E}\left[\left(\frac{\partial a}{\partial \theta_k} + \frac{\partial B}{\partial \theta_k}\alpha_i\right)' \Sigma^{-1}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y - a)\big) \Big| Q\Sigma^{-1/2}y\right]$$
$$+ \widetilde{A}\big(Q\Sigma^{-1/2}y\big)$$

for some function $\widetilde{A}(\cdot)$.

Moreover, note that $Q\Sigma^{-1/2}v_i$ and $W\Sigma^{-1/2}v_i$ are uncorrelated, hence (by normality) independent given x_i. As v_i and α_i are conditionally independent, it follows that $W\Sigma^{-1/2}(y_i - a) = W\Sigma^{-1/2}v_i$ and $Q\Sigma^{-1/2}y_i = Q\Sigma^{-1/2}a + \Sigma^{-1/2}B\alpha_i + Q\Sigma^{-1/2}v_i$ are also independent given x_i.

In particular, this implies that

$$\mathbb{E}\big(W\Sigma^{-1/2}(y_i - a) | Q\Sigma^{-1/2}y, x\big) = \mathbb{E}\big(W\Sigma^{-1/2}(y_i - a) | x\big) = 0,$$

from which it follows that

$$A(y) - \mathbb{E}\big(A(y_i) | Q\Sigma^{-1/2}y\big)$$
$$= \left(\frac{\partial a}{\partial \theta_k} + \frac{\partial B}{\partial \theta_k}\mathbb{E}\big(\alpha_i | Q\Sigma^{-1/2}y\big)\right)' \Sigma^{-1}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y - a)\big).$$

Next, we have

$$B(y) = \mathbb{E}\bigg[-\frac{1}{2}\big(\Sigma^{1/2}Q\Sigma^{-1/2}(y - a - B\alpha_i) + \Sigma^{1/2}W\Sigma^{-1/2}(y - a)\big)'$$
$$\times \frac{\partial \Sigma^{-1}}{\partial \theta_k}\big(\Sigma^{1/2}Q\Sigma^{-1/2}(y - a - B\alpha_i)$$
$$+ \Sigma^{1/2}W\Sigma^{-1/2}(y - a)\big) \Big| Q\Sigma^{-1/2}y\bigg].$$

So

$$
\begin{aligned}
&B(y)\\
&= \underbrace{-\frac{1}{2}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big)'\frac{\partial\Sigma^{-1}}{\partial\theta_k}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big)}_{=B_1(y)}\\
&\quad \underbrace{-\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big)'\frac{\partial\Sigma^{-1}}{\partial\theta_k}\big(\Sigma^{1/2}Q\Sigma^{-1/2}\big(y-a-B\mathbb{E}\big(\alpha_i|Q\Sigma^{-1/2}y\big)\big)\big)}_{=B_2(y)}\\
&\quad + \underbrace{\mathbb{E}\bigg[-\frac{1}{2}\big(\Sigma^{1/2}Q\Sigma^{-1/2}(y-a-B\alpha_i)\big)'\frac{\partial\Sigma^{-1}}{\partial\theta_k}\big(\Sigma^{1/2}Q\Sigma^{-1/2}(y-a-B\alpha_i)\big)\bigg|Q\Sigma^{-1/2}y\bigg]}_{=B_3(y)}.
\end{aligned}
$$

Note that $B_3(y)$ is a function of $Q\Sigma^{-1/2}y$, so

$$
B_3(y) - \mathbb{E}\big(B_3(y_i)|Q\Sigma^{-1/2}y\big) = 0.
$$

Note also that, by the above argument and the law of iterated expectations,

$$
\begin{aligned}
&\mathbb{E}\bigg[\big(\Sigma^{1/2}W\Sigma^{-1/2}(y_i-a)\big)'\\
&\quad \times \frac{\partial\Sigma^{-1}}{\partial\theta_k}\big(\Sigma^{1/2}Q\Sigma^{-1/2}\big(y_i-a-B\mathbb{E}\big(\alpha_i|Q\Sigma^{-1/2}y_i\big)\big)\big)\bigg|Q\Sigma^{-1/2}y\bigg] = 0.
\end{aligned}
$$

So

$$
\begin{aligned}
&B_2(y) - \mathbb{E}\big(B_2(y_i)|Q\Sigma^{-1/2}y\big)\\
&\quad = -\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big)'\\
&\qquad \times \frac{\partial\Sigma^{-1}}{\partial\theta_k}\Sigma^{1/2}Q\Sigma^{-1/2}\big(y-a-B\mathbb{E}\big(\alpha_i|Q\Sigma^{-1/2}y\big)\big).
\end{aligned}
$$

Last, we have

$$
\begin{aligned}
&\mathbb{E}\bigg[(y_i-a)'\Sigma^{-1/2}W\Sigma^{1/2}\frac{\partial\Sigma^{-1}}{\partial\theta_k}\Sigma^{1/2}W\Sigma^{-1/2}(y_i-a)\bigg|Q\Sigma^{-1/2}y\bigg]\\
&\quad = \mathbb{E}\bigg[v_i'\Sigma^{-1/2}W\Sigma^{1/2}\frac{\partial\Sigma^{-1}}{\partial\theta_k}\Sigma^{1/2}W\Sigma^{-1/2}v_i\bigg|Q\Sigma^{-1/2}y\bigg]\\
&\quad = \mathbb{E}\bigg[v_i'\Sigma^{-1/2}W\Sigma^{1/2}\frac{\partial\Sigma^{-1}}{\partial\theta_k}\Sigma^{1/2}W\Sigma^{-1/2}v_i\bigg]
\end{aligned}
$$

$$= \mathrm{Tr}\left[\Sigma^{1/2} W \Sigma^{-1/2} \Sigma \Sigma^{-1/2} W \Sigma^{1/2} \frac{\partial \Sigma^{-1}}{\partial \theta_k} \Sigma^{1/2} W \Sigma^{-1/2}\right]$$
$$= \mathrm{Tr}\left(\Sigma^{1/2} W \Sigma^{1/2} \frac{\partial \Sigma^{-1}}{\partial \theta_k}\right),$$

where we have used that $W\Sigma^{-1/2}v_i$ and $Q\Sigma^{-1/2}y_i$ are independent given x_i, and that $W^2 = W$.

Hence

$$\begin{aligned} & B_1(y) - \mathbb{E}\big(B_1(y_i)|Q\Sigma^{-1/2}y\big) \\ &\quad = -\frac{1}{2}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big)' \frac{\partial \Sigma^{-1}}{\partial \theta_k}\big(\Sigma^{1/2}W\Sigma^{-1/2}(y-a)\big) \\ &\qquad + \frac{1}{2}\,\mathrm{Tr}\left(\Sigma^{1/2}W\Sigma^{1/2}\frac{\partial \Sigma^{-1}}{\partial \theta_k}\right). \end{aligned}$$

The expression of $\mathbb{S}_k^*$ then follows from combining the results. *Q.E.D.*

Several comments are in order. To start with, consider the case where $B(x, \theta)$ does not depend on θ (Case I). Proposition S1 shows that, in this case, the bound coincides with the standard bound for exponential family models (Hahn (1997)).

In the general case (Case II), the efficient functional differencing restrictions comprise five terms. Mean and covariance restrictions in quasi-differences are

$$\mathbb{E}\left[\left(\frac{\partial a}{\partial \theta_k}\right)' \Sigma^{-1/2} W \Sigma^{-1/2}(y_i - a)\right] = 0$$

and

$$\begin{aligned} \mathbb{E}\Bigg[& (y_i - a)'\Sigma^{-1/2} W \Sigma^{1/2} \frac{\partial \Sigma^{-1}}{\partial \theta_k} \Sigma^{1/2} W \Sigma^{-1/2}(y_i - a) \\ & - \mathrm{Tr}\left(\Sigma^{1/2} W \Sigma^{1/2} \frac{\partial \Sigma^{-1}}{\partial \theta_k}\right)\Bigg] = 0, \end{aligned}$$

respectively, whereas orthogonality restrictions between quasi-differences and levels are

$$\mathbb{E}\left[(y_i - a)'\Sigma^{-1/2} W \Sigma^{1/2} \frac{\partial \Sigma^{-1}}{\partial \theta_k} \Sigma^{1/2} Q \Sigma^{-1/2}(y_i - a)\right] = 0.$$

In the dynamic model, these restrictions combine those proposed by Arellano and Bond (1991) and Ahn and Schmidt (1995). Note that, in this model, no stationarity restriction is imposed, so the extra moments of Arellano and Bover (1995) are invalid here. Imposing these restrictions would require constraining the distribution of α_i given y_{i0}.

The last two terms involve the conditional mean $\mathbb{E}(\alpha_i|Q\Sigma^{-1/2}y_i)$. It is interesting to compare the optimal mean restrictions in the normal model,

$$\mathbb{E}\left[\left(\frac{\partial a}{\partial\theta_k}+\frac{\partial B}{\partial\theta_k}\mathbb{E}\big(\alpha_i|Q\Sigma^{-1/2}y_i\big)\right)'\Sigma^{-1/2}W\Sigma^{-1/2}(y_i-a)\right]=0,$$

with the optimal restrictions obtained by Chamberlain (1992), in a model where second- and higher-order moments of errors are left unrestricted,

$$\mathbb{E}\left[\left(\frac{\partial a}{\partial\theta_k}+\frac{\partial B}{\partial\theta_k}\mathbb{E}(\alpha_i|x_i)\right)'\Sigma^{-1/2}W\Sigma^{-1/2}(y_i-a)\right]=0.$$

The difference between the two sets of moments is that, under normality, $Q\Sigma^{-1/2}y_i=Q\Sigma^{-1/2}a+\Sigma^{-1/2}B\alpha_i+Q\Sigma^{-1/2}v_i$ is statistically independent of $W\Sigma^{-1/2}(y_i-a)=W\Sigma^{-1/2}v_i$ given covariates x_i. This implies the existence of additional instruments (namely, $Q\Sigma^{-1/2}y_i$) in the quasi-differenced equations, in addition to x_i.

Note that efficient estimation of θ_0 in the normal model requires a nonparametric estimate of $\mathbb{E}(\alpha_i|Q\Sigma^{-1/2}y_i)$. This is similar to efficient estimation of common parameters in Chamberlain (1992). A regularized estimate of the optimal instruments may be based on a (semi-) parametric random-effects specification, as discussed in the text.

S1.4. *Uniform Fourier Convergence in the Random Coefficients Model*

We consider Example 1A, where, in addition, we assume that Σ is *known*. We also assume that $\operatorname{rank}(B)=\dim\alpha\equiv q$, that is, that $L_{\theta,x}$ is injective.

Let us take $\pi_\alpha=1$, and $\pi_y(y)=\exp[-\frac{1}{2}\eta y'\Sigma^{-1}y]$, where $\eta>0$. Let $Q=\Sigma^{-1/2}B[\Sigma^{-1/2}B]^{\dagger}$, and define V a $T\times q$ matrix such that $Q=VV'$ and $V'V=I_q$. Let also $W=I_T-Q$, and define U a $T\times(T-q)$ matrix such that $W=UU'$, and $U'U=I_{T-q}$.

Let us define $\mathcal{H}$ the Hilbert space of functions $\psi:\mathbb{R}^q\to\mathbb{R}$ such that

$$\int_{\mathbb{R}^q}\psi(\mu)^2\exp\left[-\frac{1}{2}\eta\mu'\mu\right]d\mu<\infty,$$

endowed with its canonical scalar product. Last, let $L_{\mathcal{H}}:\mathcal{H}\to\mathcal{H}$ be the integral operator such that, for all $\psi\in\mathcal{H}$,

$$[L_{\mathcal{H}}\psi](z)=\int_{\mathbb{R}^q}\exp\left[-\frac{1}{4}(z-\mu)'(z-\mu)\right]\times\exp\left[-\frac{1}{2}\eta\mu'\mu\right]\psi(\mu)\,d\mu$$

for all $z\in\mathbb{R}^q$.

We note that $L_{\mathcal{H}}$ is Hilbert–Schmidt, so it admits a singular value decomposition, and that $L_{\mathcal{H}}$ is self-adjoint.

We have the following result.

PROPOSITION S2: *The left singular functions of the operator* $L_{\theta,x}:\mathcal{G}_\alpha \to \mathcal{G}_y$ *are given by*

$$\phi_{j,\theta}(y) = C(\theta) H_j\big(V'\Sigma^{-1/2}y\big)\exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right], \tag{S9}$$

where H_j, $j=1,2,\dots$ *are the singular functions of the self-adjoint operator* $L_{\mathcal{H}}$, *and where* $C(\theta)$ *is a positive constant, uniformly bounded on* Θ *provided* $a(\cdot)$ *is continuous in* θ *and* Θ *is compact*.

PROOF: Let $\mathcal{Y}=\mathbb{R}^T$, and $\mathcal{A}=\mathbb{R}^q$. We have

$$\begin{aligned}
&[L_{\theta,x}L^*_{\theta,x}h](y) \\
&\quad = \int_{\mathcal{Y}}\int_{\mathcal{A}} f_{y|x,\alpha}(y|x,\alpha;\theta) f_{y|x,\alpha}(\widetilde{y}|x,\alpha;\theta)\pi_y(\widetilde{y})h(\widetilde{y})\,d\alpha\,d\widetilde{y} \\
&\quad = \int_{\mathcal{Y}} \underbrace{\left\{\int_{\mathcal{A}} f_{y|x,\alpha}(y|x,\alpha;\theta) f_{y|x,\alpha}(\widetilde{y}|x,\alpha;\theta)\,d\alpha\right\}}_{k(y,\widetilde{y})} \pi_y(\widetilde{y})h(\widetilde{y})\,d\widetilde{y}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
f_{y|x,\alpha}(y|x,\alpha;\theta) \propto \exp\Big[&-\frac{1}{2}\big(V'\Sigma^{-1/2}(y-a)-V'\Sigma^{-1/2}B\alpha\big)' \\
&\times \big(V'\Sigma^{-1/2}(y-a)-V'\Sigma^{-1/2}B\alpha\big)\Big] \\
&\times \exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right],
\end{aligned}$$

where $A\propto B$ denotes the fact that A and B are equal up to a multiplicative constant (possibly dependent on θ, x).

Using the change of variables $\beta = V'\Sigma^{-1/2}B\alpha$, and noting that $V'\Sigma^{-1/2}B$ is nonsingular, we obtain

$$\begin{aligned}
k(y,\widetilde{y}) &= \int_{\mathcal{A}} f_{v|x}(y-a-B\alpha;\theta) f_{v|x}(\widetilde{y}-a-B\alpha;\theta)\,d\alpha \\
&\propto \int_{\mathcal{A}} \exp\Big[-\frac{1}{2}\big(V'\Sigma^{-1/2}(y-a)-\beta\big)'\big(V'\Sigma^{-1/2}(y-a)-\beta\big) \\
&\qquad -\frac{1}{2}\big(V'\Sigma^{-1/2}(\widetilde{y}-a)-\beta\big)'\big(V'\Sigma^{-1/2}(\widetilde{y}-a)-\beta\big)\Big]\,d\beta
\end{aligned}$$

$$\times \exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right.$$
$$\left.-\frac{1}{2}(\widetilde{y}-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(\widetilde{y}-a)\right].$$

So, from the usual decomposition of quadratic forms,

$$k(y,\widetilde{y}) \propto \exp\left[-\frac{1}{4}\big(V'\Sigma^{-1/2}(y-\widetilde{y})\big)'(V'\Sigma^{-1/2}(y-\widetilde{y}))\right]$$
$$\times \exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right.$$
$$\left.-\frac{1}{2}(\widetilde{y}-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(\widetilde{y}-a)\right].$$

As the left singular function $\phi_{j,\theta}$ belongs to the range of $L_{\theta,x}$, there exists a function h_j such that

$$\phi_{j,\theta}(y) = h_j\big(V'\Sigma^{-1/2}y\big)\exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right].$$

The function $\phi_{j,\theta}$ satisfies

$$[L_{\theta,x}L^*_{\theta,x}\phi_{j,\theta}](y) \propto \phi_{j,\theta}(y).$$

This is equivalent to

$$h_j\big(V'\Sigma^{-1/2}y\big) \propto \int_y \left\{\exp\left[-\frac{1}{4}\big(V'\Sigma^{-1/2}(y-\widetilde{y})\big)'\big(V'\Sigma^{-1/2}(y-\widetilde{y})\big)\right]\right.$$
$$\times \exp\left[-\frac{1}{2}(\widetilde{y}-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(\widetilde{y}-a)\right]$$
$$\left.\times\, \pi_y(\widetilde{y})h_j\big(V'\Sigma^{-1/2}\widetilde{y}\big)\right\} d\widetilde{y}.$$

Then we note that, as $VV' + UU' = I_T$,

$$\pi_y(\widetilde{y}) = \exp\left[-\frac{1}{2}\eta\widetilde{y}'\Sigma^{-1}\widetilde{y}\right]$$
$$= \exp\left[-\frac{1}{2}\eta\big(V'\Sigma^{-1/2}\widetilde{y}\big)'V'\Sigma^{-1/2}\widetilde{y}\right]$$
$$\times \exp\left[-\frac{1}{2}\eta\widetilde{y}'\Sigma^{-1/2}UU'\Sigma^{-1/2}\widetilde{y}\right].$$

We thus obtain, using the change in variables $(\mu,\nu)=(V'\Sigma^{-1/2}\widetilde{y},U'\Sigma^{-1/2}\widetilde{y})$,

$$h_j(V'\Sigma^{-1/2}y)\propto\int_{\mathbb{R}^q}\exp\left[-\frac{1}{4}(V'\Sigma^{-1/2}y-\mu)'(V'\Sigma^{-1/2}y-\mu)\right]$$
$$\times\exp\left[-\frac{1}{2}\eta\mu'\mu\right]h_j(\mu)\,d\mu.$$

So, (S9) follows. Last, as $\|\phi_{j,\theta}\|=1$, the proportionality constant $C(\theta)$ satisfies

$$\begin{aligned}\frac{1}{C(\theta)^2}&=\int_y\left(H_j(V'\Sigma^{-1/2}y)\exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right]\right)^2\\&\quad\times\exp\left[-\frac{1}{2}\eta y'\Sigma^{-1}y\right]dy\\&=|\Sigma|^{1/2}\int_{\mathbb{R}^q}H_j(\mu)^2\exp\left[-\frac{1}{2}\eta\mu'\mu\right]d\mu\\&\quad\times\int_{\mathbb{R}^{T-q}}\exp\left[-(\nu-U'\Sigma^{-1/2}a)'(\nu-U'\Sigma^{-1/2}a)\right]\\&\quad\times\exp\left[-\frac{1}{2}\eta\nu'\nu\right]d\nu\\&=|\Sigma|^{1/2}\left(\frac{2\pi}{2+\eta}\right)^{(T-q)/2}\exp\left[-\frac{\eta}{2+\eta}a'\Sigma^{-1/2}UU'\Sigma^{-1/2}a\right],\end{aligned}$$

where we have used that $\|H_j\|=1$. As $a(\cdot)$ is continuous in θ and Θ is compact, and as $W=UU'$ is a projector, $a'\Sigma^{-1/2}UU'\Sigma^{-1/2}a$ is bounded. So, $C(\theta)$ is uniformly bounded.

The result follows. *Q.E.D.*

Using the expression for the left singular functions, we then verify uniform Fourier convergence for model (2).

COROLLARY S1: *The following condition is satisfied for all $h\in\mathcal{G}_y$, a.s. in x*:

$$\text{(S10)}\qquad \sup_{\theta\in\Theta}\left(\sum_{j>J}\langle\phi_{j,\theta},h\rangle^2\right)\overset{J\to\infty}{\to}0.$$

PROOF: We start by checking condition (S10) when h is a polynomial. It is enough to check the result for h of the form $(\Sigma^{-1/2}y)^{(k)}$, where $y^{(k)}=y_1^{k_1}\times\cdots\times y_T^{k_T}$. Let $(\mu,\nu)=(V'\Sigma^{-1/2}y,U'\Sigma^{-1/2}y)$. We have

$$(\Sigma^{-1/2}y)^{(k)}=(VV'\Sigma^{-1/2}y+UU'\Sigma^{-1/2}y)^{(k)}=(V\mu+U\nu)^{(k)}.$$

We note that $(V\mu + U\nu)^{(k)}$ is a polynomial in μ and ν, the coefficients of which are uniformly bounded as U and V are orthogonal matrices. So it is sufficient to check the result for h of the form $(V'\Sigma^{-1/2}y)^{(m)}(U'\Sigma^{-1/2}y)^{(\ell)}$.

For such an h, we have

$$\begin{aligned}
\langle\phi_{j,\theta}, h\rangle = C(\theta) \int_y \Big\{ & \big(V'\Sigma^{-1/2}y\big)^{(m)} \big(U'\Sigma^{-1/2}y\big)^{(\ell)} H_j\big(V'\Sigma^{-1/2}y\big) \\
& \times \exp\left[-\frac{1}{2}(y-a)'\Sigma^{-1/2}UU'\Sigma^{-1/2}(y-a)\right]\pi_y(y)\Big\}\, dy \\
= C(\theta)|\Sigma|^{1/2} & \int_{\mathbb{R}^q} \mu^{(m)} H_j(\mu) \exp\left[-\frac{1}{2}\eta\mu'\mu\right] d\mu \\
& \times \int_{\mathbb{R}^{T-q}} \nu^{(\ell)} \exp\left[-\frac{1}{2}\big(\nu - U'\Sigma^{-1/2}a\big)'\big(\nu - U'\Sigma^{-1/2}a\big)\right] \\
& \times \exp\left[-\frac{1}{2}\eta\nu'\nu\right] d\nu,
\end{aligned}$$

where we have factored π_y as in the proof of Proposition S2, and where we have used the change in variables $(\mu, \nu) = (V'\Sigma^{-1/2}y, U'\Sigma^{-1/2}y)$.

Now, as $\mu^{(m)}$ belongs to $\mathcal{H}$,

$$\sum_{j>J} \left(\int_{\mathbb{R}^q} \mu^{(m)} H_j(\mu) \exp\left[-\frac{1}{2}\eta\mu'\mu\right] d\mu \right)^2 \overset{J\to\infty}{\to} 0.$$

In addition,

$$\begin{aligned}
& \left| \int_{\mathbb{R}^{T-q}} \nu^{(\ell)} \exp\left[-\frac{1}{2}\big(\nu - U'\Sigma^{-1/2}a\big)'\big(\nu - U'\Sigma^{-1/2}a\big)\right] \exp\left[-\frac{1}{2}\eta\nu'\nu\right] d\nu \right| \\
& \quad \le \int_{\mathbb{R}^{T-q}} |\nu|^{(\ell)} \exp\left[-\frac{1}{2}\eta\nu'\nu\right] d\nu < \infty.
\end{aligned}$$

This shows uniform Fourier convergence for polynomial h.

Last, let $h \in \mathcal{G}_y$, and fix $\varepsilon > 0$. We start by noting that polynomials are dense in $\mathcal{G}_y$. For example, when $T = 1$, the (generalized) Hermite polynomials form an orthogonal basis of the weighted L^2 space $\mathcal{G}_y$. So, there exists a polynomial $\widetilde{h}$ such that $\|h - \widetilde{h}\|^2 < \frac{\varepsilon}{4}$.

For this $\widetilde{h}$, and by the previous result, there exists a J_1 such that, for all $J \ge J_1$,

$$\sup_{\theta\in\Theta} \sum_{j>J} \langle\phi_{j,\theta}, \widetilde{h}\rangle^2 < \frac{\varepsilon}{4}.$$

Therefore,

$$
\begin{aligned}
\sup_{\theta\in\Theta}\sum_{j>J}\langle\phi_{j,\theta},h\rangle^2 &\le \sup_{\theta\in\Theta}\sum_{j>J}2\big(\langle\phi_{j,\theta},\widetilde{h}\rangle^2+\langle\phi_{j,\theta},h-\widetilde{h}\rangle^2\big)\\
&\le 2\times\sup_{\theta\in\Theta}\sum_{j>J}\langle\phi_{j,\theta},\widetilde{h}\rangle^2+2\times\|h-\widetilde{h}\|^2\\
&< 2\times\frac{\varepsilon}{4}+2\times\frac{\varepsilon}{4}\\
&=\varepsilon,
\end{aligned}
$$

and the corollary is proved. *Q.E.D.*

S2. EXAMPLES: NUMERICAL EVIDENCE

In this section, we discuss several numerical issues. We start by describing an approach to numerically compute semiparametric information bounds.

S2.1. *Computing Information Bounds*

General Approach

Let $\pi_\alpha = 1/f_{\alpha|x}$, and $\pi_y = 1/f_{y|x}$. Here $f_{\alpha|x}$ and $f_{y|x}$ are assumed known. The efficient moment restrictions are given by (30). Moreover, the information bound, conditional on $x_1,\dots,x_N$, is given by

$$
\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\pi_y(y_i)^2\left[W_{\theta_0,x_i}\frac{\partial L_{\theta_0,x_i}f_{\alpha|x}}{\partial\theta}\right](y_i)\left[W_{\theta_0,x_i}\frac{\partial L_{\theta_0,x_i}f_{\alpha|x}}{\partial\theta'}\right](y_i)\,\Big|\,x_i\right). \tag{S11}
$$

Note that the bound depends on parameter values θ_0 and $f_{\alpha|x}$. Note also that the information bound simplifies relative to the general formula for the asymptotic variance of GMM, due to information equality.

Let

$$
\begin{aligned}
\mathcal{E}_{\theta,x} = \mathbb{E}\Bigg(&\pi_y(y_i)^2\left[W_{\theta_0,x_i}\frac{\partial L_{\theta_0,x_i}f_{\alpha|x}}{\partial\theta}\right](y_i)\\
&\times\left[W_{\theta_0,x_i}\frac{\partial L_{\theta_0,x_i}f_{\alpha|x}}{\partial\theta'}\right](y_i)\,\Big|\,x_i=x\Bigg).
\end{aligned}
$$

We will compute $\mathcal{E}_{\theta,x}$ by simulation, drawing N_y values $\underline{y}_s$ $(s=1,\dots,N_y)$ from $f_{y|x}(\cdot|x)$. That is,

$$
\mathcal{E}_{\theta,x}\approx\frac{1}{N_y}\sum_{s=1}^{N_y}\pi_y(\underline{y}_s)^2\left[W_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta}\right](\underline{y}_s)\left[W_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta'}\right](\underline{y}_s). \tag{S12}
$$

Let us write $W_{\theta_0,x} = I_y - Q_{\theta_0,x}$, where $Q_{\theta_0,x}$ is the orthogonal projection operator onto $\overline{\mathcal{R}(L_{\theta_0,x})}$. Equation (S12) may equivalently be written as

$$\begin{aligned}\mathcal{E}_{\theta,x} \approx \frac{1}{N_y}\sum_{s=1}^{N_y}\Bigg(&\frac{\partial \ln[L_{\theta_0,x}f_{\alpha|x}](\underline{y}_s)}{\partial\theta} \\ &- \frac{1}{[L_{\theta_0,x}f_{\alpha|x}](\underline{y}_s)}\left[Q_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta}\right](\underline{y}_s)\Bigg) \\ &\times\left(\frac{\partial \ln[L_{\theta_0,x}f_{\alpha|x}](\underline{y}_s)}{\partial\theta} - \frac{1}{[L_{\theta_0,x}f_{\alpha|x}](\underline{y}_s)}\left[Q_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta}\right](\underline{y}_s)\right)'.\end{aligned}$$

Now let us define the $N_\alpha \times 1$ and $N_y \times 1$ vectors

$$\underline{f}^{(y)}_{\theta_0,x} = \big[(f_{y|x,\alpha}(y|x,\underline{\alpha}_n;\theta_0))_n\big],$$

$$\frac{\partial \underline{L}_{\theta_0,x}\underline{f}_{\alpha|x}}{\partial\theta_k} = \left[\left(\sum_{n=1}^{N_\alpha}\frac{\partial f_{y|x,\alpha}(\underline{y}_s|x,\underline{\alpha}_n;\theta_0)}{\partial\theta_k}\right)_s\right],$$

and the $N_y \times N_\alpha$ and $N_y \times N_y$ matrices

$$\underline{L}_{\theta_0,x} = \big[(f_{y|x,\alpha}(\underline{y}_s|x,\underline{\alpha}_n;\theta_0))_{s,n}\big],$$

$$\underline{D}_{\theta_0,x} = \operatorname{diag}\left[\left(\sum_{n=1}^{N_\alpha} f_{y|x,\alpha}(\underline{y}_s|x,\underline{\alpha}_n;\theta_0)\right)_s\right].$$

We shall adopt a discretization strategy closely related to that of Section 6 in the paper, using N_y draws $\underline{y}_s$ from $f_{y|x}$, and N_α draws $\underline{\alpha}_n$ from $\overline{\pi} = f_{\alpha|x}$. We start by noting that it follows from the choice of π_α that $\overline{\pi}\cdot\pi_\alpha = 1$, and that $[(\mu_s(y))_s] \approx \frac{1}{N_\alpha}\underline{L}_{\theta,x}\underline{f}^{(y)}_{\theta,x}$. We will thus approximate $[Q_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta_k}](y)$ in the family of functions

$$\left[Q_{\theta_0,x}\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta_k}\right](y) \approx \sum_{n=1}^{N_\alpha} b_n \underbrace{f_{y|x,\alpha}(y|x,\underline{\alpha}_n;\theta_0)}_{\equiv \nu_n(y)}.$$

The projection yields

$$\begin{aligned}[(b_n)_n] \approx{}& \left[\left(\int_{\mathcal{Y}}\nu_{n_1}(y)\nu_{n_2}(y)\pi_y(y)\,dy\right)_{n_1,n_2}\right]^{\dagger} \\ &\times\left[\left(\int_{\mathcal{Y}}\nu_{n_1}(y)\left[\frac{\partial L_{\theta_0,x}f_{\alpha|x}}{\partial\theta_k}\right](y)\pi_y(y)\,dy\right)_{n_1}\right],\end{aligned}$$

that is, using importance sampling with N_y draws from $f_{y|x} = 1/\pi_y$,

$$[(b_n)_n] \approx \left[\left(\sum_{s=1}^{N_y} \frac{1}{f_{y|x}(\underline{y}_s|x)^2} \nu_{n_1}(\underline{y}_s)\nu_{n_2}(\underline{y}_s) \right)_{n_1,n_2} \right]^{\dagger} \times \left[\left(\sum_{s=1}^{N_y} \frac{1}{f_{y|x}(\underline{y}_s|x)^2} \nu_{n_1}(\underline{y}_s) \left[\frac{\partial L_{\theta_0,x} f_{\alpha|x}}{\partial \theta_k} \right](\underline{y}_s) \right)_{n_1} \right].$$

Combining, and noting that $[\frac{\partial L_{\theta_0,x} f_{\alpha|x}}{\partial \theta_k}](\underline{y}_s) \approx \frac{1}{N_\alpha} \sum_{n=1}^{N_\alpha} \frac{\partial f_{y|x,\alpha}(\underline{y}_s|x,\underline{\alpha}_n;\theta_0)}{\partial \theta_k}$, we obtain

$$\begin{aligned}
&\left[Q_{\theta_0,x} \frac{\partial L_{\theta_0,x} f_{\alpha|x}}{\partial \theta_k} \right](y) \\
&\quad \approx \frac{1}{N_\alpha} \big(\underline{f}^{(y)}_{\theta_0,x}\big)' \big(\underline{L}'_{\theta_0,x} \underline{D}^{-2}_{\theta_0,x} \underline{L}_{\theta_0,x}\big)^{\dagger} \underline{L}'_{\theta_0,x} \underline{D}^{-2}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta_k} \\
&\quad = \frac{1}{N_\alpha} \big(\underline{f}^{(y)}_{\theta_0,x}\big)' \big(\underline{D}^{-1}_{\theta_0,x} \underline{L}_{\theta_0,x}\big)^{\dagger} \underline{D}^{-1}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta_k}.
\end{aligned}$$

Using this approximation, we thus have[46]

$$\begin{aligned}
\mathcal{E}_{\theta,x} \approx{}& \frac{1}{N_y} \left(\underline{D}^{-1}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta'} \right)' \big[I_{N_y} - \big(\underline{D}^{-1}_{\theta_0,x} \underline{L}_{\theta_0,x}\big)\big(\underline{D}^{-1}_{\theta_0,x} \underline{L}_{\theta_0,x}\big)^{\dagger} \big] \\
&\times \left(\underline{D}^{-1}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta'} \right).
\end{aligned} \tag{S13}$$

This formula is intuitive: not knowing $f_{\alpha|x}$ results in a loss of information, which can be seen by comparing the fixed-effects information bound $\mathcal{E}_{\theta,x}$ with the parametric information bound for $f_{\alpha|x}$ known,

$$\mathcal{E}^{ML}_{\theta,x} \approx \frac{1}{N_y} \left(\underline{D}^{-1}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta'} \right)' \left(\underline{D}^{-1}_{\theta_0,x} \frac{\partial \underline{L}_{\theta_0,x} \underline{f}_{\alpha|x}}{\partial \theta'} \right).$$

Illustrative Calculations

Figure S1 shows the results of a computation of the inverse information bound for three models, using the formula (S13). We use two different approaches to compute (S13). The first one is based on the J-modified pseudo-inverse of the matrix $\underline{L}_{\theta_0,x}$, which we mentioned in the text (Section 6).

[46]Note that this formula differs slightly from the one that appears in the proof of Theorem 1. This is because in the above derivations we have used N_y draws from $f_{y|x}$ (while the proof of the theorem relies on draws from a discrete uniform distribution).

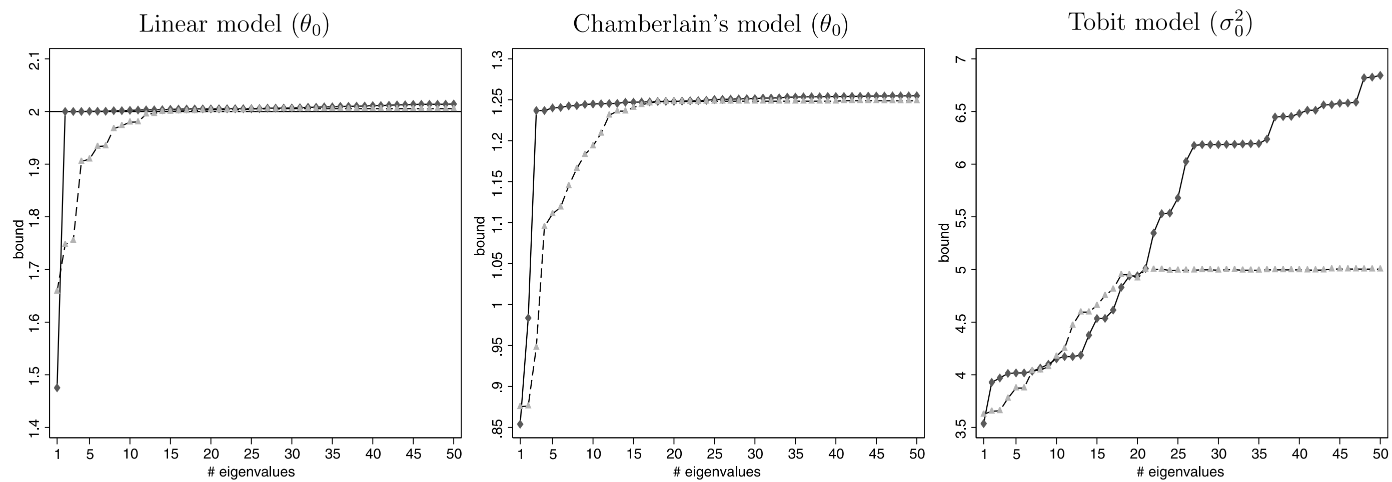

FIGURE S1.—Inverse information bounds ($T = 2$). The curves show the inverse information bound of the model, for a given number of singular values (or columns of $\underline{L}_\theta$) J used in the computation. Diamonds (solid line) correspond to the SVD-based approach; triangles (dashed line) correspond to the QR-based approach.

We refer to this approach as singular value decomposition (SVD) -based. As a second approach, we use a least-squares (LS) regression to compute $(\underline{D}_{\theta_0,x}^{-1}\underline{L}_{\theta_0,x})^{\dagger}\underline{D}_{\theta_0,x}^{-1}\frac{\partial \underline{L}_{\theta_0,x}\underline{f}_{\alpha|x}}{\partial \theta_k}$, using only the first J columns of the matrix $\underline{L}_{\theta_0,x}$. We refer to this method as QR-based.[47]

The left panel shows the results for a simple linear model $y_{it} = \alpha_i + \theta_0(t-1) + v_{it}$, $t = 1, 2$, where v_{i1} and v_{i2} are independent standard normals. We take $\theta_0 = 1$, and α_i standard normal. In this case, the first-differenced estimator reaches the information bound (e.g., Hahn (1997)), the inverse of which is equal to 2. This provides a convenient benchmark against which to compare our discretization strategy. We take $N_y = 10{,}000$, and $N_\alpha = 1000$.

We see on the left panel that both methods converge to the theoretical bound. The SVD-based method gets faster to the solution. However, when J gets large, the SVD-based method tends to deviate from the theoretical value, while the LS-based method remains stable. The difference is accentuated when we increase J beyond 50 (not shown).

The central panel shows the results for Chamberlain's model, with a similar pattern for the two methods. This calculation suggests that the inverse information bound is ≈ 1.25. That is, when $N = 100$ or 500, the corresponding standard deviations are $\approx\sqrt{1.25/100} = .112$ and $\approx\sqrt{1.25/500} = .050$, respectively.

For the Tobit model, we see that the numerical problems in the computation of singular vectors arise earlier. In particular, the SVD-based method starts to diverge from the QR-based method when $J \geq 20$. We verified that, for small singular values, the numerical computation of singular vectors starts being very imprecise. In contrast, the QR-based method gives more stable results. This computation exercise shows that precise calculation of information bounds may be difficult, due to errors caused by finite machine precision. Nevertheless, the evidence obtained suggests that the inverse bound is ≈ 5.0, and that the corresponding standard deviations for σ_0 (obtained using the delta method) are $\approx\sqrt{5/100}/2 = .112$ and $\approx\sqrt{5/500}/2 = .050$, respectively.[48]

S2.2. *Checking the Non-Surjectivity Condition*

A similar discretization approach may also be used to provide numerical evidence on (non-) surjectivity in a given model. To proceed, let us return to the setup of Section 6, where π_y is integrable, and integrals with respect to α are approximated using importance sampling based on a density $\overline{\pi}$. Fix $x \in \mathcal{X}$, and $\theta \in \Theta$. For any $h \in \mathcal{G}_y$, the squared norm of $W_{\theta,x}h$ may be computed as

$$\|W_{\theta,x}h\|^2 \approx \frac{1}{N_y}\underline{h}'\left[I_{N_y} - \underline{L}_{\theta,x}\underline{L}_{\theta,x}^{\dagger}\right]\underline{h}, \tag{S14}$$

[47] In Matlab, linear regression based on the QR decomposition can be done using the backslash operator.

[48] The fact that the bounds are numerically equal in the two models is due to chance.

where $\underline{h} = [(h(\underline{y}_s))_s]$ is an $N_y \times 1$ vector and where the expression of the $N_y \times N_\alpha$ matrix $\underline{L}_{\theta,x}$ is given in Section 6.

To illustrate the practical usefulness of this type of calculation, we computed (S14) in the Tobit model and in the simple random coefficients model, for the special choice $h(y) = 1$. Note that, given that π_y is integrable, h belongs to $\mathcal{G}_y$ and its norm is equal to 1. We take $\theta = 1$ ($\sigma = 1$ for tobit), and we show the results for $T = 1$, $T = 2$, and $T = 3$. We also compare the SVD-based and QR-based approaches to approximate the projection, as we explained above.

The results presented in Figure S2 clearly show that $W_\theta h \neq 0$ when $T = 2$ or $T = 3$. This provides numerical evidence on the fact that the operator L_θ is not surjective for that value of common parameters. In contrast, the graphs for $T = 1$ suggest that $W_\theta h = 0$, consistently with the fact that L_θ is surjective when only one period of data is used. The figure thus illustrates the fact that the availability of panel data is essential to the success of the functional differencing approach.

S2.3. *Varying the Number of Singular Values*

In the two upper panels of Figure S3, we show the mean of $\widehat{\sigma}$ and $\widehat{\theta}$, as well as the mean $\pm$ two standard deviations, across 1000 simulations, for a sample size $N = 100$. In both models, we set $h_r(y) = \phi(y - \mu_r)$, where ϕ is the standard normal p.d.f. and where μ_r takes 49 different values in $\mathbb{R}^2$:

$$\{(0,0), (0,1), (0,-1), (0,2), (0,-2), (0,3), (0,-3), \dots, (-3,-3)\}.$$

On the x-axis of the figure, we report the number of singular values J used in the numerical computation of the discretized version of the within projection operator (using the SVD-based approach). We see that the results quickly stabilize around the true value ($\sigma_0 = 1$ and $\theta_0 = 1$, respectively). This result is consistent with the absence of ill-posedness in the estimation of common parameters.[49]

S2.4. *Numerical Evidence on Uniform Fourier Convergence*

Here our aim is to provide some numerical evidence on uniform Fourier convergence in the two models that we used as illustration in Section 6 in the paper. In Section 5, we assumed uniform Fourier convergence to show root-N consistency and asymptotic normality of common parameter estimates. In

[49] We also compared the SVD-based approach with the QR-based approach which we mentioned above, and found little difference on the simulated data.

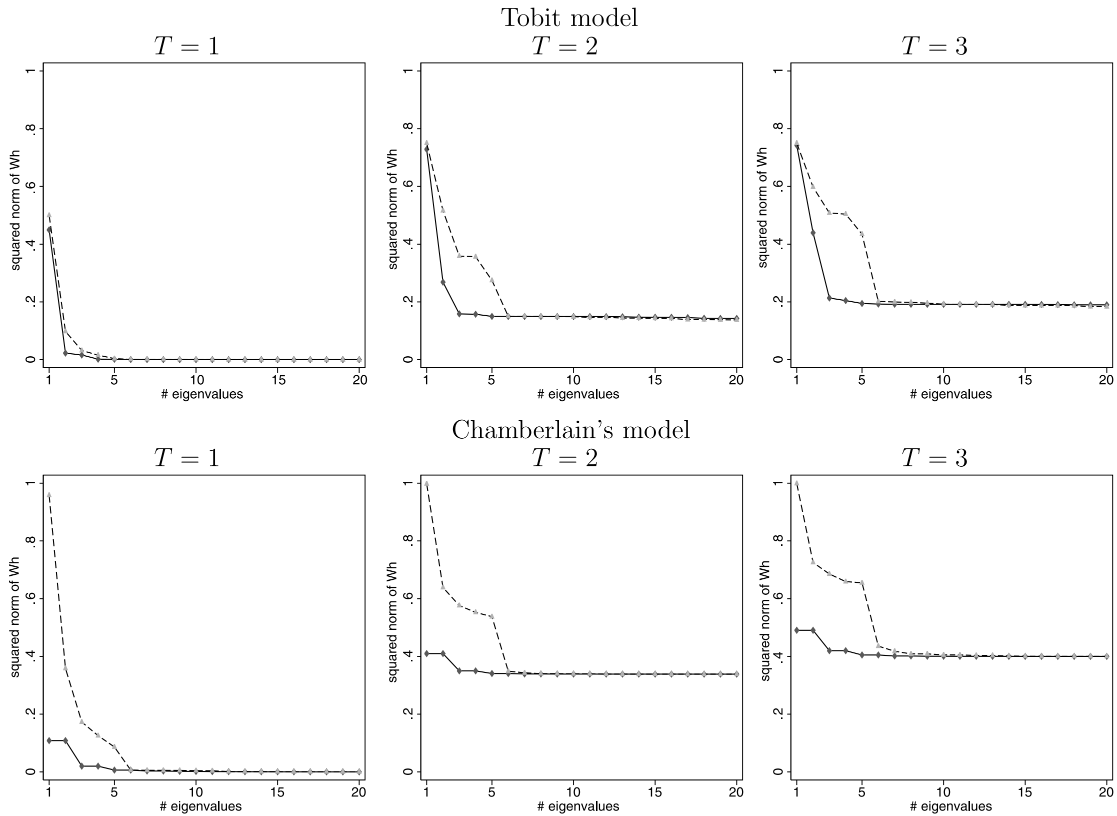

FIGURE S2.—Numerical evidence on non-surjectivity ($T = 1, 2, 3$). The curves show estimates of the squared norm of $W_\theta h$, where $h(y) = 1$ and $(\theta, \sigma) = (1, 1)$. The results are plotted against the number of singular values (or columns of $\underline{L}_\theta$) J used in the computation. Diamonds (solid line) correspond to the SVD-based approach; triangles (dashed line) correspond to the QR-based approach.

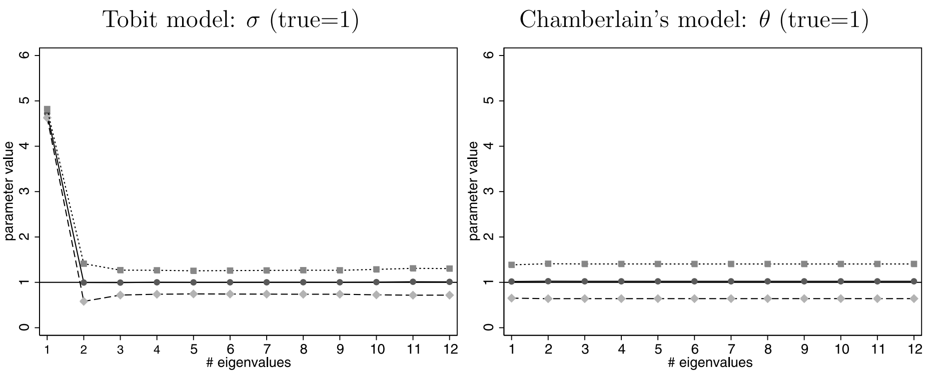

FIGURE S3.—Parameter estimates ($N = 100$, $T = 2$). On the x-axis we report the number of singular values used in estimation, while the y-axis shows parameter estimates. The functions used to construct moment functions are $\phi(\cdot - \mu_r)$, $r = 1, \dots, 49$, where the set of values for μ_r is indicated in the text. The solid and discontinuous lines show the mean estimate and the mean $\pm$ 2 standard deviations, respectively. The thin solid line indicates the true parameter value.

Figure S4, we report the sum $\sum_{j>J} \langle \phi_{j,\theta}, f_y \rangle^2$, for various J and for common parameters (θ and σ) in a grid of values ranging between .5 and 1.5.[50]

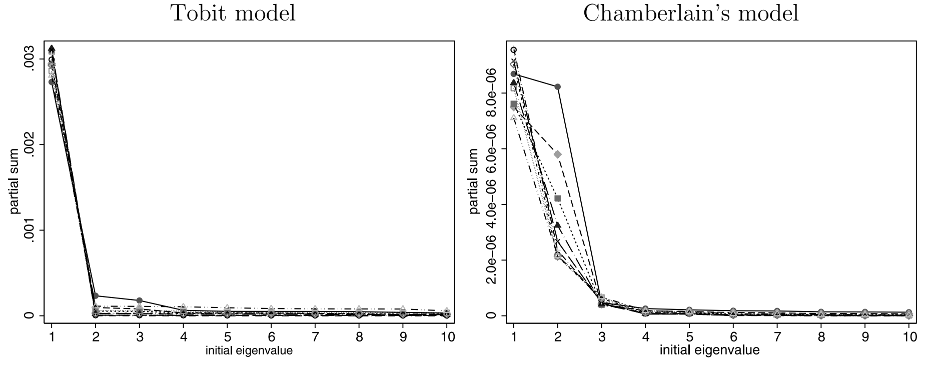

FIGURE S4.—Uniform Fourier convergence ($T = 2$). We report the quantity $\sum_{j>J} \langle \phi_{j,\theta}, f_y \rangle^2$, where $(J + 1)$ is shown on the x-axis. The various curves correspond to different parameters θ (σ on the left panel), which belong to a grid $\{.5, .6, \dots, 1.5\}$.

[50]In our experiments, we observed that estimates of singular vectors associated with very small singular values were affected by numerical error. In Chamberlain's model, the sum $\sum_{j=1}^{J} \langle \phi_{j,\theta}, f_y \rangle^2$ increased steadily with J and seemed to reach a plateau after a few singular values, yet the sum jumped after the 19th singular value (and actually became $\gg \|f_y\|^2$). For this reason, we discarded the singular values $\lambda_{j,\theta}$, $j \geq 19$ in the sum. For the Tobit model, this phenomenon occurred after the 14th singular value, and we proceeded similarly. This is additional evidence of the difficulty of computing singular vectors associated with small singular values.

Figure S4 shows that the Fourier coefficients tend quickly to zero, and there is visual evidence that the convergence is uniform over the set of parameters that we have considered. This provides numerical support for uniform Fourier convergence in those two models.

S3. SPECIFICATION TEST

In applied work, a common approach is to assume a parametric model for the individual effects. Here we show how to use the functional differencing restrictions for the purpose of specification testing.

Let

$$f_{y|x}(y|x) = \int_{\mathcal{A}} f_{y|x,\alpha}(y|x, \alpha; \theta_0) f_{\alpha|x}(\alpha|x; \eta_0)\, d\alpha$$

be a complete parametric specification of the distribution of the data, which includes a parametric model for the individual effects. A popular choice is to let $f_{\alpha|x}(\alpha|x; \eta_0)$ be a Gaussian density, with means and variances that are parsimonious functions of covariates x_i (Chamberlain (1984)).

We wish to test the null hypothesis that $f_{\alpha|x}$ is correctly specified. For this, we consider the random-effects maximum likelihood estimator (MLE) of θ_0, which solves

$$\widetilde{\theta} = \arg\max_{\theta} \left[\arg\max_{\eta} \sum_{i=1}^{N} \ln\left(\int_{\mathcal{A}} f_{y|x,\alpha}(y_i|x_i, \alpha; \theta) f_{\alpha|x}(\alpha|x_i; \eta)\, d\alpha \right) \right].$$

Then, we define the statistic

$$S = \frac{1}{N} \sum_{i=1}^{N} \varphi(y_i, x_i, \widetilde{\theta}),$$

where $\varphi = (\varphi_1, \ldots, \varphi_R)'$ is given by (35). The statistic S is simply an empirical counterpart of the functional differencing moment restrictions, evaluated at the random-effects MLE.

PROPOSITION S3: *Under the null of correct specification, and under regularity conditions given in Section 5 and standard regularity assumptions on the MLE,*

$$\sqrt{N} S \xrightarrow{d} N[0, V_S],$$

where the expression of V_S is provided in equation (S15) below.

PROOF: Let us denote $\ell_i(\theta, \eta) = \ln[\int_{\mathcal{A}} f_{y|x,\alpha}(y_i|x_i, \alpha; \theta) f_{\alpha|x}(\alpha|x_i; \eta)\, d\alpha]$, and $L_{\theta\theta} = \mathbb{E}[\frac{\partial^2 \ell_i(\theta_0, \eta_0)}{\partial\theta\partial\theta'}]$, with a similar notation for the three other components of the Hessian: $L_{\theta\eta}$, $L_{\eta\theta}$, and $L_{\eta\eta}$. Then, under standard regularity conditions and under the null of correct specification,

$$\sqrt{N}(\widetilde{\theta}-\theta_0)\xrightarrow{d}N[0,V_{\widetilde{\theta}}],$$

where $V_{\widetilde{\theta}}=[L_{\theta\theta}-L_{\theta\eta}L_{\eta\eta}^{-1}L_{\eta\theta}]^{-1}$.

Let $\varphi_i(\theta)=\varphi(y_i,x_i,\theta)$. It is easy to show that, under the null, and under the regularity conditions of Theorem 4 and standard regularity assumptions on the MLE (see Arellano (1991)),

$$\sqrt{N}S\xrightarrow{d}N[0,V_S],$$

where

$$V_S=\mathbb{E}\big[(\varphi_i(\theta_0)-GV_{\widetilde{\theta}}s_i)(\varphi_i(\theta_0)-GV_{\widetilde{\theta}}s_i)'\big], \tag{S15}$$

with $s_i=\frac{\partial\ell_i(\theta_0,\eta_0)}{\partial\theta}-L_{\theta\eta}L_{\eta\eta}^{-1}\frac{\partial\ell_i(\theta_0,\eta_0)}{\partial\eta}$, and $G=\mathbb{E}[\frac{\partial\varphi_i(\theta_0)}{\partial\theta'}]$.

A consistent estimator of V_S is then obtained as

$$\widehat{V}_S=\widehat{\mathbb{E}}\big[(\varphi_i(\widetilde{\theta})-\widehat{G}\widehat{V}_{\widetilde{\theta}}\widehat{s}_i)(\varphi_i(\widetilde{\theta})-\widehat{G}\widehat{V}_{\widetilde{\theta}}\widehat{s}_i)'\big],$$

where $\widehat{V}_{\widetilde{\theta}}$ is a consistent estimator of $V_{\widetilde{\theta}}$, $\widehat{s}_i=\frac{\partial\ell_i(\widetilde{\theta},\widetilde{\eta})}{\partial\theta}-\widehat{L}_{\theta\eta}\widehat{L}_{\eta\eta}^{-1}\frac{\partial\ell_i(\widetilde{\theta},\widetilde{\eta})}{\partial\eta}$, with $\widehat{L}_{\theta\eta}$ and $\widehat{L}_{\eta\eta}$ consistent estimators of $L_{\theta\eta}$ and $L_{\eta\eta}$, respectively, and $\widehat{G}$ is given by (49) with $\widetilde{\theta}$ in place of $\widehat{\theta}$. *Q.E.D.*

Let us assume that V_S is nonsingular. In particular, this requires that the vector of moment functions φ is not identically zero, thus restricting the model to be non-surjective. As N tends to infinity, we then have, under the null of correct specification,

$$NS'\widehat{V}_S^{-1}S\xrightarrow{d}\chi^2_R, \tag{S16}$$

where $\widehat{V}_S$ is a consistent estimator of V_S. Thus, (S16) provides a simple way to test the validity of random-effects specifications in non-surjective models. This provides an analog of the Hausman test (Hausman (1978)) in a nonlinear context.

REFERENCES

AHN, S., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27. [8]

ARELLANO, M. (1991): "Moment Testing With Non-ML Estimators," Unpublished Manuscript, CEMFI. [23]

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297. [8]

ARELLANO, M., AND O. BOVER (1995): "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51. [9]

CARRASCO, M., AND J. P. FLORENS (2009): "Spectral Methods for Deconvolving a Density," *Econometric Theory*, 27, 546–581. [3]

CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: Elsevier Science. [22]

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567–596. [9]

HAHN, J. (1997): "A Note on the Efficient Semiparametric Estimation of Some Exponential Panel Models," *Econometric Theory*, 13, 583–588. [8,18]

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1272. [23]

YOSHIDA, K. (1971): *Functional Analysis*. New York: Springer Verlag. [3]

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain; bonhomme@cemfi.es.

Manuscript received May, 2010; final revision received October, 2011.