Functional Indirect Inference

Monica Billio* & **Alain Monfort****

**University of Venice*

***CREST - INSEE*

Abstract

The class of parametric dynamic latent variable models is becoming more and more popular in economics and finance. Dynamic disequilibrium models, latent factor models, switching regimes models, stochastic volatility models are only few examples of this class of models. Inference in this class may be difficult since the computation of the likelihood function requires to integrate out the unobservable components and to calculate very high dimensional integrals. We propose an estimation procedure which could be applied to any dynamic latent model. The approach is based on the Indirect Inference principle and considers as binding functions conditional expectations of functions of the endogenous variable, given past values of this variable. These conditional expectations are estimated by a ine

Keywords: Dynamic Latent Variable Models, Indirect Inference, Binding Functions, Nonparametric Kernel Estimation.

1 Introduction

The class of parametric dynamic latent variable (DLV) models (also called factor models or hidden variable models or hierarchical models) is becoming more and more popular. Unfortunately, inference in this class of models may be difficult, because the likelihood function appears as a multivariate integral the size of which is equal to the number of observations multiplied by the size of the latent variables.

It is well known that, in some particular cases, such as linear state space models (see Kalman (1960)) or markovian switching models (see Hamilton (1989, 1990)), this difficulty can be overcome and the likelihood function can be computed recursively, or evaluated by numerical methods (see Kitagawa (1987)) or from approximated models (see Harvey, Ruiz and Shephard (1994), Diebold and Nerlove (1989)).

More recently, the simulation based inference methods allowed for the statistical treatment of new DLV models. Simulated likelihood techniques based on importance sampling methods have been applied to stochastic volatility models (see Danielsson and Richard (1993), Danielsson (1994)), to non markovian switching regime models (see Billio and Monfort (1998)) or to dynamic disequilibrium models (see Lee (1997)). The simulated EM method has been used for stochastic volatility models and for partial gaussian state space models (see Shephard (1993,1994)). Simulated Pseudo Maximum Likelihood Methods have been applied to dynamic disequilibrium models (see Laroque and Salanié (1993)). Bayesian methods based on the data augmentation principle have been proposed for the same kind of models (see Albert and Chib (1993), Jacquier, Polson and Rossi (1994)). Several of these classical or bayesian simulation based methods rest on Gibbs and Metropolis algorithms. The Indirect Inference Methods have also been used in some special cases: estimation of continuous time processes from discrete data (see Gouriéroux, Monfort and Renault (1993) , Clement (1994), Broze, Scaillet and Zakoïan (1995a) (1995b), Bianchi and Cleur (1996), Calzolari, Di Iorio and Fiorentini (1998)), stochastic volatility models (see Monfardini (1998)), non linear macroeconomic models (see Smith (1993)) and semi non parametric models (see Gallant and Tauchen (1996)). Finally the Method of Simulated Moments, which can be seen as a special case of the Indirect Inference Method, has also been used in the context of continuous time models (see Duffie and Singleton (1993)).

The aim of this paper is to propose estimation and testing procedures which could be applied to any DLV model. The Simulated Method of Moments and the Simulated Pseudo Maximum Likelihood Method already have this property of wide applicability; however since, in the kind of models considered here, it is in general impossible to simulated in the conditional distribution of the observable endogenous variables given the observed past values of these variables (and the present and past values of the exogenous variables) it is not possible to base these methods on conditional moments and, therefore, marginal moments are used. This limitation may imply some difficulty to capture the dynamic features of the model.

The method proposed here is based on the general Indirect Inference principle in which the binding functions are conditional expectations of functions of the endogenous variables, given their past values. These conditional expectations are estimated by non parametric kernel techniques (see Härdle (1991), Robinson (1983)). It turns out that, in spite of this non parametric feature, the convergence rate of the estimator thus obtained is arbitrarily close to the classical parametric rate and since the asymptotic variance-covariance matrix can be chosen small by increasing the number of conditional expectations considered, the method has potentially good finite sample properties, as confirmed by some examples. Moreover, we propose a scoring method which, given a preliminary estimation, provides a way to select the best binding functions.

The paper is organised as follows. Section 2.2 describes the general dynamic latent variable model and the difficulties arising from the computation of its likelihood function. Section 2.3 briefly recalls the main features of the Indirect Inference principle and presents the Functional Indirect Inference approach. In Section 2.4 the asymptotic properties of the proposed approach are derived and Section 2.5 presents the scoring method to choose the best binding functions. Some Monte Carlo experiments are performed in Section 2.6 and Section 2.7 conlcudes.

2 The Model

We define the general dynamic latent variable (DLV) model as:

$$\left\{ \begin{array}{ll} y_t & = r_t(y^{t-1}, y^{*t}, \varepsilon_t; \theta) \\ y_t^* & = r_t^*(y^{t-1}, y^{*t-1}, \varepsilon_t^*; \theta) \end{array} \right. \tag{1}$$

$t = 1, ..., T$, where y_t is a vector of observable variables, y_t^* is a vector of (partially) unobservable or latent variables, y^{t-1} is a notation for $(y_1', ..., y_{t-1}')'$, $\{\varepsilon_t\}$, $\{\varepsilon_t^*\}$ are two independent white noises with known distributions and θ is a vector of unknown parameters. Note that the assumption of a known distribution for ε_t and ε_t^* can be made without reql loss of generality in the parametric case, since in most cases the parameters appearing in the distributions of ε_t and ε_t^* can be easily incorporated in θ. It is also possible to consider that r_t and r_t^* depend on exogenous variables.

Model (1) contains as special cases all the models mentioned in the introduction (stochastic volatility models, dynamic disequilibrium models, non-linear state space models and in particular switching state space models) and many other models like: dynamic factor models, ARCH factor models, switching regression or ARMA models, deformed time models and so on.

Since ε_t has a known probability density function it is, in principle, possible to derive the conditional p.d.f. of y_t given y^{t-1} and y^{*t}, denoted by $f(y_t/y^{t-1}, y^{*t}; \theta)$, as the p.d.f. of the image of the probability of ε_t by $r_t(y^{t-1}, y^{*t}, \cdot; \theta)$ and, similarly, we get the p.d.f. of y_t^* given y^{t-1} and y^{*t-1}, denoted by $f(y_t^*/y^{t-1}, y^{*t-1}; \theta)$, from the second equation of (1). Therefore we can, in principle, compute the p.d.f. of (y^T, y^{*T}) by:

$$f(y^T, y^{*T}; \theta) = \prod_{t=1}^{T} f(y_t/y^{t-1}, y^{*t}; \theta) f(y_t^*/y^{t-1}, y^{*t-1}; \theta)$$

The likelihood function is then obtained by integrating out y^{*T} in this formula, which requires a $T \times p^*$ dimensional integral (p^* being the size of y_t^*). This integral is in general untractable.

For a given value of θ (and given values of the exogenous variables) it is easy to simulate paths $\tilde{y}^T(\theta)$ and $\tilde{y}^{*T}(\theta)$. In such paths each components $\tilde{y}_t(\theta)$ (or each set of components $(\tilde{y}_{t1}(\theta), ..., \tilde{y}_{tk}(\theta))$) is drawn in its *marginal* distribution; however for a given

θ it is in general much more difficult to draw y_t in its *conditional* distribution given the *observed* values of y^{t-1} and, therefore, Simulated Method of Moments or Simulated Pseudo Maximum Likelihood Methods based on conditional moments are not available. The method proposed below tries to overcome this difficulty and to capture the dynamic of the models through the evaluation of conditional expectations.

3 The Method

3.1 The general indirect inference method

The indirect inference method is based on auxiliary parameters β appearing in an auxiliary criterion function $\Psi_T(y^T, \beta)$ (for sake of simplicity we ignore exogenous variables).

The maximization of $\Psi_T(y^T, \beta)$ with respect to β gives the estimator $\hat{\beta}_T$ and the maximization of $\Psi_T(\tilde{y}^T(\theta), \beta)$ (where $\tilde{y}^T(\theta)$ is a simulated path for a given value θ of the parameter) with respect to β gives $\tilde{\beta}_T(\theta)$. When T goes to infinity $\tilde{\beta}_T(\theta)$ converges a.s. to the binding function:

$$b(\theta) = \operatorname{Arg}\max_{\beta} \Psi_\infty(\theta, \beta)$$

where $\Psi_\infty(\theta, \beta) = \lim_{T\to\infty} \Psi_T(y^T, \beta)$. Note that $\hat{\beta}_T$ converges a.s. to the pseudo true value of β denoted by $\beta_0 = b(\theta_0)$, where θ_0 is the true value of θ.

An indirect inference estimator $\hat{\theta}_{ST}(\Omega)$ of θ is obtained by minimizing:

$$\left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_{sT}(\theta)\right]' \Omega_T \left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_{sT}(\theta)\right]$$

where Ω_T is a symmetric positive definite matrix converging to a deterministic matrix Ω, and the $\tilde{\beta}_{sT}(\theta)$ are obtained from different simulated paths $\tilde{y}_s^T(\theta)$, $s = 1, ..., S$. Note that in the previous minimization $\frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_{sT}(\theta)$ can be replaced by $\tilde{\tilde{\beta}}_{ST}(\theta)$ obtained as

$$\tilde{\tilde{\beta}}_{ST}(\theta) = \operatorname{Arg}\max_{\beta} \sum_{s=1}^{S} \Psi(\tilde{y}_s^T(\theta), \beta)$$

In this general framework, and under regularity conditions, it can be shown that $\hat{\theta}_{ST}(\Omega)$ is consistent, when T goes to infinity and S is fixed, that $\sqrt{T}\left[\hat{\theta}_{ST}(\Omega) - \theta_0\right]$ is asymptotically normal and that an optimal matrix Ω can be chosen and a two step optimal procedure proposed (see Gouriéroux, Monfort and Renault (1993)).

3.2 The Functional Indirect Inference method

The main problem in the general approach described above is to choose a relevant auxiliary criterion $\Psi(y^T, \beta)$ and, therefore, relevant binding functions $b(\theta)$. Here in order to capture the dynamics of the model, we propose to consider binding functions of the form:

$$b(\theta) = [b_m(\theta)]_{m=1,\dots,M} = [E_\theta\left(g_m(y_t)/y_{mt} = \mathcal{Y}_m\right)]_{m=1,\dots,M} \tag{2}$$

where g_m, $m = 1, \dots, M$, are known functions, y_{mt} are lagged components of y_t and $\mathcal{Y}_m$ are given values. These binding functions can be associated with the auxiliary criterion:

$$\Psi_T(y^T, \beta) = -\frac{1}{2Th_T} \sum_{m=1}^{M} \sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right) [g_m(y_t) - \beta_m]^2 \tag{3}$$

where K is a kernel and h_T is a sequence of bandwidths, i.e. real positive numbers converging to zero as T goes to infinity. Indeed, when T goes to infinity $\Psi_T(\tilde{y}^T(\theta), \beta)$ converges to $-\frac{1}{2}\sum_{m=1}^{M} E_\theta\left[(g_m(y_t) - \beta_m)^2/y_{mt} = \mathcal{Y}_m\right] f_{y_{mt}}(\mathcal{Y}_m)$ (where $f_{y_{mt}}(\cdot)$ is the marginal density of y_{mt}), the maximum of which is obviously obtained for $\beta_m = b_m(\theta)$, $m = 1, \dots, M$.

The maximization of the criterion function (3) with respect to β can be done explicitly and we obtain the kernel based estimators:

$$\hat{\beta}_{T,m} = \frac{\sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right) g_m(y_t)}{\sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)} \qquad m = 1, \dots, M$$

Similarly, from simulated data $\tilde{y}^{sT}(\theta)$, $s = 1, \dots, S$, we obtain:

$$\tilde{\beta}_{sT,m} = \frac{\sum_{t=1}^{T} K\left(\frac{\tilde{y}^s_{mt} - \mathcal{Y}_m}{h_T}\right) g_m(\tilde{y}^s_t)}{\sum_{t=1}^{T} K\left(\frac{\tilde{y}^s_{mt} - \mathcal{Y}_m}{h_T}\right)} \qquad m = 1, \dots, M$$

Then the Functional Indirect Inference estimator of θ is obtained by:

$$\hat{\theta}_{ST}(\Omega) = \operatorname{Arg}\min_{\theta} \left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_{sT}(\theta)\right]' \Omega_T \left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_{sT}(\theta)\right] \tag{4}$$

with $\tilde{\beta}_{sT} = (\tilde{\beta}_{sT,1}, \dots, \tilde{\beta}_{sT,M})'$.

Let us now consider the asymptotic properties of this method and the optimal choice of Ω.

4 Asymptotic properties

Under regularity conditions, including for instance stationarity and strong mixing conditions on the process and the bounded support of the kernel, we can prove several asymptotic properties.

Proposition 1

If $T \to +\infty$, $h_T \to 0$, $Th_T \to +\infty$, and if the $\mathcal{Y}_m$, $m = 1, ..., M$ are different, then

$$\sqrt{Th_T}\,(\hat{\beta}_T - \beta_0 - F^{-1}B_T(\theta_0)) \xrightarrow{D} \mathcal{N}(0, W(\theta_0)) \qquad (5)$$

where $\beta_0 = b(\theta_0)$ and

$$B_T(\theta_0) = [B_{1T}, ..., B_{MT}]'$$

$$B_{mT} = \int_{I\!R} K(x) E_{\theta_0}\{[g_m(y_t) - \beta_{0m}]/y_{mt} = \mathcal{Y}_m + xh_T\} f_{y_{mt}}(\mathcal{Y}_m + xh_T)dx$$

$$F = \begin{bmatrix} f_{y_{1t}}(\mathcal{Y}_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{y_{Mt}}(\mathcal{Y}_M) \end{bmatrix}$$

$$W(\theta_0) = [w_{ij}]_{i,j=1,...,M}$$

$$w_{ij} = \begin{cases} \dfrac{E_{\theta_0}\{(g_i(y_t) - \beta_{0i})^2/y_{it} = \mathcal{Y}_i\}}{f_{y_{it}}(\mathcal{Y}_i)} \displaystyle\int_{I\!R} K^2(x)dx & if\ i = j \\ 0 & if\ i \neq j \end{cases}$$

and $f_{y_{mt}}(\mathcal{Y}_m)$ is the marginal density of y_{mt}.

Proof: see the appendix.

Proposition 2

If $T \to +\infty$, $h_T \to 0$, $Th_T \to +\infty$, S is fixed, and if the $\mathcal{Y}_m$, $m = 1, ..., M$ are different, then

$$\sqrt{Th_T}\left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta_0)\right] \xrightarrow{D} \mathcal{N}\left(0, \left(1+\frac{1}{S}\right) W(\theta_0)\right) \tag{6}$$

Proof: see the appendix.

Proposition 2 shows that there is no bias term in the asymptotic behaviour of the difference $\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta_0)$ and, since the asymptotic distribution of $\sqrt{Th_T}(\hat{\theta}_{ST}(\Omega_T) - \theta_0)$ is based on the asymptotic distribution of this difference, we obtain the following result.

Proposition 3

If $T \to +\infty$, $h_T \to 0$, $Th_T \to +\infty$, S is fixed, the $\mathcal{Y}_m$, $m = 1, ..., M$ are different, and Ω_T converges to Ω then

$$\sqrt{Th_T}(\hat{\theta}_{ST}(\Omega_T) - \theta_0) \xrightarrow{D} \mathcal{N}(0, \Sigma(\Omega)) \tag{7}$$

where

$$\Sigma(\Omega) = \left(1+\frac{1}{S}\right)\left[\frac{\partial b'}{\partial \theta}(\theta_0)\Omega\frac{\partial b}{\partial \theta'}(\theta_0)\right]^{-1}\frac{\partial b'}{\partial \theta}(\theta_0)\ \Omega\ W(\theta_0)\ \Omega\ \frac{\partial b}{\partial \theta'}(\theta_0)\left[\frac{\partial b'}{\partial \theta}(\theta_0)\Omega\frac{\partial b}{\partial \theta'}(\theta_0)\right]^{-1}$$

Proof: see the appendix.

According to Proposition 3 the convergence rate of the Functional Indirect Inference estimator $\hat{\theta}_{ST}(\Omega_T)$ is arbitrarily close to the usual $T^{1/2}$ convergence rate, and there is no asymptotic bias.

Using a standard argument (based on the optimality of a GLS estimator), the optimal choice of Ω is

$$\Omega^* = W^{-1}(\theta_0)$$

and, if Ω_T^* is a consistent estimator of Ω^*, we obtain a two stage optimal estimator

$\hat{\theta}^*_{ST} = \hat{\theta}_{ST}(\Omega^*_T)$, whose asymptotic behaviour is:

$$\sqrt{Th_T}(\hat{\theta}^*_{ST} - \theta_0) \xrightarrow{D} \mathcal{N}\left(0, \left(1+\frac{1}{S}\right)\left[\frac{\partial b'}{\partial \theta}(\theta_0)\; W^{-1}(\theta_0)\; \frac{\partial b}{\partial \theta'}(\theta_0)\right]^{-1}\right) \tag{8}$$

Noting that:

$$\frac{1}{Th_T}\sum_{t=1}^{T} K^2\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)\;[g_m(y_t)-\hat{\beta}_T]^2$$
$$\xrightarrow{T\to\infty} E_{\theta_0}\{(g_m(y_t)-\beta_{0m})^2/y_{mt}=\mathcal{Y}_m\} f_{y_{mt}}(\mathcal{Y}_m)\int_{I\!R} K^2(x)dx$$
$$\frac{1}{Th_T}\sum_{t=1}^{T} K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right) \xrightarrow{T\to\infty} f_{y_{mt}}(\mathcal{Y}_m)$$

it is clear that the asymptotic variance-covariance matrix of the estimated auxiliary parameters $\hat{\beta}_T$, $W(\theta_0)$, may be consistently estimated by the diagonal matrix with diagonal entries:

$$\hat{W}_{mT} = Th_T \frac{\sum_{t=1}^{T} K^2\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)\;[g_m(y_t)-\hat{\beta}_T]^2}{\left(\sum_{t=1}^{T} K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)\right)^2} \qquad m=1,...,M$$

Let us consider the simple scalar case, i.e. θ is a scalar. In this case the asymptotic variance of the estimator is simply:

$$\left(1+\frac{1}{S}\right)\left[\sum_{j=1}^{M}\frac{1}{w_{jj}}\left(\frac{\partial b_j}{\partial \theta}(\theta_0)\right)^2\right]^{-1}$$

given the diagonal structure of W. We can make this variance arbitrarily small by simply increasing the number of binding functions taken into account.

It is also possible to obtain an estimator $\hat{\theta}^{**}_{ST}$, which is asymptotically equivalent to $\hat{\theta}^*_{ST}$, from the Generalized Least Square formula (see Gouriéroux and Monfort (1995) chap. 9)

$$\hat{\theta}^{**}_{ST} = \hat{\theta}_{ST} + S\quad\left[\sum_{s=1}^{S}\frac{\partial\tilde{\beta}'_{sT}}{\partial\theta}(\hat{\theta}_{ST})\hat{W}_T^{-1}\sum_{s=1}^{S}\frac{\partial\tilde{\beta}_{sT}}{\partial\theta'}(\hat{\theta}_{ST})\right]^{-1}$$
$$\sum_{s=1}^{S}\frac{\partial\tilde{\beta}'_{sT}}{\partial\theta}(\hat{\theta}_{ST})\hat{W}_T^{-1}\left[\hat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\hat{\theta}_{ST})\right] \tag{9}$$

This approach is particularly interesting because it allows to improve the estimator $\hat{\theta}_{ST}$ without performing another optimization step, but simply by computing the first order derivatives of $\tilde{\beta}_{sT}$.

Alternatively, according to the proposal of Gallant and Tauchen (1996), it is possible to implement the indirect inference procedure by calibrating the parameter of interest θ through the score function, i.e. byminimizing a quadratic form on the score vector

$$\tilde{\theta}_{ST} = \operatorname{Arg}\min_{\theta} \Delta\Psi_T V_T \Delta\Psi_T' \tag{10}$$

where

$$\begin{aligned}
\Delta\Psi_T &= \sum_{s=1}^{S} \frac{\partial \Psi_T}{\partial \beta}\left[\tilde{y}^{sT}(\theta), \hat{\beta}_T\right] \\
&= \left[\frac{1}{Th_T}\sum_{s=1}^{S}\sum_{t=1}^{T} K\left(\frac{\tilde{y}^s_{mt} - \mathcal{Y}_m}{h_T}\right)\left(g_m(\tilde{y}^s_t) - \hat{\beta}_{mT}\right)\right]_{m=1,\dots,M}
\end{aligned}$$

and V_T converges to a positive definite matrix V. For $V^* = (FW(\theta_0)F)^{-1}$, we obtain an estimator asymptotically equivalent to $\hat{\theta}^*_{ST}$. Such a matrix V^* can be consistently estimated by the diagonal matrix with diagonal entries:

$$\left[\frac{1}{Th_T}\sum_{t=1}^{T} K^2\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)\ [g_m(y_t) - \hat{\beta}_T]^2\right]^{-1}$$

Finally the procedure can be generalized to the case where the y_{mt} are multivariate. In this case, the binding functions $b(\theta)$ can be associated with the auxiliary criterion:

$$\mathbf{\Psi}_T(y^T, \beta) = -\frac{1}{2Th_T^d}\sum_{m=1}^{M}\sum_{t=1}^{T} \mathbb{K}\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)[g_m(y_t) - \beta_m]^2$$

where d is the dimension of y_{mt} and $\mathbb{K}$ is a d-dimensional kernel on $\mathbb{R}^d$.

The asymptotic results still hold and Propositions 1 to 3 need only minor evident modifications.

5 Scoring the binding functions

Given a first consistent parameter estimator $\hat{\theta}_T$, obtained with some binding functions, we are interested in looking for some other binding functions which can improve the asymptotic efficiency of the Functional Indirect Inference estimator.

An estimator of the asymptotic variance-covariance matrix of the optimal Functional Indirect Inference estimator associated with M given binding functions $b_m(\theta)$, $m = 1, ..., M$, is (up to the factor $1 + \frac{1}{S}$):

$$\begin{aligned} \Sigma_M &= \left[\frac{\partial b'}{\partial \theta}(\hat{\theta}_T) \; \hat{W}_T^{-1} \; \frac{\partial b}{\partial \theta'}(\hat{\theta}_T) \right]^{-1} \\ &= \left[\sum_{j=1}^{m} \hat{w}_{jj}^{-1} \frac{\partial b_j}{\partial \theta}(\hat{\theta}_T) \frac{\partial b_j}{\partial \theta'}(\hat{\theta}_T) \right]^{-1} \\ &= \left[\sum_{j=1}^{m} c_j c_j' \right]^{-1} \end{aligned} \tag{11}$$

where $c_j = \hat{w}_{jj}^{-1/2} \dfrac{\partial b_j}{\partial \theta}(\hat{\theta}_T)$.

If we take into account a further binding function $b_{M+1}(\cdot)$, we obtain the following estimated asymptotic variance-covariance matrix

$$\begin{aligned} \Sigma_{M+1} &= \left[\sum_{j=1}^{M+1} c_j c_j' \right]^{-1} \\ &= \left[\Sigma_M^{-1} + c_{M+1} c_{M+1}' \right]^{-1} \\ &= \Sigma_M^{1/2} \left[I + \Sigma_M^{1/2} \; c_{M+1} c_{M+1}' \; \Sigma_M^{1/2} \right]^{-1} \Sigma_M^{1/2} \\ &= \Sigma_M^{1/2} \left[I - \frac{\Sigma_M^{1/2} \; c_{M+1} c_{M+1}' \; \Sigma_M^{1/2}}{1 + c_{M+1}' \; \Sigma_M \; c_{M+1}} \right] \Sigma_M^{1/2} \\ &= \Sigma_M - \frac{\Sigma_M \; c_{M+1} c_{M+1}' \; \Sigma_M}{1 + c_{M+1}' \; \Sigma_M \; c_{M+1}} \end{aligned} \tag{12}$$

Then given the initial estimator $\hat{\theta}_T$ and an estimation of its variance-covariance matrix we can choose $b_{M+1}(\cdot)$ in such a way to maximize a score based on some characteristics of Σ_{M+1}, for instance the trace or the determinant.

Proposition 4

If we want to minimize the trace of Σ_{M+1}, we have to select b_{M+1} in order to maximize

$$\frac{c'_{M+1}\ \Sigma_M^2\ c_{M+1}}{1 + c'_{M+1}\ \Sigma_M\ c_{M+1}} \tag{13}$$

while if we want to minimize the determinant of Σ_{M+1}, we have to maximize

$$c'_{M+1}\ \Sigma_M\ c_{M+1} \tag{14}$$

Proof: see the appendix.

We thus obtain an iterative selection procedure of the binding functions, if at each step:

- we consider a set of candidates,
- we choose the next binding function by selecting the one with the best score,
- we recompute the estimation of θ, in order to improve the estimation of Σ.

We can also reconsider the choice of the first M binding functions in order to improve the initial estimator. We can, for instance, use a sequential procedure, in which at each stage we maximize the product of the non zero eigenvalues of $\left[\sum_{j=1}^{m} c_j c'_j\right]$. Let us define C_m as the matrix with c_j as the j^{th} column, then the non zero eigenvalues of $C_m C'_m = \left[\sum_{j=1}^{m} c_j c'_j\right]$ are the same as the ones of $C'_m C_m$ and then we can easily choose the next binding function, since the determinant of $C'_{m+1} C_{m+1}$, where $C_{m+1} = (C_m, c_{m+1})$, is simply

$$|C'_{m+1} C_{m+1}| = |C'_m C_m|\ (c'_{m+1} c_{m+1} - c'_{m+1} C_m (C'_m C_m)^{-1} C'_m c_{m+1}) \tag{15}$$

6 Applications

In order to assess the performance of the proposed approach, we carry out some Monte Carlo experiments.

First, we apply the Functional Indirect Inference estimator to a simple moving average model of order one and compare the results with the maximum likelihood estimates. This experiment is thus a benchmark to evaluate the efficiency of the proposed method.

Next we consider a gaussian factor ARCH, for which it is difficult to derive its exact likelihood function. We present some experiments and compare the performance of the Functional Indirect Inference with that of the others methods, such as the Method of Simulated Moments and the Indirect Inference method.

All the optimization problems has been implemented numerically using the application *Optmum* of Gauss 3.2. The BFGS method has been used, which is a quasi-Newton method[1], and numerical computation of the derivatives of the objective function has been exploited. As starting values for the algorithm, the true parameter values has been chosen throughout the experiment[2]. These good starting points allow a considerable time reduction in the length of the experiments, as they ensure that the algorithm will start from a point close enough to the minimum (or maximum) of the criterion function.

Anyway, it is important to stress that the aim of these Monte Carlo experiments is to assess the feasibility and the general applicability of the proposed approach.

6.1 The MA(1) model

We consider simulated samples of size T=1000 drawn from the following MA(1) process:

$$y_t = \varepsilon_t + \alpha \varepsilon_{t-1}$$

$t = 1, ..., T$, where $\varepsilon_t \sim II\mathcal{N}(0, \sigma^2)$ and where the true values of the parameters are $\alpha_1 = 0.5$ and $\sigma_0 = 1$.

As far as the kernel choice is concerned, we consider a truncated gaussian kernel with standard deviation equal to the empirical standard deviation of the conditioning variable and we fix h_T equal to one.

[1]It uses both first order and second order derivatives information, but relies on approximation of the Hessian matrix.

[2]We performed some sensitivity analysis and verified that perturbing the starting values did not affect the outcome of the optimization problem.

We consider a first estimator $\hat{\theta}_{ST}$ obtained with the following binding functions

$$E(y_t/y_{t-1} = 1.2)$$
$$E(y_t^2/y_{t-2} = 0)$$

where we choose as conditioning values the empirical mean and standard deviation of the contidioning variable. Then, in order to select the auxiliary parameters, we take into account the following binding functions and choose among them by maximizing their score as indicated in Proposition 4:

$$\begin{array}{lll} E(y_t^u/y_{t-\ell} = \mathcal{Y}) & u = 1,2,3,4 & \ell = 1,2,3,4 \\ E(exp(uy_t)/y_{t-\ell} = \mathcal{Y}) & u = \pm 1, \pm 2, \pm 3 & \ell = 1,2,3,4 \end{array} \quad (16)$$

for $\mathcal{Y} \in [-2.5, 2.5]$. The final choice is

$$\begin{array}{ll} E(y_t/y_{t-1} = \mathcal{Y}) & \text{with } \mathcal{Y} = \pm 1.9, 1.9 \\ E(y_t^2/y_{t-2} = \mathcal{Y}) & \text{with } \mathcal{Y} = 0, \pm 0.1, \pm 0.25 \end{array}$$

We calculate the optimal estimator of $\theta = (\alpha, \sigma)$, $\hat{\theta}^*_{ST}$, with $S = 10$, by numerical minimization of the criterion function (4), for the optimal choice of Ω_T, estimated from the first estimator $\hat{\theta}_{ST}$. We also compute the asymptotically equivalent estimator $\hat{\theta}^{**}_{ST}$ by the Generalized Least Square formula. The ML estimates have been obtained by an unconditional Maximum Likelihood approach, performed by the Kalman filter.

Table 1 displays mean and standard deviation of the estimated parameters over 1000 replications of the Monte Carlo experiment. FII_1 indicates the first step Functional Indirect Inference estimator, FII the Functional Indirect Inference estimator, FII_{GLS} the Functional Indirect Inference estimator obtained by the GLS formula and ML indicates the Maximum Likelihood estimator.

The Functional Indirect Inference estimator performs well and the efficiency is improved by taking into account a larger number of binding functions (FII is better than FII_1). Clearly it is not possible to outperform the ML approach but the result seems satisfying, given the fact that only one lagged variable has been considered. Its performance can certainly be improved if we take into account conditional expectations given more than one lagged endogenous variable. It is also worth noting that the GLS formula performs quite well and is very fast to compute because it does not demand a further optimization step.

FII_1	mean	standard deviation	root mean square error
$\hat{\alpha}$	0.512	0.117	0.117614
$\hat{\sigma}$	0.994	0.0529	0.053235
FII	mean	standard deviation	root mean square error
$\hat{\alpha}$	0.512	0.0916	0.092385
$\hat{\sigma}$	0.994	0.0443	0.044699
FII_{GLS}	mean	standard deviation	root mean square error
$\hat{\alpha}$	0.492	0.0917	0.092049
$\hat{\sigma}$	1.006	0.0461	0.046487
ML	mean	standard deviation	root mean square error
$\hat{\alpha}$	0.501	0.0273	0.027313
$\hat{\sigma}$	0.999	0.0224	0.022428

Table 1: MA(1) model. True values: $\alpha_1 = 0.5$, $\sigma_0 = 1$. $T = 1000$, $S = 10$.

6.2 The factor ARCH model

Model (1) contains as special case the gaussian factor ARCH (F-ARCH) model, which is defined as follows:

$$\begin{cases} y_t^* &= (\alpha_1 + \alpha_2 y_{t-1}^{*2})^{1/2} \varepsilon_t^*, \\ y_t &= \beta y_t^* + \varepsilon_t, \end{cases} \tag{17}$$

$t = 1, ..., T$, where $\varepsilon_t^* \overset{iid}{\sim} N(0,1)$, $\varepsilon_t \overset{iid}{\sim} N(0,\Sigma)$ independently and (β, α_1) satisfy some identifying condition (for instance, the first component of β is 1 or $\alpha_1 = 1$).

This model is a good alternative to multivariate ARCH models, which contain a large number of parameters, and then require to introduce some constraints in order to make this number smaller. To introduce some unobserved factors is a natural approach, compatible with the needs of financial theory and with some features of financial series which often have common evolution in the volatilities (Diebold and Nerlove (1989), Engle, Ng and Rothschild (1990), King, Sentana and Wadhwani (1994)). The advantage of introducing an unobserved component has as counterpart the difficult computation of the likelihood function, which requires a T-dimensional integral. Diebold and Nerlove (1989) propose to apply the extended Kalman filter, which leads to some approximations, while Gouriéroux, Monfort and Renault (1993) suggest the indirect inference approach.

For simplicity and identifiability reasons, the original representation (17) is reduced to:

$$\begin{cases} y_t^* &= (\alpha_1 + \alpha_2 y_{t-1}^{*2})^{1/2} \varepsilon_t^* \qquad \varepsilon_t^* \sim N(0,1), \\ y_t &= \beta y_t^* + \varepsilon_t \qquad \varepsilon_t \sim N_p(0, \sigma^2 I_p), \end{cases} \tag{18}$$

with $\beta_1 = 1$ and the dimension of y_t is $p = 2$. We consider simulated samples of size T=500 with $\theta_0 = (\alpha_{10}, \alpha_{20}, \sigma_0, \beta_{20})' = (0.2, 0.7, 0.5, -0.5)'$. As in the MA(1) example we consider a truncated gaussian kernel with standard deviation equal to the empirical standard deviation of the conditioning variable and we fix h_T equal to one[3].

We consider a first step estimator (FII_1) obtained by considering six auxiliary parameters and then the choice of the auxiliary parameters has been done among the following ones:

$$\begin{array}{llll} E(y_{i,t}^u / y_{j,t-\ell} = \mathcal{Y}) & i,j = 1,2 \quad u = 1,2,3,4 & \ell = 1,2,3,4 & \mathcal{Y} \in [-2.5, 2.5] \\ E(y_{i,t}^2 / y_{j,t-\ell}^2 = \mathcal{Y}) & i,j = 1,2 & \ell = 1,2,3,4 & \mathcal{Y} \in [-2.5, 2.5] \end{array}$$

according to their score. It is worth noting that we take only one conditioning variable. We choose eleven auxiliary parameters and consider the Functional Indirect Inference estimator associated with $\Omega = I$ (FII) and with the optimal choice of Ω_T ($OIFF$); we also compute the asymptotically equivalent estimator by the Generalized Least Square formula (FII_{GLS}).

For comparison purposes, we consider some alternative estimation methods. In particular we consider the Method of Simulated Moments (MSM) and an application of the Indirect Inference method. As far as the Method of Simulated Moments is concerned, we consider the marginal version of the moments used as auxiliary parameters for the Functional Indirect Inference method and we also consider the optimal choice of the matrix Ω^* ($OMSM$). As auxiliary model for the Indirect Inference method, we consider a vectorial autoregressive model (VAR) for y_t^2:

$$y_t^2 = \sum_{i=1}^{q} A_i y_{t-i}^2 + u_t$$

[3]We performed some sensitivity analysis and then chose $h_T = 1$.

and take the A_i's as auxiliary parameters. In particular we consider two versions of this estimator with $q = 2, 3$ (II_{VAR2} and II_{VAR3}).

For all these methods, we consider $S = 2$ and perform 200 replications of the Monte Carlo experiment. Table 2 reports the non convergence[4] cases of the optimization algorithm for the different estimation methods. The Functional Indirect Inference estimator demonstrates to be quite robust, since, contrary to the other methods, it always converges. The Mehtod of Simulated Moments appears to perform poorly, because it is very unstable and converges only 3 times over 5.

	FII_1	FII	$OFII$	FII_{GLS}	MSM	$OMSM$	II_{VAR2}	II_{VAR3}
Number of cases	0	0	0	2	75	85	5	18
Percentage	0%	0%	0%	1%	37.5%	42.5%	2.5%	9%

Table 2: F-ARCH(1) model. Non convergence cases over 200 replications of the Monte Carlo experiment.

Tables 3-4 present mean and standard deviation of the estimated parameters[5] for all the considered methods (for each method only the convergence cases are considered) and figures 2-3 show the empirical distribution of the different estimators: $\hat{\alpha}_2$ is on the x-axis and $\hat{\beta}_2$ on the y-axis.

As already noted in the literature, the use of the optimal metric does not improve the finite sample properties of the estimator[6]: this is true for the Functional Indirect Inference method, while for the Method of Simulated Moments there is a bigger number of non convergence cases, even if the root mean square error decreases.

Despite the small number of simulations taken into account (S=2), the methods based on the indirect inference principle appear to be satisfactory in terms of both bias of the estimates and standard deviations.

[4]We consider that the convergence is not reached after 500 iterations or when the algorithm converges to abnormal values of the parameters.

[5]In the optimization algorithm we imposed the following constraints: $0 < \alpha_1 < 2$, $0 < \alpha_2 < 2$ and $\sigma > 0$.

[6]For this reason we do not consider the optimal metric in the Indirect Inference case.

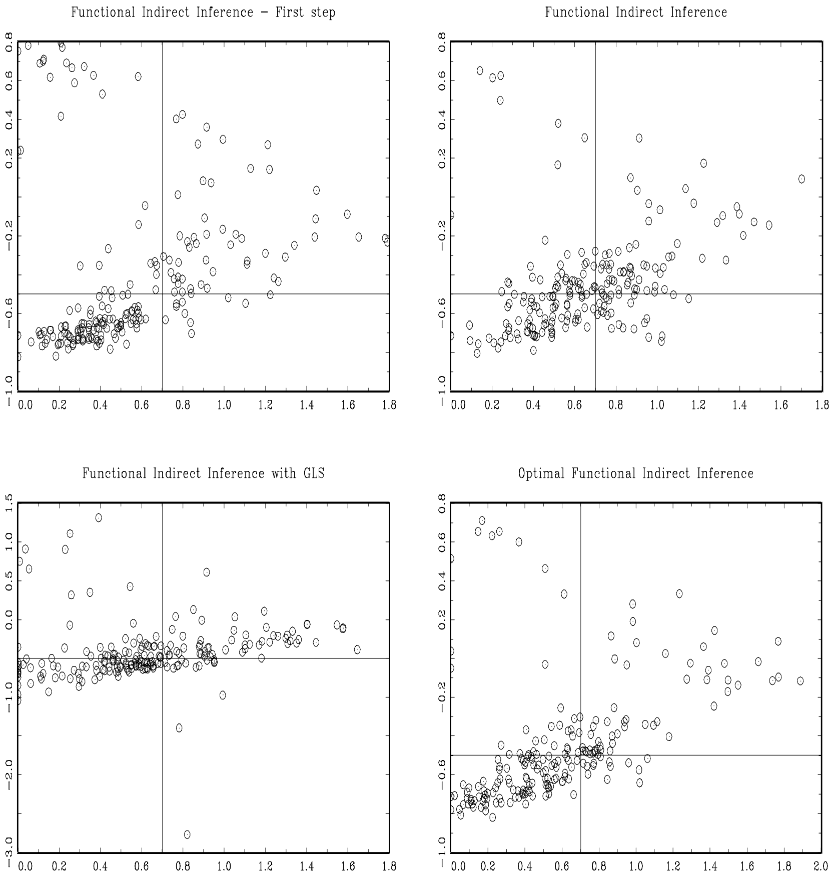

Figure 1: F-ARCH(1) model. Empirical distribution of the different Functional Indirect Inference estimators: $\hat{\alpha}_2$ is on the x-axis and $\hat{\beta}_2$ on the y-axis (the lines indicate the true parameter values).

In table 5, we present the root mean square error of each method standardized by the corresponding root mean square error of the Functional Indirect Inference estimator. Clearly, the approach proposed in this paper outperforms the other methods in both root mean square error and computational[7] terms. It is important to underline how the

[7] As already noted, for the Functional Indirect Inference method the convergence is always reached

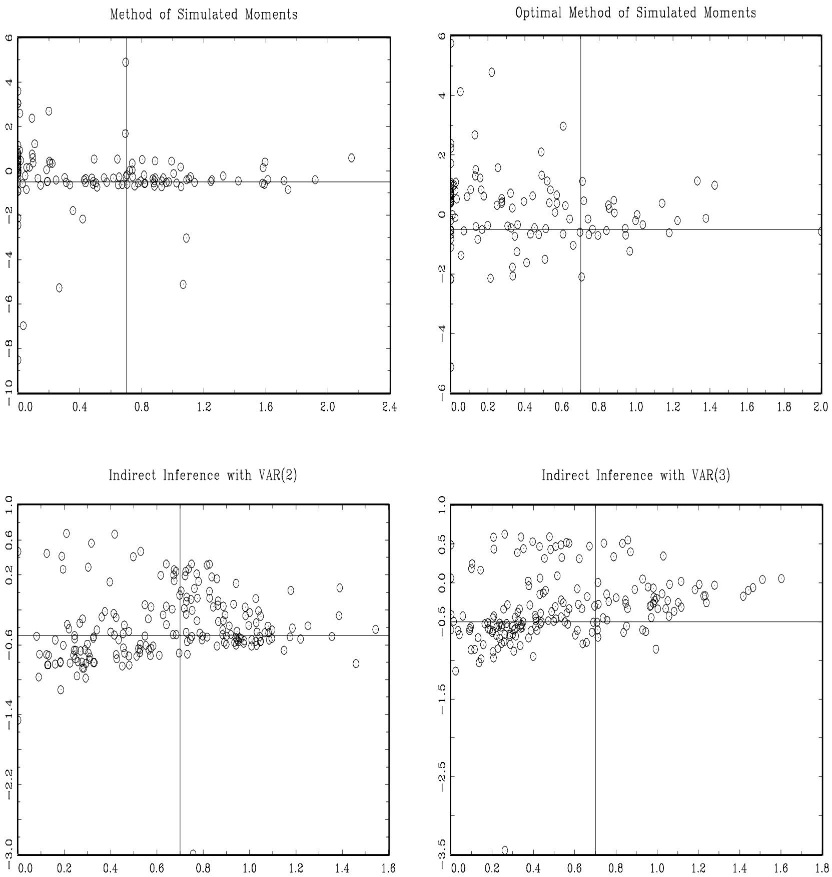

Figure 2: F-ARCH(1) model. Empirical distribution of the other estimators: $\hat{\alpha}_2$ is on the x-axis and $\hat{\beta}_2$ on the y-axis (the lines indicate the true parameter values).

root mean square error decreases by taking into account a bigger number of auxiliary parameters.

It is also worth noting that the *GLS* formula performs very well. This approach is very fast to compute, because it does not demand a further optimization step, and it performs better than the Functional Indirect Inference method with optimal metric

and often within the first 20 iterations.

Mean	α_1	α_2	σ	β_2	replications retained
FII_1	0.334684	0.556001	0.314295	-0.40757	200
FII	0.237223	0.659036	0.461409	-0.4454	200
$OFII$	0.286644	0.595822	0.409314	-0.44694	200
FII_{GLS}	0.252978	0.629196	0.481302	-0.43527	198
MSM	0.476785	0.541089	0.405163	-0.22395	125
$OMSM$	0.341201	0.374154	0.371498	0.185235	115
II_{VAR2}	0.260345	0.643855	0.569553	-0.40173	195
II_{VAR3}	0.243918	0.542239	0.583113	-0.32346	182

Table 3: F-ARCH(1) model. Estimator means (for each method, only the convergence cases are considered). True values: $\alpha_1 = 0.2$, $\alpha_2 = 0.7$, $\sigma = 0.5$, $\beta_2 = -0.5$.

Standard deviation	α_1	α_2	σ	β_2	replications retained
FII_1	0.196715	0.362206	0.246679	0.412412	200
FII	0.135208	0.306117	0.135336	0.263301	200
$OFII$	0.182849	0.391374	0.19595	0.312538	200
FII_{GLS}	0.15717	0.371635	0.145953	0.383932	198
MSM	0.603442	0.514904	0.262414	1.553827	125
$OMSM$	0.414472	0.405703	0.197529	1.349192	115
II_{VAR2}	0.210344	0.333679	0.484585	0.414968	195
II_{VAR3}	0.124789	0.35593	0.805028	0.463508	182

Table 4: F-ARCH(1) model. Estimator standard deviations (for each method, only the convergence cases are considered).

($OFII$).

These results are very encouraging and, given the great generality of the approach (no specific feature of the model has been used), the obtained results are promising for

Relative root mean square error	α_1	α_2	σ	β_2	replications retained
FII_1	1.699997	1.262055	2.194021	1.571730	200
FII	1	**1**	**1**	**1**	200
$OFII$	1.442823	1.311340	1.534257	1.178901	200
FII_{GLS}	1.182694	1.224946	1.045582	1.447921	198
MSM	4.734036	1.744779	1.982686	5.868874	125
$OMSM$	3.122291	1.684840	1.674464	5.627421	115
II_{VAR2}	1.560411	1.095593	3.478630	1.585870	195
II_{VAR3}	**0.943337**	1.260583	5.750737	1.844494	182

Table 5: F-ARCH(1) model. Relative root mean square errors (for each method, only the convergence cases are considered). Each column is standardized by the root mean square error of the Functional Indirect Inference estimator ($\text{RMSE}_{FII}(\alpha_1) = 0.140238$, $\text{RMSE}_{FII}(\alpha_2) = 0.308846$, $\text{RMSE}_{FII}(\sigma) = 0.140731$, $\text{RMSE}_{FII}(\beta_2) = 0.268903$). Bold figures indicate the best performances.

a large class of models.

7 Concluding remarks

We propose an estimation procedure which could be applied to any dynamic latent variable model. This procedure takes into account the dynamic features of the models since it is based on a general Indirect Inference Method using as binding functions some conditional expectations of functions of the endogenous variables, given their past values. Even if non parametric kernel techniques are considered, it turns out that the convergence rate of the proposed estimators is arbitrarily close to the classical one. Moreover scoring methods is proposed in order to select the best binding functions. Some Monte Carlo experiments show the feasibility and the good performance of the approach.

Appendix

Proof of Proposition 1. For $y_{1t} = y_{2t} = ... = y_{Mt} = y_{t-\ell}$ with a fixed ℓ, i.e. with the same conditioning variable, we refer to the Proposition 1.3.10 of Tenreiro (1995), which demonstrates that if $\mathcal{Y}_m$, $m = 1, ..., M$, are distinct points of $\mathbb{R}$, then

$$\sqrt{Th_T}\left(\hat{\beta}_T - \beta_0 - F^{-1}B_T(\theta_0)\right) \xrightarrow{D} \mathcal{N}(0, W(\theta_0))$$

when T goes to infinity.

If the conditioning variable y_{mt} changes with m, i.e. if it is not equal to the same lagged variable $y_{t-\ell}$ for fixed ℓ, with the same hypothesis of Proposition 1.3.10 of Tenreiro (1995), the more general result of Proposition 1 still holds and it can be proved by similar arguments.

Recalling that

$$\hat{\beta}_{T,m} = \frac{\sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right) g_m(y_t)}{\sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)}$$

we can write:

$$\frac{1}{\sqrt{Th_T}} \sum_{t=1}^{T} \left[K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m}) - E\left(K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m})\right)\right]$$

$$-\sqrt{Th_T}\left(\hat{\beta}_{T,m} - \beta_{0m}\right)\frac{1}{Th_T}\sum_{t=1}^{T} K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)$$

$$+\frac{1}{\sqrt{Th_T}}\sum_{t=1}^{T} E\left(K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m})\right) \qquad = 0$$

and then

$$\sqrt{Th_T}\left[\left(\hat{\beta}_{T,m} - \beta_{0m}\right) f_{y_{mt}}(\mathcal{Y}_m) - \frac{1}{h_T}E\left(K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m})\right)\right]$$

$$= \frac{1}{\sqrt{Th_T}}\sum_{t=1}^{T}\left[K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m}) - E\left(K\left(\frac{y_{mt} - \mathcal{Y}_m}{h_T}\right)(g_m(y_t) - \beta_{0m})\right)\right] + o_p(1)$$

$$\sqrt{Th_T}\left[\hat{\beta}_{T,m} - \beta_{0m} - \frac{1}{h_T\, f_{y_{mt}}(\mathcal{Y}_m)} E\left(K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right)\right]$$

$$= \frac{1}{f_{y_{mt}}(\mathcal{Y}_m)\sqrt{Th_T}}\sum_{t=1}^{T}\left[K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m}) - E\left(K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right)\right] + o_p(1)$$

As far as the bias term is concerned, we have:

$$\frac{1}{h_T} \quad E_{\theta_0}\left(K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right) =$$

$$= \int_{\mathbb{R}^2}\frac{1}{h_T}K\left(\frac{x-\mathcal{Y}_m}{h_T}\right)(g_m(y)-\beta_{0m})f_{y_{mt}y_t}(x,y)dx\,dy$$

$$= \int_{\mathbb{R}}\frac{1}{h_T}K\left(\frac{x-\mathcal{Y}_m}{h_T}\right)\int_{\mathbb{R}}(g_m(y)-\beta_{0m})\frac{f_{y_{mt}y_t}(x,y)}{f_{y_{mt}}(x)}dy f_{y_{mt}}(x)dx$$

$$= \int_{\mathbb{R}}\frac{1}{h_T}K\left(\frac{x-\mathcal{Y}_m}{h_T}\right)E_{\theta_0}\{(g_m(y)-\beta_{0m})/y_{mt}=x\}f_{y_{mt}}(x)dx$$

$$= \int_{\mathbb{R}}K(x)E_{\theta_0}\{(g_m(y)-\beta_{0m})/y_{mt}=\mathcal{Y}_m+xh_T\}f_{y_{mt}}(\mathcal{Y}_m+xh_T)dx$$

$$= B_{mT}$$

Compared with results in Tenreiro (1995), the only new feature is the computation of the asymptotic covariance between $\sqrt{Th_T}\left[\hat{\beta}_{T,m}-\beta_{0m}-\frac{1}{f_{y_{mt}}(\mathcal{Y}_m)}B_{mT}\right]$ and $\sqrt{Th_T}\left[\hat{\beta}_{T,j}-\beta_{0j}-\frac{1}{f_{y_{jt}}(\mathcal{Y}_j)}B_{jT}\right]$. In this computation the relevant terms typically are (up to a proportionality constant):

$$\frac{1}{Th_T}\sum_{t=1}^{T-k}\left[K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m}) - E\left(K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right)\right]$$

$$\times\left[K\left(\frac{y_{j,t+k}-\mathcal{Y}_j}{h_T}\right)(g_j(y_{t+k})-\beta_{0j}) - E\left(K\left(\frac{y_{jt}-\mathcal{Y}_j}{h_T}\right)(g_j(y_t)-\beta_{0j})\right)\right]$$

and we have that:

$$E\left(K\left(\frac{y_{jt}-\mathcal{Y}_j}{h_T}\right)(g_j(y_t)-\beta_{0j})\right)\frac{1}{Th_T}\sum_{t=1}^{T-k}\left[K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right]$$
$$\sim \frac{1}{T}B_{jT}\sum_{t=1}^{T-k}\left[K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right]\overset{T\to\infty}{\longrightarrow} 0$$

$$\frac{1}{h_T}E\left(K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)(g_m(y_t)-\beta_{0m})\right)E\left(K\left(\frac{y_{jt}-\mathcal{Y}_j}{h_T}\right)(g_j(y_t)-\beta_{0j})\right)$$
$$\sim h_T B_{mT}B_{jT}\overset{T\to\infty}{\longrightarrow} 0$$

$$\frac{1}{Th_T}\sum_{t=1}^{T-k}\left[K\left(\frac{y_{mt}-\mathcal{Y}_m}{h_T}\right)K\left(\frac{y_{jt}-\mathcal{Y}_j}{h_T}\right)(g_m(y_t)-\beta_{0m})\,(g_j(y_t)-\beta_{0j})\right]$$
$$\sim h_T\ E\{(g_m(y_t)-\beta_{0m})(g_j(y_{t+k})-\beta_{0j})/y_{mt}=\mathcal{Y}_m, y_{jt+k}=\mathcal{Y}_j\}\overset{T\to\infty}{\longrightarrow} 0$$

Proof of Proposition 2. The random variables $\hat{\beta}_T$, $\tilde{\beta}_{sT}(\theta_0)$, $s=1,...,S$, are independent and identically distributed. Consequently, when T goes to infinity, they have the same bias term which disappears in the difference (6). The variance-covariance matrix follows from the independence among the different drawings and data.

Proof of Proposition 3. As already noted when T goes to infinity

$$\text{(A.1)}\qquad \begin{aligned}\lim_{T\to\infty}\Psi_T(\tilde{y}^T(\theta),\beta) &= -\frac{1}{2}\sum_{m=1}^{M}E_\theta\left[(g_m(y_t)-\beta_m)^2/y_{mt}=\mathcal{Y}_m\right]f_{y_{mt}}(\mathcal{Y}_m)\\ &=\Psi_\infty(\theta,\beta)\end{aligned}$$

(see Proposition 1.3.9 in Tenreiro (1995), in which a mean squared convergence and therefore a convergence in probability is demonstrated) and $\Psi_\infty(\theta,\beta)$ has its maximum at $\beta=b(\theta)$. Given that $\hat{\beta}_T$ converges to $b(\theta_0)$ and $\tilde{\beta}_{sT}(\cdot)$ converges to the binding function $b(\cdot)$, if the equation

$$\text{(A.2)}\qquad \beta=b(\theta)$$

admits a unique solution in θ, we have that

$$\hat{\theta}_{ST}(\Omega) = \text{Arg}\min_\theta \left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta)\right]' \Omega_T \left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta)\right]$$

$$\rightarrow \text{Arg}\min_\theta [b(\theta_0) - b(\theta)]' \Omega [b(\theta_0) - b(\theta)]$$

$$= \{\theta : \ b(\theta) = b(\theta_0)\} \quad \text{(as soon as } \Omega \text{ is positive definite)}$$

$$= \theta_0 \qquad\qquad \text{(from A.2)}$$

Let us consider the first order conditions of the minimization problem (4)

$$\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\beta}'_{sT}}{\partial\theta}(\hat{\theta}_{ST})\right]\Omega_T\left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\hat{\theta}_{ST})\right] = 0$$

If we consider an expansion around the value θ_0 we have

$$\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\beta}'_{sT}}{\partial\theta}(\theta_0)\right]\Omega_T\sqrt{Th_T}\left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta_0) - \frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\beta}'_{sT}}{\partial\theta}(\theta_0)\left(\hat{\theta}_{ST} - \theta_0\right)\right] = o_p(1)$$

then

$$\sqrt{Th_T}\left(\hat{\theta}_{ST} - \theta_0\right) = \left[\frac{\partial b'}{\partial\theta}(\theta_0)\Omega\frac{\partial b}{\partial\theta'}(\theta_0)\right]^{-1}\frac{\partial b'}{\partial\theta}(\theta_0)\Omega\sqrt{Th_T}\left[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_{sT}(\theta_0)\right] + o_p(1)$$

and the result follows from Proposition 2.

Proof of Proposition 4. From equations (12)

$$tr(\Sigma_{M+1}) = tr(\Sigma_M) - tr\left(\frac{\Sigma_M\ c_{M+1}c'_{M+1}\ \Sigma_M}{1 + c'_{M+1}\ \Sigma_M\ c_{M+1}}\right)$$

$$= tr(\Sigma_M) - \frac{c'_{M+1}\ \Sigma_M^2\ c_{M+1}}{1 + c'_{M+1}\ \Sigma_M\ c_{M+1}}$$

and

$$|\Sigma_{M+1}| = \frac{|\Sigma_M|}{|I + \Sigma_M^{1/2}\ c_{M+1}c'_{M+1}\ \Sigma_M^{1/2}|}$$

$$= \frac{|\Sigma_M|}{1 + c'_{M+1}\ \Sigma_M\ c_{M+1}}$$

since $I + \Sigma_M^{1/2}\ c_{M+1}c'_{M+1}\ \Sigma_M^{1/2}$ is a symmetric positive definite matrix, which has $M-1$ eigenvalues equal to 1 and the last one equal to $(1 + c'_{M+1}\ \Sigma_M\ c_{M+1})$. The result follows.

Bibliography

Albert J.H. and Chib S. (1993), Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics* 11/1, 1-15.

Bierens H.J. (1987), Kernel Estimators of Regression Functions, in *Advances in Econometrics*, Vol. 1, ed. Dewley T.F., Cambridge University Press, 99-144.

Bianchi C. and Cleur E.M. (1996), Indirect estimation of stochastic differential equation models: some computational experiments, *Comput. Econ.*, 9, 257-274.

Billio M. and Monfort A. (1998), Switching state space models: likelihood function, filtering and smoothing, *Journal of Statistical Planning and Inference*, 68/1, 65-103 (chap. 3).

Bosq D. and Lecoutre J.P. (1987), *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

Broze L., Scaillet O. and Zakoïan J.M. (1995a), Testing for continuous-time models of the short-term interest rates, *Journal of Empirical Finance*, 2, 199-223.

Broze L., Scaillet O. and Zakoïan J.M. (1995b), Quasi indirect inference for diffusion processes, *Document de Travail du CREST* n. 9511. *Econometric Theory*, forthcoming.

Calzolari, G., Di Iorio, F. and Fiorentini, G. (1998), Control variates for reduction in indirect inference: Interest rate models in continuous time, *Econometrics Journal*, 1, 100-112.

Clément E. (1994), Inférence statistique des processus de diffusion, *Document de Travail du CREST* n. 9404.

Danielsson J. (1994), Stochastic volatility in asset prices: estimation with simulated maximum likelihood, *Journal of Econometrics*, 64, 375-400.

Danielsson J. and Richard J.F. (1993), Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics*, 8, S153-S174.

Diebold F.X. and Nerlove M. (1989), The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics* 4, 1-21.

Duffie D. and Singleton K. (1993), Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929-952.

Engle R., Ng V. and Rothschild M. (1990), Asset pricing with a factor ARCH variance structure: empirical estimates for Treasury Bills, *Journal of Econometrics*, 45, 213-237.

Gallant A.R. and Tauchen G. (1996), Which moments to match?, *Econometric Theory*, 12, 657-681.

Gouriéroux C. and Monfort A. (1995), *Statistics and econometric models.* Cambridge University Press.

Gouriéroux C., Monfort A. and Renault E. (1993), Indirect Inference, *Journal of Applied Econometrics*, 8, S85-S118.

Gouriéroux C., Monfort A. and Tenreiro C.M.R. (1995), Kernel M-Estimators and Functional Residuals Plots, *Document de Travail du CREST* n. 9546.

Hamilton J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57/2, 357-384.

Hamilton J.D. (1990), Analysis of time series subject to changes in regime, *Journal of Econometrics*, 45, 39-70.

Härdle W. (1991), *Applied Nonparametric Regression.* Cambridge University Press.

Harvey A.C., Ruiz E. and Shephard N. (1994), Multivariate stochastic variance models, *Review of Economic Studies*, 61, 247-264.

Jacquier J., Polson N.G. and Rossi P.E. (1994), Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics*, 12, 371-417.

Kalman R.E. (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, Transactions ASMA, Series D*, 82, 35-45.

King M.A., Sentana E. and Wadhwani S.B. (1994), Volatility and links between national stock markets, *Econometrica*, 62, 901-933.

Kitagawa G. (1987), Non-Gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association*, 82, 1032-1063.

Laroque G. and Salanié B. (1993), Simulation based estimation of models with lagged latent variables, *Journal of Applied Econometrics*, 8, S119-S133.

Lee L.F. (1997), Simulation estimation of dynamic switching regression and dynamic disequilibrium models - some Monte Carlo results, *Journal of Econometrics*, 78, 179-204.

Monfardini C. (1998), Estimating stochastic volatility models through indirect inference, *Econometrics Journal*, 1, 113-128.

Robinson P.M. (1983), Nonparametric estimators for time series, *Journal of Time Series Analysis*, 4, 185-207.

Shephard N. (1993), Fitting non-linear time series models, with applications to stochastic variance models, *Jornal of Applied econometrics* 8, S135-S152.

Shephard N. (1994), Partial non-Gaussian state space, *Biometrika*, 81/1, 115-131.

Smith A. (1993), Estimating nonlinear time series models using simulated vector autoregressions, *Journal of Applied Econometrics* 8, S63-S84.

Tenreiro C.M.R. (1994), Estimation non paramétrique d'une fonction de moments conditionnels sous des conditions locales, *C. R. Acad. Sci. Paris*, Ser. I, 318, 1149-1152.

Tenreiro C.M.R. (1995), Estimation fonctionnelle: applications aux tests d'adéquation et de paramètre constant, *Ph.D. Dissertation*, Department of Mathematics, University of Coimbra, Portugal.