Série des Documents de Travail

n° 2016-32

Composite Indirect Inference with Application to Corporate Risks

C.Gouriéroux[1]

A.Monfort[2]

[1] CREST and University of Toronto. E-mail: christian.gouriéroux@ensae.fr

[2] CREST. E-mail: alain.monfort@ensae.fr

Composite Indirect Inference with Application to Corporate Risks

C., GOURIEROUX [(1)] and A., MONFORT [(2)]

October, 23, 2016

The authors gratefully acknowledge support of the chair LCL : "New Challenges for New Data", and the chair ACPR : "Regulation and Systemic Risks".

[1]CREST and University of Toronto.
[1]CREST and Banque de France.

Composite Indirect Inference with Application to Corporate Risks

Abstract

It is frequent to deal with parametric models which are difficult to analyze, due to the large number of data and/or parameters, complicated nonlinearities, or unobservable variables. The aim of this paper is to explain how to analyze such models by means of a set of simplified models, called instrumental models, and how to combine these instrumental models in an optimal way. In this respect our paper provides a bridge between the econometric literature on indirect inference and the statistical literature on composite likelihood. The composite indirect inference principle is illustrated by an application to the analysis of corporate risks.

Keywords : Indirect Inference, Composite Likelihood, Instrumental Model, Pseudo Maximum Likelihood, Yule-Walker Equation, Corporate Risk, Basel Regulation, Asymptotic Single Risk Factor.

1 Introduction

It is frequent to deal with parametric models which are difficult to analyze due to the large number of data and/or parameters, complicated nonlinearities, or unobservable variables. The aim of our paper is to explain how to analyze such models by means of a set of simplified models, called instrumental models, and how to combine these instrumental models in an optimal way. In this respect our paper provides a bridge between the econometric literature on indirect inference [Gourieroux, Monfort, Renault (1993), Smith (1993), Gallant, Tauchen (1996), Gourieroux, Monfort (1996) a], and the statistical literature on composite likelihood [Lindsay (1988), Varin, Vidoni (2005), Varin et al. (2011)]. The Composite Indirect Inference (CII) approach is described in Section 2, where we also derive and discuss the asymptotic properties of the CII estimator of the true parameter value. We discuss the informational content on this true value of an additional instrumental model. This notion is used to introduce overidentification tests in the spirit of Szroeter (1983). The implementation of the CII methodology is illustrated in Section 3 for the monitoring of corporate risks with a dynamic extension of the Asymptotic Single Risk Factor (ASRF) model introduced by Vasicek (1991) and currently the core of Basel regulation for credit risk. Section 4 concludes. Appendices derive and discuss the asymptotic properties of CII estimators.

2 The Approach

We describe in this section the principles of the composite indirect inference approaches and the properties of the associated composite indirect inference (CII) estimators. The derivation of the asymptotic properties of these CII estimators is given in Appendix 1 for models without exogenous variables, and in Appendix 2 when exogenous variables are introduced.

2.1 The Data Generating Process (DGP)

For expository purpose, we consider i.i.d. observations $y_t, t = 1, \ldots, T$ (see Section 2.5 and Appendices 1-2 for various extensions). We assume that the observations have been generated by a true distribution f_0, belonging to a parametric model, $f(y;\theta)$ with parameter $\theta, \theta \in \Theta$, and with $f_0 = f(y;\theta_0)$, where θ_0 denotes the true parameter value. There exist different parametric models containing the true distribution. The selected one is used for two purposes: i) to define an estimation method like the maximum likelihood, ii) to simulate artificial datasets. Due to this second use this selected well-specified model is often called the Data Generating Process (DGP). This terminology is used later on.

In principle, parameter θ_0 can be estimated by maximum likelihood (ML), that is, by computing :

$$\hat{\theta}_T = \arg\max_{\theta} \sum_{t=1}^{T} \log f(y_t; \theta). \tag{2.1}$$

However, the computation of the ML estimate can be difficult and/or numerically inaccurate when the number of observations becomes large (i.e. T and/or $n = dim(y_t)$ large).

2.2 The Instrumental Model (IM)

An instrumental model is a simplified version of the DGP, which is easier to analyze. There exist different ways to simplify the DGP, such as the linearization of nonlinear features with respect to variables, or to parameters, the omission of some dependence existing between the components of y, the aggregation of outcomes of y, or a diminution of the number of parameters.

We denote by $IM_k = [g_k(y_t; \beta_k), \beta_k \in B_k], k = 1, \dots, K.$ the instrumental models. The parametric model IM_k depends on a parameter β_k, whose dimension can be smaller, equal, or larger than the dimension of the initial parameter θ.

Each instrumental model IM_k can be estimated by the pseudo-maximum likelihood (PML). The PML estimator of β_k is :

$$\hat{\beta}_{k,T} = \arg\max_{\beta_k} \sum_{t=1}^{T} \log g_k(y_t; \beta_k). \tag{2.2}$$

These PML estimators converge (almost surely) to the pseudo-true value $b_k(\theta_0)$ of β_k defined as the solution of the asymptotic maximization problem:

$$b_k(\theta_0) = \arg\max_{\beta_k} E_{\theta_0} \log g_k(y; \beta_k).$$

It also satisfies the first-order condition :

$$
\begin{aligned}
& E_{\theta_0}\left[\frac{\partial}{\partial \beta_k}\log g_k(y_t;b_k(\theta_0))\right] = 0 \\
\iff & \int \frac{\partial}{\partial \beta_k}\log g_k(y_t;b_k(\theta_0)) f(y_t;\theta_0) dy_t = 0.
\end{aligned} \tag{2.3}
$$

Asymptotically the PML estimators are such that :

$$
\sqrt{T}[\hat{\beta}_{k,T} - b_k(\theta_0)] \simeq \left[E_{\theta_0}\left(\frac{-\partial^2 \log g_k(y_t;b_k(\theta_0))}{\partial \beta_k \partial \beta_k'}\right)\right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial \beta_k}. \tag{2.4}
$$

2.3 How to "Aggregate" the Instrumental Models ?

By considering K instrumental models, we are implicitly replacing the initial observations $(y_t, t = 1, \ldots, T)$ by a smaller set of summary statistics $\hat{\beta}_{k,T}, k = 1, \ldots, K$. Introducing the binding function $b(\theta) = [b_1'(\theta), \ldots, b_K'(\theta)]'$, these summary statistics asymptotically satisfy the model[2] :

$$
\hat{\beta}_T' = \left[\hat{\beta}_{1,T}', \ldots, \hat{\beta}_{K,T}'\right]' \equiv b(\theta) + U_T, \tag{2.5}
$$

where U_T is multivariate Gaussian with :

$$
E(U_T) = 0, V(U_T) = \Sigma_0 = J_0^{-1}\Omega_{g,0}J_0^{-1}, \tag{2.6}
$$

where J_0 is the block-diagonal matrix with diagonal blocks :

$$
J_{\beta_k\beta_k} = E_{\theta_0}\left[-\frac{\partial^2 \log g_k(y, b_k(\theta_0))}{\partial \beta_k \partial \beta_k'}\right],
$$

and the blocks of $\Omega_{g,0} = V_{\theta_0}[\frac{\partial \log g}{\partial \beta}(y; b(\theta_0))]$ are :

[2]Note that $\hat{\beta}_T$ can be alternatively derived from a single optimization problem :

$$
\hat{\beta}_T = \arg\max_{\beta} \sum_{k=1}^{K}\sum_{t=1}^{T} \log g_k(y_t;\beta_k) \equiv \arg\max_{\beta} \sum_{t=1}^{T} \log g(y_t;\beta),
$$

with respect to $\beta' = (\beta_1', \ldots, \beta_K')'$. When the parameter size and the number K of instrumental models are large, it can be more convenient computationally to perform the maximizations (2.2) in parallel than to solve the single optimization above.

$$Cov_{\theta_0}\left[\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial \beta_k},\frac{\partial \log g_l(y_t;b_l(\theta_0))}{\partial \beta_l}\right]$$

$$= \quad E_{\theta_0}\left[\frac{\partial \log g_k(y_t;b_k(\theta_o))}{\partial \beta_k}\frac{\partial \log g_l(y_t;b_l(\theta_0))}{\partial \beta_l'}\right],$$

where $\frac{\partial \log g}{\partial \beta'}(y,b(\theta_0)) \equiv [\frac{\partial \log g_1}{\partial \beta_1'}(y;b_1(\theta_0)),\ldots,\frac{\partial \log g_K}{\partial \beta_K'}(y;b_K(\theta_0))]$.

i) Known binding function

When the binding functions $b_k : \theta \to b_k(\theta), k = 1,\ldots,K$, are known in closed form, these summary statistics can be combined in an optimal way by Asymptotic Least Squares (ALS)[see e.g. Gourieroux, Monfort, Trognon (1986), Gourieroux, Monfort (1996)b, Kodde, Palm, Pfann (1990)] to get an estimator of θ_0. This estimator, called composite indirect inference (CII) estimator, is the solution of :

$$\tilde{\theta}_T = \arg\min_{\theta}[\hat{\beta}_T - b(\theta)]'\hat{\Sigma}_T^{-1}[\hat{\beta}_T - b(\theta)], \tag{2.7}$$

where $\hat{\Sigma}_T$ is a consistent estimator of Σ_0 given in (2.6).

Under appropriate identification conditions (see Appendix 1 for the order and rank conditions for identification), this estimator is consistent :

$$plim_{T\to\infty}\tilde{\theta}_T = \theta_0,$$

asymptotically normal and such that :

$$V_{as}[\sqrt{T}(\tilde{\theta}_T - \theta_0)] = (\frac{db'(\theta_0)}{d\theta}\Sigma_0^{-1}\frac{db(\theta_0)}{d\theta'})^{-1} = [\Omega_{fg,0}\Omega_{g,0}^{-1}\Omega_{gf,0}]^{-1},$$

where :

$$\begin{aligned}\Omega_{fg,0} &= E_{\theta_0}[\frac{\partial \log f(y;\theta_0)}{\partial \theta}\frac{\partial \log g(y;b(\theta_0))}{\partial \beta'}] \\ &= Cov_0\left[\frac{\partial \log f(y;(\theta_0))}{\partial \theta},\frac{\partial \log g(y;b(\theta_0))}{\partial \beta}\right], \\ \Omega_{g,0} &= V_{\theta_0}[\frac{\partial \log g(y;b(\theta_0))}{\partial \beta}].\end{aligned} \tag{2.8}$$

It is possible to evaluate the loss of efficiency due to the use of instrumental models, by comparing with the optimal ML estimator which has an asymptotic variance-covariance matrix equal to $\Omega_{f,0}^{-1}$, where $\Omega_{f,0}$ is the Fisher information matrix for one observation, i.e. $\Omega_{f,0} = V_{\theta_0}\left[\frac{\partial \log f(y,\theta_0)}{\partial\theta}\right]$.

Indeed we have :

$$V_{\theta_0}\left[\begin{array}{c}\frac{\partial \log f(y,\theta_0)}{\partial\theta}\\ \frac{\partial \log g(y,b(\theta_0))}{\partial\beta}\end{array}\right] = \left[\begin{array}{cc}\Omega_{f,0} & \Omega_{fg,0}\\ \Omega_{gf,0} & \Omega_{g,0}\end{array}\right].$$

By using the multivariate Cauchy-Schwarz inequality :

$$\Omega_{f,0} >> \Omega_{fg,0}\Omega_{g,0}^{-1}\Omega_{gf,0}$$

where $>>$ denotes the standard ordering on symmetric matrices, we get :

$$Vas\ [\sqrt{T}(\hat{\theta}_T - \theta_0)] = \Omega_{f,0}^{-1} << (\Omega_{fg,0}\Omega_{g,0}^{-1}\Omega_{gf,0})^{-1} = Vas[\sqrt{T}(\tilde{\theta}_T - \theta_0)].$$

Moreover $\Omega_{fg,0}\Omega_{g,0}^{-1}\Omega_{gf,0}$ is the variance-covariance matrix of the vector obtained by projecting in L_2 each component of $\frac{\partial \log f(y;\theta_0)}{\partial\theta}$ on the space G spanned by the components of $\frac{\partial \log g(y;b(\theta_0))}{\partial\beta}$, that is the variance-covariance matrix of $LE_{\theta_0}[\frac{\partial \log f(y;\theta_0)}{\partial\theta}|\frac{\partial \log g(y;b(\theta_0))}{\partial\beta}]$, where LE_{θ_0} denotes the theoretical linear regression (since all the variables involved are zero-mean). Therefore, if the components of $\frac{\partial \log f(y,\theta_0)}{\partial\theta}$ are close to G, the approximation by instrumental models will be good. In the extreme case where these components belong to $G, \tilde{\theta}_T$ is asymptotically efficient.

ii) Unknown binding functions

The approach above can be extended to unknown binding functions by applying indirect inference. Let us assume that we know how to simulate artificial data from the DGP, that is, for any given value θ, we know how to draw $y_t^s(\theta), t = 1, \ldots, ST$, in $f(y_t;\theta)$. Then we can introduce the PML estimator of β_k based on this artificial data set :

$$\hat{\beta}_k^s(\theta) = \arg\max_{\beta_k} \sum_{t=1}^{ST} \log g_k(y_t^s(\theta);\beta_k). \tag{2.9}$$

This estimator tends to $b_k(\theta)$, when T tends to infinity, and can be used to estimate the unknown binding functions.

The composite indirect inference (CII) estimator of θ is now the solution of :

$$\tilde{\theta}_T^s = \arg\min_{\theta} [\hat{\beta}_T - \hat{\beta}_T^s(\theta)]' \hat{\Sigma}_T^{-1} [\hat{\beta}_T - \hat{\beta}_T^s(\theta)], \tag{2.10}$$

where $\hat{\beta}_T^s(\theta) = [\hat{\beta}_1^s(\theta)', \ldots, \hat{\beta}_K^s(\theta)']'$.

It is consistent, asymptotically normal, with asymptotic variance :

$$V_{as}[\sqrt{T}(\tilde{\theta}_T^s - \theta_0)] = (1 + \frac{1}{S}) V_{as}[\sqrt{T}(\tilde{\theta}_T - \theta_0)], \tag{2.11}$$

with a simple effect of the number S of replications when simulating the artificial data (see Appendix 1 ix) and the references therein). Of course for $S \to \infty$, we are back to the case of known binding functions.

2.4 Alternative CII Estimators

Alternative indirect inference estimators have been proposed in the literature, either based on an appropriate minimization of pseudo-scores [Gallant, Tauchen (1996)], or on the optimization of composite likelihood functions [see the survey in Varin et al. (2011)]. We describe these alternative approaches and explain why the based pseudo-scores CII estimator are asymptotically equivalent to the CII estimators introduced above, whereas the CII estimators based on a composite likelihood are consistent, but generally less efficient.

i) Score based CII Estimators

Let us consider the score of the instrumental model IM_k evaluated at the simulated data set and PML estimates :

$$\sum_{t=1}^{ST} \frac{\partial \log g_k(y_t^s(\theta); \hat{\beta}_{kT})}{\partial \beta_k}.$$

It can be shown [see e.g. Gallant, Tauchen (1996), Gourieroux, Monfort (1996)a, Chapter 4, Appendix A2], that, for θ close to the true value θ_0, we have :

$$\frac{1}{\sqrt{ST}} \sum_{t=1}^{ST} \frac{\partial \log g_k(y_t^s(\theta); \hat{\beta}_{kT})}{\partial \beta_k} \simeq J_{\beta_k \beta_k} \sqrt{T} [\hat{\beta}_{kT} - \hat{\beta}_{kT}^s(\theta)]. \tag{2.12}$$

The equivalence (2.12) is used to define an alternative CII estimator of θ by :

$$\hat{\theta}_T^s = \arg\min_\theta \sum_{t=1}^{ST} \frac{\partial \log g(y_t^s(\theta); \hat{\beta}_T)}{\partial \beta'} [\hat{V}(\frac{\partial \log g_k}{\partial \beta_k})]^{-1} \sum_{t=1}^{ST} \frac{\partial \log g(y_t^s(\theta); \hat{\beta}_T^s)}{\partial \beta}, \tag{2.13}$$

where $\frac{\partial \log g(y_t^s(\theta), \hat{\beta}_T)}{\partial \beta'} = [\frac{\partial \log g_1(y_t^s(\theta); \hat{\beta}_{1T})}{\partial \beta_1'}, \ldots, \frac{\partial \log g_K(y_t^s(\theta); \hat{\beta}_{KT})}{\partial \beta_K'}]$.

This estimator has the same asymptotic behaviour as the CII estimator $\tilde{\theta}_T^s$.

ii) Likelihood Based CII Estimators

Instead of aggregating the PML estimators (see Section 2.3), or the pseudo-scores (see Section 2.4 i)), it has also been suggested in the literature to aggregate directly the pseudo log-likelihood functions of the instrumental models [see e.g. Varin et al. (2011)]. Let us introduce a set of weights : $\alpha_k, k = 1, \ldots, K$, with $\alpha_k > 0, \forall k, \sum_{k=1}^{K} \alpha_k = 1$. The associated estimators are :

$$\theta_T^* = \arg\max_\theta \sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_k \log g_k(y_t; b_k(\theta)), \tag{2.14}$$

for known binding functions and :

$$\theta_T^{*s} = \arg\max_\theta \sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_k \log g_k(y_t; \hat{\beta}_{kT}^s(\theta)), \tag{2.15}$$

when the binding functions are estimated by simulations.

It is easily checked that such estimators are consistent. However, these estimators are less efficient than the estimators $\tilde{\theta}_T, \tilde{\theta}_T^s$ introduced above (see Appendix 3). Indeed the weights correspond to a diagonal metric, with the same weights for the different parameters of a given instrumental model, and this diagonal metric is generally different from the metric based on optimal weights.

2.5 Extensions

The asymptotic expansions used to derive the CII estimators and their asymptotic behaviours are valid in more complicated frameworks including time series, lagged endogenous variables, and instrumental objective functions (see Appendix 1).[3]. In this extended framework, we assume that the process (y_t) is stationary satisfying mixing conditions to allow for the derivation of asymptotic results.

The instrumental models are replaced by instrumental objective functions of the type :

$$\sum_{t=1}^{T} \log g_k(y_t, \beta_k), k = 1, \dots, K,$$

which are optimized to derive estimators $\hat{\beta}_{kT}$, but are not necessarily interpretable as the log-likelihood functions of a misspecified model.

Let us illustrate these extensions by the case of a Markov process of order p, (y_t^*), say. The DGP is parametric and characterized by the conditional density of y_t^* given $y_{t-1}^*, \dots, y_{t-p}^*$, or equivalently by the stationary distribution of $y_t = (y_t^*, y_{t-1}^*, \dots, y_{t-p}^*)'$.

How to analyze by instrumental objective functions the parameters characterizing the dynamics of (y_t^*)? Different objective functions have been suggested in the literature. If the binding functions corresponding to these objective functions are known in closed form the method of Section 2.3.i can be applied, otherwise indirect inference techniques of Sections 2.3 ii or 2.4.i are required.

i) Marginal pseudo-likelihood (see e.g. Cox, Reid (2004), Varin (2008)).

[3]The results can also be extended to semi-parametric models with exogenous variables (see the discussion in Appendix 2), or to spatial processes.

The instrumental objective function is :

$$\sum_{t=1}^{T} \log f(y_t^*; \beta), \tag{2.16}$$

where f is the marginal pdf of y_t^*, that is the stationary density of the process. The parameter β differs from the parameter θ of the DGP, since it includes only the element of θ with marginal interpretations. This objective function corresponds to the log-likelihood function of a misspecified instrumental model, in which the $y_t^{*\prime} s$ have been assumed serially independent.[4]

ii) Pairwise "pseudo-likelihood" at lag h [see e.g. Gourieroux, Monfort, Trognon (1984), p338].

The instrumental objective functions are :

$$\sum_{t=1}^{T} \log f_h(y_t^*, y_{t-h}^*; \beta_h), h = 1, \ldots, H, \tag{2.17}$$

where f_h is the marginal joint distribution of (y_t^*, y_{t-h}^*). Despite its name this objective function can no longer be interpreted as the log-likelihood function of a misspecified model (see e.g. the discussion in Section 3.1).

By considering these pairwise pseudo-likelihoods, we implicitly extend to nonlinear dynamics the Yule-Walker estimation approach proposed for linear ARMA models.

2.6 Additional Instrumental Model

Let us now discuss the effect of introducing in the analysis an additional instrumental model IM_{K+1}, say. We consider the case of known binding functions for expository purpose, but the analysis is easily extended to unknown binding functions.

We can compute the CII estimator $\hat{\theta}_{K,T}$ based on the first K instrumental models and the CII estimator $\hat{\theta}_{K+1,T}$ based on the $K+1$ instrumental models. These estimators are both consistent and asymptotically jointly normal. Moreover from the interpretation in terms of projections given in Section 2.3

[4] This explains the alternative terminology "independence likelihood" used in Chandler, Bate (2007).

i), we deduce that $\hat{\theta}_{K+1,T}$ is more accurate than $\hat{\theta}_{K,T}$. The gain in accuracy can be measured by means of :

$$G_{K+1|K} = V_{as}^{-1}[\sqrt{T}(\hat{\theta}_{K+1,T} - \theta_0)] - V_{as}^{-1}[\sqrt{T}(\hat{\theta}_{K,T} - \theta_0)]. \tag{2.18}$$

Given the interpretation of these quantities mentioned in Section 2.3, this gain is equal to the variance-covariance matrix of :

$$LE_{\theta_0}[\frac{\partial \log f(y;\theta_0)}{\partial \theta} | \frac{\partial \log g_{K+1}(y; b_{K+1}(\theta))}{\partial \beta_{K+1}} - LE_{K,\theta_0}(\frac{\partial \log g_{K+1}(y; b_{K+1}(\theta_0))}{\partial \beta_{K+1}})],$$

where $LE(.|.)$ denotes the theoretical linear regression (in L^2) of a variable on the conditioning ones, and LE_{K,θ_0} the linear regression on the first K pseudo-scores.

Therefore we have :

$$G_{K+1|K} = Q_{f,g_{K+1}|\underline{g_K}} Q^{-1}_{g_{K+1}|\underline{g_K}} Q'_{f,g_{K+1}|\underline{g_K}}, \tag{2.19}$$

where $Q_{f,g_{K+1}|\underline{g_K}}$ (resp. $Q_{g_{K+1}|\underline{g_k}}$) is the covariance matrix of the true score $\frac{\partial \log f(y;\theta_0)}{\partial \theta}$ and the $(K+1)^{th}$ pseudo-score $\frac{\partial \log g_{K+1}(y; b_{K+1}(\theta_0))}{\partial \beta_{K+1}}$ (resp. the variance-covariance matrix of the $(K+1)^{th}$ pseudo-score) given the K previous pseudo-scores $\frac{\partial \log g_k(y; b_k(\theta_0))}{\partial \beta_k}, k = 1, \ldots, K$. This gain is a partial covariance matrix. This gain measures the additional information on θ given by the $(K+1)^{th}$ instrumental model.

3 Analysis of Corporate Risks

Let us now illustrate how to implement composite indirect inference to corporate risk analysis. We consider the rating history of a given corporate evolving between investment or speculative grade [see e.g. Gourieroux, Jasiak (2001), Gagliardini, Gourieroux (2005) for the modelling of the migration between ratings]. We first describe the standard approach suggested in Basel regulation and its interpretation in terms of composite pseudo-likelihood. Then we extend the estimation approach i) to account for more lags, ii) to allow for non Gaussian latent model, iii) to account for systematic risk.

In this application, the process y_t is not i.i.d. and, therefore, the relevant asymptotic behavior of the estimators is the one described in Appendix 1.

3.1 The standard model

The standard approach in Basel 2 defines the rating history $y_t = 1$ (investment grade) at date t, $y_t = 0$, (speculative grade), from the values of a quantitative latent variable y_t^*, interpreted in terms of log asset/liability ratio [see e.g. Merton (1974)]. More precisely, the basic (dynamic) model is [see e.g. Merton (1974), Crouhy et al. (2000)] :

$$y_t^* = m + \rho(y_{t-1}^* - m) + \sigma u_t, \tag{3.1}$$

where (u_t) is a sequence of independent standard normal variables, $u_t \sim N(0,1)$, and (y_t^*) is the stationary solution of this autoregressive equation. Then the observable dichotomous variable is defined by :

$$y_t = 1, \text{if } y_t^* > c, y_t = 0, \text{ otherwise.} \tag{3.2}$$

Since the inequality $y_t^* > c$ is equivalent to $\dfrac{y_t^* - m}{\sigma} > \dfrac{c-m}{\sigma}$, all latent parameters are not identifiable and we can introduce as identification restrictions $c = 0, \sigma = 1$.

Then the marginal distribution of y_t is characterized by :

$$\begin{aligned} P(y_t = 1) &= P(y_t^* > 0) = P(m + (1/\sqrt{1-\rho^2})U > 0) \\ &= \Phi\left(m\sqrt{1-\rho^2}\right), \end{aligned} \tag{3.3}$$

where U is standard normal $N(0,1)$, with c.d.f. Φ.

The marginal log p.d.f. of y_t is :

$$\log f_1(y_t;\delta) = y_t \log \Phi(\delta) + (1-y_t)\log\Phi(-\delta)$$

where $\delta = m\sqrt{1-\rho^2}$.

In the joint distribution of (y_t, y_{t-h}) we have :

$$
\begin{aligned}
& P(y_t = 1, y_{t-h} = 1) = P(y_t^* > 0, y_{t-h}^* > 0) \\
= \;& P(-\sqrt{1-\rho^2}(y_t^* - m) < m\sqrt{1-\rho^2}, -\sqrt{1-\rho^2}(y_{t-h}^* - m) < m\sqrt{1-\rho^2}) \\
= \;& \psi(m\sqrt{1-\rho^2}, m\sqrt{1-\rho^2}, \rho^h), \qquad (3.4) \\
= \;& \psi(\delta, \delta, \rho^h),
\end{aligned}
$$

where $\psi(a, b, \rho)$ is the bivariate c.d.f. $\psi(a, b, \rho) = P[U < a, V < b]$ of a bivariate Gaussian vector $(U, V)' \sim N[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}]$.

Therefore the joint log p.d.f. of (y_t, y_{t-h}) is :

$$
\begin{aligned}
\log f_{2,h}(y_t, y_{t-h}; \delta, \rho) \;=\;& y_t y_{t-h} \log \psi(\delta, \delta, \rho^h) \\
+\;& [y_t(1 - y_{t-h}) + (1 - y_t)y_{t-h}] \log[\Phi(\delta) - \psi(\delta, \delta, \rho^h)] \\
+\;& (1 - y_t)(1 - y_{t-h}) \log[1 - 2\Phi(\delta) + \psi(\delta, \delta, \rho^h)].
\end{aligned}
$$

i) The recursive estimation method

In the standard approach for Basel regulation the relations (3.3)-(3.4) are used to calibrate parameters m and ρ by solving the system :

$$
\begin{cases}
\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_t = \Phi\left(m\sqrt{1-\rho^2}\right) = \Phi(\delta), \text{ with } \delta = m\sqrt{1-\rho^2}, \\[2ex]
\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_t y_{t-1} = \psi\left(m\sqrt{1-\rho^2}, m\sqrt{1-\rho^2}, \rho\right) = \psi(\delta, \delta, \rho).
\end{cases} \qquad (3.5)
$$

This provides the estimator of δ as :

$$
\hat{\delta} = \Phi^{-1}(\frac{1}{T}\sum_{t=1}^{T} y_t),
$$

and then $\hat{\rho}$ is deduced by solving :

$$\frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} = \psi(\hat{\delta}, \hat{\delta}, \rho).$$

This approach is simply a recursive pseudo-likelihood approach, based on the marginal likelihood and the pairwise likelihood at horizon 1 :

$$\sum_{t=1}^{T} \log f_1(y_t; \delta) \text{ and } \sum_{t=1}^{T} f_{2,1}(y_t, y_{t-1}; \delta, \rho), \text{ respectively.}$$

ii) Alternative estimation methods

These recursive estimators differ from the estimators obtained by maximizing the composite pseudo-likelihood defined as the sum of these marginal and pairwise pseudo-likelihoods. They also differ from the optimal CII estimator based on these functions. These three estimation approaches are all consistent, but have different finite sample properties.

To compare these finite sample properties, we perform below a Monte Carlo study. The number of observations is $T = 100$, and the values of the parameters are $\delta = 0, \rho = 0, .25, .50, .75$. The number of replications is 500. We consider three estimation methods. The first one is the recursive approach (R) described above. The second one is a composite pseudo-likelihood method (CPML1) based on the maximisation of :

$$\sum_{t=1}^{T} [\log f_1(y_t; \delta) + \log f_{2,1}(y_t, y_{t-1}; \delta, \rho)],$$

which provides consistent estimators. The third one is the composite indirect inference method (CII) based on the objective functions $\sum_{t=1}^{T} \log f_1(y_t; \beta_1)$ and $\sum_{t=1}^{T} \{\log f_1(y_t, \beta_{21}) + \log f_{21}(y_t, y_{t-1}; \beta_{21}, \beta_{22}\}$ and the known binding functions $b_1(\delta, \rho) = \delta, b_{21}(\delta, \rho) = \delta, b_{22}(\delta, \rho) = \rho$.

The results are given of the estimators R, CPML1 and CII of Table 1, showing the RMSE (Root Mean Square Error).

	R		CPML1		CII		CPML2	
	δ	ρ	δ	ρ	δ	ρ	δ	ρ
$\rho = 0$	.124	.151	.123	.151	.123	.148	.122	.148
$\rho = .25$	.150	.150	.149	.150	.149	.149	.148	.149
$\rho = .50$	.203	.142	.202	.141	.202	.140	.201	.136
$\rho = .75$	.286	.116	.285	.115	.285	.114	.284	.109

TABLE 1 : RMSE of the estimators of $\delta(\delta = 0)$ and ρ.

As expected the RMSE of the estimators of δ is increasing with ρ, whereas the RMSE of the estimators of ρ is decreasing with ρ.

The performance of the CII is always equal or slightly better than the performance of the other estimators. The very modest superiority of the CII estimator in finite sample is reminiscent of the finite sample performance of the optimal GMM estimators. Indeed in both cases the optimal weights are based on first step estimators impacting the precision of the second step estimators.

In the previous methods we only consider one lag. Additional information could be obtained by introducing more than one lag. Indeed, although the latent process is Markov of order 1, it is not the case for the observed process y_t. Let us for instance consider the composite pseudo likelihood estimator involving two lags (CPML2), i.e. the estimator obtained by maximizing :

$$\sum_{t=1}^{T}[\log f_1(y_t, \delta) + \sum_{h=1}^{2} f_{2,h}(y_t, y_{t-h}; \delta, \rho^h)].$$

The last columns of Table 1 show that, as far as the estimation of ρ is concerned, the difference between the RMSE of the CPML1 and CPML2 methods could be of approximately 5% for large values of ρ. We have checked that introducing much more lags do not significantly improve the estimation.

3.2 NonGaussian Latent Model

Let us now assume that the true autoregressive model is :

$$y_t^* = m + \rho(y_{t-1}^* - m) + u_t^*, \tag{3.6}$$

where the $u_t^{*'}s$ are i.i.d. with a given distribution f_0^*. Then the composite pseudo-likelihood approach can no longer be used. Indeed, to compute the marginal pseudo-likelihood we need the expression of the stationary density of u_t^*, which has no closed from expression except in special cases such as :

the Gaussian case : $(y_t^* - m)\sqrt{1-\rho^2} \sim N(0,1)$.

the Cauchy case : $(y_t^* - m)(1-|\rho|) \sim$ Cauchy.

However, we can now apply to model (3.6) composite indirect inference based on either a Gaussian autoregressive process, or a Cauchy autoregressive process. Since the distribution of the innovations u_t^* is misspecified, the PML estimators are not consistent and have to be adjusted by indirect inference. Since the binding functions are unknown, they must be computed by simulation techniques.

Different CII approaches can be considered. They depend :

- on the maximum lag H;
- on the selected pseudo-distribution of the innovation;
- on the constraints introduced on the parameters of the pseudo-distribution.

More precisely the pairwise Gaussian pseudo log-likelihood at horizon h can be written as : $\Sigma_t \log g_{2,h}(y_t, y_{t-h}; \delta, \beta^h)$, or as $\Sigma_t \log g_{2,h}(y_t, y_{t-h}, \delta, \beta_h)$, where $g_{2,h}$ is deduced from formula (3.4). That is, we can constrain the β_h parameters to be entirely compatible with the (misspecified) Gaussian autoregressive model, or left them unconstrained.

The same remark can be done for the Cauchy autoregressive process, where the new $g_{2,h}$ function is constructed from :

$$P[y_t^* > 0, y_{t-h}^* > 0] = \tilde{\psi}(m(1-(|\rho|), m(1-|\rho|), \rho^h),$$

where : $\tilde{\psi}(a,b,\rho) = P[\rho U + (1-|\rho|)W < a, U < b)$, and U, W are independent Cauchy variables.

As an illustration let us consider the finite sample properties of the CII estimator when $m = 0., \rho = .75$ and the u_t^* in (3.6) have a Student distribution with 8 degrees of freedom (and a unit variance). We consider the Composite Indirect Inference using the objective functions $\sum_{t=1}^{T} \log f_1(y_t, \beta_1)$ and

$\sum_{t=1}^{T}[\log f_1(y_t;\beta_{21}) + \log f_{21}(y_t, y_{t-1};\beta_{21},\beta_{22})]$, where f_1 and f_{21} are the Gaussian p.d.f. given above. We have to first maximize these functions when the y_t are the observations: we get $\hat{\beta}_{1,T}$ and $(\hat{\beta}_{21,T}, \hat{\beta}_{22,T})$. Then we have to estimate the binding functions using simulated paths $y_t^s(\delta,\rho), t = 1,\ldots,ST$ and maximizing the objective functions :

$\sum_{t=1}^{ST} \log f_1[y_t^s(\delta,\rho),\beta_1]$ with respect to β_1, giving $\hat{\beta}_1^s(\delta,\rho)$, and

$\sum_{t=1}^{T}[\log f_1[y_t^s(\delta,\rho);\beta_{21}]+\log f_{21}[y_t^s(\delta,\beta);\beta_{21},\beta_{22}]$ with respect to $\beta_2 = (\beta_{21},\beta_{22})$ giving $\hat{\beta}_2^s(\delta,\rho)$.

Since functions $\log f_1$ and $\log f_{21}$ are sum of terms which are products of functions of $y_t^s(\delta,\rho)$ by functions of (β_1,β_2), in the objective functions above the terms depending on the simulated path $y_t^s(\delta,\rho), t = 1,\ldots,ST$, can be computed independently of β_1 and β_2. Therefore we can take S large without computational cost ($S = 100$ in the application). In other words the $\hat{\beta}_1^s(\delta,\rho)$ and $\hat{\beta}_2^s(\delta,\rho)$ can be very close to the unknown binding functions $b_1(\delta,\rho)$ and $b_2(\delta,\rho)$. In the final step we compute the Indirect Inference estimators $\tilde{\delta}_T^s$ and $\tilde{\rho}_T^s$ of δ and ρ obtained by minimizing the Euclidian distance between the vectors $(\hat{\beta}_{1,T}, \hat{\beta}_{21,T}, \hat{\beta}_{22,T})$ and $[\hat{\beta}_1^s(\delta,\rho), \hat{\beta}_{21}^s(\delta,\rho), \hat{\beta}_{22}^s(\delta,\rho)]$. Note that the simulated paths $y_t^s(\delta,\rho), t = 1,\ldots,ST$ for different values of (δ,ρ) are obtained from the same ST drawings of u_t^* in the Student distribution.

The estimators $\hat{\beta}_{1T}$ and $\hat{\beta}_{21,T}$ of δ give approximately the same RMSE : .301 and .305, respectively. They are significantly larger than the RMSE given in Table 1, in agreement with the misspecified pseudo distributions. The RMSE of the estimator $\hat{\beta}_{22,T}$ of ρ is equal to .113, that is a value slightly larger than the RMSE of the CPML2 method given in Table 1 (i.e. .109). The RMSE of the Composite Indirect Inference estimators $\hat{\delta}_T^s$ and $\hat{\beta}_{22,T}^s$ of δ and ρ are respectively .276 and .111; the improvement in the precision is particularly clear for the estimation of δ.

3.3 Systematic risk

The previous approach could be extended to jointly analyze the risk of several corporates. In the basic model for an homogenous segment of corporates and the possibility of migration risk dependence by means of a common factor, the latent model becomes :

$$\left\{\begin{array}{lcl} y^*_{i,t} & = & m + \gamma F_t + \sqrt{1-\gamma^2} u^*_{i,t}, i = 1,\ldots,n, \\ \\ F_t & = & \rho F_{t-1} + \sqrt{1-\rho^2} v_t, \end{array}\right. . \tag{3.7}$$

where the shocks $u^*_{i,t}, i = 1,\ldots,n, t = 1,\ldots,T$ are assumed independent, the $u^*_{i,t}$'s with a common distribution f^*_0 and the $v'_t s$ with a common distribution g^*_0 such that $E(u^*_{it}) = 0, E(v_t) = 0, V(u^*_{it}) = 1, V(v_t) = 1$ and where $-1 < \rho < 1, 0 < \gamma < 1$ (the sign of γ is not identifiable). The observable variables are :

$$y_{i,t} = 1, \text{ if } y^*_{i,t} > 0, y_{it} = 0, \text{ otherwise.} \tag{3.8}$$

When f^*_0, g^*_0 are assumed standard normal, the parameters have the following interpretations : m is the expectation of the latent variables, γ^2 is the spatial correlation and ρ the serial correlation of the systematic factor.

The likelihood function involve T dimensional integrals and is untactable.

a) The standard approach

Let us assume standard normal errors u^*_{it}. Conditional on the factor path $\underline{F_t}$, the variables $y_{it}, i = 1,\ldots,n$ are independent, valued in $\{0,1\}$ and such that :

$$P(y_{it} = 1|\underline{F_t}) = P[y^*_{it} > 0|\underline{F_t}] = \Phi\left(\frac{m+\gamma F_t}{\sqrt{1-\gamma^2}}\right). \tag{3.9}$$

Let us denote by $\widehat{PD}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, the default frequency at date t. If the size of the population of corporates is large, we have approximately :

$$\widehat{PD}_t \simeq \Phi\left(\frac{m+\gamma F_t}{\sqrt{1-\gamma^2}}\right),$$

$$\text{or } \Phi^{-1}(\widehat{PD}_t) \simeq \frac{m}{\sqrt{1-\gamma^2}} + \frac{\gamma}{\sqrt{1-\gamma^2}} F_t, \tag{3.10}$$

that is an approximate relationship between $\Phi^{-1}(\widehat{PD}_t)$ and F_t. Since F_t is zero-mean with unit variance, a filtered value of the common factor is :

$$\begin{aligned}\hat{F}_t &= \frac{\Phi^{-1}(\widehat{PD}_t) - \hat{\mu}}{\hat{\sigma}} \quad \text{with} \\ \hat{\mu} &= \frac{1}{T}\sum_{t=1}^{T}\Phi^{-1}(\widehat{PD}_t), \hat{\sigma}^2 = \{\frac{1}{T}\sum_{t=1}^{T}[\Phi^{-1}(\widehat{PD}_t) - \hat{\mu}]^2\}.\end{aligned}$$

$\hat{F}_t$ is a consistent estimator of F_t is both n and T are large.
we deduce the estimator of m and γ^2 by solving the system :

$$\frac{\hat{m}}{\sqrt{1-\hat{\gamma}^2}} = \hat{\mu}, \ \frac{\hat{\gamma}^2}{1-\hat{\gamma}^2} = \hat{\sigma}^2,$$

We get $\hat{\gamma}^2 = \dfrac{\hat{\sigma}^2}{1+\hat{\sigma}^2}, \hat{m} = \dfrac{\hat{\mu}}{(1+\hat{\sigma}^2)^{1/2}}$,
and the estimator of parameter ρ by regressing $\hat{F}_t$ on $\hat{F}_{t-1}, t = 1, \ldots, T$.

This approach is known as the Asymptotic Single Risk Factor (ASRF) model [Vasicek (1991)] and is a special application of granularity theory [Gagliardini, Gourieroux (2015)].

b) Composite pseudo likelihood

If the errors v_t are also standard normal, the pairs of variables $(y^*_{it}, y^*_{i,t-1})$ and $(y^*_{it}, y^*_{jt}), i \neq j$, are bivariate normal. Therefore we can apply a composite pseudo likelihood approach based on pairwise p.d.f.'s with respect to time, and with respect to corporates. With clear notations the composite pseudo likelihood includes the following terms :

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\log f_2^s(y_{it}, y_{i,t-1}; m, \gamma, \rho) \text{ and } \sum_{t=1}^{T}\sum_{i<j}\log f_2^w(y_{it}, y_{jt}; m, \gamma),$$

where f_2^s (resp. f_2^w) denote the serial (resp. within) pairwise p.d.f..

These terms are equal, respectively, to :

$$N_{11}\log\psi(m, m, \gamma^2\rho) + (N_{10} + N_{01})\log[\Phi(m) - \psi(m, m, \gamma^2\rho)]$$

$$+N_{00}\log[1 - 2\Phi(m) + \psi(m, m, \gamma^2\rho)],$$

where N_{kl} is equal to the number of observations $(y_{i,t}, y_{i,t-1})$ such that $y_{i,t} = k, y_{i,t-1} = l$ [with $N_{11} + N_{10} + N_{01} + N_{00} = n(T-1)$],
and to

$$\tilde{N}_{11} \log \psi(m, m, \gamma^2) + (\tilde{N}_{10} + \tilde{N}_{01}) \log[\Phi(m) - \psi(m, m, \gamma^2)]$$

$$+\tilde{N}_{00} \log[1 - 2\Phi(m) + \psi(m, m, \gamma^2)],$$

where $\tilde{N}_{k,l}$ is the number of observations $(y_{i,t}, y_{j,t}, i < j)$ such that $y_{i,t} = k, y_{j,t} = l$ [with $\tilde{N}_{11} + \tilde{N}_{10} + \tilde{N}_{01} + \tilde{N}_{00} = \frac{n(n-1)T}{2}$].

The limit, when $T \to \infty$, of the objective function divided by T, is easily seen to be maximized at $m = m_0, \gamma^2 = \gamma_0^2, \rho = \rho_0$, the true value (using Kullback inequality). The composite pseudo-likelihood method is, therefore, consistent when $T \to \infty$ and n is fixed.

As an illustration let us consider a Monte-Carlo exercise, with $m = 0, \gamma^2 = .4, \rho = .5, T = 100$ and $n = 20, 50, 100$, and with 500 replications.

Table 2 gives the RMSE of the standard estimator (S) and of the composite maximum likelihood estimator (C). The composite pseudo maximum likelihood estimator performs always at least as well as the standard estimate. As expected its comparative advantage is larger when the segment size n is small. In particular for $n = 20$ the RMSE of the standard estimator of γ^2 is more than twice that of the composite pseudo maximum likelihood estimator.

n		m	γ^2	ρ
20	S	.114	.111	.141
	C	.111	.053	.130
50	S	.114	.058	.108
	C	.111	.048	.108
100	S	.114	.050	.102
	C	.111	.050	.102

TABLE 2 : RMSE of the Standard (S) and Composite PML (C) estimators

Figure 1 presents the empirical distribution of both estimators of γ^2 for $n = 20$. It shows that the composite PML estimator is better than the standard one both in terms of bias ($-.013$ against $.096$) and of standard error (0.051 against 0.055).

Figure 2 displays the empirical distribution of the estimators of ρ for $n = 20$. It shows that the composite PML is better in terms of bias ($-.022$ against $-.109$), but worse in terms of standard error (.128 against .091). However as mentioned in Table 2 the composite PML estimator is better in terms of RMSE (.130 against .141).

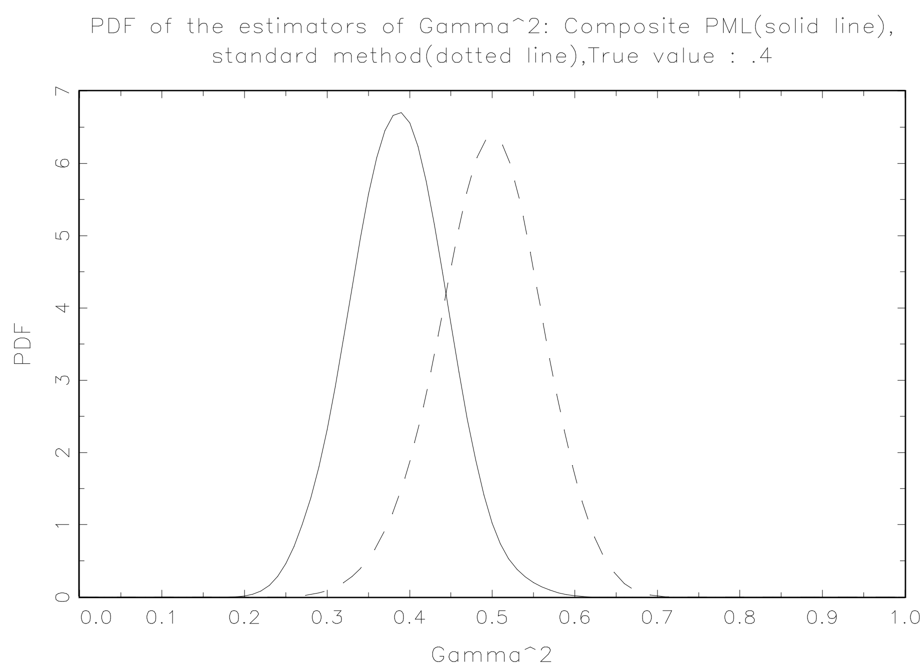

FIGURE 1

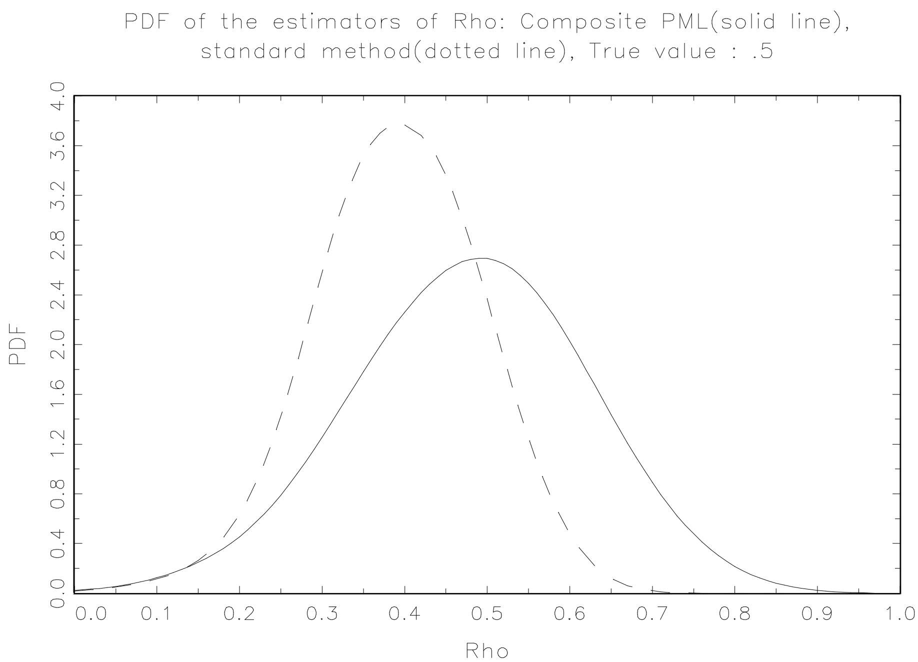

FIGURE 2

4 Concluding Remarks

In complicated nonlinear dynamic models with latent factors, composite indirect inference estimators provide consistent results and are easy to implement with a reasonable efficiency loss. Their practical usefulness has been illustrated by the application to the estimation of corporate risk models considered in the current Basel regulation.

References

Chandler, R., and S., Bate (2007) : "Inference for Clustered Data Using the Independence Loglikelihood", Biometrika, 94, 167-183.

Cox, D., and N., Reid (2004) : "A Note on Pseudolikelihood Constructed from Marginal Densities", Biometrika, 91, 729-737.

Crouhy, M., Galai, D., and R., Mark (2000) : "A Comparative Analysis of Current Credit Risk Models", Journal of Banking and Finance, 29, 59-117.

Gagliardini, P., and C., Gourieroux (2005) : "Stochastic Migration Models with Application to Corporate Risks", Journal of Financial Econometrics, 3, 188-226.

Gagliardini, P., and C., Gourieroux (2015) : "Granularity Theory with Applications to Finance and Insurance", Cambridge University Press.

Gallant, R., and G., Tauchen (1996) : "Which Moments to Match?", Econometric Theory, 12, 657-681.

Gourieroux, C., and J., Jasiak (2001) : "Financial Econometrics", Princeton Univ. Press.

Gourieroux, C., and A., Monfort (1996)a : "Simulation Based Econometric Methods", Oxford University Press, 174 pages.

Gourieroux, C., and A., Monfort (1996)b :"Statistic and Econometric Models", Cambridge Univ. Press, Vol 1, 504 pages.

Gourieroux, C., Monfort, A., and E., Renault (1993) : "Indirect Inference", Journal of Applied Econometrics, 8, 85-118.

Gourieroux, C., Monfort, A., and A., Trognon (1984) : "A General Approach to Serial Correlation", Econometric Theory, 1, 315-340.

Gourieroux, C., Monfort, A., and A., Trognon (1986) : "Moindres Carrés Asymptotiques", Annales de l'INSEE, 58, 91-122.

Kodde, D., Palm, F., and G., Pfann (1990) : "Asymptotic Least Squares Estimation : Efficiency Considerations and Application", Journal of Applied Econometrics, 5, 229-243.

Lindsay, B. (1988) : "Composite Likelihood Methods", Contempory Mathematica, 80, 220-231.

Merton, R. (1974) : "On the Pricing of Corporate Debt : The Risk Structure of Interest Rates", Journal of Finance, 20, 449-470.

Smith, A. (1993) : "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", Journal of Applied Econometrics, 8, 63-84.

Szroeter, J (1983) : "Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", Econometrica, 51, 335-354.

Varin, C. (2008) : "On Composite Marginal Likelihoods", Adv. Statist. Anals. 92, 1-28.

Varin, C., Reid, N., and D., Firth (2011) : "An Overview of Composite Likelihood Methods", Statistica Sinica, 21, 5-42.

Varin, C., and P., Vidoni (2005) : "A Note on Composite Likelihood Inference and Model Selection", Biometrika, 92, 519-528.

Vasicek, O. (1991) : "Limiting Loan Loss Probability Distribution", KMV Corporation.

Appendix 1

Asymptotic Properties of the CII Estimators (no exogenous variables)

The derivations are sketched under the assumption of stationary observations (y_t) satisfying geometric mixing conditions for ergodicity and asymptotic normality.

i) The DGP

We assume that the marginal distribution of y_t is characterized by a p.d.f. $f(y_t;\theta)$, with parametric form, and corresponds to a true value θ_0 of the parameter. As seen in Subsection 2.5, the process (y_t) can stack the current and lagged values of another observed process (y_t^*) of a smaller dimension. Therefore the marginal distribution of y_t gives information on the dynamics of y_t^*.

ii) The instrumental objective functions

We consider K instrumental objective functions, which are used to define the intermediate summary statistics by optimization. We have :

$$\hat{\beta}_{kT} = \arg\max_{\beta_k} \sum_{t=1}^{T} \log g_k(y_t;\beta_k). \tag{a.1}$$

As mentioned in Section 2.5 the objective function is not necessarily the logarithm of a misspecified likelihood.

iii) Consistency of $\hat{\beta}_{kT}$

The statistic :

$$\hat{\beta}_{kT} = \arg\max_{\beta_k} \frac{1}{T} \sum_{t=1}^{T} \log g_k(y_t;\beta_k),$$

will tend to a solution of the asymptotic optimization problem corresponding to $T \to \infty$. By the stationarity assumption and the ergodicity property we get :

$$\lim_{T\to\infty} \hat{\beta}_{kT} = \beta_{k,\infty} = \arg\max_{\beta_k} E_0[\log g_k(y_t;\beta_k)], \tag{a.2}$$

$$\text{where :} \quad E_0[\log g_k(y_t;\beta_k)] = \int \log g_k(y_t;\beta_k) f(y_t;\theta_o) dy_t. \tag{a.3}$$

Therefore the solution of this asymptotic optimization problem is a function of the unknown true value :

$$\beta_{k,\infty} \equiv b_k(\theta_0) \quad \text{, say.} \tag{a.4}$$

This defines the binding function b_k. Let us now focus on the sensitivity of the binding function with respect to parameter θ.

iv) Sensitivity of the binding function

The binding function is the solution of optimization problem (a.2)-(a.3) and satisfies the first-order condition :

$$\frac{\partial}{\partial\theta} E_0[\log g_k(y_t;b_k(\theta_0))] = 0 \Longleftrightarrow \int \frac{\partial \log g_k(y_t, b_k(\theta_0))}{\partial \beta_k} f(y_t;\theta_0) dy_t = 0. \tag{a.5}$$

We can differentiate with respect to θ the implicit equation (a.5) to get :

$$\int f(y_t;\theta_0) \frac{\partial^2 \log g_k(y_t, b_k(\theta_0))}{\partial\beta_k \partial\beta_k'} \frac{\partial b_k(\theta_0)}{\partial\theta'} dy_t$$

$$+ \quad \int \frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k} \frac{\partial f(y_t;\theta_o)}{\partial\theta'} dy_t = 0,$$

or equivalently :

$$E_0[-\frac{\partial^2 \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k\partial\beta_k'}] \frac{\partial b_k(\theta_o)}{\partial\theta'} + E_0[\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k} \frac{\partial \log f(y_t;\theta_0)}{\partial\theta'}] = 0.$$

Since the instrumental pseudo-score $\frac{\partial \log g_k(y_t;b_K(\theta_0))}{\partial\theta}$ is zero mean (see eq. (a.5)), we get :

$$-J_{\beta_k\beta_k}\frac{\partial b_k(\theta_0)}{\partial\theta'} + Cov_0[\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k}, \frac{\partial \log f(y_t;\theta_0)}{\partial\theta}] = 0, \quad \text{(a.6)}$$

where $J_{\beta_k\beta_k} = E_0[\frac{-\partial^2 \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k\partial\beta_k'}]$ denotes the instrumental pseudo-Hessian.

v) Asymptotic normality of the instrumental estimators

The instrumental estimator $\hat{\beta}_{kT}$ satisfies the first-order condition :

$$\sum_{t=1}^{T}\frac{\partial \log g_k[y_t;\hat{\beta}_{kT})]}{\partial\beta_k} = 0. \quad \text{(a.7)}$$

The first-order condition can be expanded around the limit value $b_k(\theta_0)$. We get :

$$\sum_{t=1}^{T}\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k} + \sum_{t=1}^{T}\frac{\partial^2 \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k\partial\beta_k'}[\hat{\beta}_{kT} - b_k(\theta_0)] \simeq 0.$$

After appropriate standardizations by $\sqrt{T}$ and T, and the application of the Law of Large Numbers, we get :

$$\sqrt{T}[\hat{\beta}_{kT} - b_k(\theta_0)] \simeq [J_{\beta_k\beta_K}]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k}, \quad \text{(a.8)}$$

or by stacking the different instrumental estimators :

$$\sqrt{T}[\hat{\beta}_T - b(\theta_0)] \simeq J_0^{-1}\frac{1}{\sqrt{T}}\left[\left(\sum_{t=1}^{T}\frac{\partial \log g_k(y_t;b_k(\theta_0)]}{\partial\beta_k}\right)\right]_{k=1,\dots,K}. \quad \text{(a.9)}$$

where $J_0 = diag[J_{\beta_k,\beta_k}]$.

For a geometrically mixing process (y_t) the vector :

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}\frac{\partial \log g_k(y_t;b_k(\theta_0))}{\partial\beta_k}\right)_{k=1,\dots,K}$$

is asymptotically normal :

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}\frac{\partial \log g_k(g_k; b_k(\theta_0))}{\partial \beta_k}\right) \approx N(0,V), \tag{a.10}$$

where :

$$\begin{aligned} V_0 &= \lim_{T\to\infty} V_0\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log g_k(y_t; b_k(\theta_0))}{\partial \beta_k}\right)_{k=1,\dots,K} \\ &= \sum_{h=-\infty}^{\infty} Cov_0\left[\left(\frac{\partial \log g(y_t; b(\theta_0))}{\partial \beta}\right), \left(\frac{\partial \log g(y_{t-h}; b(\theta_0))}{\partial \beta}\right)\right] \end{aligned} \tag{a.11}$$

$$\text{with} \qquad \frac{\partial \log g(y_t; b(\theta_0))}{\partial \beta} = \left[\frac{\partial \log g_1(y_t; b_1(\theta_0))}{\partial \beta_1'}, \dots, \frac{\partial \log g_K(y_t; b_K(\theta_0))}{\delta \beta_K'}\right]' \tag{a.12}$$

The asymptotic distribution of $\sqrt{n}[\hat{\beta}_T - b(\theta_0)]$ is $N(0,\Sigma_0)$ with $\Sigma_0 = J_0^{-1}V_0J_0^{-1}$.

vi) The asymptotic model

This asymptotic result can be rewritten as :

$$\hat{\beta}_T \simeq b(\theta_0) + J_0^{-1/2}\frac{V_0^{1/2}}{\sqrt{T}}[diag(J_{\beta_k\beta_k})]^{1/2}U, \tag{a.13}$$

where U is a standard normal vector of dimension $q = \sum_{k=1}^{K} q_k$.

The system (a.13) looks like a nonlinear regression model with heteroscedasticity. An optimal estimator of θ_0 derived from this model is the asymptotic least squares estimator of θ solution of :

$$\tilde{\theta}_T = \arg\min_{\theta}[\hat{\beta}_T - b(\theta)]'\hat{J}\hat{V}^{-1}\hat{J}[\hat{\beta}_T - b(\theta)], \tag{a.14}$$

where $\hat{J}$ (resp. $\hat{V}$) is a consistent estimate of J_0 (resp. V_0).

This estimator can be used when the binding functions $b_k, k = 1,\dots,K$, are known.

vii) Consistency of $\tilde{\theta}_T$

This estimator tends to a solution of the asymptotic optimization problem corresponding to $(a.14)$, that is, to a value θ_∞ such that:

$$b(\theta_\infty) = b(\theta_0).$$

Therefore it is consistent of θ_0, iff function b is one-to-one. This is an identification condition, which requires :

- the order condition $q = \sum_{k=1}^{K} q_k \geq p$;
- the rank condition :

$$\begin{aligned} & Rk[\frac{\partial b(\theta_0)}{\partial \theta'}] = p \\ \iff \quad & Rk\ Cov_0 \left(\left[\frac{\partial \log g_k(y_t, b_k(\theta_0))}{\partial \beta_k} \right], \frac{\partial \log f(y_t;\theta_0)}{\partial \theta} \right) = p. \end{aligned}$$

viii) Asymptotic distribution of $\tilde{\theta}_T$

It is easily obtained from the properties of a nonlinear GLS or ALS estimator :

$$\begin{aligned} \sqrt{T}(\tilde{\theta}_T - \theta_0) \quad \approx \quad & N\left(0, \{Cov_0 \left[\frac{\partial \log f(y_t;\theta_0)}{\partial \theta}, \left(\frac{\partial \log g_k(y_t; b_k(\theta_0))}{\partial \beta_k} \right) \right] . \right. \\ & \left. V_0^{-1} Cov_0 \left[\left(\frac{\partial \log g_k(y_t; b_k(\theta_0))}{\partial \beta_k} \right), \frac{\partial \log f(y_t;\theta_0)}{\partial \theta} \right] \}^{-1} \right). \end{aligned} \tag{a.15}$$

The asymptotic variance-covariance matrix is also equal to $\left[\frac{db'(\theta_0)}{d\theta} \Sigma_0^{-1} \frac{db(\theta_0)}{d\theta'} \right]^{-1}$.

ix) The CII estimator

When the binding function is unknown, it can be estimated by simulations from the DGP with a path of length ST, when S is the number of replications. The CII estimator is the solution

$$\tilde{\theta}_T^S = \arg\min_{\theta}[\hat{\beta}_T - \hat{\beta}_T^s(\theta)]'\hat{J}\hat{V}^{-1}\hat{J}[\hat{\beta}_T - \hat{\beta}_T^s(\theta)]. \tag{a.16}$$

It is the optimal nonlinear GLS estimator based on the asymptotic model such that [see Gourieroux, Monfort (1986)a, Chapter 4] :

$$\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^s(\theta_0)) \approx N(0, (1 + \frac{1}{S})J_0^{-1}V_0J_0^{-1}. \tag{a.17}$$

The expression of the variance-covariance matrix is due to the fact that the drawing by the Nature in the true distribution (taken into account in $\hat{\beta}_T$ and the drawing by the econometrician (taken into account in $\hat{\beta}_T^s(\theta)$) are independent. The scale factor $1 + \frac{1}{S}$ follows, as well as the similar effect on the asymptotic distribution of $\tilde{\theta}_T^S$.

Appendix 2

Model with Exogenous Variables

The approach can be extended to model with exogenous variables x_t. We will just focus on the differences between the cases with and without exogenous variables, and refer to Gourieroux, Monfort, Renault (1993) for details about the asymptotic expansions.

i) The DGP

The model involves two types of variables : endogenous variables y_t^*, say, and exogenous variables x_t^*. say. For expository purpose, we consider that the joint process is Markov of order 1 and that its transition can be decomposed into :

$$f(y_t^*|y_{t-1}^*, x_t^*)\pi(x_t^*|x_{t-1}^*), \text{ say.}$$

The fact that the second term depends on x^*_{t-1} only, and not also on y^*_{t-1}, is the exogeneity condition.

Now we can distinguish parametric and semi-parametric DGP's. In the parametric case, the elements of the decomposition depend on a parameter θ, with true value θ_0 :

$$f(y^*_t | y^*_{t-1}, x^*_t; \theta_0) \pi(x^*_t | x^*_{t-1}; \theta_0), \tag{a.18}$$

and we are interested in estimating θ_0

In the semi-parametric framework the first component depends on a parameter θ, but the second component is not a priori constrained. The model becomes :

$$f(y^*_t | y^*_{t-1}, x^*_t; \theta_0) \pi_0(x^*_t | x^*_{t-1}),$$

with true parameters θ_0 and π_0, the latter one being a functional parameter.

ii) The Instrumental Models and the Binding Functions

Let us denote $y_t = (y^{*'}_t, y^{*'}_{t-1})', x_t = (x^{*'}_t, x^{*'}_{t-1})'$, and introduce instrumental objective functions $g_k(y_t, x_t; \beta_k), k = 1, \ldots, K$. The associated summary statistics :

$$\hat{\beta}_{kT} = \arg\max_{\beta_k} \sum_{t=1}^{T} \log g_k(y_t, x_t; \beta_k),$$

will tend to limits $\beta_{k\infty}$ depending on the true distribution.

In the parametric case, this limit is a function of θ_0 only :

$$\beta_{k\infty} = b_k(\theta_0).$$

In the semi-parametric case the limit depends on both θ_0 and functional parameter π_0 :

$$\beta_{k\infty} = b_k(\theta_0, \pi_0).$$

iii) Parametric DGP

The approach developed in Appendix 1 directly applies to the parametric case. Typically, the unknown binding functions can be estimated from **joint simulations** of x_t^*, y_t^* in the DGP corresponding to value θ of the parameter and the identification condition is fulfilled when $q = \sum_{k=1}^{K} q_k \geq p$.

iv) Semi-Parametric DGP

The situation is significantly different in the semi-parametric framework. Indeed the total number of parameters is infinite due to the unknown distribution π_0 of the x. Thus we cannot invert the relation $(\theta_0, \pi_0) \to \beta_\infty$. However this difficulty can be circumvented if the partial mapping.

$$\theta_0 \to b(\theta_0, \pi_0),$$

is one-to-one. The idea is to consider simulated values of $y_1^s(\theta), \ldots, y_T^s(\theta)$ drawn in the DGP conditional on the observed values of the exogenous variables $x_1, \ldots, x_T$, and to draw independently S sets of such values.

Then we can compute :

$$\hat{\beta}_{kT}^s(\theta) = \arg\max_{\beta_k} \sum_{t=1}^{T} \log g_k(y_t^s(\theta); \beta_k),\ s = 1, \ldots, S,$$

and match in an appropriate way the summary statistics computed from the observations, i.e. $\hat{\beta}_{kT}, k = 1, \ldots, K$, and the average of the summary statistics computed from the simulations, i.e. $\frac{1}{S} \sum_{s=1}^{S} \hat{\beta}_{k,T}(\theta), k = 1, \ldots, K$.

The asymptotic properties of the CII estimator are slightly modified compared to the results derived in Appendix 1 in order to account for the conditional simulations. The difference is in the expression of the matrix V in (a.10), which has now to be replaced by [see Gourieroux, Monfort, Renault (1993), Prop 3 and eq. (19)].

$$\tilde{V} = \lim_{T\to\infty} E_0 V_0 \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g_k(y_t, x_t; b_k(\theta_0, \pi_0))}{\partial \beta_k} | x_1, \ldots, x_T \right\}. \quad \text{(a.19)}$$

Appendix 3

Asymptotic behaviour of the likelihood based CII estimator

This estimator is defined by :

$$\theta^*_T = \arg\max_\theta \sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_k \log g_k(y_t, b_k(\theta)).$$

The asymptotic problem is :

$$\arg\max_\theta \sum_{k=1}^{K} \alpha_k E_{\theta_0} \log g_k(y, b_k(\theta)).$$

Since $b_k(\theta_0)$ maximizes $E_{\theta_0} \log g_k(y, \beta)$ with respect to β, θ_0 is a maximum of this asymptotic problem.

The first-order conditions are :

$$\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_k \frac{db'_k(\theta^*_T)}{d\theta} \frac{\partial \log g_k[y_t, b_k(\theta^*_T)]}{\partial \beta_k} = 0.$$

A first-order expansion around θ_0 gives :

$$\sum_{k=1}^{K} \alpha_k \frac{db'_k(\theta_0)}{\partial\theta} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g_k[y_t, b_k(\theta_0)]}{\partial \beta_k}$$

$$- \sum_{k=1}^{K} \alpha_k \frac{\partial b'_k(\theta_0)}{\partial\theta} J_{\beta_k\beta_k} \frac{\partial b_k((\theta_0)}{\partial\theta} \sqrt{T}(\hat{\theta}^*_T - \theta_0) \simeq 0,$$

(since $E_{\theta_0}\left[\frac{\partial \log g_k}{\beta_k}(y, b_k(\theta_0))\right] = 0$),

$$\text{or } \sqrt{T}(\theta^*_T - \theta_0) = \left[\frac{\partial b'(\theta_0)}{\partial\theta} J_{\alpha,0} \frac{\partial b(\theta_0)}{\partial\theta'}\right]^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} J_{\alpha,0}\sqrt{T}[\hat{\beta}_T - b(\theta_0)],$$

where :

$$J_{\alpha,0} \quad = \quad -diag\left[\alpha_k J_{\beta_k\beta_k}\right],$$

using the fact that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g_k(y_t, b_k(\theta_0))}{\partial \beta k}$ is asymptotically equivalent to $- - J_{\beta_k \beta_k} \sqrt{T}(\hat{\beta}_{k,T} - b_k(\theta_0))$.

On the other hand we know, from the ALS theory, that this estimator of θ_0 is less efficient than $\tilde{\theta}_T$, since $\sqrt{T}(\tilde{\theta}_T - \theta_0)$ is asymptotically equivalent to the best linear function of $\hat{\beta}_T - b(\theta_0)$, namely :

$$\left[\frac{db'(\theta_0)}{d\theta} \Sigma_0^{-1} \frac{db(\theta_0)}{d\theta'} \right]^{-1} \frac{db'(\theta)}{d\theta} \Sigma_0^{-1} \sqrt{T}[\hat{\beta}_T - b(\theta_0)].$$

In the expression of $\sqrt{T}(\theta_T^* - \theta_0), \Sigma_0^{-1}$ is replaced by $J_{\alpha,0}$ and, therefore θ_T^* is less efficient.

In other words θ_T^* is asymptotically equivalent to the Nonlinear Generalized Least Squares estimator in the asymptotic model :

$$\hat{\beta}_T = b(\theta) + u_T, \text{ with } E(u_T) = 0, V(u_T) = \Sigma_0,$$

using the suboptimal metric $J_{\alpha,0} = J_0[diag\ \alpha_k]J_0$, instead of the optimal metric : $\Sigma_0^{-1} = J_0 V_0^{-1} J_0$.