Supplementary Material to:
Identification by Laplace Transforms
in Nonlinear Time Series and Panel Models
with Unobserved Stochastic Dynamic Effects

Patrick Gagliardini
Università della Svizzera Italiana (USI, Lugano) and Swiss Finance Institute (patrick.gagliardini@usi.ch)

Christian Gouriéroux
CREST and University of Toronto (christian.gourieroux@ensae.fr)

Abstract

In Section B.1 we provide additional results on the linear Gaussian model with common factor and contagion. In Section B.2 we give additional identification results for the semiparametric setting. Finally, in Section B.3 we derive the GMM semiparametric efficiency bound from cross-differencing and illustrate our findings for the Poisson count model with stochastic time effects. Equation numbers (n) and $(a.n)$ for $n = 1, 2, ...$ refer to the main body and Appendix A of the paper, respectively.

B.1 Additional results for the linear Gaussian model with common factor and contagion

In this section we provide results for the structural VAR model in Example 2 concerning the link between the nonlinear moment restrictions and the Yule-Walker equations (Subection B.1.1) and model parameters estimation in the homoskedastic setting (Subsection B.1.2).

B.1.1 Link with Yule-Walker equations

We show that the nonlinear moment restrictions (a.1)-(a.3) are well-chosen (parameter-dependent) linear combinations of the Yule-Walker equations for the structural VAR model (4.1). The Yule-Walker equations involving the covariance function $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$, for $h = 0, 1, 2, ...$ of the observable process are derived by multiplying the first equation in system (4.1) by y_t', y_{t-1}', y_{t-2}', ..., respectively, and computing the expectation on both sides. We get:

$$\Gamma_0(0) = B\Delta_0(0) + C\Gamma_0(1)' + \Sigma, \quad \text{(b.1)}$$

$$\Gamma_0(1) = B\Delta_0(1) + C\Gamma_0(0), \quad \text{(b.2)}$$

$$\Gamma_0(h) = B\Delta_0(h) + C\Gamma_0(h-1), \quad h \geq 2, \quad \text{(b.3)}$$

where $\Delta_0(h) = Cov_0(f_t, y_{t-h})$ is the cross-covariance function of processes (f_t) and (y_t). These covariances are not directly nonparametrically identifiable from the data. However, some linear combinations of the Yule-Walker equations (b.1)-(b.3) do not involve the unidentifiable covariances $\Delta_0(h)$. To see this, let us multiply the first equation in system (4.1) by f_{t+h}', for $h = 0, 1, 2, ...$ and compute the expectation, to get:

$$\Delta_0(h)' = B(\Phi')^h + C\Delta_0(h+1)', \quad h = 0, 1, 2, ... \quad \text{(b.4)}$$

Then, by considering equation (b.1), subtracting equation (b.2) post-multiplied by C', and using equation (b.4) with $h = 0$, we get:

$$\Gamma_0(0) - \Gamma_0(1)C' = C\Gamma_0(1)' - C\Gamma_0(0)C' + BB' + \Sigma,$$

i.e., the first-order nonlinear moment restriction (a.1). Similarly, by considering equation (b.2), subtracting equation (b.3) for $h = 2$ post-multiplied by C', and using equation (b.4) with $h = 1$, we get:

$$\Gamma_0(1) - \Gamma_0(2)C' = C\Gamma_0(0) - C\Gamma_0(1)C' + B\Phi B',$$

i.e., the second-order nonlinear moment restriction (a.2). The third-order nonlinear moment restrictions are obtained in an analogous manner as a linear combination of the equations in (b.3) for $h = 2$ and $h = 3$.

B.1.2 Consistent estimation from first-order nonlinear restrictions in the homoskedastic case

Let us consider the model with static factor, i.e. $\Phi = 0$, and assume $\Sigma_0 = \sigma_0^2 Id_n$ in (4.1), i.e. the innovations of the measurement equation are independent and homoskedastic. In Section 4.1 we show that the parameters are identifiable from the identified set $\mathcal{E}_0$. The proof of identifiability suggests a consistent estimation method. Let us denote $G_T(\sigma^2, B, C)$ a criterion based on moment restrictions (4.3) such that G_T tends to a nondegenerate limit G_∞, say, when the sample size T tends to infinity. Due to the lack of first-order identifiability (Corollary 1), the moment estimator defined by:

$$(\hat{\sigma}_T^2, \hat{B}_T, \hat{C}_T) = \arg\min_{\sigma^2, B, C} G_T(\sigma^2, B, C),$$

where the minimization is such that matrix $B'B$ is diagonal, converges for T large towards the set $\mathcal{E}_0$, but not necessarily to the true value of the parameter (σ_0^2, B_0, C_0). Corollary 2 suggests how to modify the optimization criterion to recover consistency. Let us recall that two symmetric matrices of the same dimension A_0 and A_1 are ordered $A_0 \succeq A_1$ if, and only if, their ranked eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_K(A)$ are such that: $\lambda_j(A_0) \geq \lambda_j(A_1)$, for all $j = 1, ..., K$. Thus, to get the consistency, the estimation criterion has to be penalized:

$$\begin{aligned}(\hat{\sigma}_T^2, \hat{B}_T, \hat{C}_T) &= \arg\min_{\sigma^2, B, C} \left\{G_T(\sigma^2, B, C) + \alpha_T \left[\sigma^2 + \lambda_1(BB') + \cdots + \lambda_K(BB')\right]\right\} \\ &= \arg\min_{\sigma^2, B, C} \left\{G_T(\sigma^2, B, C) + \alpha_T \left[\sigma^2 + Tr(BB')\right]\right\},\end{aligned}$$

where the minimization is such that matrix $B'B$ is diagonal and the positive weight α_T tends to 0 at an appropriate speed when T tends to infinity. The study of the properties of such an estimator is beyond the scope of this paper.

B.2 Additional results for semi-parametric identification

In this section we first consider the multivariate Poisson model with stochastic intensity, and provide a discussion of Assumption 1 (Subsection B.2.1) and show how semi-parametric identification can be achieved by cross-differencing (Subsection B.2.2). Then, we provide identification results from third-order nonlinear moment restrictions in a general semi-parametric setting (Subsection B.2.3).

B.2.1 Discussion of Assumption 1 in the multivariate Poisson model with stochastic intensity

Let us consider Assumption 1 in the multivariate Poisson model with stochastic intensity. We show two properties stated in Section 5.2 of the main body. Recall that $\psi_f(w,\tilde{w}) := \log E_0[\exp(w' f_t + \tilde{w}' f_{t-1})]$.

Property B.1: *When the factor is one-dimensional, and if function* $\dfrac{\partial^2 \psi_f}{\partial w \partial \tilde{w}}$ *is analytic in a neighbourhood of* 0 *in the complex domain, then Assumption 1 is equivalent to: Process* (f_t) *is not i.i.d.*

Proof: For $K = 1$, Condition (5.7) in Assumption 1 is equivalent to: $\dfrac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}} \neq 0$, for some $w, \tilde{w} \in \mathcal{W}$, for any neighbourhood $\mathcal{W}$ of 0. The negation of this statement is: $\dfrac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}} = 0$, for all $w, \tilde{w} \in \mathcal{W}$, in a neighbourhood $\mathcal{W}$ of 0. Now, we use that a complex analytic function vanishing on an open domain vanishes everywhere. Hence, $\dfrac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}} = 0$, for all $w, \tilde{w}$. This is equivalent to the fact that function ψ_f is additive in its two arguments: $\psi_f(w,\tilde{w}) = \psi_f(w,0) + \psi_f(0,\tilde{w})$. In turn, this is equivalent to f_t and f_{t-1} being independent, i.e., to Markov process (f_t) being i.i.d.

Property B.2: *When the factor is multidimensional, and if function* $\dfrac{\partial^2 \psi_f}{\partial w \partial \tilde{w}'}$ *is analytic in a neighbourhood of* 0 *in the complex domain, then Assumption 1 is implied by the following condition:*

$$\textit{If } \eta' f_t \textit{ and } \eta' f_{t-1} \textit{ are independent, for a } \eta \in \mathbb{R}^K\textit{, then } \eta = 0. \tag{b.5}$$

Proof: Assume condition (b.5) holds. We have to prove that Assumption 1 is valid. Obviously, (f_t) cannot be an i.i.d. process. So, we only have to show condition (5.7) in Assumption 1. Let $\mathcal{W}$ be a neighbourhood of 0. To show (5.7), we prove the implication in reversed order of the negative statements, i.e., let us suppose that $\eta \in \mathbb{R}^K$ is not the zero vector and show that then $\eta' \dfrac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}'} \eta \neq 0$ for some $w, \tilde{w} \in \mathcal{W}$. Suppose the latter property is not true, i.e.

$$\eta' \frac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}'} \eta = 0, \quad \forall w, \tilde{w} \in \mathcal{W}. \tag{b.6}$$

Consider the process $g_t := \eta' f_t$. The log joint Laplace transform of (g_t, g_{t-1}) is

$$\psi_g^\eta(u,v) := \log E[\exp(u g_t + v g_{t-1})] = \psi_f(u\eta, v\eta). \tag{b.7}$$

Thus:

$$\frac{\partial^2\psi_g^\eta(u,v)}{\partial u\partial v} = \eta' \frac{\partial^2\psi_f(u\eta, v\eta)}{\partial w\partial \tilde{w}'}\eta. \tag{b.8}$$

From (b.6) it follows that $\frac{\partial^2\psi_g^\eta(u,v)}{\partial u\partial v} = 0$ for u, v in a small neighbourhood of 0. By the arguments in the proof of Property B.1, it follows that $g_t = \eta' f_t$ and $g_{t-1} = \eta' f_{t-1}$ are independent. However, by condition (b.5), this is not possible and thus (b.6) is not true.

B.2.2 Identification of the multivariate Poisson model with stochastic intensity via cross-differencing

The first-order nonlinear cross-differencing approach presented in Section 6 for a panel data framework is applicable to multivariate time series (corresponding to a single "individual" only), where the heterogeneity of the factor sensitivities is not observable and involves unknown parameters. To illustrate this fact, let us consider the multivariate Poisson model with common stochastic intensity introduced in Section 5.2, with a single unobservable factor. Equation (6.2) reads in this case:

$$E\left[\exp(uy_{i,t})|\underline{y_{t-1}}, \underline{f_t}\right] = \exp\left\{(e^u - 1)\beta_i f_t\right\} \quad \forall i,\ \forall u \in \mathcal{U}.$$

With a change of variable $v = (e^u - 1)\beta_i$, we get:

$$E\left[\exp\{\log(1 + v/\beta_i)y_{i,t}\}|\underline{y_{t-1}}, \underline{f_t}\right] = \exp\left\{v f_t\right\} \quad \forall i,\ \forall v \in \mathcal{V}.$$

Assuming that the factor loading for the first individual is not zero, we set $\beta_1 = 1$ as a normalization condition. Thus, after applying the cross-differencing approach and integrating out the latent factor from the conditioning set, we get for any pair $(i, 1)$, with $i = 2, 3, ..., n$, the conditional moment restrictions:

$$E\left[\exp\{\log(1 + v/\beta_i)y_{i,t}\} - \exp\{\log(1 + v)y_{1,t}\}|\underline{y_{t-1}}\right], \quad \forall i = 2, 3, ..., n,\ \forall v \in \mathcal{V}. \tag{b.9}$$

This continuum of nonlinear moment restrictions can be used to identify the loadings parameters.

B.2.3 Identification from third-order nonlinear restrictions

Let us consider the general semi-parametric framework of Assumptions A.1-A.3. If the regression parameters and the transition p.d.f. of the latent factor are not second-order nonlinearly identifiable, the information in third-order restrictions can be exploited for identification. An approach similar to the one presented in Section 5.1 can be pursued. We present here an alternative approach relying on the implication of the Markovianity assumption on the latent factor (conditional on the covariates). For any given value (B, C, θ) of the regression parameters, the system of first-order nonlinear moment restrictions (3.3) allows to compute by Fourier inversion the conditional distribution [1] of f_t given $\underline{x_t}$. Similarly, for any given (B, C, θ), the second- and third-order moment restrictions (3.6) and (3.8) allow to compute by Fourier inversion the conditional distributions of (f_t, f_{t-1}) given $\underline{x_t}$, and of (f_t, f_{t-1}, f_{t-2}) given $\underline{x_t}$, respectively. In particular, we can compute the conditional transition of the unobserved component at horizon 1: $g_1(f_t|f_{t-1}, \underline{x_t}; B, C, \theta)$, and the conditional transition at horizon 2: $g_2(f_t|f_{t-2}, \underline{x_t}; B, C, \theta)$, say, for given B, C, θ. When evaluated for the true value of the parameters (B_0, C_0, θ_0), these functions are equal to the true conditional transition functions of the latent factor process at horizons 1 and 2, respectively. By the Markov Assumption A.3, we get the Kolmogorov relationship:

$$g_2(f|\tilde{f}, \underline{x_t}; B, C, \theta) = \int g_1(f|f_{t-1}, \underline{x_t}; B, C, \theta) g_1(f_{t-1}|\tilde{f}, \underline{x_t}; B, C, \theta) df_{t-1}, \quad \forall f, \tilde{f}, \ \forall \underline{x_t}, \tag{b.10}$$

for the true value of the regression parameters B, C, θ. This relatioship yields an infinite number of nonlinear restrictions indexed by the admissible values of f, $\tilde{f}$, $\underline{x_t}$. Under the assumption that (b.10) holds for the true value of the regression parameters B_0, C_0, θ_0 only, this condition can be used to identify B_0, C_0, θ_0.

B.3 GMM efficiency bounds and cross-differencing

In semi-parametric panel data models with cross-differencing (Section 6), the parameters of interest are identified by a continuum of conditional moment restrictions:

$$E[h_t(\beta, u)|\underline{y_{t-1}}, \underline{x_t}] = 0, \ \ \forall u \in U, \text{ say}, \tag{b.11}$$

where function h_t, with dimension $\dim(h_t) = d$, depends on the observable variables $\underline{y_t}$, $\underline{x_t}$, vector β with dimension $\dim(\beta) = p$ includes the parameters of the affine nonlinear regression model B, C, θ, and U

[1] When (B, C, θ) is the true parameter value (B_0, C_0, θ_0), the resulting function is the true conditional density of f_t given $\underline{x_t}$. When (B, C, θ) is not the true parameter value, the resulting function $l(f_t|\underline{x_t}; B, C, \theta)$, say, may not be a density. The same remark applies for functions $l(f_t, f_{t-1}|\underline{x_t}; B, C, \theta)$ and $l(f_t, f_{t-1}, f_{t-2}|\underline{x_t}; B, C, \theta)$ obtained using second- and third-order restrictions. This fact is not a problem for our identification strategy. In fact, we construct functions g_1 and g_2 in equation (b.10) by the standard rules, e.g. $g_1(f_t|f_{t-1}, \underline{x_t}; B, C, \theta) = l(f_t, f_{t-1}|\underline{x_t}; B, C, \theta) / \int l(f_t, f_{t-1}|\underline{x_t}; B, C, \theta) df_t$.

is the set of admissible arguments in the Laplace transform. [2] In this section, we first define the semi-parametric efficiency bound for estimating parameter β from the conditional moment restrictions (b.11), and explain how it can be reached from above by sequences of finite grids on argument u (Subsection B.3.1). Then, we illustrate the patterns of the efficiency bounds for the Poisson count panel data model with stochastic time effect introduced in Example 7 (Subsection B.3.2). The proofs of the results are provided in Subsection B.3.3.

B.3.1 GMM efficiency bounds for a continuum of conditional moment restrictions

Let us introduce the conditional variance-covariance matrix:

$$\Omega_t(u,\tilde{u}) = Cov\left(h_t(\beta_0,u), h_t(\beta_0,\tilde{u})|\underline{y_{t-1}}, \underline{x_t}\right), \tag{b.12}$$

where β_0 denotes the true parameter value, and define the associated conditional covariance operator A_t by:

$$A_t\varphi(u) = \int_U \Omega_t(u,\tilde{u})\varphi(\tilde{u})d\pi(\tilde{u}),$$

for any admissible function φ in $L^2(U,\pi)$, where $L^2(U,\pi)$ is the Hilbert space of d-variate square integrable functions of argument $u \in U$ equipped with the inner product $\langle\varphi,\tilde{\varphi}\rangle = \int_U \varphi(u)'\tilde{\varphi}(u)d\pi(u)$ for measure π on U.

Assumption SM.1. *The process $(y_t', x_t')'$ is strictly stationary and ergodic.*

Assumption SM.2. *The conditional moment function h_t is square integrable w.r.t. the product measure $P \otimes \pi$ for data $\underline{y_t}$, $\underline{x_t}$ and argument u, i.e., $\int_U E[\|h_t(\beta,u)\|^2]d\pi(u) < \infty$, for any β.*

When the set U is unbounded, the choice of measure π can accommodate conditional moment functions whose second-order moments are bounded away from zero for all u. Assumption SM.2 implies in particular that i) the moment function $h_t(\beta,\cdot)$ and the conditional expectation $E[h_t(\beta,\cdot)|\underline{y_{t-1}},\underline{x_t}]$ are in $L^2(U,\pi)$, P-a.s., for any β, and ii) the conditional variance-covariance matrix Ω_t is well-defined, P-a.s.

[2] The conditional moment restrictions for parametric identification considered in Section 4, or those for semi-parametric identification in Section 5, do not contain the lagged endogenous variables $\underline{y_{t-1}}$ in the conditioning set. Thus, process $h_t(\beta_0,u)$, for given argument u, is not a martingale difference sequence w.r.t. information $\underline{y_{t-1}}$, $\underline{x_t}$. The GMM efficiency bound could be derived in such frameworks as well, but at the cost of additional complexity. We do not consider those frameworks in this section.

Assumption SM.3. *The operator A_t maps $L^2(U,\pi)$ to $L^2(U,\pi)$, is injective and compact, P-a.s.*

The operator A_t is self-adjoint. By the compactness property in Assumption SM.3, there exists an orthonormal basis of $L^2(U,\pi)$ consisting of eigenfunctions $\varphi_{t,j}$, $j = 1,2,...$ of operator A_t, with associated eigenvalues $\lambda_{t,j}$, $j = 1,2,...$ [see e.g. Kress (1999) for the spectral decomposition of compact operators]. The eigenvalues are such that:

$$\lambda_{t,j} = \langle \varphi_{t,j}, A_t \varphi_{t,j} \rangle = V\left(\int_U \varphi_{t,j}(u)' h_t(\beta_0, u) d\pi(u) | \underline{y_{t-1}}, \underline{x_t} \right) > 0,$$

where the strict inequality follows from the injectivity of operator A_t. Moreover, we can rank the eigenvalues in decreasing order $\lambda_{t,1} \geq \lambda_{t,2} \geq ...$, and we have $\lambda_{t,j} \to 0$ as $j \to \infty$, P-a.s.

Define the functions:

$$g_{t,j}(\beta) = \int_U \varphi_{t,j}(u)' h_t(\beta, u) d\pi(u) = \langle \varphi_{t,j}, h_t(\beta, \cdot) \rangle, \quad j = 1,2,...,$$

that are inner products of the moment function with the eigenfunctions of the covariance operator. From (b.11), we get the countable set of conditional moment restrictions:

$$E[g_{t,j}(\beta) | \underline{y_{t-1}}, \underline{x_t}] = 0, \ \ j = 1,2,... \tag{b.13}$$

Given that the functions $\varphi_{t,j}$ build a basis of $L^2(U,\pi)$, P-a.s., under Assumption SM.2 this countable set of conditional moment restrictions is equivalent to the original continuum of conditional moment restrictions (b.11).

Let us denote by Σ_J, say, the GMM efficiency bound for estimating parameter β from the conditional moment restrictions in (b.13) for $j = 1,...,J$, with given integer J, and $T \to \infty$ [Hansen (1985), Chamberlain (1987)]. We have:

$$\Sigma_J = \left(E\left[E\left[\frac{\partial G_t^J(\beta_0)'}{\partial \beta} | \underline{y_{t-1}}, \underline{x_t} \right] V\left[G_t^J(\beta_0) | \underline{y_{t-1}}, \underline{x_t} \right]^{-1} E\left[\frac{\partial G_t^J(\beta_0)}{\partial \beta'} | \underline{y_{t-1}}, \underline{x_t} \right] \right] \right)^{-1}, \tag{b.14}$$

where $G_t^J(\beta) = [g_{t,1}(\beta), ..., g_{t,J}(\beta)]'$. The GMM efficiency bound Σ_J can be written in terms of the spectral decomposition of operator A_t. More precisely, let us define:

$$D_t(u) = E\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta'} | \underline{y_{t-1}}, \underline{x_t} \right]. \tag{b.15}$$

We have [see Section B.3.3 i)]:

$$\Sigma_J = \left(E\left[\sum_{j=1}^J \frac{1}{\lambda_{t,j}} \langle D_t, \varphi_{t,j} \rangle \langle \varphi_{t,j}, D_t \rangle \right] \right)^{-1}, \tag{b.16}$$

where $\langle \varphi_{t,j}, D_t \rangle$ is the row vector with components $\langle \varphi_{t,j}, D_{t,k} \rangle$, for $k = 1, ..., p$, with $D_{t,k}(u) = E\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta_k} | \underline{y_{t-1}}, \underline{x_t}\right]$, and $\langle D_t, \varphi_{t,j} \rangle = \langle \varphi_{t,j}, D_t \rangle'$.

Definition 5. *The GMM efficiency bound for estimating parameter β from the continuum of conditional moment restrictions (b.11) is the limit:*

$$\Sigma = \lim_{J \to \infty} \Sigma_J,$$

when this limit exists and is a positive definite matrix.

The existence and the positive-definiteness of the limit is guaranteed by the next assumption.

Assumption SM.4. *i) If $D_t \xi = 0$ in $L^2(U, \pi)$, P-a.s., for $\xi \in \mathbb{R}^p$, then $\xi = 0$.*
ii) We have:

$$E\left[\sum_{j=1}^{\infty} \frac{1}{\lambda_{t,j}} \| \langle D_t, \varphi_{t,j} \rangle \|^2 \right] < \infty.$$

Assumption SM.4 i) is the counterpart of the usual full-rank condition for local identification.

Proposition 9. *Under Assumptions SM.1-SM.4, the GMM efficiency bound Σ exists and is equal to:*

$$\Sigma = \left(E\left[\sum_{j=1}^{\infty} \frac{1}{\lambda_{t,j}} \langle D_t, \varphi_{t,j} \rangle \langle \varphi_{t,j}, D_t \rangle \right]\right)^{-1} = \left(E\left[\langle D_t, A_t^{-1} D_t \rangle \right]\right)^{-1}.$$

Proof: See Sections B.3.3 ii) and iii).

Even if the inverse of operator A_t is not defined on the whole vector space $L^2(U, \pi)$, under Assumptions SM.1-SM.4 the function

$$z_t = A_t^{-1} D_t \tag{b.17}$$

exists in $L^2(U, \pi)$ [see Section B.3.3 iii)]. Proposition 9 extends the formula for the asymptotic variance of efficient GMM estimators with a continuum of moment restrictions in Carrasco, Florens (2000, 2014) to a setting with conditional information. Carrasco et al. (2007) consider a dynamic setting with unobservable components. They show the asymptotic efficiency of a GMM estimator based on a continuum of moment restrictions induced by the joint characteristic function of the observable component, under a Markov assumption for the latter. We do not assume Markovianity of the observable component.

The GMM efficiency bound in Proposition 9 can also be derived by an optimal choice of p instruments corresponding to function z_t defined in equation (b.17). Indeed, let us consider the function:

$$g_t(\beta) = \int_U z_t(u)' h_t(\beta, u) d\pi(u) = \langle z_t, h_t(\beta, \cdot) \rangle. \tag{b.18}$$

From (b.11) and the Law of Iterated Expectation, function $g_t(\beta)$ defines an exactly identified unconditional moment restriction $E[g_t(\beta_0)] = 0$. We show in Subsection B.3.3 iv) that the asymptotic variance of the GMM estimator based on the unconditional moment restriction $E[g_t(\beta_0)] = 0$ is:

$$E\left[\frac{\partial g_t(\beta_0)}{\partial \beta'}\right]^{-1} E[g_t(\beta_0)g_t(\beta_0)']E\left[\frac{\partial g_t(\beta_0)'}{\partial \beta}\right]^{-1} = \Sigma, \tag{b.19}$$

i.e. the GMM efficiency bound. Thus, z_t is the optimal instrument.

Finally, we show that the GMM efficiency bound Σ can be approximated from above by the GMM efficiency bound from the conditional moment restrictions based on a fine grid of values for the argument of the Laplace transform [see e.g. Singleton (2001) for GMM estimation with the conditional characteristic function evaluated on a grid]. For expository purpose, let us assume that set U is a hyperrectangular domain in $\mathbb{R}^N$, i.e. $U = [a_1, b_1) \times ... \times [a_N, b_N)$. We consider multi-dimensional grids that are obtained by partitioning each interval $[a_l, b_l)$ in subintervals, and building the Cartesian products of these subintervals. We get in this way a partition of set U in non-overlapping subrectangles U_m, for $m = 1, ..., M$, say, whose union is U. Moreover, let $u_m \in U_m$ for any $m = 1, ..., M$. We refer to the set of subrectangles U_m and points u_m as a "multi-dimensional grid". The diameter Δ_M of such a grid is defined as the largest of the diameters of the subrectangles, where the diameter of a subrectangle U_m is the largest length of the intervals whose Cartesian product generates U_m.

Assumption SM.5. *For any multi-dimensional grid corresponding to a partition of U in M non-overlapping subrectangles, i) the (dM, dM) variance-covariance matrix Ω_t^M with blocks $\Omega_t(u_m, u_{m'})$ is positive definite, P-a.s., and ii) the (dM, p) matrix $D_t^M = (D_t(u_1)', ..., D_t(u_M)')'$ is such that $D_t^M \xi = 0$, P-a.s., for $\xi \in \mathbb{R}^p$, implies $\xi = 0$.*

Assumption SM.6. *Functions $h_t(\beta_0, u)$, $\partial h_t(\beta_0, u)/\partial \beta'$ and $z_t(u)$ are continuous w.r.t. argument u, P-a.s.*

For any multi-dimensional grid, let Σ_M^{gr} denote the GMM efficiency bound from the conditional moment restrictions (b.11) corresponding to arguments u_m, with $m = 1, ..., M$, that is:

$$\Sigma_M^{gr} = \left(E\left[D_t^{M\,\prime}(\Omega_t^M)^{-1}D_t^M\right]\right)^{-1}. \tag{b.20}$$

We have $\Sigma \leq \Sigma_M^{gr}$ in the ordering of positive-definite matrices, since the efficiency bound Σ_M^{gr} is based on a subset of the information. We have the following Proposition:

Proposition 10. *Under Assumption SM.1-SM.6, if the number of subrectangles M tends to infinity and the diameter of the grid Δ_M tends to zero, then Σ_M^{gr} tends to the GMM efficiency bound:*

$$\Sigma_M^{gr} \to \Sigma, \ \text{ as } M \to \infty \text{ and } \Delta_M \to 0.$$

Proof: See Section B.3.3 v).

B.3.2 An illustration

In this subsection we present an illustration with the Poisson count panel data model with stochastic time effects introduced in Example 7. This is a semi-parametric model. We compare the GMM efficiency bounds obtained from two sets of conditional moment restrictions: the continuum set of nonlinear cross-differencing restrictions (6.5) and the linear cross-differencing restrictions (6.6). We also investigate the informational content of the nonlinear cross-differencing restrictions for different values of the argument in the Laplace transform.

i) The Data Generating Process (DGP)

The individual count histories are independent conditionally on the common latent factor (f_t), with conditional Poisson distribution $y_{i,t} \sim \mathcal{P}(f_t + x_{i,t}\alpha + y_{i,t-1}c)$. The regressor $x_{i,t}$ is scalar. The Markov processes $(x_{i,t})$, $i = 1, ..., n$, and (f_t) are exogenous and mutually independent. The former follow identical ARG processes with scale parameter $\delta_x > 0$, degree of freedom parameter $\nu_x > 0$ and first-order autocorrelation $\rho_x < 1$ [see equation (4.10)]. The latter process follows an ARG process with parameters $\delta_f > 0$, $\nu_f > 0$ and $\rho_f < 1$. We set the number of individuals as $n = 25$, that is a realistic choice for instance in view of applications to corporate default count data aggregated per industrial sectors.

Throughout the numerical experiments of this section, we set the ARG parameters of the exogenous covariate processes as $\rho_x = 0.5$, $\nu_x = 0.5$ and $\delta_x = 1$, and the parameters of the latent process as $\rho_f = 0$, $\nu_f = 0.5$ and $\delta_f = 2$. This choice is such that the first two unconditional moments of the exogenous covariates and the latent factor are $E(x_{i,t}) = E(f_t) = 1$ and $V(x_{i,t}) = V(f_t) = 2$. [3] Finally, we set $\alpha = 0.5$ and consider several values for the autoregressive coefficient c.

ii) GMM efficiency bounds based on first-order nonlinear cross-differencing

[3] The links of model parameters with unconditional moments are $E[x_{i,t}] = \frac{\nu_x \delta_x}{1-\rho_x}$, $V[x_{i,t}] = \nu_x \left(\frac{\delta_x}{1-\rho_x}\right)^2$, and similarly for process (f_t).

a) The nonlinear restrictions with cross-differencing (6.5) yield a continuum of conditional moment restrictions as in (b.11), where the vector conditional moment function is:

$$h_t(\beta, u) = [H_{i,t}(\beta, u) - H_{1,t}(\beta, u)]_{i=2,3,\ldots,n}, \tag{b.21}$$

with dimension $d = n - 1$ and function $H_{i,t}(\beta, u) = \exp\{uy_{i,t} + (1 - \exp u)(x_{i,t}\alpha + y_{i,t-1}c)\}$, and the parameter is $\beta = (\alpha, c)'$. For any admissible value of the scalar argument u, we consider $n-1$ conditional moment restrictions only, since the remaining restrictions can be written as linear combinations of them. [4]

We compute the GMM efficiency bound Σ for parameter β when $T \to \infty$ and n is fixed using the results in Section B.3.1. [5] We use the explicit expressions for function D_t and conditional variance matrix Ω_t [see Section B.3.3 vi)]:

$$D_t(u) = -(e^u - 1)\Psi_t^f(e^u - 1)\Delta z_t, \tag{b.22}$$

and:

$$\Omega_t(u, \tilde{u}) = \omega_{1,t}(u, \tilde{u})\iota_{n-1}\iota_{n-1}' + diag\{\omega_{i,t}(u, \tilde{u}), i = 2, 3, ..., n\}, \tag{b.23}$$

where ι_{n-1} is the $(n-1, 1)$ vector of ones,

$$\omega_{i,t}(u, \tilde{u}) = \Psi_t^f(e^{u+\tilde{u}} - 1)e^{(e^u-1)(e^{\tilde{u}}-1)z_{i,t}'\beta_0} - \Psi_t^f(e^u + e^{\tilde{u}} - 2),$$

the rows of $(n-1, 2)$ matrix Δz_t are $(z_{i,t} - z_{1,t})'$ for $i = 2, 3, ..., n$, with $z_{i,t} = (x_{i,t}, y_{i,t-1})'$, and $\Psi_t^f(u) = E\left[\exp(uf_t)|\underline{y_{t-1}}, \underline{x_t}\right]$ is the conditional Laplace transform of the stationary distribution of the ARG process (f_t) given $\underline{y_{t-1}}$, $\underline{x_t}$. Since the latent factor process is i.i.d. under the parameter choice $\rho_f = 0$ for the DGP, and is independent of the covariate processes, the conditional Laplace transform boils down to the stationary Laplace transform:

$$\Psi_t^f(u) = \Psi^f(u) = \frac{1}{(1 - \delta_f u)^{\nu_f}}.$$

Computing function Ψ_t^f for non-zero autocorrelation of the latent process would require a nonlinear filtering approach. Since the study of the impact of parameter ρ_f on the efficiency bound is not the main focus of our illustration, we limit ourselves to the choice $\rho_f = 0$. From equation (b.23), the conditional second moment of the moment function $h_t(\beta_0, u)$ exists if, and only if, $\Psi^f(e^{2u} - 1) < \infty$, i.e. $u < u_{max} = \frac{1}{2}\log(\frac{1-\rho_f}{\delta_f} + 1)$. For our DGP we get $u_{\max} = 0.2027$.

[4] The efficiency bound is independent from the selected set of linearly independent conditional moment restrictions.

[5] The semiparametric efficiency bound when both n and T tend to infinity can be derived from the results in Gagliardini and Gourieroux (2014).

For any finite grid, we compute the asymptotic variance Σ_M^{gr} in (b.20) by using equations (b.22)-(b.23) and approximating the expectation by a sample average over a long simulated path of the process under the DGP. We compute the GMM efficiency bound Σ via the approximation for a fine grid as in Proposition 10.

b) For comparison purposes we also consider the GMM efficiency bound for the linear cross-differencing restrictions (6.6). They yield the $n-1$ conditional moment restrictions:

$$E[h_t(\beta)|\underline{y_{t-1}}, \underline{x_t}] = 0, \tag{b.24}$$

where $h_t(\beta) = [H_{i,t}(\beta) - H_{1,t}(\beta)]_{i=2,...,n}$ and $H_{i,t}(\beta) = (y_{i,t} - x_{i,t}\alpha - y_{i,t-1}c) - (y_{1,t} - x_{1,t}\alpha - y_{1,t-1}c)$. The asymptotic GMM efficiency bound from linear cross-differencing is:

$$\Sigma^* = (E[\Delta z_t' \Omega_t^{-1} \Delta z_t])^{-1},$$

where $\Omega_t = \omega_{1,t}\iota_{n-1}\iota_{n-1}' + diag(\omega_{i,t}, i = 2, 3, ..., n)$ and $\omega_{i,t} = 1 + z_{i,t}'\beta_0$.

c) Let us now discuss the values and patterns of the GMM efficiency bounds for the DGP of paragraph i). For the sake of conciseness, we focus our analysis on the autoregressive parameter c. We start by setting the value of this parameter in the DGP as $c = 0.5$. For the purpose of interpreting the results, we report the asymptotic efficiency bounds in terms of standard deviations which are scaled in order to correspond to a sample of size $T = 100$. Such a sample size is realistic e.g. for an application with monthly data on corporate default counts. We find that the efficiency bound for estimating parameter c from the nonlinear cross-differencing restrictions is such that $\sqrt{\Sigma_{cc}/T} = 0.0165$, where Σ_{cc} denotes the lower-right element of the $(2, 2)$ matrix Σ, and the efficiency bound from the linear cross-differencing restrictions is such that $\sqrt{\Sigma_{cc}^*/T} = 0.0175$. The informational content of nonlinear restrictions reduces the standard deviation for estimating parameter c of more than 6 percent in the considered DGP.

In Figure 1 we investigate the efficiency loss incurred when a single argument is used in the cross-differencing nonlinear restrictions instead of the continuum set. We display the scaled standard deviation $\sqrt{\Sigma_{1,cc}^{gr}(u)/T}$ as a function of the argument u in the admissible set, where $\Sigma_1^{gr}(u)$ is the GMM asymptotic variance for the nonlinear cross-differencing restrictions with a single argument u. The corresponding curve is U-shaped, lies above the horizontal line for value $\sqrt{\Sigma_{cc}/T}$ corresponding to the GMM efficiency bound, and has a minimum near $u = -0.1$. The loss of information when the optimal argument u is adopted is small. However, this optimal argument depends on the DGP and is unknown to the econometrician. In contrast, the loss of information can be large for other argument values, especially at the boundaries of the admissible set. The curve $\sqrt{\Sigma^{gr}_{1,cc}(u)/T}$ intersects the horizontal line defined by the efficiency bound $\sqrt{\Sigma^{*}_{cc}/T}$ from linear cross-differencing in argument $u = 0$. In fact, the asymptotic variance $\Sigma^{gr}_{1}(u)$ is not defined for argument $u = 0$ because the moment condition for that argument is degenerate equal to zero. However, $\Sigma^{gr}_{1}(u)$ tends to Σ^{*} as $u \to 0$ because the first-order expansion of the nonlinear cross-differencing restrictions for small u yields the linear cross-differencing restrictions as seen in Section 6.2.

In Figure 2 we display the scaled asymptotic standard deviation $\sqrt{\Sigma^{gr}_{2,cc}(0,u)/T}$, which corresponds to the joint use of the linear cross-differencing restrictions (i.e., argument 0 in the limit sense) and the nonlinear cross-differencing restrictions for argument u, as a function of u. This analysis is useful to understand which argument values in the nonlinear cross-differencing restrictions are most informative when used *additionally* to the linear cross-differencing restrictions. We find that values close to $u = -0.20$ are the most informative in this incremental measure, and the corresponding GMM asymptotic variance is very close to the efficiency bound Σ_{cc}. Moreover, this finding shows that a small number of arguments for the cross-differencing restrictions - if well chosen - convey most part of the information in the continuum of conditional moment restrictions, at least for the considered GDP. The curve in Figure 2 reaches the horizontal line corresponding to the efficiency bound from linear cross-differencing restrictions for argument u tending to 0, or u tending to the boundaries of the admissible set u_{max} and $-\infty$ (not shown), because in those cases the additional information provided by argument u is small.

In Figure 3 we study the patterns of the GMM efficiency bounds for estimating the autoregressive coefficient c, as functions of the value of parameter c in the DGP. Specifically, we display the values of the scaled efficiency bounds $\sqrt{\Sigma_{cc}/T}$ and $\sqrt{\Sigma^{*}_{cc}/T}$ from nonlinear and linear cross-differencing restrictions, respectively, as functions of parameter value c in the DGP. Both curves feature an inverted U-shape, which means that the GMM asymptotic variance for estimating parameter c gets smaller either for values of c close to 0, or for values close to 1. A naive analogy with the linear autoregressive process without unobservable effects offers an interpretation for the fact that the accuracy for estimating the autoregressive coefficient c increases when the process is more persistent. The difference between the two curves in Figure 3 is bigger for small values of c. For those DGPs, the contribution of the nonlinear cross-differencing restrictions is more important. For instance, for values of c below 0.20, say, the nonlinear cross-differencing restrictions reduce the asymptotic standard deviation of about 15%, or more, compared to the linear restrictions.

To summarize, the main findings of our numerical illustrations are: (i) the continuum of nonlinear cross-differencing restrictions has a significant contribution for asymptotic efficiency compared to the finite set of linear cross-differencing restrictions, (ii) the incremental informational content of nonlinear restrictions is bigger for argument values close to - but different from - 0, and (iii) the GMM efficiency bound is well approximated even with a small number of well-chosen argument values, such as two or three.

B.3.3 Proofs

i) Proof of equation (b.16)

By the orthonormality of the eigenfunction basis, we have:

$$Cov\left(g_{t,j}(\beta_0), g_{t,k}(\beta_0)|\underline{y_{t-1}}, \underline{x_t}\right) = \int_U \int_U \varphi_{t,j}(u)'\Omega_t(u,\tilde{u})\varphi_{t,k}(\tilde{u})d\pi(u)d\pi(\tilde{u}) = \langle \varphi_{t,j}, A_t\varphi_{t,k}\rangle = \lambda_{t,j}\delta_{j,k},$$

where $\delta_{j,k} = 1$ if $j = k$, and $= 0$, otherwise, and:

$$E\left[\frac{\partial g_{t,j}(\beta_0)}{\partial \beta'}|\underline{y_{t-1}}, \underline{x_t}\right] = \int_U \varphi_{t,j}(u)'E\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta'}|\underline{y_{t-1}}, \underline{x_t}\right]d\pi(u) = \langle \varphi_{t,j}, D_t\rangle.$$

Then, equation (b.16) follows from equation (b.14).

ii) Proof of Proposition 9, first equality

The first equality in Proposition 9 follows from the definition of Σ (Definition 5), the expression of Σ_J in equation (b.16), and the fact that we can apply the Lebesgue's dominated convergence Theorem for $J \to \infty$ in equation (b.16) under Assumption SM.4. Matrix $E\left[\sum_{j=1}^{J}\frac{1}{\lambda_{t,j}}\langle D_t, \varphi_{t,j}\rangle\langle\varphi_{t,j}, D_t\rangle\right]$ is positive definite. Indeed, if $\xi' E\left[\sum_{j=1}^{J}\frac{1}{\lambda_{t,j}}\langle D_t, \varphi_{t,j}\rangle\langle\varphi_{t,j}, D_t\rangle\right]\xi = 0$ for a vector $\xi \in \mathbb{R}^p$, then we get $\langle\varphi_{t,j}, D_t\xi\rangle = 0$ for j, P-a.s. This implies $D_t\xi = 0$ in $L^2(U, \pi)$, P-a.s, and hence $\xi = 0$ by Assumption SM.4 i). Thus, the inverse of $E\left[\sum_{j=1}^{J}\frac{1}{\lambda_{t,j}}\langle D_t, \varphi_{t,j}\rangle\langle\varphi_{t,j}, D_t\rangle\right]$ exists and is positive definite.

iii) Proof of Proposition 9, second equality

Assumption SM.4 ii) implies that:

$$\sum_{j=1}^{\infty}\frac{1}{\lambda_{t,j}}\langle\varphi_{t,j}, D_{t,k}\rangle^2 < \infty, \quad P - a.s, \tag{b.25}$$

for any $k = 1, ..., p$. Since $\mathcal{N}(A_t^*)^\perp = \mathcal{N}(A_t)^\perp = L^2(U, \pi)$ by the self-adjoint property of operator A_t and Assumption SM.3, where $\mathcal{N}(\cdot)$ denotes the null space of a matrix, the summability condition in (b.25) implies that $D_{t,k}$ is in the range of operator A_t, P-a.s., for any k, by the Picard theorem [see e.g. Kress (1999)]. The range of operator A_t is a subspace of vector space $L^2(U, \pi)$ known as the Reproducing Kernel Hilbert Space (RKHS). Under Assumption SM.4, $A_t^{-1} D_t$ exists and is equal to $A_t^{-1} D_t = \sum_{j=1}^{\infty} \frac{1}{\lambda_{t,j}} \langle \varphi_{t,j}, D_t \rangle \varphi_{t,j}$. Thus, we have:

$$\langle D_t, A_t^{-1} D_t \rangle = \sum_{j=1}^{\infty} \frac{1}{\lambda_{t,j}} \langle D_t, \varphi_{t,j} \rangle \langle \varphi_{t,j}, D_t \rangle,$$

which yields the second equality in Proposition 9.

iv) Proof of Equation (b.19)

From (b.18) we have

$$\begin{aligned} E\left[\frac{\partial g_t(\beta_0)}{\partial \beta'}\right] &= E\left[\int_U z_t(u)' E\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta'} | \underline{y_{t-1}}, \underline{x_t}\right] d\pi(u)\right] \\ &= E[\langle A_t^{-1} D_t, D_t \rangle] = E[\langle D_t, A_t^{-1} D_t \rangle], \end{aligned} \tag{b.26}$$

since A_t^{-1} is self-adjoint on the RKHS, and:

$$\begin{aligned} E[g_t(\beta_0) g_t(\beta_0)'] &= E\left[\int_U \int_U z_t(u)' E[h_t(\beta_0, u) h_t(\beta_0, \tilde{u})' | \underline{y_{t-1}}, \underline{x_t}] z_t(\tilde{u}) d\pi(u) d\pi(\tilde{u})\right] \quad \text{(b.27)} \\ &= E\left[\int_U \int_U z_t(u)' \Omega_t(u, \tilde{u}) z_t(\tilde{u}) d\pi(u) d\pi(\tilde{u})\right] = E[\langle z_t, A_t z_t \rangle] = E[\langle D_t, A_t^{-1} D_t \rangle]. \end{aligned}$$

Thus, from equations (b.26) and (b.27) and Proposition 9, the asymptotic variance of the GMM estimator based on the unconditional moment restriction $E[g_t(\beta_0)] = 0$ is:

$$E\left[\frac{\partial g_t(\beta_0)}{\partial \beta'}\right]^{-1} E[g_t(\beta_0) g_t(\beta_0)'] E\left[\frac{\partial g_t(\beta_0)'}{\partial \beta}\right]^{-1} = (E[\langle D_t, A_t^{-1} D_t \rangle])^{-1} = \Sigma,$$

which proves equation (b.19).

v) Proof of Proposition 10

We use an argument similar to Singleton (2001), Section 5. Let us use grid u_m, $m = 1, ..., M$, to approximate the integral in (b.18) and define the function:

$$g_t^M(\beta) = \frac{1}{M} \sum_{m=1}^{M} z_t(u_m)' h_t(\beta, u_m) v_m,$$

where v_m is the multi-dimensional volum of subrectangle U_m, i.e., the product of the corresponding subintervals lengths. Let V_M denote the asymptotic variance of the GMM estimator that uses the exactly identified unconditional moment function g_t^M, that is:

$$V_M = E\left[\frac{\partial g_t^M(\beta_0)}{\partial \beta'}\right]^{-1} E[g_t^M(\beta_0)g_t^M(\beta_0)']E\left[\frac{\partial g_t^M(\beta_0)'}{\partial \beta}\right]^{-1}. \tag{b.28}$$

Then, we have:

$$\Sigma \leq \Sigma_M^{gr} \leq V_M, \tag{b.29}$$

in the order of symmetric matrices, where the second inequality holds because asymptotic variance V_M corresponds to a GMM estimator that deploys the conditional moment restrictions for argument values u_m, with $m = 1, ..., M$, using a (in general) non-optimal instrument. Now, when the number of grid points M tends to infinity and the grid diameter Δ_M tends to 0, we have:

$$V_M \to \Sigma, \tag{b.30}$$

because of the convergence of the Riemann sums in (b.28) to the corresponding integrals in (b.26) and (b.27) under the continuity condition in Assumption SM.6 and an application of the Lebesgue theorem, and using equation (b.19). Then, (b.29) and (b.30) imply $\Sigma_M \to \Sigma$, which concludes the proof.

vi) Proof of equations (b.22) and (b.23)

Let us first prove the expression of function D_t given in equation (b.22). We use the definition of D_t in (b.15) and the $(i-1)$-th component of the gradient of the moment function in (b.21):

$$\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta'}\right]_{i-1} = \exp\{uy_{i,t} + (1-e^u)z_{i,t}'\beta_0\}(1-e^u)z_{i,t}' - \exp\{uy_{1,t} + (1-e^u)z_{1,t}'\beta_0\}(1-e^u)z_{1,t}',$$

where $i = 2, 3, ..., n$. From equation (6.4) evaluated at the true parameter value we have:

$$E[\exp(uy_{i,t})|\underline{f_t}, \underline{y_{t-1}}, \underline{x_t}] = \exp\{(f_t + z_{i,t}'\beta_0)(e^u - 1)\}. \tag{b.31}$$

Then, we get:

$$E\left[\left(\frac{\partial h_t(\beta_0, u)}{\partial \beta'}\right)_{i-1} |\underline{f_t}, \underline{y_{t-1}}, \underline{x_t}\right] = -\exp\{f_t(e^u - 1)\}(e^u - 1)(z_{i,t}' - z_{1,t}'),$$

i.e. in vector notation:

$$E\left[\frac{\partial h_t(\beta_0, u)}{\partial \beta'} |\underline{f_t}, \underline{y_{t-1}}, \underline{x_t}\right] = -\exp\{f_t(e^u - 1)\}(e^u - 1)\Delta z_t.$$

By the Law of Iterated Expectation, we get:

$$D_t(u) \;=\; E\left[E\left[\frac{\partial h_t(\beta_0,u)}{\partial \beta'}|\underline{f_t},\underline{y_{t-1}},\underline{x_t}\right]|\underline{y_{t-1}},\underline{x_t}\right] = -(e^u-1)E\left[\exp\{f_t(e^u-1)\}|\underline{y_{t-1}},\underline{x_t}\right]\Delta z_t,$$

which yields equation (b.22) using the definition of the conditional Laplace transform Ψ_t^f.

Let us now prove the expression for matrix $\Omega_t(u,\tilde{u})$ in equation (b.23). From the definition in (b.12), the $(i-1,j-1)$ element of matrix $\Omega_t(u,\tilde{u})$ is given by:

$$\begin{aligned}
[\Omega_t(u,\tilde{u})]_{i-1,j-1} &= E\left[(h_t(\beta_0,u))_{i-1}(h_t(\beta_0,\tilde{u}))_{j-1}|\underline{y_{t-1}},\underline{x_t}\right] \\
&= E\left[H_{i,t}(\beta_0,u)H_{j,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right] - E\left[H_{i,t}(\beta_0,u)H_{1,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right] \\
&\quad -E\left[H_{1,t}(\beta_0,u)H_{j,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right] + E\left[H_{1,t}(\beta_0,u)H_{1,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right],
\end{aligned}$$

for $i,j=2,3,...,n$. To compute these conditional expectations, we use:

$$\begin{aligned}
E\left[H_{i,t}(\beta_0,u)H_{i,t}(\beta_0,\tilde{u})|\underline{f_t},\underline{y_{t-1}},\underline{x_t}\right] &= E[\exp\{(u+\tilde{u})y_{i,t}\}|\underline{f_t},\underline{y_{t-1}},\underline{x_t}]\exp\{(2-e^u-e^{\tilde{u}})z_{i,t}'\beta_0\} \\
&= \exp\{(f_t+z_{i,t}'\beta_0)(e^{u+\tilde{u}}-1)+(2-e^u-e^{\tilde{u}})z_{i,t}'\beta_0\} \\
&= \exp\{(e^u-1)(e^{\tilde{u}}-1)z_{i,t}'\beta_0+f_t(e^{u+\tilde{u}}-1)\},
\end{aligned}$$

for all i, where the second equality follows from (b.31), and:

$$\begin{aligned}
E\left[H_{i,t}(\beta_0,u)H_{j,t}(\beta_0,\tilde{u})|\underline{f_t},\underline{y_{t-1}},\underline{x_t}\right] &= E\left[H_{i,t}(\beta_0,u)|\underline{f_t},\underline{y_{t-1}},\underline{x_t}\right]E\left[H_{j,t}(\beta_0,\tilde{u})|\underline{f_t},\underline{y_{t-1}},\underline{x_t}\right] \\
&= \exp\{f_t(e^u+e^{\tilde{u}}-2)\},
\end{aligned}$$

for $i \neq j$. Then, by the Law of Iterated Expectation, we get:

$$E\left[H_{i,t}(\beta_0,u)H_{i,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right] \;=\; \exp\{(e^u-1)(e^{\tilde{u}}-1)z_{i,t}'\beta_0\}\Psi_t^f(e^{u+\tilde{u}}-1),$$

for all i, and:

$$E\left[H_{i,t}(\beta_0,u)H_{j,t}(\beta_0,\tilde{u})|\underline{y_{t-1}},\underline{x_t}\right] \;=\; \Psi_t^f(e^u+e^{\tilde{u}}-2),$$

for $i \neq j$. Thus, the diagonal terms of matrix $\Omega_t(u,\tilde{u})$ are:

$$\begin{aligned}
[\Omega_t(u,\tilde{u})]_{i-1,i-1} &= \Psi_t^f(e^{u+\tilde{u}}-1)\left(\exp\{(e^u-1)(e^{\tilde{u}}-1)z_{i,t}'\beta_0\}+\exp\{(e^u-1)(e^{\tilde{u}}-1)z_{1,t}'\beta_0\}\right) \\
&\quad -2\Psi_t^f(e^u+e^{\tilde{u}}-2) \\
&= \omega_{i,t}(u,\tilde{u})+\omega_{1,t}(u,\tilde{u}),
\end{aligned}$$

and the out-of-diagonal terms are:

$$[\Omega_t(u,\tilde{u})]_{i-1,j-1} = \Psi_t^f(e^{u+\tilde{u}}-1)\exp\{(e^u-1)(e^{\tilde{u}}-1)z_{1,t}'\beta_0\} - \Psi_t^f(e^u+e^{\tilde{u}}-2) = \omega_{1,t}(u,\tilde{u}).$$

Using the matrix notation, we get equation (b.23).

References

Carrasco, M., Chernov, M., Florens, J.P., and E., Ghysels (2007): "Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions", Journal of Econometrics, 140, 529-573.

Carrasco, M., and J.-P., Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions", Econometric Theory, 16, 797-834.

Carrasco, M., and J.-P., Florens (2014): "On the Asymptotic Efficiency of GMM", Econometric Theory, 30, 372-406.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", Journal of Econometrics 34, 305-334.

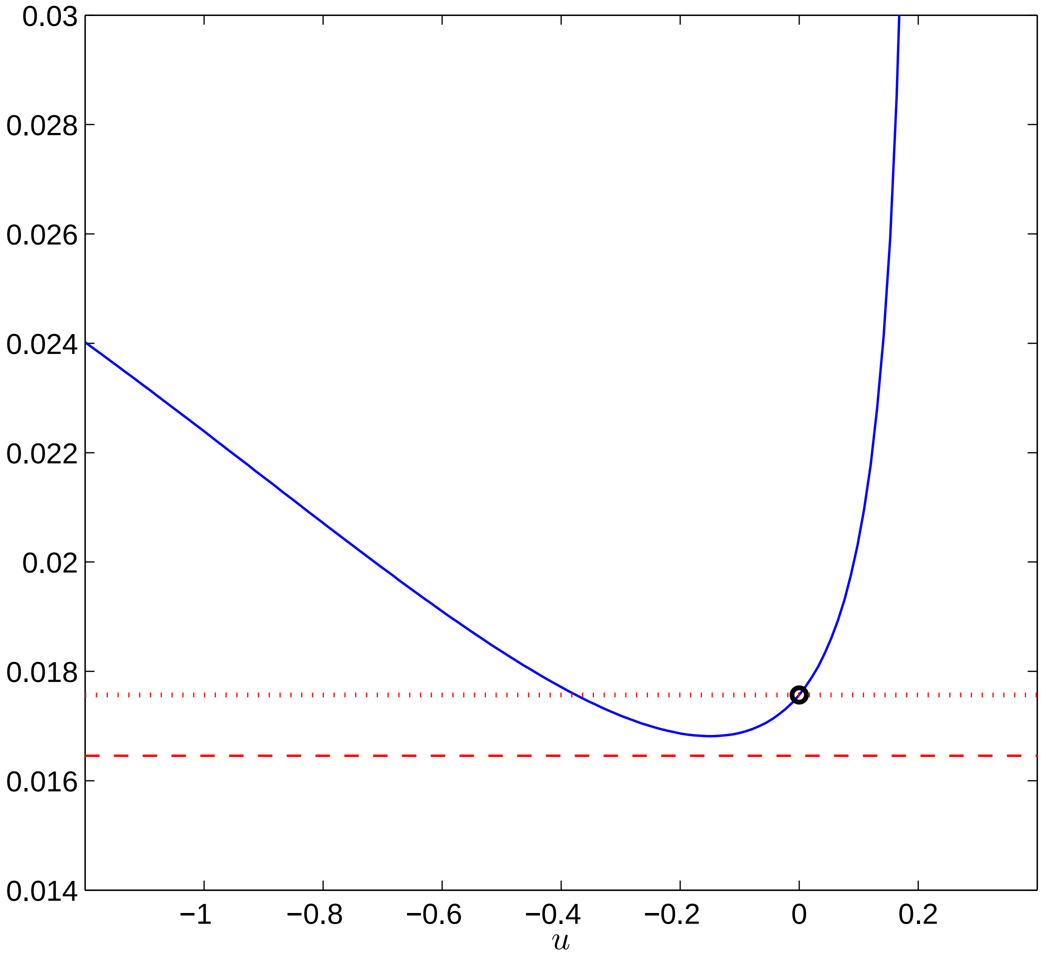
Gagliardini, P., and C., Gourieroux (2014): "Efficiency in Large Dynamic Panel Models with Common Factors", Econometric Theory, 30, 961-1020.

Hansen, L. (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators", Journal of Econometrics, 30, 203-238.

Kress, R. (1999): *Linear Integral Equations*, Springer Verlag.

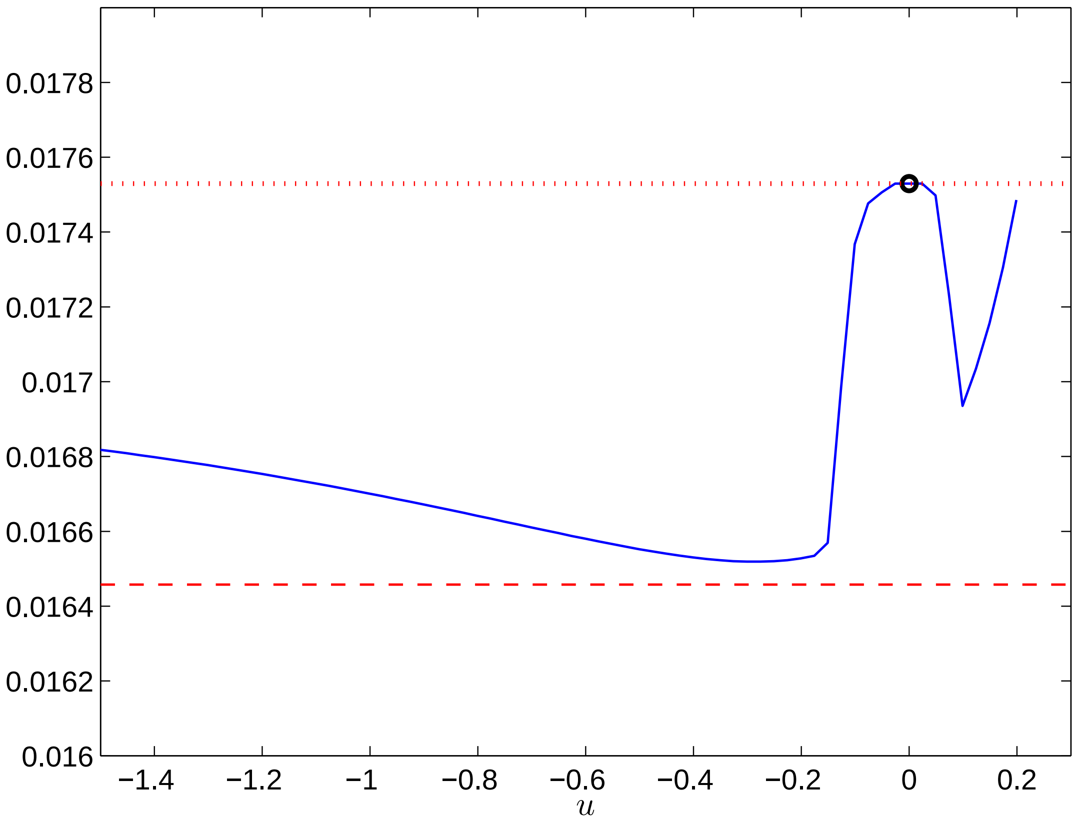
Singleton, K. (2001): "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function", Journal of Econometrics, 102, 111-141.

Figure 1: Efficiency loss for nonlinear cross-differencing with a single argument.



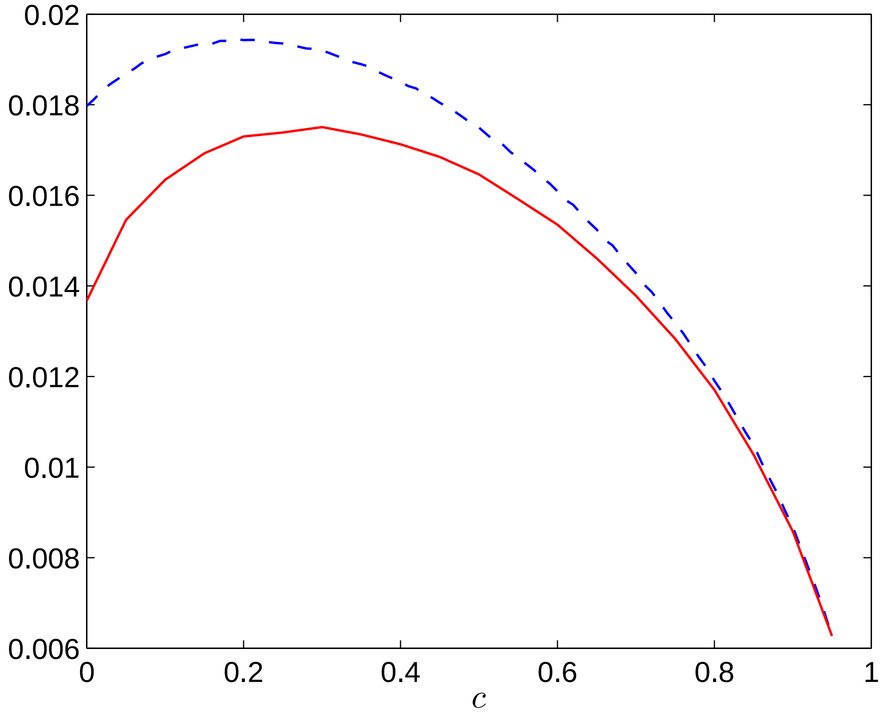
The blue curve displays the scaled asymptotic standard deviation $\sqrt{\Sigma_{1,cc}^{gr}(u)/T}$ for estimating the autoregressive parameter c in the Poisson count panel data model with stochastic time effect as a function of argument u, where $\Sigma_{1,cc}^{gr}(u)$ is the lower-right element in the asymptotic variance-covariance matrix of the efficient GMM estimator for parameter $\beta = (\alpha, c)'$ using the nonlinear cross-differencing conditional moment restrictions (6.5) with a single argument u, and $T = 100$. The horizontal red dashed line at value $\sqrt{\Sigma_{cc}/T}$ corresponds to the GMM efficiency bound for estimating parameter c using the nonlinear cross-differencing restrictions (6.5) for all admissible real arguments u. The horizontal red dotted line at value $\sqrt{\Sigma_{cc}^{*}/T}$ corresponds to the GMM efficiency bound for the linear cross-differencing conditional moment restrictions (6.6). The DGP is as in Section B.3.2 i) with $c = 0.5$ for $n = 25$.

Figure 2: Incremental informational content of nonlinear cross-differencing restrictions.



The blue curve displays the scaled asymptotic standard deviation $\sqrt{\Sigma^{gr}_{2,cc}(0,u)/T}$ for estimating the autoregressive parameter c in the Poisson count panel data model with stochastic time effect as a function of argument u, where $\Sigma^{gr}_{2,cc}(0,u)$ is the lower-right element in the asymptotic variance-covariance matrix of the efficient GMM estimator for parameter $\beta = (\alpha, c)'$ using jointly the linear cross-differencing restrictions (6.6) and the nonlinear cross-differencing restrictions (6.5) with a single argument u, and $T = 100$. The horizontal red dashed line at value $\sqrt{\Sigma_{cc}/T}$ corresponds to the GMM efficiency bound for estimating parameter c using the nonlinear cross-differencing conditional moment restrictions (6.5) for all admissible real arguments u. The horizontal red dotted line at value $\sqrt{\Sigma^*_{cc}/T}$ corresponds to the GMM efficiency bound for the linear cross-differencing conditional moment restrictions (6.6). The DGP is as in Section B.3.2 i) with $c = 0.5$ for $n = 25$.

Figure 3: Patterns of the efficiency bounds as functions of the autoregressive parameter c.



This Figure displays the patterns of the GMM efficiency bounds for estimating the autoregressive coefficient c in the Poisson count panel data model with stochastic time effect from linear and nonlinear cross-differencing restrictions, respectively, as a function of the parameter value c in the DGP. The solid red line displays $\sqrt{\Sigma_{cc}/T}$ as a function of parameter value c in the DGP, where Σ_{cc} is the lower-right element of the GMM efficiency bound matrix for estimating parameter $\beta = (\alpha, c)'$ using the continuum of nonlinear cross-differencing conditional moment restrictions in (6.5), and $T = 100$. The dashed blue line displays $\sqrt{\Sigma^*_{cc}/T}$ as a function of c, where Σ^*_{cc} is the lower-right element of the GMM efficiency bound matrix using the linear cross-differencing conditional moment restrictions in (6.6). The DGP is as in Section B.3.2 i) for $n = 25$.