

# EFFICIENT DERIVATIVE PRICING BY EXTENDED METHOD OF MOMENTS

P., GAGLIARDINI\*, C., GOURIEROUX† and E., RENAULT‡

(First version: January 2002)  
(Revised version: July 2004)

---

\*University of St. Gallen and University of Lugano.

†CREST, CEPREMAP (Paris) and University of Toronto.

‡CIRANO-CIREQ (Montreal) and University of North Carolina at Chapel Hill.

## Efficient Derivative Pricing by Extended Method of Moments

### Abstract

In this paper we consider an incomplete market framework and explain how to use jointly observed prices of the underlying asset and of some derivatives written on this asset for an efficient pricing of other derivatives. This question involves two types of moment restrictions, which can be written either for a given value of the conditioning variable or can be uniform with respect to this conditioning variable. This distinction between local and uniform conditional moment restrictions leads to an extension of the Generalized Method of Moments (GMM), a method in which all restrictions are assumed uniform. The Extended Method of Moments (XMM) provides estimators of the parameters with different rates of convergence: the rate is the standard parametric one for the parameters which are identifiable from the uniform restrictions, whereas the rate can be nonparametric for the risk premium parameters. We derive the (kernel) nonparametric efficiency bounds for estimating a conditional moment of interest and prove the asymptotic efficiency of XMM. To avoid misleading arbitrage opportunities in estimated derivative prices, an XMM estimator based on an information criterion is introduced. The general results are applied in a stochastic volatility model to get efficient derivative prices, to measure the uncertainty of estimated prices and to estimate the risk premium parameters.

**Keywords:** Generalized Method of Moments, Weak Instruments, Information Based Estimation, Nonparametric Efficiency, Limited and Full Information, Derivative Pricing, Stochastic Volatility, Risk Premium.

**JEL number:** C13, C14, G12.

# Valorisation efficace de produits dérivés par la Méthode des Moments Étendue

## Résumé

Dans cet article nous supposons les marchés incomplets et expliquons comment utiliser conjointement des observations sur les prix d'un actif support et de certains de ses dérivés pour valoriser de façon efficace d'autres dérivés. Cette question fait intervenir deux types de conditions de moments, qui peuvent soit être écrites pour une valeur donnée de la variable conditionnante, soit être uniformes par rapport aux valeurs possibles de cette variable. Cette distinction entre les conditions de moments conduit à étendre la méthode des moments généralisée (GMM), dans laquelle toutes les restrictions sont uniformes. La méthode des moments étendue (XMM) fournit des estimateurs des paramètres avec divers taux de convergence. Ce taux est paramétrique pour les paramètres caractérisant la distribution historique des prix, qui sont identifiables à partir des restrictions uniformes; il peut être non paramétrique pour les paramètres définissant les primes de risque. Nous dérivons les bornes d'efficacité non paramétrique pour ces paramètres et montrons l'efficacité asymptotique de la méthode des moments étendue. Ces résultats sont utilisés pour obtenir des valorisations efficaces de dérivés, mesurer l'incertitude des prix estimés et analyser les primes de risque dans un modèle à volatilité stochastique.

**Mots clés:** Méthode des moments généralisée, instruments faibles, efficacité non paramétrique, information complète, information limitée, valorisation d'options, volatilité stochastique, prime de risque.

**Classification JEL:** C13, C14, G12

# 1 Introduction

The Generalized Method of Moments (GMM) has been initially introduced by Hansen (1982), Hansen, Singleton (1982) to estimate parameters defined by Euler conditions. Typically in a Consumption based CAPM [Lucas (1978)] the moment restrictions at date  $t$  are:

$$p_{i,t} = E_t [p_{i,t+1} \delta (q_t/q_{t+1}) U'(C_{t+1}; \gamma)/U'(C_t; \gamma)], \quad i = 1, \dots, n, \quad (1)$$

where  $U$  is a utility function,  $p_{i,t}$  the observed prices of the  $n$  financial assets,  $q_t$  the price of the consumption good,  $C_t$  the consumption level and  $E_t$  denotes the conditional expectation given the available information including the lagged values of prices and income. The parameters of interest are the preference parameter  $\gamma$  and the psychological discount rate  $\delta$ . The model is semi-parametric. GMM focuses on the estimation of  $\theta = (\gamma', \delta)'$  and disregards the nuisance parameter, that is the joint conditional distribution of prices  $p_{i,t+1}, i = 1, \dots, n$ , and consumption  $C_{t+1}$ . Recently different approaches, called empirical likelihood, minimum chi-square or information based approach, have been proposed to simplify the derivation of a GMM parameter and to improve its finite sample properties<sup>1</sup>. The basic idea is to estimate jointly the structural parameter  $\theta$  and the nuisance infinite dimensional parameter under the moment restrictions.

However the Euler conditions are not only useful to estimate the preference parameters or test a structural equilibrium model. They are also used in Finance for pricing derivatives. More precisely the Euler condition is considered as a pricing formula :

$$p_{i,t} = E_t [M_{t,t+1}(\theta) p_{i,t+1}], \quad i = 1, \dots, n, \quad \forall t, \quad (2)$$

where  $M_{t,t+1}(\theta) = \delta (q_t/q_{t+1}) U'(C_{t+1}; \gamma)/U'(C_t; \gamma)$  is a parameterized stochastic discount factor (sdf) [see e.g. Hansen, Richard (1987), Hansen, Jagannathan (1991), Bansal, Viswanathan (1993), Cochrane (2001)]. This pricing formula is assumed to be valid also for the other assets, whose payoffs are

---

<sup>1</sup>See e.g. Owen (1991), (2001), Qin, Lawless (1994), Hansen, Heaton, Yaron (1996), Kitamura, Stutzer (1997), Imbens (1997), Smith (1997), Imbens, Spady, Johnson (1998), Baggerly (1998), Kitamura (2001), Kitamura, Tripathi, Ahn (2004), Ai, Chen (2003), Newey, Smith (2004), Bonnal, Renault (2004).

written on  $p_{1,t}, \dots, p_{n,t}$  and whose current prices are not observed. For instance the price at date  $t_0$  of a European call, written on  $p_1$ , with strike  $K$  and residual maturity 1 is:

$$c_{t_0}(1, K) = E_{t_0} [M_{t_0, t_0+1}(\theta)(p_{1, t_0+1} - K)^+]. \quad (3)$$

It is naturally estimated by :

$$\hat{c}_{t_0}(1, K) = \hat{E}_{t_0} [M_{t_0, t_0+1}(\hat{\theta})(p_{1, t_0+1} - K)^+], \quad (4)$$

where  $\hat{\theta}$  is a GMM estimator of  $\theta$  and  $\hat{E}_{t_0}$  is a (functional) estimator of the conditional expectation. For the application to derivative pricing the interest is focused on the conditional moment  $c_{t_0}(1, K) = E_{t_0}(a)$  of the function  $a = M_{t_0, t_0+1}(p_{1, t_0+1} - K)^+$ , which is the product of the sdf by the derivative payoff. Clearly this problem requires the joint estimation of the parameter  $\theta$  and of the conditional distribution.

However the problem becomes much more complicated if we want to account for the observed prices of assets which are less actively traded (such as derivatives) when estimating  $\theta$  and the conditional distribution. Typically we can observe the prices of the short term zero-coupon bond:

$$B(t, t+1) = E_t [M_{t, t+1}(\theta)], \quad t = 1, \dots, T, \quad (5)$$

the prices of the underlying asset:

$$p_{1,t} = E_t [M_{t, t+1}(\theta) p_{1, t+1}], \quad t = 1, \dots, T, \quad (6)$$

and for instance the at-the money call price at date  $t_0$ :

$$c_{t_0}(1, p_{1, t_0}) = E_{t_0} [M_{t_0, t_0+1}(\theta)(p_{1, t_0+1} - p_{1, t_0})^+]. \quad (7)$$

In this situation the structural parameter  $\theta$  is subject to two types of moment restrictions, which can be satisfied either for multiple environments [uniform moment restrictions, equations (5) and (6)] or only for a given one [local moment restrictions, equation (7)]. The two types of moment restrictions are difficult to take into account jointly. This explains in practice (but also in the academic literature) the approaches which have been followed:

i) The observations of derivative prices can be neglected, the parameter  $\theta$  estimated by a standard GMM method and the derivative price of interest

approximated by (4). The drawback of this approach is that the risk premium parameters are generally non identifiable and are fixed a priori to zero.

ii) An alternative is the so-called cross-sectional approach which is based on derivative prices at date  $t_0$  only. However the convergence of the estimators requires a large number of liquid derivatives and the condition is far to be satisfied on derivative markets.

The aim of this paper is to use jointly observed prices of the underlying asset and of some derivatives written on this asset for an efficient pricing of other derivatives. Technically we explain how to estimate conditional moments under both types of conditional moment restrictions, which are either uniform or local with respect to the conditioning variable. In particular we derive the nonparametric efficiency bound of the conditional moment(s) of interest and explain how to reach this bound. In Section 2 we study the set of moment estimators for both structural parameters and the conditional moment of interest. The two types of moment restrictions are carefully studied, since they have different consequences concerning the identifiability of structural parameters and the accuracy of the estimator of the conditional moment of interest. In particular, the linear combinations of structural parameters which are identifiable from uniform moment restrictions converge at a parametric rate, whereas the other linear combinations have a nonparametric rate of convergence. We show that there exists an optimal moment method, called extended method of moments (XMM), for the conditional moment of interest, that is which minimizes its asymptotic variance. This minimal variance defines the so-called (kernel) nonparametric efficiency bound. We derive the explicit expression of the efficiency bound in the general framework. Finally we consider the special cases of limited information, when all constraints are conditional on a given environment, and of full information, when all constraints are uniform with respect to the environment.

By definition the extended method of moments is (kernel) nonparametrically efficient. However it does not provide in general a coherent estimator of the whole conditional distribution. The aim of Section 3 is to consider an information based approach to estimate jointly the structural parameter and the conditional distribution. The associated information based estimators of the moments of interest are also (kernel) nonparametrically efficient.

Section 4 is concerned with the application to efficient derivative pricing from both observed underlying asset prices and derivative prices. We discuss in detail the moment restrictions for this problem and distinguish these

constraints depending whether they are uniform or local with respect to the conditioning variable. The approach is applied in Section 5 to a stochastic volatility model. We discuss carefully the identifiability of the different parameters from the uniform restrictions. Based on the (kernel) nonparametric efficiency bounds, we provide the patterns of the confidence bands on derivative prices according to maturity and strike. This allows us to measure the uncertainty on estimated derivative prices, when the sole informational content of no-arbitrage is taken into account. We discuss the finite sample properties of the estimated derivative prices and of the structural parameters by Monte-Carlo. The estimators of the structural parameters which correspond to the risk premium on volatility converge at a nonparametric rate, whereas the estimators of the other components of the structural parameter converge at a parametric rate. These different behaviours are consequences of market incompleteness and the lack of liquidity on derivative markets. Section 6 concludes. The proofs are gathered in appendices. Since the focus of the paper is mainly on the structural interpretations of moment restrictions and on the application to derivative pricing, the proofs of asymptotic results are sketched with the purpose of deriving the forms of the limiting distributions and of the efficiency bounds. Detailed technical results including regularity conditions are available upon request.

## 2 Extended Method of Moments

In this section we consider the estimation of conditional moments  $E_0[a(Y; \theta_0)|X = x]$  under moment restrictions  $E_0[g(Y, \theta_0)|X = x] = 0$  from a sample of observations  $(x_t, y_t)$ ,  $t = 1, \dots, T$ , where process  $(X_t, Y_t)$  is assumed stationary. In this framework it is important to discuss carefully the set of estimating constraints.

i) Firstly we can be interested in a conditional moment  $E_0(a|x_0) = E_0[a(Y; \theta_0)|X = x_0]$  for a given value  $x_0$ , under the set of constraints  $E_0[g(Y; \theta_0)|X = x] = 0, \forall x$ . The moment to be estimated has a local interpretation, whereas parameter  $\theta_0$  is defined uniformly in  $x$ . Equivalently we can consider that we are interested in a conditional moment  $E_0(a|x_0)$  under marginal moment restrictions  $E_0[g_1(Y, X; \theta_0)] = 0$ , where  $g_1$  is derived by multiplying function  $g(Y; \theta)$  by appropriate instrumental variables. This explains the different rates of convergence for the different parameters, that are

a parametric rate for the estimator of  $\theta$  (based on marginal moments) and a nonparametric rate for the estimator of the conditional moment  $E_0(a|x_0)$ . As a consequence the asymptotic accuracy of the estimated moment of interest is not influenced by the estimation of  $\theta_0$ .

ii) Secondly we can be interested in a conditional moment  $E_0(a|x_0) = E_0[a(Y; \theta_0)|X = x_0]$ , given the constraints  $E_0[g(Y; \theta_0)|X = x_0] = 0$ . Both the moment of interest  $E_0(a|x_0)$  and the parameter  $\theta_0$  have local interpretations. The rates of convergence are nonparametric for both parameter  $\theta_0$  and conditional moment  $E_0(a|x_0)$ . The asymptotic accuracy of the estimated conditional moment  $E_0(a|x_0)$  will take into account the estimation of  $\theta_0$ .

These two cases are said with full- and limited information, respectively. In Subsection 2.1 we consider the general framework in which the structural parameter  $\theta$  is subject to both types of moment restrictions, that are uniform or local restrictions. We study the set of moment estimators of the conditional moment of interest, look for an optimal one and compute the nonparametric efficiency bound. In Subsection 2.2 the result is applied to the pure limiting cases of full and limited information, respectively.

## 2.1 General framework

Let us consider a general framework with both uniform and local constraints:

$$\begin{aligned} E[g(Y; \theta) | X = x] &= 0, \quad \forall x, \\ E[\tilde{g}(Y; \theta) | X = x_0] &= 0, \end{aligned} \tag{8}$$

where  $\theta$  is an unknown structural parameter with dimension  $p$ . As usual in GMM approach, we assume in a first step that the uniform restrictions have been replaced by a set of marginal restrictions, by introducing a finite number of appropriate instrumental variables. Then in a second step we discuss the optimal choice of the instruments.

### 2.1.1 Efficiency bound for given instruments

#### i) Identification condition

Let us introduce instruments  $Z = H(X)$  and let function  $g_1$  define the corresponding marginal restrictions:  $E_0[Z \cdot g(Y; \theta_0)] = E_0[g_1(Y, X; \theta_0)] = 0$ .



Therefore structural parameter  $\theta$  satisfies both marginal and conditional (local) restrictions:

$$E_0 [g_1(Y, X; \theta_0)] = 0, \quad E_0 [g_2(Y; \theta_0) | X = x_0] = 0, \quad (9)$$

where  $g_2 = (\tilde{g}', g')'$  is obtained by gathering all conditional restrictions for environment  $x_0$ . Intuitively there are different situations concerning the identifiability of parameter  $\theta$ .

- i) If  $\theta$  is identifiable from the marginal restrictions only, the conditional ones  $E_0 [g_2(Y; \theta_0) | X = x_0] = 0$  provide a negligible additional information, and the efficient estimator of  $\theta$  will converge at a parametric rate.
- ii) If  $\theta$  is not identifiable from the marginal restrictions only, but is identifiable from both types of restrictions, we can expect different parametric, or nonparametric rates of convergence according to the function (component) of parameter  $\theta$  which is considered. This will be the general situation.

More precisely, the identification assumptions are the following.

**Assumption A.1:** *Parameter  $\theta$  is globally identifiable from marginal and conditional moment restrictions, that is the application:*

$$\theta \rightarrow \left( E_0 [g_1(Y, X; \theta)]', E_0 [g_2(Y; \theta) | X = x_0]' \right)' \quad \text{is one-to-one.}$$

**Assumption A.2:** *Parameter  $\theta$  is locally identifiable from marginal and conditional moment restrictions, that is the matrix:*

$$\begin{pmatrix} E_0 \left[ \frac{\partial g_1}{\partial \theta'} (Y, X; \theta_0) \right] \\ E_0 \left[ \frac{\partial g_2}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] \end{pmatrix} \text{ has full column rank.}$$

The above rank condition implies the order condition  $K_1 + K_2 \geq p$ , where  $K_1$  [resp.  $K_2$ ] denotes the number of marginal restrictions (resp. conditional restrictions). If matrix  $E_0 \left[ \frac{\partial g_1}{\partial \theta'} (Y, X; \theta_0) \right]$  has full column rank, then parameter  $\theta$  is locally identifiable from the marginal restrictions only and is said to be full information identifiable.

Assumptions A.1 and A.2 provide the identification conditions for structural parameter  $\theta$ . However, the parameter of primary interest for our purposes is the conditional moment:

$$\beta_0 = E_0 [a(Y; \theta_0) | X = x_0],$$

where  $a$  is a function of dimension  $L$ . At this step it is important to discuss the interpretation of parameter of interest  $\beta_0$ , which corresponds to a set of derivative prices at some given date in the application. More precisely we have to distinguish the mapping:

$$x \longmapsto E_0 [a(Y; \theta_0) | X = x],$$

which is a function, usually interpreted in terms of prediction, and its value at a given point  $x_0$ , that is  $\beta_0 = E_0 [a(Y; \theta_0) | X = x_0]$ , which can be considered as a standard parameter. The latter interpretation is used for developing our estimation approach. To this end, the parameters to be estimated can be written in an extended vector  $\theta^* = (\theta', \beta')'$  [see e.g. Back, Brown (1992)], whose true value  $(\theta'_0, \beta'_0)'$  satisfies the extended set of moment restrictions:

$$\begin{pmatrix} E_0 [g_1(Y, X; \theta_0)] \\ E_0 [g_2(Y; \theta_0) | X = x_0] \\ E_0 [a(Y; \theta_0) - \beta_0 | X = x_0] \end{pmatrix} = 0.$$

Since the dimension of  $\beta$  is equal to the number of moments of interest, that is the dimension of  $a$ , the extended parameter  $\theta^*$  is also globally or locally identified under Assumptions A.1, A.2. Note that the extended problem always involves restrictions conditional on a given value of the conditioning variable (the restriction defining  $\beta_0$ ), even if  $\theta_0$  is defined by means of uniform restrictions only.

## ii) Moment estimator

We will now consider moment estimators for  $\theta^*$  based on the approximated moment restrictions:

$$\begin{pmatrix} \widehat{E} [g_1(Y, X; \theta)] \\ \widetilde{E} [g_2(Y; \theta) | x_0] \\ \widetilde{E} [a(Y; \theta) - \beta | x_0] \end{pmatrix} \simeq 0,$$

where  $\widehat{E}$  and  $\widetilde{E}[\cdot|x_0]$  denote a sample average and a kernel estimator of the conditional moment, respectively. More precisely let us introduce a kernel estimator<sup>2</sup> of the conditional density  $f_0(y|x_0)$ . For expository purpose we assume that processes  $X_t$  and  $Y_t$  have identical dimension  $d$ , say, which is generally the case in applications to derivative pricing. The kernel density estimator is defined by:

$$\widehat{f}(y|x_0) = \frac{1}{h_T^d} \sum_{t=1}^T K\left(\frac{y_t - y}{h_T}\right) K\left(\frac{x_t - x_0}{h_T}\right) / \sum_{t=1}^T K\left(\frac{x_t - x_0}{h_T}\right), \quad (10)$$

where  $K$  is the  $d$ -dimensional kernel and  $h_T$  the bandwidth. The kernel  $K$  is a non-negative symmetric function satisfying:

$$\int_{\mathbb{R}^d} K(u)du = 1, \quad w^2 = \int_{\mathbb{R}^d} K^2(u)du < \infty.$$

The kernel density estimator can be used to approach a conditional moment  $E_0(g_2|x_0) = E_0[g_2(Y; \theta)|X = x_0]$  by:

$$\widetilde{E}(g_2|x_0) = \int g_2(y; \theta) \widehat{f}(y|x_0) dy \simeq \sum_{t=1}^T g_2(y_t; \theta) K\left(\frac{x_t - x_0}{h_T}\right) / \sum_{t=1}^T K\left(\frac{x_t - x_0}{h_T}\right).$$

Under standard regularity conditions including the bandwidth conditions:  $Th_T^d \rightarrow \infty$ ,  $(Th_T^d)^{1/2} h_T^2 \rightarrow 0$  as  $T \rightarrow \infty$ , the estimator  $\widetilde{E}(g_2|x_0)$  is consistent and asymptotically normal:

$$\sqrt{Th_T^d} \left( \widetilde{E}(g_2|x_0) - E_0(g_2|x_0) \right) \xrightarrow{d} N\left(0, w^2 V_0(g_2|x_0) / f_X(x_0)\right),$$

where  $f_X$  denotes the marginal density of process  $(X_t)$ . In particular the different estimated moments have different rates of convergence, that are  $\sqrt{T}$  for  $\widehat{E}$ ,  $\sqrt{Th_T^d}$  for  $\widetilde{E}(\cdot|x_0)$ .

**Definition 1:** A (kernel) moment estimator  $\widehat{\theta}^* = \left( \widehat{\theta}', \widehat{\beta}' \right)'$  of parameter  $\theta^*$

---

<sup>2</sup>Another type of nonparametric estimator of the conditional density could have been used.

is defined by:

$$\widehat{\theta}^*(\Omega) = \arg \min_{\theta^*=(\theta', \beta)'} \left( \sqrt{T} \widehat{E} [g_1(Y, X; \theta)]', \sqrt{Th_T^d} \widetilde{E} [g_2(Y; \theta)|x_0]', \sqrt{Th_T^d} \widetilde{E} [a(Y; \theta) - \beta|x_0]' \right) \\ \Omega \left( \sqrt{T} \widehat{E} [g_1(Y, X; \theta)]', \sqrt{Th_T^d} \widetilde{E} [g_2(Y; \theta)|x_0]', \sqrt{Th_T^d} \widetilde{E} [a(Y; \theta) - \beta|x_0]' \right)',$$

where  $\Omega$  is a weighting matrix.

Under standard regularity conditions, the associated (kernel) moment estimator of parameter  $\beta$  is consistent, converges at a nonparametric rate  $\sqrt{Th_T^d}$ , and is asymptotically normal with a variance-covariance matrix  $w^2 V_Z(\Omega)$  depending on the weighting matrix  $\Omega$ .

### iii) Semi-parametric efficiency bound

**Definition 2:** The (kernel) semi-parametric efficiency bound  $\mathcal{B}_Z(x_0, a)$  for  $\beta_0 = E_0(a|x_0)$  and given instruments  $Z$  is the minimal asymptotic variance  $V_Z(\Omega)$  corresponding to the optimal choice of  $\Omega$ .

The main result of this subsection is that the (kernel) semi-parametric efficiency bound for  $\beta_0$  depends on the selected instrument  $Z$  only through the local identification content of the corresponding marginal restrictions:

$$g_1(Y, X, \theta_0) = Z \cdot g(Y, \theta_0).$$

More precisely, what really matters is the null space of matrix  $J_1^Z = E_0 \left[ \partial g_1(Y, X; \theta_0) / \partial \theta' \right]$ . If  $s_Z$  denotes the rank of this matrix, its null space is characterized by a  $p \times (p - s_Z)$  matrix  $R_Z$  such that:

$$E_0 \left[ \frac{\partial g_1}{\partial \theta'}(Y, X; \theta_0) \right] R_Z = 0.$$

The columns of  $R_Z$  generate the null space of matrix  $J_1^Z$ . Moreover let us denote by  $\widetilde{R}$  a  $p \times s_Z$  matrix whose columns complete those of  $R_Z$  to a basis of  $\mathbb{R}^p$ . Then, the  $p \times p$  matrix  $R_1 = \left( \widetilde{R}, R_Z \right)$  is non singular and allows to define a new parameterization:

$$\eta = R_1^{-1} \theta = \left( \eta'_1, \eta'_2 \right)'.$$

The vector  $\eta_1$  defines  $s_Z$  linear combinations of structural parameters  $\theta_0$  which are identified from the marginal restrictions, while  $\eta_2$  corresponds to  $p - s_Z$  linear combinations for which the marginal restrictions are not informative, since:

$$E_0 \left[ \frac{\partial g_1}{\partial \eta_2'} (Y, X; \theta_0) \right] = E_0 \left[ \frac{\partial g_1}{\partial \theta'} (Y, X; \theta_0) \right] R_Z = 0.$$

This implies that parameters  $\eta_1$  can be estimated at a standard parametric rate, whereas  $\eta_2$  features a nonparametric rate of convergence induced by the local conditional moment restrictions. Then the intuition of the main result below is the following: as far as the semi-parametric efficiency bound for  $\beta_0$  is concerned, parameters  $\eta_1$  can be considered as known without estimation error, since they are actually estimated with a parametric rate of convergence, which is infinitely faster than the non-parametric rate of convergence for estimation of  $\beta_0$ . This is why the efficiency bound for  $\beta_0$  depends on the instrument  $Z$  only through the information matrix  $I_{0,Z}$ :

$$I_{0,Z} = f_X(x_0) \begin{pmatrix} E_0 \left( \frac{\partial g_2}{\partial \eta_2'} \right) & 0 \\ E_0 \left( \frac{\partial a}{\partial \eta_2'} \right) & -Id_L \end{pmatrix}' \begin{pmatrix} V_0(g_2) & Cov_0(g_2, a) \\ Cov_0(a, g_2) & V_0(a) \end{pmatrix}^{-1} \begin{pmatrix} E_0 \left( \frac{\partial g_2}{\partial \eta_2'} \right) & 0 \\ E_0 \left( \frac{\partial a}{\partial \eta_2'} \right) & -Id_L \end{pmatrix}, \quad (11)$$

where all moments are conditional on  $X = x_0$ . Note that  $I_{0,Z}^{-1}$  is similar to a standard GMM efficiency bound for estimation of parameters  $(\eta_2', \beta)'$  from moment restrictions based on functions  $(g_2', a' - \beta)'$ , but, by contrast with the standard setting [Hansen (1982), Back, Brown (1992)], both true unknown values of parameters and restrictions are defined conditional on the given value  $x_0$  of  $X$ .

**Proposition 1** : *Let instruments  $Z$  satisfying Assumptions A.1 and A.2 be given and the associated information matrix  $I_{0,Z}$  be defined by (11). Then, the (kernel) semi-parametric efficiency bound  $\mathcal{B}_Z(x_0, a)$  for conditional moment*

$\beta_0 = E_0(a|x_0)$  is the lower diagonal  $L \times L$  block of matrix  $I_{0,Z}^{-1}$ , that is:

$$\begin{aligned} \mathcal{B}_Z(x_0, a) &= \frac{1}{f_X(x_0)} \left\{ V_0 a - \text{Cov}_0(a, g)(V_0 g)^{-1} \text{Cov}_0(g, a) \right. \\ &\quad + \left[ E_0 \left( \frac{\partial a}{\partial \theta'} \right) R_Z - \text{Cov}_0(a, g)(V_0 g)^{-1} E_0 \left( \frac{\partial g}{\partial \theta'} \right) R_Z \right] \\ &\quad \left[ R'_Z E_0 \left( \frac{\partial g'}{\partial \theta} \right) (V_0 g)^{-1} E_0 \left( \frac{\partial g}{\partial \theta'} \right) R_Z \right]^{-1} \\ &\quad \left. \left[ R'_Z E_0 \left( \frac{\partial a'}{\partial \theta} \right) - R'_Z E_0 \left( \frac{\partial g'}{\partial \theta} \right) (V_0 g)^{-1} \text{Cov}_0(g, a) \right] \right\}, \end{aligned}$$

where all moments are conditional on  $X = x_0$  and evaluated at the true parameter value  $\theta_0$ .

**Proof.** See Appendix 1. ■

The matrix  $R_Z$  is not unique, but is defined up to a post-multiplication by a non-singular square matrix. The above semi-parametric efficiency bound is not modified by such a post-multiplication.

#### iv) Interpretation in terms of weak instruments

The problem considered above is related to the use of weak instruments. More precisely, the marginal moment restrictions in (9) are obtained by introducing standard instruments satisfying the usual conditions. At the contrary, the moment restrictions corresponding to a given value of the conditioning variable can be approximately written as:

$$E_0 [g_2(Y; \theta_0) | X = x_0] \simeq E_0 \left[ \frac{1}{f_X(x_0)h_T} K \left( \frac{X - x_0}{h_T} \right) g_2(Y; \theta_0) \right].$$

Thus they correspond to a marginal moment restriction constructed from instrument  $K \left( \frac{X - x_0}{h_T} \right) / [h_T f_X(x_0)]$ . This instrument admits a local interpretation which explains the different rate of convergence of the structural parameters when this "weak" instrument is used<sup>3</sup>. Despite this interpretation, the problem considered above is not a special case of the standard

---

<sup>3</sup>The corresponding sample moment  $\sqrt{Th_T^d} \tilde{E} [g_2(\theta)|x_0]$  in Definition 1 of the estimator is of order  $\sqrt{Th_T^d}$  for  $\theta \neq \theta_0$ , which is lower than  $\sqrt{T}$ , as in the standard case of weak instruments [see Stock, Wright (2000)].

literature on weak instruments [see e.g. Stock, Wright (2000)]. For instance the functions of the parameters which are weakly [resp. strongly] identified are not known a priori, and the asymptotic properties, especially the rates of convergence, of the associated GMM estimator differ from the rates of convergence obtained in the other types of applications which have been considered earlier in the literature.

## 2.1.2 Nonparametric efficiency bound

### i) Optimal instruments

The main lesson of the previous subsection is that the instrument  $Z$  only matters for estimation of  $\beta_0$  through the null space of matrix  $J_1^Z = E_0 \left[ \partial g_1(Y, X; \theta_0) / \partial \theta' \right]$ . Larger this null space, larger is the vector  $\eta_2$  of structural parameters which are non-identified from marginal restrictions and must be estimated at a non-parametric rate jointly with  $\beta$ , leading to the asymptotic joint covariance matrix  $I_{0,Z}^{-1}$ . Therefore, if  $Z$  and  $W$  are two alternative sets of instruments such that the null space of  $J_1^Z$  is included in the null space of  $J_1^W$ , the semi-parametric efficiency bound  $\mathcal{B}_Z(x_0, a)$  cannot be larger than  $\mathcal{B}_W(x_0, a)$ .

It turns out that there are many ways to choose instruments  $Z$  in order to get a minimal null set for  $J_1^Z$ . Let us define the subspace  $N_0$  of vectors  $v$  of  $\mathbb{R}^p$  such that:

$$E_0 \left[ \frac{\partial g}{\partial \theta'}(Y; \theta_0) \mid X = x \right] v = 0, \quad (12)$$

almost surely for the marginal distribution  $P^X$  of  $X$ . Vectors in  $N_0$  correspond to linear combinations of parameters  $\theta$  that cannot be identified from the uniform restrictions. Subspace  $N_0$  is included in the null space of  $J_1^Z$  for any choice of the instruments  $Z$ . Therefore, this null space will be minimal as soon as it coincides with  $N_0$ . Let us consider in particular:

$$Z = E_0 \left( \frac{\partial g'}{\partial \theta}(Y; \theta_0) \mid X \right) W(X), \quad (13)$$

where  $W(X)$  is ( $P^X$ -almost surely) a positive definite matrix. Then, for  $v$  in the null space of  $J_1^Z$ , we have:

$$E_0 \left[ E_0 \left( \frac{\partial g'}{\partial \theta}(Y; \theta_0) \mid X \right) W(X) E_0 \left( \frac{\partial g}{\partial \theta'}(Y; \theta_0) \mid X \right) v \right] = 0,$$

or:

$$v' E_0 \left( \frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X) E_0 \left( \frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) v = 0, \quad P^X\text{-almost surely,}$$

or:

$$E_0 \left( \frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) v = 0, \quad P^X\text{-almost surely,}$$

that is  $v$  belongs to  $N_0$ . Therefore, the choice (13) of instruments  $Z$  provides the minimal null set  $J_1^Z$  and is optimal whenever it fulfils the identification Assumptions A.1 and A.2. Moreover, for this special choice of instruments, Assumption A.2 is clearly tantamount to the following identification assumption:

**Assumption A.2\*:** *The structural parameter  $\theta$  is locally identifiable from the conditional restrictions, that is  $v = 0$  is the only vector which fulfils jointly:*

(i) *the uniform restrictions:  $E_0 \left[ \frac{\partial g}{\partial \theta'} (Y; \theta_0) | X = x \right] v = 0$ ,  $P^X$ -almost surely,*

(ii) *the conditional restrictions:  $E_0 \left[ \frac{\partial g_2}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] v = 0$ , for the given value  $x_0$  of  $X$ .*

Thus, we have shown:

**Lemma 1:** *Under Assumption A.2\* any instrument  $Z = E_0 \left( \frac{\partial g'}{\partial \theta} | X \right) W(X)$ , where  $W(X)$  is a positive definite matrix, satisfies Assumption A.2 and is an optimal instrument for estimating  $\beta_0 = E_0(a|x_0)$ .*

Since we focus on nonparametric estimation of  $\beta_0$ , the set of optimal instruments is larger than the standard one derived by Hansen (1982) and Chamberlain (1987) for efficient estimation of the structural parameters  $\theta$ . While in the standard framework  $W(X) = [Var_0(g(Y, \theta_0)|X)]^{-1}$  is the efficient weighting of the conditionally heteroskedastic moment conditions, any choice of a positive definite matrix  $W(X)$  is valid when  $\beta$  is the parameter of interest. Moreover, the optimality result given in Lemma 1 is more general than the standard one, since it does not require full (parametric) identification of  $\theta$ .

## ii) The identification assumption



Another useful formulation of Assumption A.2\* can be derived by considering a  $p \times (p - s)$  matrix  $R$  whose columns constitute a basis of the space  $N_0$  defined in (12). Indeed any vector  $v$  satisfying Assumption A.2\* i) can be written as  $v = Rc$  for some  $(p - s)$ -dimensional vector  $c$ . Then, Assumption A.2\* ii) becomes:

$$E_0 \left[ \frac{\partial g_2}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] Rc = 0 \implies c = 0,$$

that is  $E_0 \left[ \frac{\partial g_2}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] R$  is full column rank. Thus, Assumption A.2\* can be rewritten as:

**Assumption A.2\*:** *The matrix:*

$$E_0 \left[ \frac{\partial g_2}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] R \quad \text{is full column rank,}$$

for any  $p \times (p - s)$  matrix  $R$  whose columns generate the space:

$$N_0 = \left\{ v \in \mathbb{R}^p : E_0 \left[ \frac{\partial g}{\partial \theta'} (Y; \theta_0) | X = x \right] v = 0, P^X\text{-almost surely} \right\}.$$

### iii) Kernel nonparametric efficiency bound

Let us now derive the (kernel) nonparametric efficiency bound. By the above results, the matrix  $R$  coincides with the matrix  $R_Z$  corresponding to the optimal instrument  $Z$  in Lemma 1. A new parameterization:

$$\eta = R_1^{-1} \theta = \left( \eta'_1, \eta'_2 \right)',$$

can be defined as above with  $R_1 = \left( \tilde{R}, R \right)$ , where matrix  $\tilde{R}$  completes the basis of  $\mathbb{R}^p$ . The vector  $\eta_1$  represents the maximal set of structural parameters that can be identified from uniform restrictions only. Then the information matrix  $I_0$  corresponding to parameters  $(\eta'_2, \beta')'$  is defined from (11) by:

$$I_0 = f_X(x_0) \left( \begin{array}{cc} E_0 \left( \frac{\partial g_2}{\partial \theta'} \right) R & 0 \\ E_0 \left( \frac{\partial a}{\partial \theta'} \right) R & -Id_L \end{array} \right)' \left( \begin{array}{cc} V_0(g_2) & Cov_0(g_2, a) \\ Cov_0(a, g_2) & V_0(a) \end{array} \right)^{-1} \left( \begin{array}{cc} E_0 \left( \frac{\partial g_2}{\partial \theta'} \right) R & 0 \\ E_0 \left( \frac{\partial a}{\partial \theta'} \right) R & -Id_L \end{array} \right).$$

The main result of this section is a direct consequence of Proposition 1.

**Proposition 2** : Let Assumption A.2\* be satisfied. Then the (kernel) non-parametric efficiency bound  $a \rightarrow \mathcal{B}(x_0, a)$  for conditional moment  $E_0(a|x_0)$  is the lower diagonal  $L \times L$  block of matrix  $I_0^{-1}$ , that is:

$$\begin{aligned} \mathcal{B}(x_0, a) = & \frac{1}{f_X(x_0)} \left\{ V_0 a - \text{Cov}_0(a, g)(V_0 g)^{-1} \text{Cov}_0(g, a) \right. \\ & + \left[ E_0 \left( \frac{\partial a}{\partial \theta'} \right) R - \text{Cov}_0(a, g)(V_0 g)^{-1} E_0 \left( \frac{\partial g}{\partial \theta'} \right) R \right] \\ & \left[ R' E_0 \left( \frac{\partial g'}{\partial \theta} \right) (V_0 g)^{-1} E_0 \left( \frac{\partial g}{\partial \theta'} \right) R \right]^{-1} \\ & \left. \left[ R' E_0 \left( \frac{\partial a'}{\partial \theta} \right) - R' E_0 \left( \frac{\partial g'}{\partial \theta} \right) (V_0 g)^{-1} \text{Cov}_0(g, a) \right] \right\}, \quad \forall a, \end{aligned}$$

where all moments are conditional on  $X = x_0$ , evaluated at the true parameter value  $\theta_0$  and matrix  $R$  is defined in Assumption A.2\*.

The efficiency bound of Proposition 2 is not modified by post-multiplication of the matrix  $R$  by a non-singular matrix.

## 2.2 Special cases

Proposition 2 can be applied to the limiting cases of full and limited information, respectively.

### 2.2.1 Full information identifiability

When the structural parameter  $\theta$  is full information identifiable, the space  $N_0 = \{0\}$  and the column space of matrix  $R$  in Proposition 2 is zero. We get the corollary below.

**Corollary 1:** *The full information (kernel) nonparametric efficiency bound is :*

$$\mathcal{B}(x_0, a) = \frac{1}{f_X(x_0)} \left\{ V_0(a|x_0) - \text{Cov}_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} \text{Cov}_0(g_2, a|x_0) \right\}.$$

This result is easily understood when all moment restrictions  $E_0[g(Y; \theta_0)|X = x] = 0$  are uniform and  $\theta$  is full information identifiable.

Since  $\theta$  can be estimated at a parametric rate using the marginal moment restrictions, it can be assumed known for the computation of the nonparametric efficiency bound. This explains why the second term of the decomposition of the efficiency bound involving derivatives with respect to  $\theta$  vanishes.

The same reasoning applies when  $\theta$  is full information identifiable and satisfies both uniform and local restrictions, since the additional local restrictions are not informative for the estimation of  $\theta$ . Note however that they are informative for the estimation of the moment of interest  $\beta_0 = E_0(a|x)$ . Indeed the nonparametric efficiency bound in Corollary 1 involves the whole set of constraints  $g_2 = (\tilde{g}', g')'$ .

Finally, note that the conditional moment of interest is also equal to:

$$E_0(a|x_0) = E_0[a(Y; \theta_0) - Cov_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} g_2(Y; \theta_0) | x_0].$$

The bound is nothing but the variance-covariance matrix of the residual term in the affine regression of  $a$  on  $g_2$ . A similar interpretation has already been put forward by Back and Brown (1993) in an unconditional setting and extended to a conditional framework by Bonnal and Renault (2004).

### 2.2.2 Limited information

Let us now assume that all moment restrictions are conditional on the given value  $X = x_0$ :

$$E[\tilde{g}(Y; \theta_0) | X = x_0] = 0.$$

**Corollary 2:** *The limited information (kernel) nonparametric efficiency bound is given by:*

$$\begin{aligned} \mathcal{B}(x_0, a) &= \frac{1}{f_X(x_0)} \left\{ V_0 a - Cov_0(a, \tilde{g})(V_0 \tilde{g})^{-1} Cov_0(\tilde{g}, a) \right. \\ &\quad \left. + \left[ E_0\left(\frac{\partial a}{\partial \theta}\right) - Cov_0(a, \tilde{g})(V_0 \tilde{g})^{-1} E_0\left(\frac{\partial \tilde{g}}{\partial \theta}\right) \right] \right. \\ &\quad \left[ E_0\left(\frac{\partial \tilde{g}'}{\partial \theta}\right) (V_0 \tilde{g})^{-1} E_0\left(\frac{\partial \tilde{g}}{\partial \theta'}\right) \right]^{-1} \\ &\quad \left. \left[ E_0\left(\frac{\partial a'}{\partial \theta}\right) - E_0\left(\frac{\partial \tilde{g}'}{\partial \theta}\right) (V_0 \tilde{g})^{-1} Cov_0(\tilde{g}, a) \right] \right\}, \end{aligned}$$

where all moments are conditional on  $X = x_0$  and evaluated at  $\theta_0$ .

This is the formula in Proposition 2 with  $g_2 = \tilde{g}$  and  $R = Id$ , since no linear combination of parameter  $\theta$  is full-information identifiable.

### 3 Information based estimator

The estimation of optimal instruments and the derivation of the associated optimal weighting matrix in a moment method may be difficult to implement in practice and provide rather erratic results in finite sample [see e.g. Altonji, Segal (1996), Hansen, Heaton, Yaron (1996)]. It has been proposed in the literature (see the Introduction) to derive the optimal moment estimator in a single step by optimizing with respect to both the structural parameter and the conditional pdf an appropriate measure of discrepancy between the distribution and the unconstrained kernel density subject to the moment restrictions. The discrepancy measure is usually chosen among the Cressie-Read family of divergences [Cressie, Read (1984)], leading to the so-called empirical likelihood, chi-square or Kullback-Leibler information based (KLIC) approach.

In this section we develop an information based equivalent of the XMM estimator. The goal is to correct for a structural drawback of the XMM approach as introduced in the previous section, namely lack of coherency. More precisely, conditional moments estimated by XMM are generally not consistent with an underlying estimator of the conditional pdf which satisfies the unit mass and non-negativity constraints. In derivative pricing applications, where the conditional (risk neutral) pdf is interpreted as a state price density, such a feature may imply misleading arbitrage opportunities in estimated option prices.

The whole existing literature on information based estimation considers a setting with uniform moment restrictions and assumes the full information identifiability of parameter  $\theta$ . In this section we develop an approach for the general framework with both uniform and local restrictions by combining in an appropriate way chi-square and KLIC measures. The aim of this approach is to get an estimator of the conditional pdf, which satisfies the unit mass and non-negativity restrictions, while keeping the estimator tractable.

In the first subsection, we explain why the XMM approach features a lack of coherency, and does not provide an appropriate approximation of the conditional density. The information based estimator is introduced in Section 3.2 and its (kernel) nonparametric efficiency is proved. Finally Section 3.3

considers the limiting cases of full and limited information.

### 3.1 A lack of coherency of XMM

It is well-known that a GMM approach can feature a lack of coherency, when the conditional moments of interest are multiple. More precisely, it is expected that an estimation approach for  $E_0(a|x_0) = E[a(Y; \theta_0)|X = x_0]$  provides an estimator of the type:

$$\widehat{E}(a|x_0) = \int a(y; \widetilde{\theta}) \widetilde{f}(y|x_0) dy,$$

where  $\widetilde{\theta}$  is an estimator of  $\theta$  and  $\widetilde{f}$  is an estimator of the conditional density. The XMM approach does not satisfy this requirement.

i) For instance in the full-information case with full information identifiable parameter, the XMM estimator of the moment of interest coincides with the estimator of the moment of the residual:

$$E_0 [a(Y; \theta_0) - Cov_0(a, g|x_0) V_0(g|x_0)^{-1} g(Y; \theta_0) | x_0],$$

which can be written as:

$$\int a(y; \theta_0) \widehat{f}(y|x_0) \left[ 1 - g(y; \theta_0)' V_0(g|x_0)^{-1} E_0(g|x_0) \right] dy.$$

This is an integral expression, with respect to a measure which does not depend on  $a$ , satisfies the unit mass restriction, but is not necessarily positive.

ii) Moreover in the general mixed framework such an integral representation can even not exist, since the XMM estimator of  $\theta$  depends on the moment of interest  $a$ .

Therefore, it is important to introduce an estimation method, which is both coherent and (kernel) nonparametrically efficient.

### 3.2 Information based estimator

The (unconstrained) kernel estimator  $\widehat{f}(y|x)$  is a consistent estimator of the conditional pdf. However it is not nonparametrically efficient, since it does not take into account the parameterized moment restrictions. The kernel

density estimator can be improved by looking for the pdf which is the closest to  $\widehat{f}(y|x)$  and satisfies the moment restrictions.

In this section we consider the joint estimator defined by:

$$\begin{aligned} & \left( \widehat{f}_0(\cdot|x_0), \widehat{f}_0(\cdot|x_1), \dots, \widehat{f}_0(\cdot|x_T), \widehat{\theta} \right) \\ = & \arg \min_{f^0, f^1, \dots, f^T, \theta} \frac{1}{T} \sum_{t=1}^T \int \frac{[\widehat{f}(y|x_t) - f^t(y)]^2}{\widehat{f}(y|x_t)} dy + h_T^d \int \log [f^0(y)/\widehat{f}(y|x_0)] f^0(y) dy, \\ \text{s.t.} & \int f^t(y) dy = 1, \quad t = 1, \dots, T, \\ & \int f^0(y) dy = 1, \\ & \int g(y; \theta) f^t(y) dy = 0, \quad t = 1, \dots, T, \\ & \int g_2(y; \theta) f^0(y) dy = 0. \end{aligned}$$

The objective function includes two components: a chi-square distance is used for the optimization with respect to the conditional distributions associated with the sample values of the conditioning variable, whereas an information criterion is used for the conditioning value  $x_0$  corresponding to the conditional moment of interest. Moreover two types of constraints are introduced: the uniform restrictions are written for all observations  $x_1, \dots, x_T$ , whereas the local restrictions are written for  $x_0$  only. The chi-square component allows for closed form solutions  $f^1(\theta), \dots, f^T(\theta)$  for a given  $\theta$  without ensuring positivity. Therefore the objective function is easily concentrated with respect to  $f^1, \dots, f^T$ . Then the information criterion will allow a solution  $\widehat{f}_0(\cdot|x_0)$  satisfying the unit mass and positivity restrictions<sup>4</sup>. In particular, the computation of the estimator only involves the optimization of a concentrated criterion with respect to parameter  $\theta$  and a Lagrange multiplier of dimension  $\dim(g_2)$  [see Appendix 2 i) for the concentration of the objective function].

Then the information based estimator of the conditional moment is defined by:

$$\widehat{E}(a|x_0) = \int a(y; \widehat{\theta}) \widehat{f}_0(y|x_0) dy.$$

The (kernel) nonparametric efficiency of the information based estimator of  $f_0(y|x_0)$  is established in Appendix 2.

---

<sup>4</sup>See e.g. Kitamura-Stutzer (1997).

**Proposition 3** *The estimator  $\hat{E}(a|x_0)$  is consistent, converges at rate  $\sqrt{Th_T^d}$ , is asymptotically normal and (kernel) nonparametrically efficient:*

$$\frac{\sqrt{Th_T^d}}{w}(\hat{E}(a|x_0) - E_0(a|x_0)) \xrightarrow{d} N(0, \mathcal{B}(x_0, a)),$$

for any  $a$ .

### 3.3 Special cases

#### 3.3.1 Limited information

When the moment restrictions are:

$$E_0[\tilde{g}(Y; \theta_0)|X = x_0] = \int \tilde{g}(y; \theta_0) f_0(y|x_0) dy = 0,$$

the optimization problem becomes:

$$(\hat{f}_0(\cdot|x_0), \hat{\theta}) = \arg \min_{f, \theta} \int \log \left[ \frac{f(y)}{\hat{f}(y|x_0)} \right] f(y) dy \quad (14)$$

$$\text{s.t.} \quad \int f(y) dy = 1, \int f(y) \tilde{g}(y; \theta) dy = 0.$$

The associated estimator  $\hat{E}(a|x_0) = \int a(y; \hat{\theta}) \hat{f}_0(y|x_0) dy$  is (kernel) nonparametrically efficient. Its asymptotic variance is given by the expression of  $\mathcal{B}(x_0, a)$  in Corollary 2.

#### 3.3.2 Full information

In the full information case a (kernel) nonparametrically efficient estimator can be defined by optimizing the mixed chi-square/information criterion with respect to both  $\theta$  and the conditional distribution [see Section 3.2]. Our approach extends results derived in the literature in the special case of pure uniform restrictions and i.i.d. observations. For instance Bonnal and Renault (2004) derive a result similar to Proposition 3, but without imposing positivity of the estimated conditional distribution. Kitamura, Tripathi, Ahn (2004) focus on estimation and inference about structural parameter  $\theta$  only adopting a smooth empirical likelihood approach. Smith (2004) establish the results of Kitamura, Tripathi, Ahn (2004) for the Generalized Empirical

Likelihood (GEL) methodology <sup>5</sup>. Section 3.2 extends the results of these papers to a general setting with both uniform and local moment restrictions where the structural parameter is possibly full information unidentifiable, which is the relevant setting for derivative pricing applications.

In the full information case, a (kernel) efficient estimator of the moment of interest can also be derived in a two step approach. Indeed the structural parameter  $\theta$  can be estimated consistently (and efficiently) by means of the uniform restrictions only. This allows to separate the estimation of  $\theta$  and the estimation of the conditional pdf of interest  $f(y|x_0)$ . A two step estimator is defined by:

$$\widehat{E}(a|x_0) = \int a(y; \widehat{\theta}) \widehat{f}_0(y|x_0) dy,$$

where:

$$\begin{aligned} \widehat{f}_0(\cdot|x_0) &= \arg \min_f \int \log[f(y)/\widehat{f}(y|x_0)] f(y) dy, \\ \text{s.t.} \quad &\int f(y) dy = 1, \int f(y) g_2(y; \widehat{\theta}) dy = 0, \end{aligned} \quad (15)$$

and  $\widehat{\theta}$  is any estimator of  $\theta$  converging at a parametric rate. This estimator can be a consistent (but possibly inefficient) moment estimator, a GMM estimator, or a continuously updated estimator [see Hansen, Heaton, Yaron (1996)]. Insofar as  $\widehat{\theta}$  is consistent and root- $T$  asymptotically normal,  $\widehat{E}(a|x_0)$  reaches the (kernel) nonparametric efficiency bound in Corollary 1. By contrast with standard GMM, the two-step procedure does not imply any efficiency loss since the rates of convergence are different in the two steps.

## 4 Derivative pricing

Usually derivative pricing formulas involve two types of parameters characterizing the dynamic of the underlying asset returns and the risk premia, respectively. The parameters can be finite dimensional or functional, leading in practice to parametric or nonparametric pricing methods.

i) When the markets are complete, the only parameter concerns the dynamics of the underlying asset returns and can be estimated from return

---

<sup>5</sup>See also Ai-Chen (2003).



data. When the parameter is finite dimensional, it is usually estimated by maximum likelihood<sup>6</sup>. Alternatively the estimation of some parameters can be based on derivative data only. For instance, if the Black-Scholes model is well-specified, the volatility can be estimated by an implied volatility computed from an observed option price. However the drift is not identifiable from option prices observed at a given data, that is by a cross-sectional approach.

ii) In the incomplete market framework, the model includes in general both dynamic and risk premia parameters. These parameters can be estimated by an appropriate use of both historical and cross-sectional data. Such approaches have already been considered in the literature for parametric models [see e.g. Garcia, Luger, Renault (2003) for estimation, De Munnik, Schotman (1994), Bams (1998) for specification tests]. Some of the parameters can also be estimated by a pure cross-sectional approach using option data only, the typical example being the parametric fit of Black-Scholes implied volatility surfaces. Different nonparametric approaches have also been considered in the literature. They are based for instance on the nonparametric approximation of the implied volatility surface [see e.g. Hutchinson, Lo, Poggio (1994) for using neural networks, and Ait-Sahalia, Lo (1998), who use a kernel approach and deduce a nonparametric estimator of the state price density]. An alternative approach relies on maximum entropy risk neutral densities for given maturity, derived by using both asset and option data [see e.g. Rubinstein (1994), Jackwerth, Rubinstein (1996), Buchen, Kelly (1996), Stutzer (1996), Jondeau, Rockinger (2000)].

In this section, we consider a semi-nonparametric approach, in which the historical parameter is functional and the risk premia parameter is finite dimensional. We explain how to use jointly underlying returns and derivative prices for efficient pricing of other derivatives.

## 4.1 The estimating constraints

For expository purpose let us consider European calls written on an underlying asset with geometric return  $y_t$ . We assume that the return series  $(Y_t)$  is a Markov process of order one under the historical probability and that the price at  $t$  of a European call with moneyness strike  $s$  and residual maturity

---

<sup>6</sup>This approach can be extended to functional parameter, leading for instance to indirect spline estimation of the state price density [see e.g. Gouriéroux, Monfort (2001)].

one can be written as<sup>7</sup> :

$$c_t(s) = E_t [m(Y_{t+1}; \theta)(\exp Y_{t+1} - s)^+], \quad (16)$$

where  $m(Y_{t+1}; \theta)$  is the stochastic discount factor. The finite dimensional parameter  $\theta$  characterizes the risk premia, whereas the historical conditional distribution of  $Y_{t+1}$  given  $Y_t$  is left unspecified.

Let us now assume observations of a finite number of derivative prices  $c_{t_0}(s_k)$ ,  $k = 1, \dots, K$ , at a given date  $t_0$  and observations of underlying asset returns for earlier dates  $t = t_0 - T + 1, \dots, t_0$ . Then the moment restrictions are twofold. Some constraints concern the derivatives and are given by :

$$c_{t_0}(s_k) = E [m(Y_{t+1}; \theta)(\exp Y_{t+1} - s_k)^+ | Y_t = y_{t_0}], k = 1, \dots, K. \quad (17)$$

Other constraints concern the pricing formula for the underlying asset and the riskfree asset. They are :

$$\begin{aligned} E[m(Y_{t+1}; \theta) | Y_t = y_t] &= 1, \forall y_t, \\ E[m(Y_{t+1}; \theta) \exp Y_{t+1} | Y_t = y_t] &= 1, \forall y_t, \end{aligned} \quad (18)$$

assuming for simplicity a deterministic zero risk-free rate.

The second subset of constraints on  $\theta$  are uniform with respect to the conditioning value, whereas the conditioning value is fixed in the first subset. The distinction between the two types of moment restrictions is due to the lack of liquidity of some assets. If the asset is highly liquid, its price can be observed at any date leading to uniform conditional moment restrictions (if the number of observation dates is large and the return process stationary with a continuous stationary distribution). If the asset is not very liquid the price is observed for a limited number of dates, for instance for environment  $y_{t_0}$  existing at date  $t_0$ <sup>8</sup>. This is the case for derivatives.

Different pricing formulas are derived below depending whether the second subset of moment restrictions is totally taken into account.

---

<sup>7</sup>More precisely this is the call price in percentage of the underlying asset price. Since  $(p_{t+1} - sp_t)^+ = p_t(\exp y_{t+1} - s)^+$ , the call or put written on  $p_{t+1}$  can also be written on  $\exp y_{t+1}$ .

<sup>8</sup>See the discussion in Aït-Sahalia, Lo (1998) for the evolution of the set of liquid options on S&P.

## 4.2 Derivative pricing with limited information

Let us assume that the interest is in the price at  $t_0$  of a European call with maturity 1 and strike  $s$ . Its price is equal to the conditional moment:

$$E(a|y_{t_0}) = E[m(Y_{t+1}; \theta)(\exp Y_{t+1} - s)^+ | Y_t = y_{t_0}].$$

Under limited information, the only restrictions  $E(\tilde{g}|y_{t_0})$ , which are taken into account, correspond to the same conditioning value. There is a set of  $K + 2$  restrictions:

$$\begin{aligned} E[m(Y_{t+1}; \theta)(\exp Y_{t+1} - s_k)^+ - c_{t_0}(s_k) | Y_t = y_{t_0}] &= 0, \quad k = 1, \dots, K, \\ E[m(Y_{t+1}; \theta) - 1 | Y_t = y_{t_0}] &= 0, \\ E[m(Y_{t+1}; \theta) \exp Y_{t+1} - 1 | Y_t = y_{t_0}] &= 0. \end{aligned} \tag{19}$$

Then we can apply the estimation approach described in Section 3.3.1. This approach ensures that the estimated risk neutral pdf is nonnegative, which is compatible with the no arbitrage restrictions<sup>9</sup>.

*Whereas the conditional moment restrictions concern date  $t_0$  and environment  $y_{t_0}$  only, the approach is not a pure cross-sectional approach.* Indeed the observations  $y_{t_0-T+1}, \dots, y_{t_0}$  corresponding to the other dates are used in the estimation approach of the conditional (historical) pdf. In particular the derivative prices will be consistently estimated, if the number of observations  $T$  is large, even if the number of derivatives  $K$  is rather small (but larger than the parameter size). In particular the asymptotic theory is very different and more realistic than the theory usually developed in the literature, which assumes an infinite number of liquid derivatives at the date of interest[see e.g. Ait-Sahalia, Lo (1998)].

Finally note that the limited information method differs from the entropy based approaches introduced in the literature by the choice of the benchmark risk neutral distribution. In our framework this distribution is  $m(y; \theta_0)f_0(y|x_0)$ , where  $f_0(\cdot|x_0)$  is the historical conditional pdf. In Stutzer (1996) p1639 the benchmark distribution is the historical distribution itself

---

<sup>9</sup>The conditional moments could also be estimated by XMM, but as noted in Section 3.1 this does not ensure a corresponding risk neutral density, and, if the latter exists, its positivity. Thus XMM could create misleading arbitrage opportunities in estimated option prices.

(implicitly assuming zero risk premia); a parametric benchmark such as a Black-Scholes lognormal distribution is suggested by Rubinstein (1994) and Jackwerth, Rubinstein (1996), whereas a uniform distribution has been implicitly selected in Buchen, Kelly (1996) and Jondeau, Rockinger (2000). Moreover, maximum entropy methods focus on the state price density for a given date and a given maturity, whereas our approach allows to estimate coherently state price densities at a given date for all maturities.

### 4.3 Derivative pricing with mixed limited- and full information

Let us now assume that both types of moment restrictions (17) and (18) are taken into account. Derivative pricing can be improved by considering jointly the dynamics of the underlying asset prices between  $t = t_0 - T + 1$  and  $t = t_0$  and the way some prices of European calls depend on the strike for date  $t_0$ .

Two cases have to be distinguished according to the full information identifiability of parameter  $\theta$  from underlying asset price dynamics.

#### i) Full information identifiability

If parameter  $\theta$  is identifiable from uniform moment restrictions (18), restrictions (17) can asymptotically be neglected for the estimation of  $\theta$ . A GMM estimator  $\hat{\theta}$  of  $\theta$  can be computed by using restrictions (18) only and is consistent at a parametric rate. Then we can apply the estimation method described in Section 3.3.2 with  $\hat{f}(\cdot|x_0)$  a kernel estimator of the conditional pdf given  $Y_t = y_{t_0}$  and the set of restrictions (19).

#### ii) Full information underidentifiability

As seen in Section 2.1, a part of the parameters can be identified from the asset price dynamics (uniform restrictions) and will converge at a parametric rate, whereas the remaining parameters are identified by means of the cross-sectional restriction (17) and converge at a nonparametric rate. The latter are linear combinations of parameters  $R'\theta$ , where the columns of matrix  $R$  span the null space  $N_0$  defined in (12), with moment function  $g$  corresponding to restrictions (18). In this case the estimation has to be performed with the general criterion introduced in Section 3.2.

When  $\theta$  is full information underidentified there exist a multiplicity of values of parameter  $\theta$ , that is a multiplicity of sdf, such that the no-arbitrage conditions are satisfied for both the riskfree asset and the underlying risky asset [see equation (18)]. In the incomplete market framework, the choice of a parametric specification for the sdf may be not sufficient to get a unique pricing kernel from the observation of liquid asset prices. In other words, the specification allows for some residual incompleteness and, from a financial point of view, the degree of full information underidentification is equal to the dimension of this residual incompleteness.

#### 4.4 Comparison of the limited and mixed information approaches

Let us first note that both approaches use jointly historical information [by means of the kernel estimate of the conditional pdf and possibly by uniform moment restrictions (18)] and cross sectional information by moment restrictions (17). Moreover they are consistent when  $T$  tends to infinity with  $K$  fixed, whenever  $\theta$  is identifiable from the whole set of uniform and local moment restrictions.

When  $\theta$  is identifiable from the conditional restrictions at date  $t_0$ , it is possible to use either the general approach, or the limited information approach. The limited information method is likely to be preferred in practice in a first step. Firstly the asymptotic variance is larger than the variance derived by the general approach, leading to larger prediction intervals for derivative prices (which is a drawback from a statistical point of view), but more secure risk management (which is an advantage from the financial point of view). Secondly it corresponds to the usual practice of reporting daily the implied volatilities in the Black-Scholes framework. More precisely let us assume that the pricing model is misspecified and that the stochastic discount factor is  $m(y_{t+1}, \theta(y_t))$ , in which  $\theta$  depends on the lagged value. The limited information method provides the estimate of  $\theta(y_{t_0})$ , whereas the general method provides a kind of average of  $\theta(y_t)$  on all values observed in the past, without the interpretation of an integrated risk premium. By applying the limiting information approach at several consecutive dates  $t_0, t_0 + 1, t_0 + 2, \dots$ , we can expect to detect an instability of the risk premium.

## 5 Stochastic volatility model

In this section we illustrate the extended method of moments (XMM), or its information based equivalent, for efficient derivative pricing. In Section 5.1 we describe the data generating process to get the prices of the underlying asset and derivatives. The DGP is a discrete time version of the stochastic volatility model of Heston (1993) and Ball-Roma (1994), with a risk premium introduced in the return equation<sup>10</sup>. In Section 5.2 we describe the semi-parametric model which is used for derivative pricing and discuss the identification of the risk premia parameter. The nonparametric efficiency bounds for limited- and mixed information restrictions are computed in Section 5.3 for the prices of European calls. We discuss how they depend on the strike and on the set of observed derivative prices. Finally, the finite sample properties of the estimated option prices and of the estimated structural parameters are analyzed by Monte-Carlo in Section 5.4.

### 5.1 The design

Let us consider a market with a risk-free asset, with a zero risk-free rate, and a risky asset with geometric return  $r_t = \log(p_t/p_{t-1})$  such that:

$$r_t = \gamma\sigma_t^2 + \sigma_t\varepsilon_t, \quad (20)$$

where  $(\varepsilon_t)$  is a standard Gaussian white noise,  $\sigma_t^2$  denotes the volatility and  $\gamma$  measures the magnitude of the risk premium in the expected return. The intercept is set to zero because of no-arbitrage restrictions. Indeed for zero volatility  $\sigma_t = 0$  the return becomes deterministic and has to coincide with the zero risk-free rate.

The volatility  $(\sigma_t^2)$  is stochastic, with a dynamics independent of the shocks  $(\varepsilon_t)$  on returns. It follows an autoregressive gamma process (ARG), which is the time discretized Cox-Ingersoll-Ross process [see Gouriéroux, Jasiak (2000)]. The transition distribution of the stochastic volatility is characterized by the conditional Laplace transform (moment generating function):

$$\begin{aligned} \Psi_t(u) &= E [\exp(-u\sigma_{t+1}^2) \mid \sigma_t^2] \\ &= \exp[-a(u)\sigma_t^2 - b(u)], \end{aligned} \quad (21)$$

---

<sup>10</sup>See also Gouriéroux, Sufana (2004).

where:  $a(u) = \rho \frac{u}{1+cu}$ ,  $b(u) = \delta \log(1 + cu)$ . The positive parameter  $\rho$  is the first order autocorrelation of the variance process  $(\sigma_t^2)$ , parameter  $\delta \geq 0$  describes its (conditional) over-/under-dispersion, and  $c > 0$  is a scale parameter. In this model the factors are the return and the volatility:

$$Y_t = (r_t, \sigma_t^2)' . \quad (22)$$

Model (20)-(22) is completed by the parametric specification of the stochastic discount factor for period  $(t, t + 1)$ . The sdf is specified as:

$$M_{t,t+1} = \exp(-\nu_0 - \nu_1 \sigma_{t+1}^2 - \nu_2 \sigma_t^2 - \nu_3 r_{t+1}), \quad (23)$$

where  $\nu_0, \nu_1, \nu_2, \nu_3$  are parameters. The exponential affine specification (23) is compatible with the no-arbitrage restrictions and provides simple pricing formulas.

Let us first consider the restrictions implied by no-arbitrage opportunity. They are obtained by writing the pricing formula for both the risk-free asset and the underlying asset. We get:

$$\begin{aligned} & \begin{cases} E_t(M_{t,t+1}) = 1, \\ E_t(M_{t,t+1} \exp r_{t+1}) = 1, \end{cases} \\ \Leftrightarrow & \begin{cases} E_t \exp[-\nu_0 - \nu_1 \sigma_{t+1}^2 - \nu_2 \sigma_t^2 - \nu_3 r_{t+1}] = 1, \\ E_t \exp[-\nu_0 - \nu_1 \sigma_{t+1}^2 - \nu_2 \sigma_t^2 - (\nu_3 - 1) r_{t+1}] = 1, \end{cases} \\ \Leftrightarrow & \begin{cases} E_t \exp\left[-\nu_0 - \left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2}\right) \sigma_{t+1}^2 - \nu_2 \sigma_t^2\right] = 1, \\ E_t \exp\left[-\nu_0 - \left(\nu_1 + (\nu_3 - 1) \gamma - \frac{(\nu_3 - 1)^2}{2}\right) \sigma_{t+1}^2 - \nu_2 \sigma_t^2\right] = 1, \end{cases} \\ & \text{(by integrating } r_{t+1} \text{ conditional on } \sigma_{t+1}^2) \\ \Leftrightarrow & \begin{cases} \nu_0 + a\left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2}\right) \sigma_t^2 + \nu_2 \sigma_t^2 + b\left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2}\right) = 0, \\ \nu_0 + a\left[\nu_1 + (\nu_3 - 1) \gamma - \frac{(\nu_3 - 1)^2}{2}\right] \sigma_t^2 + \nu_2 \sigma_t^2 + b\left[\nu_1 + (\nu_3 - 1) \gamma - \frac{(\nu_3 - 1)^2}{2}\right] = 0. \end{cases} \end{aligned} \quad (24)$$

Since the above conditions have to be satisfied for any admissible value of  $\sigma_t^2$ , we get the following restrictions on the parameters:

$$\begin{cases} \nu_0 + b\left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2}\right) = 0, \\ \nu_0 + b\left[\nu_1 + (\nu_3 - 1) \gamma - \frac{(\nu_3 - 1)^2}{2}\right] = 0, \\ \nu_2 + a\left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2}\right) = 0, \\ \nu_2 + a\left[\nu_1 + (\nu_3 - 1) \gamma - \frac{(\nu_3 - 1)^2}{2}\right] = 0. \end{cases}$$

Since functions  $a$  and  $b$  are one-to-one, the difference between the first two equations (resp. the last two equations) imply:

$$\nu_1 + (\nu_3 - 1)\gamma - \frac{(\nu_3 - 1)^2}{2} = \nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2},$$

that is:

$$\nu_3 = \gamma + \frac{1}{2}.$$

From the same pairs of equations we deduce:

$$\begin{aligned} \nu_0 &= -b\left(\nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2}\right) = -\delta \log [1 + c(\nu_1 + \gamma^2/2 - 1/8)], \\ \nu_2 &= -a\left(\nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2}\right) = -\rho \frac{\nu_1 + \gamma^2/2 - 1/8}{1 + c(\nu_1 + \gamma^2/2 - 1/8)}. \end{aligned} \quad (25)$$

Therefore we get the following proposition.

**Proposition 4** : *The sdf is compatible with the no-arbitrage conditions if and only if:*

$$\begin{aligned} \nu_0 &= -\delta \log [1 + c(\nu_1 + \gamma^2/2 - 1/8)], \\ \nu_2 &= -\rho \frac{\nu_1 + \gamma^2/2 - 1/8}{1 + c(\nu_1 + \gamma^2/2 - 1/8)}, \\ \nu_3 &= \gamma + 1/2. \end{aligned}$$

In particular parameter  $\nu_1$  is unrestricted. Thus in this incomplete market framework (in which the liquid assets are the riskfree asset and the underlying risky asset) the risk premium for current stochastic volatility can be fixed arbitrarily, that is the dimension of residual market incompleteness is equal to 1. This residual incompleteness is not a consequence of the specific ARG dynamics assumed for stochastic volatility, but will be the general case for instance when state variables  $Y_t$  follow an affine process. Indeed in this case the specification of a parametric exponential affine sdf generally does not select a unique pricing kernel.

The (standardized) price at  $t$  of a European call with moneyness strike  $k$  and residual maturity  $h$  is given by:

$$\begin{aligned} c_t(h, k) &= \frac{1}{p_t} E_t [M_{t,t+1} \dots M_{t+h-1,t+h} (p_{t+h} - kp_t)^+] \\ &= E_t (M_{t,t+1} \dots M_{t+h-1,t+h} [\exp(r_{t+1} + \dots + r_{t+h}) - k]^+). \end{aligned}$$



As usual in the stochastic volatility framework, the option price can be written in terms of Black-Scholes price and integrated volatility

$\sigma_{t+1}^2(h) = (\sigma_{t+1}^2 + \dots + \sigma_{t+h}^2) / h$ . We get:

$$c_t(h, k) = E_t^Q [BS(h, k, \sigma_{t+1}^2(h))],$$

where  $Q$  is the risk neutral probability and  $BS(h, k, \sigma^2)$  denotes the Black-Scholes price of a European call with moneyness strike  $k$ , residual maturity  $h$  and constant volatility  $\sigma^2$ . The derivative price is easily computed by Monte-Carlo since under the risk neutral probability the returns still follow stochastic volatility model (20)-(21) with risk premium parameter  $\gamma^* = -1/2$  and ARG volatility parameters<sup>11</sup>:

$$\rho^* = \frac{\rho}{[1 + c(\nu_1 + \gamma^2/2 - 1/8)]^2}, \quad \delta^* = \delta, \quad c^* = \frac{c}{1 + c(\nu_1 + \gamma^2/2 - 1/8)}. \quad (26)$$

To illustrate the properties of the stochastic ARG volatility model discussed above, we display in Figures 1 and 2 below a joint simulated path for the return and the volatility, and the pattern of the implied Black-Scholes volatility as function of the moneyness strike, respectively. The simulations are performed for the following set of values for the parameters<sup>12</sup>:

$\gamma = 0.5$	$\rho = 0.85$	$\delta = 1.266$	$c = 2.7 \cdot 10^{-5}$
$\nu_0 = -3.37 \cdot 10^{-4}$	$\nu_1 = 10$	$\nu_2 = -8.498$	$\nu_3 = 1$

[Insert Figure 1: Simulated returns and volatility]

[Insert Figure 2: Implied Black-Scholes volatility]

As expected, the return series features volatility clustering, with periods of high return volatility corresponding to large values of the stochastic volatility. The Black-Scholes implied volatility<sup>13</sup> admits an asymmetric smile, more pronounced for in-the-money values of the moneyness strike  $k$ .

<sup>11</sup>The proof is available from the authors.

<sup>12</sup>Note that the risk premium interpretation does not necessarily imply that all  $\nu$  parameters are positive. They have been fixed to ensure that  $\nu_3 > 0$ ,  $\nu_1 + \nu_2 > 0$  and often  $\nu_0 + \nu_1\sigma_{t+1}^2 + \nu_2\sigma_t^2 + \nu_3r_{t+1} > 0$ .

<sup>13</sup>We select a residual maturity  $h = 1$  for the European call and the relevant information at date  $t$  is the volatility  $\sigma_t^2$ , whose value is set equal to the stationary mean  $E\sigma_t^2$ .

## 5.2 The observations and the model

In the next sections we assume that the observations are  $r_{t_0-T+1}, \dots, r_{t_0}$ ,  $\sigma_{t_0-T+1}^2, \dots, \sigma_{t_0}^2$  and some derivative prices at date  $t_0$ , corresponding to monyness strikes  $s_1 = 1$ ,  $s_2 = 0.98$ ,  $s_3 = 1.02$ . The observed prices have been generated by the design of Section 5.1 with the same set of parameter values. We are now interested in an efficient estimation of some option prices.

### i) The model

For simplicity we assume that the specified sdf is compatible with the design above:

$$M_{t,t+1}(\theta) = \exp(-\nu_0 - \nu_1 \sigma_{t+1}^2 - \nu_2 \sigma_t^2 - \nu_3 r_{t+1}),$$

where  $\theta = (\nu_0, \nu_1, \nu_2, \nu_3)'$  is now an unknown parameter. Moreover the conditional distribution of  $Y_t = (r_t, \sigma_t^2)$  given  $Y_{t-1}$  is left unspecified.

### ii) Full information identifiability

Let us now discuss the identifiability of parameter  $\theta$  from the uniform conditional restrictions:

$$\begin{cases} E_t(M_{t,t+1}(\theta)) = 1, \\ E_t(M_{t,t+1}(\theta) \exp r_{t+1}) = 1, \end{cases}$$

assumed valid for any conditioning value  $y_t$ . From Proposition 4, only three independent linear combinations of parameter  $\theta$  can be identified, including parameter  $\nu_3$ . Therefore in this model the structural parameter  $\theta$  is full information underidentified.

At this step two approaches can be followed:

- i. We can consider the stochastic discount factor above without introducing additional restrictions on parameters  $\nu_0, \nu_1, \nu_2$ . Then the degree of underidentification from asset dynamics is equal to 1. The null space  $N_0$  defined in equation (12) has dimension 1, and is spanned by [see Appendix 3]:

$$R = \begin{pmatrix} -\delta \frac{c}{1+c\nu_1} \\ 1 \\ -\rho \frac{1}{(1+c\nu_1)^2} \\ 0 \end{pmatrix} = \begin{pmatrix} -3.37 \cdot 10^{-5} \\ 1 \\ -0.850 \\ 0 \end{pmatrix}. \quad (27)$$

Component  $R'\theta$  involving parameters  $\nu_0, \nu_1, \nu_2$  is full information non-identified. Typically parameters  $\nu_0, \nu_1, \nu_2$  can only be identified by the cross-sectional restrictions from observed derivative prices.

- ii. Alternatively, we can introduce an identification restriction on the risk premium, for instance  $\nu_1 = 0$ . Under this restriction  $\nu_0, \nu_1, \nu_2$  become full information identifiable from asset price dynamics, and the estimation problem will be greatly simplified. It is interesting to note that this second approach is often followed in the financial literature, at the risk of a misspecification in the identification restriction. Moreover such an approach can have some misleading consequences when considering the confidence interval for derivative prices, which will likely be too narrow.

The first approach will be considered in this paper.

### iii) Limited information identifiability and residual maturity

An additional identification problem may arise in the limited information framework since the information depends on the maturity of the observed derivative prices. For instance, if all observed derivative prices correspond to a short maturity  $h = 1$ , only parameters  $\nu_{0,t_0} = \nu_0 + \nu_2\sigma_{t_0}^2$  and  $\nu_1$  can be identified from asset dynamics and observed derivative prices. This allows to identify the prices of derivatives with the same maturity 1, but not the prices of derivatives with larger maturity. However parameters  $\nu_0$  and  $\nu_2$  can be identified separately by means of observed prices of derivatives with residual maturity larger than 1. To summarize, when the structural parameter is limited information underidentified it can be necessary to use derivative prices with different maturities to be able to estimate the derivative prices for all maturities. Note however that the derivative prices with large residual maturity are not very informative. Indeed, let us consider the geometric stochastic yield associated with the sdf, that is:

$$-\frac{1}{h} \log(M_{t,t+1} \dots M_{t+h+1,t+h}) = \nu_0 + \nu_1 \frac{1}{h} \sum_{k=1}^h \sigma_{t+k}^2 + \nu_2 \frac{1}{h} \sum_{k=0}^{h-1} \sigma_{t+k}^2 + \nu_3 \frac{1}{h} \sum_{k=1}^h r_{t+k}.$$

If the joint process  $(\sigma_t^2, r_t)$  is stationary, the geometric stochastic yield tends to the deterministic long run level  $\nu_0 + (\nu_1 + \nu_2) E\sigma_t^2 + \nu_3 E r_t$  for  $h$  tending to infinity. Thus this combination of the structural parameters is limited

information identifiable from long run derivative prices, but the structural parameters themselves are not.

#### iv) Link with the literature

Stock, Wright (2000) considered also an application to asset pricing formula, in which the sdf is deduced from the optimisation of an expected CRRA utility function. Thus the sdf is:

$$M_{t,t+1}(\theta) = \delta (C_{t+1}/C_t)^\gamma,$$

where  $C_t$  denotes the consumption<sup>14</sup>. In our framework both  $\delta$  and  $\gamma$  parameters would be full information identifiable from the observed asset prices of the basic assets. Thus the discussion differs from the discussion in Stock, Wright (2000) in which the risk aversion parameter is assumed a priori weakly identified. In our framework the weak identification can only be the consequence of some lack of observations on derivative prices and concern some special risk premium parameters.

### 5.3 Nonparametric efficiency bounds

Two cases will be distinguished according to the type of information.

#### i) Limited information

The cross-sectional restrictions are:

$$\begin{aligned} E [M_{t_0,t_0+1}(\theta) - 1 | y_{t_0}] &= 0, \\ E [M_{t_0,t_0+1}(\theta) \exp r_{t_0+1} - 1 | y_{t_0}] &= 0, \\ E [M_{t_0,t_0+1}(\theta) (\exp r_{t_0+1} - s)^+ - c_{t_0}(s) | y_{t_0}] &= 0, \quad s = 0.98, 1, 1.02. \end{aligned}$$

The conditional moments of interest are the prices of European calls at horizon 1:

$$E(a(s) | y_{t_0}) = E [M_{t_0,t_0+1}(\theta) (\exp r_{t_0+1} - s)^+ | y_{t_0}], \quad \forall s.$$

---

<sup>14</sup>In this framework the price of the consumption good is  $q_t = 1$  and the returns  $y_t$  have to be interpreted as real returns.

The identifiable parameters are  $\nu_{0,t_0} = \nu_0 + \nu_2\sigma_{t_0}^2$ ,  $\nu_1$ ,  $\nu_3$ , but are sufficient to identify the conditional moments of interest, which have the same maturity. We provide in Figure 3 the nonparametric efficiency bound  $\mathcal{B}(y_{t_0}, s)$  for  $E(a(s)|y_{t_0})$  as a function of  $s$ , computed according to Corollary 2.

[Insert Figure 3: Nonparametric efficiency bound, limited information]

The solid line corresponds to the call price  $E(a(s)|y_{t_0})$ , the dashed lines to confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ , computed for  $w^2/Th_T^2 f_X(x_0) = 1$ <sup>15</sup>. The current factor  $y_{t_0}$  corresponds to a variance  $\sigma_{t_0}^2$  equal to the stationary expectation  $E\sigma_t^2$ . The width of the confidence interval for derivative price  $E(a(s)|y_{t_0})$  depends on moneyness strike  $s$ . The interval is generally narrower for almost at-the-money (ATM) options, whereas it is wider when the derivative is deep in-the-money (ITM) or out-of-the-money (OTM). Moreover the width of the interval is zero when  $s$  corresponds to the moneyness strikes of the observed calls.

To compare the results for derivatives with longer maturity, let us consider the nonparametric efficiency bound for a European call with residual maturity  $h = 60$  days. At  $t_0$  the prices of three derivatives with same residual maturity  $h = 60$  and strikes  $s = 0.9, 1, 1.1$ , respectively, are assumed to be observed. In this case the whole parameter  $\theta$  is limited information identifiable. The efficiency bound is displayed in Figure 4 below.

[Insert Figure 4: Nonparam. eff. bound, limited information, maturity 60]

The confidence interval is larger for ITM strikes, and generally larger compared to maturity  $h = 1$ .

Note that the confidence intervals are pointwise confidence intervals. The choice of derivative prices corresponding to different strikes have in practice to be compatible with both a confidence band and also with the no arbitrage restrictions. This implies the selection of a decreasing convex function compatible with the band.

## ii) Mixed limited- and full-information

---

<sup>15</sup>We adopt this normalization to illustrate the pattern of the nonparametric efficiency bound as a function of the moneyness strike. The selection of empirically relevant sample size  $T$  and bandwidth  $h_T$  will be discussed in the next section, where we report the corresponding actual size of the efficiency bound.

Let us now consider the general approach with both uniform and local restrictions, see Proposition 2. The conditional moments of interest are still:

$$E(a(s)|y_{t_0}) = E[M_{t_0,t_0+1}(\theta)(\exp r_{t_0+1} - s)^+ | y_{t_0}], \quad \forall s.$$

We check in Appendix 3 that Assumption A.2\* is satisfied in our stochastic volatility framework. Matrix  $R$  is given in equation (27). The nonparametric efficiency bound for a European call with residual maturity  $h = 1$  is displayed in Figure 5.

[Insert Figure 5: Nonparametric efficiency bound, mixed information]

The confidence interval is similar to that obtained in the limited information framework (see Figure 3). To understand this fact, let us recall that, in the mixed case and at maturity 1, we have to estimate one parameter more than in the limited information case. This effect is approximately compensated by the contribution of the uniform restrictions.

In Figure 6 we display the mixed information nonparametric efficiency bound for a call option with residual maturity  $h = 60$ , when the price of three derivatives with same maturity and strikes  $s = 0.9, 1, 1.1$ , respectively, are observed.

[Insert Figure 6: Nonparam. eff. bound, mixed information, maturity 60]

The confidence band is more narrow than in the limited information case (see Figure 4). This is especially clear for ITM values of  $s$  far from the observed strikes. The effect on the confidence interval is less clear close to observed strikes, or for rather deep OTM strikes. Indeed in these regions the option price is tight down from the observed prices, or has to be close to zero by definition, respectively.

Finally, in Figure 7 we display the mixed information nonparametric efficiency bound for a European call with maturity  $h = 60$ , when the prices of three derivatives with maturity 20 and strikes  $s = 0.9, 1, 1.1$ , respectively, are observed.

[Insert Figure 7: Nonparam. eff. bound, mixed information, maturity 60 and 20]

In this case the maturity of the observed derivatives does not correspond with the maturity of interest. This explains why the nonparametric efficiency bound is much larger compared with Figure 6, and in particular it is different from zero for all moneyness strikes. Thus observed derivative prices at the maturity of interest have a large informational content for the estimation of other derivative prices.

## 5.4 Monte-Carlo

In this section we report the results of Monte-Carlo simulations to investigate the finite sample properties of the information based estimator. Data are generated according to the ARG stochastic volatility model described in Section 5.1. We consider the general framework with both local and uniform restrictions.

At date  $t_0$  the prices of three derivatives  $c_{t_0}(h, s_k)$  with maturity  $h = 2$  and moneyness strikes  $s_k = 0.98, 1, 1.02$  are fixed. They are computed by simulation as explained in Section 5.1 with available information  $\sigma_{t_0}^2 = E(\sigma_t^2)$ . Then we simulate  $S = 500$  paths of return-volatility process  $(r_t, \sigma_t^2)$ ,  $t = t_0 - T + 1, \dots, t_0$ , for sample size  $T = 250$ , such that the observed values at date  $t_0$  are  $r_{t_0} = 0$ ,  $\sigma_{t_0}^2 = E(\sigma_t^2)$ . Such paths are obtained by simulating the process backward. More precisely the time-discretized version of the Cox-Ingersoll-Ross process is time reversible. Therefore  $(r_t, \sigma_t^2)$  follows the same stochastic volatility process both in direct and in reversed time. The information based estimator of structural parameter  $\theta_0$  and of European call prices  $E[a(h, s)|y_{t_0}]$  at date  $t_0$  for maturity  $h = 2$  and different strikes  $s$  are computed for each simulated sample according to Section 3.2. The moment restrictions involve both the uniform no-arbitrage conditions from riskfree and underlying asset returns as well as the local restrictions from observed derivative prices at date  $t_0$ . The kernel estimator of the conditional pdf is based on a Gaussian product kernel with different bandwidths for return and volatility, which are equal to  $h_{r,T} = 0.0039$  and  $h_{\sigma,T} = 0.0025$ , respectively<sup>16</sup>. Finally the selected sample size  $T = 250$  corresponds to approximately 1 year of trading days, which is the sample length typically suggested by the regulator for risk management purposes.

### i) Derivative prices

We display in Table 1 below the mean, the median, the 95% confidence interval, as well as the 5% and 95% quantiles of the estimated European call prices for maturity  $h = 2$  and different values of the strike between  $s = 0.95$  and  $s = 1.05$ .

[Table 1: Derivative prices, maturity 2, sample size 250]

---

<sup>16</sup>These bandwidths are selected in order to get an appropriate smoothing of the joint conditional pdf of  $(r_t, \sigma_t)$  at sample size  $T = 250$ .

For comparison we also report for each moneyness strike the corresponding true derivative price and the 95% asymptotic confidence interval based on the nonparametric efficiency bound, computed according to Section 5.3. As seen previously in Figures 5 and 6, the width of the nonparametric bound varies with strike  $s$ . In particular, the information content of the moment restrictions for estimating derivative prices can be very different across strikes. For instance the width of the confidence interval is less than 1% of the true price for strike  $s = 0.96$ , whereas it amounts to about 35% for strike  $s = 1.04$ .

Let us first consider the finite sample bias of estimated derivative prices. This bias is typically positive for ITM call options, whereas prices of OTM calls are generally underestimated. The corresponding relative pricing errors are less than 0.5% for strikes below  $s = 1.03$ , whereas it is about 5% for the deep OTM strikes.

Let us now consider the finite sample accuracy of the estimated derivative prices. The 95% confidence intervals and the 5% interquantile ranges of estimated call prices feature patterns across strikes similar to the patterns of the nonparametric efficiency bound, but they are wider. For instance, the 95% interquantile range is about 4% of the median (or mean) call price for strike  $s = 0.96$ , and about 135% for the deep OTM strike  $s = 1.04$ <sup>17</sup>. In particular, these bounds are much larger than those typically reported in the literature based on fully parametric specifications. A narrow parametric bound however can be highly misleading in the presence of model misspecifications, that is when the true data generating process of underlying asset returns does not belong to the selected parametric family. In practice, the nonparametric bounds derived from the finite sample distribution of the information based estimator are likely to be preferred, since they provide more secure bounds for risk management purposes.

Finally, we display in Figure 8 the histograms of estimated derivative prices for different strikes.

[Figure 8: Histograms of estimated derivative prices]

These finite sample distributions feature non-Gaussian patterns, especially for deep ITM or deep OTM strikes, for which left, respectively right, skewed and fat tailed distributions are observed.

## ii) Structural parameter

---

<sup>17</sup>For strike  $s = 1.04$  the interquantile range is highly skewed, with the median very close to the lower bound.



Although the focus of this paper is on estimation of the conditional moments corresponding to derivative prices, it is interesting to consider also the results for the estimator of structural parameter  $\theta$ . In Table 2 we display the mean, the median, the standard deviation and the 5% interquantile range of estimator  $\hat{\theta}$ .

[Table 2: Structural parameter, maturity 2, sample size 250]

The estimator of each component is downward biased for sample size  $T = 250$ , especially for parameters  $\nu_0, \nu_1, \nu_2$ . The medians are also below, but closer to, the true values  $\theta_0$ . The estimators feature large standard deviation and wide interquantile ranges. These results are confirmed by the histograms of the estimates, which are displayed in Figure 9.

[Figure 9: Histograms of estimated structural parameters]

The finite sample distributions of parameters  $\nu_0, \nu_1, \nu_2$  are highly non-normal, in particular skewed to the left, with fat tails. The distribution of parameter  $\nu_3$ , instead, is closer to a Gaussian distribution (even if it is not very accurate). This difference in patterns of the finite sample distributions reflects the different rates of convergence of the estimators, which are the parametric rate  $T^{1/2}$  for  $\nu_3$  and the nonparametric rate  $(Th_{r,T}h_{\sigma,T})^{1/2}$  for  $\nu_0, \nu_1, \nu_2$ . These different rates of convergence are a consequence of market incompleteness, which cause parameters  $\nu_0, \nu_1, \nu_2$  related to the risk premium for stochastic volatility to be full information non-identifiable.

## 6 Concluding remarks

The aim of this paper was to explain why the standard GMM approach is not appropriate for derivative pricing in an incomplete market framework, even if the stochastic discount factor is specified parametrically. In this framework it is necessary to distinguish the moment conditions which are uniform with respect to the conditioning variable and the restrictions which are valid for a given value of this variable. The extended method of moments explains how to mix appropriately these types of moment conditions. This implies different rates of convergence for the different parameters. In the derivative pricing application, the risk premium parameters are not necessarily identified from historical data on the prices of liquid assets, and some of them can only be

deduced from less frequent cross-sectional observations on derivative prices. This implies different rates of convergence for the risk premia parameter, and approximations of derivative prices much less accurate than usually reported in both the theoretical or applied literature.

The analysis emphasizes on the notions of limited and full information identifiability, and the relationship between the degree of full information underidentifiability and the dimension of residual incompleteness in the application to derivative pricing. It differs from the standard GMM literature, which always assume full information identifiability.

## APPENDIX 1

### Semi-parametric efficiency bound

We derive below the expression of the semi-parametric efficiency bound by considering the linearization of the nonlinear constraints around the true value. The moment conditions defining the (kernel) moment estimator can be linearized around  $\theta = \theta_0$ :

$$\begin{aligned}\sqrt{T}\widehat{E}[g_1(Y, X; \theta)] &\simeq \sqrt{T}\widehat{E}[g_1(Y, X; \theta_0)] + E_0 \left[ \frac{\partial g_1}{\partial \theta'}(Y, X; \theta_0) \right] \sqrt{T}(\theta - \theta_0), \\ \sqrt{Th_T^d}\widetilde{E}[g_2(Y; \theta)|x_0] &\simeq \sqrt{Th_T^d}\widetilde{E}[g_2(Y; \theta_0)|x_0] + \sqrt{Th_T^d}E_0 \left[ \frac{\partial g_2}{\partial \theta'}(Y; \theta_0)|x_0 \right] (\theta - \theta_0), \\ \sqrt{Th_T^d}\widetilde{E}[a(Y; \theta) - \beta|x_0] &\simeq \sqrt{Th_T^d}\widetilde{E}[a(Y; \theta_0) - \beta_0|x_0] + \sqrt{Th_T^d}E_0 \left[ \frac{\partial a}{\partial \theta'}(Y; \theta_0)|x_0 \right] (\theta - \theta_0) \\ &\quad - \sqrt{Th_T^d}(\beta - \beta_0).\end{aligned}$$

Let us introduce the change of parameters:

$$\theta = \begin{pmatrix} \widetilde{R} & R_Z \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}.$$

Then we get:

$$\begin{aligned}\sqrt{T}\widehat{E}[g_1(\theta)] &\simeq \sqrt{T}\widehat{E}(g_1) + E_0 \left( \frac{\partial g_1}{\partial \theta'} \right) \widetilde{R}\sqrt{T}(\eta_1 - \eta_1^0), \\ \sqrt{Th_T^d}\widetilde{E}[g_2(\theta)|x_0] &\simeq \sqrt{Th_T^d}\widetilde{E}(g_2|x_0) + \sqrt{h_T^d}E_0 \left( \frac{\partial g_2}{\partial \theta'}|x_0 \right) \widetilde{R}\sqrt{T}(\eta_1 - \eta_1^0) \\ &\quad + E_0 \left( \frac{\partial g_2}{\partial \theta'}|x_0 \right) R_Z\sqrt{Th_T^d}(\eta_2 - \eta_2^0), \\ \sqrt{Th_T^d}\widetilde{E}[a(\theta) - \beta|x_0] &\simeq \sqrt{Th_T^d}\widetilde{E}(a - \beta_0|x_0) + \sqrt{h_T^d}E_0 \left( \frac{\partial a}{\partial \theta'}|x_0 \right) \widetilde{R}\sqrt{T}(\eta_1 - \eta_1^0) \\ &\quad + E_0 \left( \frac{\partial a}{\partial \theta'}|x_0 \right) R_Z\sqrt{Th_T^d}(\eta_2 - \eta_2^0) - \sqrt{Th_T^d}(\beta - \beta_0).\end{aligned}$$

It is known that the moment estimator corresponding to the optimal weighting matrix is asymptotically equivalent to the GLS estimator of the linear

system:

$$\begin{pmatrix} \sqrt{T}\widehat{E}(g_1) \\ \sqrt{Th_T^d}\widehat{E}(g_2|x_0) \\ \sqrt{Th_T^d}\widehat{E}(a - \beta_0|x_0) \end{pmatrix} \simeq - \begin{pmatrix} E_0\left(\frac{\partial g_1}{\partial \theta'}\right)\widetilde{R} & 0 & 0 \\ 0 & E_0\left(\frac{\partial g_2}{\partial \theta'}|x_0\right)R_Z & 0 \\ 0 & E_0\left(\frac{\partial a}{\partial \theta'}|x_0\right)R_Z & -Id_L \end{pmatrix} \begin{pmatrix} \sqrt{T}(\eta_1 - \eta_1^0) \\ \sqrt{Th_T^d}(\eta_2 - \eta_2^0) \\ \sqrt{Th_T^d}(\beta - \beta_0) \end{pmatrix} + U,$$

where the error term  $U$  is zero-mean, with variance:

$$V(U) = \begin{pmatrix} \bar{V}_1 & 0 \\ 0 & \Sigma \end{pmatrix},$$

where:

$$\bar{V}_1 = \lim_{T \rightarrow \infty} V_0 \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T g_1(Y_t, X_t; \theta_0) \right], \quad \Sigma = \frac{1}{f_X(x_0)} \begin{pmatrix} V_0(g_2|x_0) & Cov_0(g_2, a|x_0) \\ Cov_0(a, g_2|x_0) & V_0(a|x_0) \end{pmatrix},$$

and the asymptotic correlation between the sample moment and the kernel estimator is zero. Thus estimators  $\sqrt{T}(\widehat{\eta}_1 - \eta_1^0)$  and  $\sqrt{Th_T^d}(\widehat{\eta}_2 - \eta_2^0, \widehat{\beta} - \beta_0)$  are asymptotically normal, independent with asymptotic variances:

$$V_{as} \left[ \sqrt{T}(\widehat{\eta}_1 - \eta_1^0) \right] = \left[ \widetilde{R}' E_0 \left( \frac{\partial g_1'}{\partial \theta} \right) \bar{V}_1^{-1} E_0 \left( \frac{\partial g_1}{\partial \theta'} \right) \widetilde{R} \right]^{-1},$$

and:

$$V_{as} \left[ \sqrt{Th_T^d} \begin{pmatrix} \widehat{\eta}_2 - \eta_2^0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} \right] = \left[ \begin{pmatrix} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R_Z & 0 \\ E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R_Z & -Id \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R_Z & 0 \\ E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R_Z & -Id \end{pmatrix} \right]^{-1},$$

respectively. In particular the semi-parametric efficiency bound is the lower  $(L, L)$  block of matrix  $V_{as} \left[ \sqrt{Th_T^d}(\widehat{\eta}_2 - \eta_2^0, \widehat{\beta} - \beta_0) \right]$ .

## APPENDIX 2

### Information based estimator

The aim of this Appendix is to derive the asymptotic expansion of the objective function and of the estimators, in order to prove the asymptotic nonparametric efficiency of the information based estimator (Proposition 3).

#### i) Concentration with respect to functional parameter

Let us introduce Lagrange multipliers  $\lambda_0, \mu_0, \lambda_t, \mu_t, t = 1, \dots, T$ . The Lagrangian function is given by:

$$\begin{aligned} \mathcal{L} = & \frac{1}{T} \sum_{t=1}^T \int \frac{[\widehat{f}(y|x_t) - f^t(y)]^2}{\widehat{f}(y|x_t)} dy + h_T^d \int \log [f^0(y)/\widehat{f}(y|x_0)] f^0(y) dy \\ & - 2 \sum_{t=1}^T \mu_t \left( \int f^t(y) dy - 1 \right) - h_T^d \mu_0 \left( \int f^0(y) dy - 1 \right) \\ & - 2 \sum_{t=1}^T \lambda'_t \int g(y; \theta) f^t(y) dy - h_T^d \lambda'_0 \int g_2(y; \theta) f^0(y) dy. \end{aligned}$$

The first order conditions w.r.t. functional parameters  $f_t, t = 1, \dots, T, f_0$  are:

$$\begin{aligned} [f^t(y) - \widehat{f}(y|x_t)] \frac{1}{\widehat{f}(y|x_t)} - \mu_t - \lambda'_t g(y; \theta) &= 0, \quad t = 1, \dots, T, \\ 1 + \log \left( f^0(y)/\widehat{f}(y|x_0) \right) - \mu_0 - \lambda'_0 g_2(y; \theta) &= 0, \end{aligned}$$

that are:

$$f^t(y) = \widehat{f}(y|x_t) + \mu_t \widehat{f}(y|x_t) + \lambda'_t g(y; \theta) \widehat{f}(y|x_t), \quad t = 1, \dots, T, \quad (\text{A.1})$$

$$f^0(y) = \widehat{f}(y|x_0) \exp \left( \lambda'_0 g_2(y; \theta) + \mu_0 - 1 \right). \quad (\text{A.2})$$

The Lagrange multipliers are deduced by the constraints. From (A.1) we get:

$$\int f^t(y) dy = 1 \iff \mu_t = -\lambda'_t \int g(y; \theta) \widehat{f}(y|x_t) dy,$$

and:

$$\begin{aligned}
& \int g(y; \theta) f^t(y) dy = 0 \\
\iff & \int g(y; \theta) \widehat{f}(y|x_t) dy + \int g(y; \theta) g(y; \theta)' \widehat{f}(y|x_t) dy \cdot \lambda_t + \mu_t \int g(y; \theta) \widehat{f}(y|x_t) dy = 0 \\
\iff & \lambda_t = - \left[ \int g(y; \theta) g(y; \theta)' \widehat{f}(y|x_t) dy - \int g(y; \theta) \widehat{f}(y|x_t) dy \int g(y; \theta)' \widehat{f}(y|x_t) dy \right]^{-1} \\
& \cdot \int g(y; \theta) \widehat{f}(y|x_t) dy, \quad t = 1, \dots, T.
\end{aligned}$$

Similarly from (A.2) we deduce the value of Lagrange multiplier  $\mu_0$ :

$$\int f^0(y) dy = 1 \iff \exp(1 - \mu_0) = \int e^{\lambda_0' g_2(y; \theta)} \widehat{f}(y|x_0) dy.$$

Thus from (A.1), (A.2)  $\mu_0, \lambda_t, \mu_t, t = 1, \dots, T$ , can be eliminated to get the concentrated functional parameters:

$$\begin{aligned}
f^t(y; \theta) &= \widehat{f}(y|x_t) - \widetilde{E}(g(\theta)|x_t)' \widetilde{V}(g(\theta)|x_t)^{-1} \left[ g(y; \theta) - \widetilde{E}(g(\theta)|x_t) \right] \widehat{f}(y|x_t), \\
t &= 1, \dots, T, \\
f^0(y; \theta, \lambda_0) &= \frac{\exp \lambda_0' g_2(y; \theta)}{\widetilde{E}(\exp \lambda_0' g_2(\theta)|x_0)} \widehat{f}(y|x_0), \tag{A.3}
\end{aligned}$$

where  $\widetilde{E}(\cdot|x)$  and  $\widetilde{V}(\cdot|x)$  denote the conditional expectation and the conditional variance w.r.t. the kernel density, respectively. The concentrated objective function becomes:

$$\begin{aligned}
\mathcal{L}_c(\theta, \lambda_0) &= \frac{1}{T} \sum_{t=1}^T \widetilde{E}(g(\theta)|x_t)' \widetilde{V}(g(\theta)|x_t)^{-1} \widetilde{E}(g(\theta)|x_t) \\
&\quad - h_T^d \log \widetilde{E}(\exp \lambda_0' g_2(\theta)|x_0).
\end{aligned}$$

Then the information based estimator is such that  $\widehat{\theta}$  is solution of the saddle point problem [see Kitamura-Stutzer (1997) in a marginal framework]:

$$\widehat{\theta} = \arg \min_{\theta} \mathcal{L}_c(\theta, \lambda_0(\theta)),$$

where:

$$\lambda_0(\theta) = \arg \max_{\lambda_0} \mathcal{L}_c(\theta, \lambda_0) \iff \tilde{E} \left( g_2(\theta) \exp \lambda_0(\theta)' g_2(\theta) | x_0 \right) = 0,$$

and the conditional density estimators are:

$$\begin{aligned} \hat{f}_0(\cdot | x_t) &= f^t(\cdot; \hat{\theta}), \quad t = 1, \dots, T, \\ \hat{f}_0(\cdot | x_0) &= f^0(\cdot; \hat{\theta}, \hat{\lambda}_0), \quad \hat{\lambda}_0 = \lambda_0(\hat{\theta}). \end{aligned}$$

## ii) Asymptotic expansion of the concentrated objective function

Since the conditional moment restrictions are satisfied asymptotically, we have  $\hat{\lambda}_0 \rightarrow 0$  when  $T \rightarrow \infty$ . Therefore we can consider the second order asymptotic expansion of function  $\mathcal{L}_c(\theta, \lambda_0)$  in a neighbourhood of  $\theta = \theta_0$ ,  $\lambda_0 = 0$ . Let us first consider the expansion w.r.t.  $\lambda_0$ . We have:

$$\begin{aligned} &\log \tilde{E} \left( \exp \lambda_0' g_2(\theta) | x_0 \right) \\ &\simeq \log \left[ 1 + \lambda_0' \tilde{E} (g_2(\theta) | x_0) + \frac{1}{2} \lambda_0' \tilde{E} \left( g_2(\theta) g_2(\theta)' | x_0 \right) \lambda_0 \right] \\ &\simeq \lambda_0' \tilde{E} (g_2(\theta) | x_0) + \frac{1}{2} \lambda_0' \tilde{V} (g_2(\theta) | x_0) \lambda_0. \end{aligned}$$

Therefore we can asymptotically concentrate w.r.t.  $\lambda_0$ :

$$\lambda_0 \simeq -\tilde{V} (g_2(\theta) | x_0)^{-1} \tilde{E} (g_2(\theta) | x_0), \quad (\text{A.4})$$

and the asymptotic expansion of the concentrated objective function becomes:

$$\begin{aligned} \mathcal{L}_c(\theta) &\simeq \frac{1}{T} \sum_{t=1}^T \tilde{E} (g(\theta) | x_t)' \tilde{V} (g(\theta) | x_t)^{-1} \tilde{E} (g(\theta) | x_t) \\ &\quad + \frac{1}{2} h_T^d \tilde{E} (g_2(\theta) | x_0)' \tilde{V} (g_2(\theta) | x_0)^{-1} \tilde{E} (g_2(\theta) | x_0). \end{aligned}$$

Let us now consider the expansion around  $\theta = \theta_0$ . We have:

$$\tilde{E} (g(\theta) | x_t) \simeq \tilde{E} (g(\theta_0) | x_t) + E_0 \left( \frac{\partial g}{\partial \theta} (\theta_0) | x_t \right) (\theta - \theta_0),$$

$$\tilde{V}(g(\theta)|x_t) \simeq V_0(g(\theta_0) | x_t),$$

and similarly for the expectations of function  $g_2$ . Thus we get:

$$\begin{aligned} \mathcal{L}_c(\theta) &\simeq \frac{1}{T} \sum_{t=1}^T \left\{ \tilde{E}(g|x_t) + E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) (\theta - \theta_0) \right\}' V_0(g | x_t)^{-1} \\ &\quad \cdot \left\{ \tilde{E}(g|x_t) + E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) (\theta - \theta_0) \right\} \\ &\quad + \frac{1}{2} h_T^d \left\{ \tilde{E}(g_2|x_0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) (\theta - \theta_0) \right\}' V_0(g_2 | x_0)^{-1} \\ &\quad \cdot \left\{ \tilde{E}(g_2|x_0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) (\theta - \theta_0) \right\}, \end{aligned}$$

where functions  $g, g_2$  are evaluated at  $\theta_0$ .

### iii) Asymptotic expansion of $\hat{\theta}$

In order to derive the asymptotic expansion of  $\hat{\theta}$ , we have to carefully distinguish the directions of  $\theta$  converging at a parametric rate [respectively at a nonparametric rate]. Let us introduce the change of parameter:

$$\eta = R_1^{-1} \theta = \left( \eta'_1, \eta'_2 \right)',$$

where  $R_1 = \begin{pmatrix} \tilde{R} & R \end{pmatrix}$  and  $R$  is a matrix whose columns span the null space  $N_0$  [see Section 2.1.2]. Then we have:

$$E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) (\theta - \theta_0) = E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) \tilde{R} (\eta_1 - \eta_1^0).$$

We get:

$$\begin{aligned} &\mathcal{L}_c(\eta) \\ &\simeq \frac{1}{T} \sum_{t=1}^T \left\{ \tilde{E}(g|x_t) + E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) \tilde{R} (\eta_1 - \eta_1^0) \right\}' V_0(g|x_t)^{-1} \\ &\quad \cdot \left\{ \tilde{E}(g|x_t) + E_0 \left( \frac{\partial g}{\partial \theta'} \mid x_t \right) \tilde{R} (\eta_1 - \eta_1^0) \right\} \\ &\quad + \frac{1}{2} h_T^d \left\{ \tilde{E}(g_2|x_0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) \tilde{R} (\eta_1 - \eta_1^0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) R (\eta_2 - \eta_2^0) \right\}' V_0(g_2|x_0)^{-1} \\ &\quad \cdot \left\{ \tilde{E}(g_2|x_0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) \tilde{R} (\eta_1 - \eta_1^0) + E_0 \left( \frac{\partial g_2}{\partial \theta'} \mid x_0 \right) R (\eta_2 - \eta_2^0) \right\}. \end{aligned}$$



The asymptotic expansion of  $\widehat{\eta}_1$  can be obtained from the maximization of the first term in  $\mathcal{L}_c(\eta)$  since the contribution of the second term is asymptotically negligible. We get:

$$\begin{aligned} \sqrt{T} (\widehat{\eta}_1 - \eta_1^0) &\simeq - \left[ \frac{1}{T} \sum_{t=1}^T \widetilde{R}' E_0 \left( \frac{\partial g'}{\partial \theta} |x_t \right) V_0 (g|x_t)^{-1} E_0 \left( \frac{\partial g}{\partial \theta'} |x_t \right) \widetilde{R} \right]^{-1} \\ &\quad \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^T \widetilde{R}' E_0 \left( \frac{\partial g'}{\partial \theta} |x_t \right) V_0 (g|x_t)^{-1} \int g(y; \theta_0) \widehat{f}(y|x_t) dy. \end{aligned}$$

Thus  $\widehat{\eta}_1$  converges at a parametric rate.

The asymptotic expansion of  $\widehat{\eta}_2$  can be deduced by the maximization of the second component of  $\mathcal{L}_c(\eta)$ . Estimator  $\widehat{\eta}_2$  converges at a nonparametric rate and thus terms involving  $(\widehat{\eta}_1 - \eta_1^0)$  can be neglected. We get:

$$\begin{aligned} (\widehat{\eta}_2 - \eta_2^0) &\simeq - \left[ R' E_0 \left( \frac{\partial g_2'}{\partial \theta} |x_0 \right) V_0 (g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} |x_0 \right) R \right]^{-1} \\ &\quad \cdot R' E_0 \left( \frac{\partial g_2'}{\partial \theta} |x_0 \right) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy. \end{aligned} \tag{A.5}$$

#### iv) Asymptotic expansion of $\widehat{f}_0(\cdot|x_0)$

Let us consider the expansion of  $f^0(y; \theta, \lambda_0)$  in (A.3) around  $\lambda_0 = 0$ . We have:

$$\begin{aligned} f^0(y; \theta, \lambda_0) &\simeq \frac{1 + \lambda_0' g_2(y; \theta)}{1 + \lambda_0' \widetilde{E}(g_2(\theta)|x_0)} \widehat{f}(y|x_0) \\ &\simeq \left[ 1 + \lambda_0' \left( g_2(y; \theta) - \widetilde{E}(g_2(\theta)|x_0) \right) \right] \widehat{f}(y|x_0) \\ &\simeq \widehat{f}(y|x_0) \\ &\quad - \widetilde{E}(g_2(\theta)|x_0)' \widetilde{V}(g_2(\theta)|x_0)^{-1} \left( g_2(y; \theta) - \widetilde{E}(g_2(\theta)|x_0) \right) \widehat{f}(y|x_0), \end{aligned}$$

from (A.4). Thus we get:

$$\begin{aligned}
& \widehat{f}_0(y|x_0) \\
&= f_0(y; \widehat{\theta}, \widehat{\lambda}_0) \\
&\simeq \widehat{f}(y|x_0) - \widetilde{E} \left( g_2(\widehat{\theta})|x_0 \right)' \widetilde{V} \left( g_2(\widehat{\theta})|x_0 \right)^{-1} \left( g_2(y; \widehat{\theta}) - \widetilde{E} \left( g_2(\widehat{\theta})|x_0 \right) \right) \widehat{f}(y|x_0) \\
&\simeq \widehat{f}(y|x_0) - \widetilde{E} \left( g_2(\widehat{\theta})|x_0 \right)' V_0(g_2|x_0)^{-1} g_2(y; \theta_0) f(y|x_0). \tag{A.6}
\end{aligned}$$

Moreover:

$$\begin{aligned}
\widetilde{E} \left( g_2(\widehat{\theta})|x_0 \right) &\simeq \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy + E_0 \left( \frac{\partial g_2}{\partial \theta'} |x_0 \right) (\widehat{\theta} - \theta_0) \\
&\simeq \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy + E_0 \left( \frac{\partial g_2}{\partial \theta'} |x_0 \right) R (\widehat{\eta}_2 - \eta_2^0) \\
&\quad \text{(since the contribution of } \widehat{\eta}_1 - \eta_1^0 \text{ is asymptotically negligible)} \\
&= (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy,
\end{aligned}$$

from (A.5), where:

$$\begin{aligned}
M &= E_0 \left( \frac{\partial g_2}{\partial \theta'} |x_0 \right) R \left[ R' E_0 \left( \frac{\partial g_2'}{\partial \theta} |x_0 \right) V_0(g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} |x_0 \right) R \right]^{-1} \\
&\quad \cdot R' E_0 \left( \frac{\partial g_2'}{\partial \theta} |x_0 \right) V_0(g_2|x_0)^{-1},
\end{aligned}$$

is an orthogonal projector for the inner product  $V_0(g_2|x_0)^{-1}$ . After replacing in (A.6) we get:

$$\widehat{f}_0(y|x_0) \simeq \widehat{f}(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V_0(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy. \tag{A.7}$$

## v) Asymptotic expansion of the moment of interest

We have:

$$\begin{aligned}
& \widehat{E}(a|x_0) \\
&= \int a(y; \widehat{\theta}) \widehat{f}_0(y|x_0) dy \\
&\simeq \int a(y; \theta_0) f(y|x_0) dy + \int \frac{\partial a}{\partial \theta'}(y; \theta_0) f(y|x_0) dy (\widehat{\theta} - \theta_0) \\
&\quad + \int a(y; \theta_0) [\widehat{f}_0(y|x_0) - f(y|x_0)] dy \\
&\simeq E(a|x_0) + E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R (\widehat{\eta}_2 - \eta_2^0) \\
&\quad + \int a(y; \theta_0) \left\{ \widehat{f}(y|x_0) - f(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V_0 (g_2|x_0)^{-1} \right. \\
&\quad \left. (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \right\} dy \quad [\text{from (A.7)}] \\
&= E(a|x_0) - E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R \left[ R' E_0 \left( \frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]^{-1} \\
&\quad \cdot R' E_0 \left( \frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \quad [\text{from (A.5)}] \\
&\quad + \int a(y; \theta_0) [\widehat{f}(y|x_0) - f(y|x_0)] dy \\
&\quad - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy.
\end{aligned}$$

Thus we get:

$$\begin{aligned}
& \widehat{E}(a|x_0) - E(a|x_0) \\
&\simeq \int a(y; \theta_0) \delta \widehat{f}(y|x_0) dy - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy \\
&\quad - \left[ E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right] \\
&\quad \cdot \left[ R' E_0 \left( \frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]^{-1} \\
&\quad \cdot R' E_0 \left( \frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy, \tag{A.8}
\end{aligned}$$

where  $\delta \widehat{f}(y|x_0) = \widehat{f}(y|x_0) - f(y|x_0)$ .

**vi) Asymptotic distribution of the estimator**

Let us finally derive the asymptotic distribution of the conditional moment estimator  $\widehat{E}(a|x_0)$ . In the asymptotic expansion (A.8), the first two terms involve the residual of the regression of  $\int a(y; \theta_0) \delta \widehat{f}(y|x_0) dy$  on  $\int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy$ . This residual is asymptotically independent of the third term. Thus we get:

$$\frac{\sqrt{Th_T}}{w} \left[ \widehat{E}(a|x_0) - E(a|x_0) \right] \xrightarrow{d} N(0, W(x_0)/f_X(x_0)),$$

where the asymptotic variance is such that:

$$\begin{aligned} W(x_0) &= V_0(a|x_0) - Cov_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} Cov_0(g_2, a|x_0) \\ &\quad + \left[ E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R - Cov_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right] \\ &\quad \cdot \left[ R' E_0 \left( \frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0(g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]^{-1} \\ &\quad \cdot \left[ E_0 \left( \frac{\partial a}{\partial \theta'} | x_0 \right) R - Cov_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} E_0 \left( \frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]'. \end{aligned}$$

Since  $W(x_0)/f_X(x_0)$  corresponds to the (kernel) nonparametric efficiency bound  $\mathcal{B}(x_0, a)$  [see Proposition 2], the (kernel) nonparametric efficiency of the information based estimator is proved.

## APPENDIX 3

### Identifiability in stochastic volatility model

In this Appendix we consider the identifiability of structural parameter  $\theta$  in the stochastic volatility model. Specifically we check Assumption A.2\* and provide the expression of matrix  $R$  defining the directions of full information underidentification.

#### i) Computation of matrix $R$

The null space  $N_0$  associated with the uniform restrictions is the linear space of vectors  $v \in \mathbb{R}^4$  such that:

$$E_0 \left( \left( \begin{array}{c} 1 \\ \exp r_{t+1} \end{array} \right) \frac{\partial M_{t,t+1}(\theta_0) | y_t}{\partial \theta'} v = 0, \quad \forall y_t. \quad (\text{A.9})$$

We know that  $\theta_0$  satisfies the no-arbitrage restrictions:

$$E_0 \left( M_{t,t+1}(\theta_0) \left( \begin{array}{c} 1 \\ \exp r_{t+1} \end{array} \right) | y_t \right) = \left( \begin{array}{c} 1 \\ 1 \end{array} \right), \quad \forall y_t.$$

We deduce that any  $\theta = \theta_0 + v\varepsilon$ , where  $\varepsilon$  is small and  $v$  satisfies (A.9), is also such that:

$$E_0 \left( M_{t,t+1}(\theta) \left( \begin{array}{c} 1 \\ \exp r_{t+1} \end{array} \right) | y_t \right) = \left( \begin{array}{c} 1 \\ 1 \end{array} \right), \quad \forall y_t,$$

at first order in  $\varepsilon$ . Therefore the vectors of  $N_0$  are the directions  $d\theta = \theta - \theta_0$  of parameter changes, which are compatible with no-arbitrage. From Proposition 4 and equations (25) the parameters  $\theta$  which are compatible with no-arbitrage are characterized by the nonlinear restrictions:

$$\begin{aligned} \nu_0 &= -b(\nu_1 + \nu_3\gamma - \nu_3^2/2), \\ \nu_2 &= -a(\nu_1 + \nu_3\gamma - \nu_3^2/2), \\ \nu_3 &= \gamma + 1/2, \end{aligned}$$

where  $\gamma$  is a parameter of the DGP considered as fixed. In particular,  $\gamma = 1/2$  for the DGP considered in Section 5. Therefore:

$$\begin{aligned} \nu_0 &= -b(\nu_1), \\ \nu_2 &= -a(\nu_1), \\ \nu_3 &= 1. \end{aligned}$$

Thus the tangent set is spanned by the vector:

$$v = \begin{pmatrix} d\nu_0/d\nu_1 \\ d\nu_1/d\nu_1 \\ d\nu_2/d\nu_1 \\ d\nu_3/d\nu_1 \end{pmatrix} = \begin{pmatrix} -db(\nu_1)/d\nu_1 \\ 1 \\ -da(\nu_1)/d\nu_1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\delta \frac{c}{1+c\nu_1} \\ 1 \\ -\rho \frac{1}{(1+c\nu_1)^2} \\ 0 \end{pmatrix}.$$

Therefore matrix  $R$  is given by:

$$R = \begin{pmatrix} -\delta \frac{c}{1+c\nu_1} \\ 1 \\ -\rho \frac{1}{(1+c\nu_1)^2} \\ 0 \end{pmatrix}. \quad (\text{A.10})$$

## ii) Check of Assumption A.2\*

Let us now verify that Assumption A.2\* is satisfied when the conditional restrictions include the observed price of a European call. We have to prove that:

$$E_0 \left( \frac{\partial M_{t,t+1}}{\partial \theta'} (\theta_0) (\exp r_{t+1} - s)^+ \mid y_{t_0} \right) R \neq 0, \quad \forall s > 0.$$

In fact we have:

$$\begin{aligned} & E_0 \left( \frac{\partial M_{t,t+1}}{\partial \theta'} (\theta_0) (\exp r_{t+1} - s)^+ \mid y_{t_0} \right) R \\ &= -E_0 \left( M_{t,t+1} (\theta_0) (\exp r_{t+1} - s)^+ (1, \sigma_{t+1}^2, \sigma_t^2, r_{t+1}) R \mid y_{t_0} \right) \\ &= \left[ \delta \frac{c}{1+c\nu_1} + \rho \frac{1}{(1+c\nu_1)^2} \sigma_t^2 \right] E_0 \left( M_{t,t+1} (\theta_0) (\exp r_{t+1} - s)^+ \mid y_{t_0} \right) \\ &\quad - E_0 \left( M_{t,t+1} (\theta_0) (\exp r_{t+1} - s)^+ \sigma_{t+1}^2 \mid y_{t_0} \right). \end{aligned}$$

From (26) we have:

$$\delta \frac{c}{1+c\nu_1} + \rho \frac{1}{(1+c\nu_1)^2} \sigma_t^2 = \rho^* \sigma_t^2 + \delta^* c^* = \overset{Q}{E}_t [\sigma_{t+1}^2],$$

where  $Q$  denotes the risk neutral distribution, whereas from the Hull-White formula:

$$\begin{aligned} E_0 \left( M_{t,t+1} (\theta_0) (\exp r_{t+1} - s)^+ \mid y_{t_0} \right) &= \overset{Q}{E}_{t_0} [BS(1, s, \sigma_{t+1}^2)], \\ E_0 \left( M_{t,t+1} (\theta_0) (\exp r_{t+1} - s)^+ \sigma_{t+1}^2 \mid y_{t_0} \right) &= \overset{Q}{E}_{t_0} [\sigma_{t+1}^2 BS(1, s, \sigma_{t+1}^2)]. \end{aligned}$$

Thus we get:

$$E_0 \left( \frac{\partial M_{t,t+1}}{\partial \theta'} (\theta_0) (\exp r_{t+1} - s)^+ \mid y_{t_0} \right) R = -\mathit{Cov}_t^Q [\sigma_{t+1}^2, BS(1, s, \sigma_{t+1}^2)],$$

which is negative since the Black-Scholes price is an increasing function of volatility.

## REFERENCES

Ai, C., and X., Chen (2003) : "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", *Econometrica*, 71, 1795-1843.

Aït-Sahalia, Y. (1996) : "Nonparametric Pricing of Interest Rate Derivative Securities", *Econometrica*, 64, 527-560.

Aït-Sahalia, Y., and A., Lo (1998) : "Nonparametric Estimation of State Price Densities Implicit in Financial Asset Prices", *Journal of Finance*, 53, 499-547.

Altonji, J. G., and L. M., Segal (1996): "Small Sample Bias in GMM Estimation of Covariance Structures", *Journal of Business and Economic Statistics*, 14, 353-366.

Back, K., and D., Brown (1992) : "GMM, Maximum Likelihood and Nonparametric Efficiency", *Economics Letters*, 39, 23-28.

Back, K., and D., Brown (1993) : "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.

Baggerly, K. (1998) : "Empirical Likelihood as a Goodness of Fit Measure", *Biometrika*, 85, 535-574.

Ball, C., and A., Roma (1994) : "Stochastic Volatility Option Pricing", *Journal of Financial and Quantitative Analysis*, 29, 589-607.

Bams, D. (1998) : "An Empirical Comparison of Time Series and Cross-Sectional Information in the Longstaff-Schwartz Term Structure Models", Maastricht University, D.P.

Bansal, R., and S., Viswanathan (1993) : "No Arbitrage and Arbitrage Pricing : A New Approach", *Journal of Finance*, 48, 1231-1262.

Bickel, P., Klaassen, A., Ritov, Y. and J., Wellner (1993) : "Efficient and Adaptive Estimation for Semi-Parametric Models", Baltimore, MD : The John Hopkins University Press.



Bonnal, H., and E., Renault (2004) : "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood", Working Paper, University of Montreal.

Brown, B. (1998): "Efficient Semiparametric Estimation of Expectations", *Econometrica*, 66, 453-464.

Buchen, P., and M., Kelly (1996) : "The Maximum Entropy Distribution of an Asset Inferred from Option Prices", *Journal of Financial and Quantitative Analysis*, 31, 143-159.

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

Cochrane, J. H. (2001) : "Asset Pricing", Princeton University Press.

Cressie, N., and T., Read (1984): "Multinomial Goodness of Fit Tests", *Journal of the Royal Statistical Society*, 8, 46, 440-464.

De Munnik, J., and P., Schotman (1994) : "Cross-Sectional Versus Time Series Estimation of Term Structure Models : Empirical Results for the Dutch Bond Market", *Journal of Banking and Finance*, 18, 997-1025.

Donald, S., Imbens, G., and W., Newey (2003) : "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions", *Journal of Econometrics*, 117, 55-93.

Donald, S., and W., Newey (2000) : "A Jackknife Interpretation of the Continuous Updating Estimator", *Economics Letters*, 67, 239-243.

Donoho, D., and R., Liu, (1991): "Geometrizing Rate of Convergence II", *Annals of Statistics*, 19, 633-667.

Fan, J., (1993): "Local Linear Regression Smoothers and their Minimax Efficiency", *Annals of Statistics*, 21, 196-216.

Garcia, R., Luger, R. and E., Renault (2003) : "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables", *Journal of Econometrics*, 116, 49-83.

Gouriéroux, C., and J., Jasiak (2000) : "Autoregressive Gamma Process", CREST-DP.

Gourieroux, C., and A., Monfort (2001) : "Pricing with Splines", CREST-DP.

Gourieroux, C., and R., Sufana (2004): "Derivative Pricing with Multivariate Stochastic Volatility", HEC Montreal DP.

Hansen, L. (1982) : "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L., Heaton, P., and A., Yaron (1996) : "Finite Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics*, 14, 262-280.

Hansen, L., and R., Jagannathan (1991) : "Implications of Security Market Data for Models of Dynamic Economics", *Journal of Political Economy*, 99, 225-262.

Hansen, L., and S., Richard (1987) : "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", *Econometrica*, 55, 587-613.

Hansen, L., and K., Singleton (1982) : "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286.

Heston, S., (1993) : "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *Review of Financial Studies*, 6, 327-343.

Hutchinson, J., Lo, A., and T., Poggio (1994) : "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks", *Journal of Finance*, 49, 771-818.

Imbens, G. (1997): "One-Step Estimators for Over-identified Generalized Method of Moments Models", *Review of Economic Studies*, 64, 359-383.

Imbens, G., Spady, R., and P., Johnson (1998): "Information Theoretic

Approaches to Inference in Moment Conditions Models", *Econometrica*, 66, 333-357.

Jackwerth, J. C., and M., Rubinstein (1996): "Recovering Probability Distributions from Option Prices", *Journal of Finance*, 51, 1611-1631.

Jondeau, E., and M., Rockinger (2000) : "Entropy Densities", Working Paper, Banque de France, Paris.

Kitamura, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions", *Econometrica*, 69, 1661-1672.

Kitamura, Y., and M., Stutzer (1997) : "An Information Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.

Kitamura, Y., Tripathi, G. and H., Ahn (2004) : "Empirical Likelihood Based Inference in Conditional Moment Restriction Models", *Econometrica*, 72, 1667-1714.

Lucas, R. (1978) : "Asset Prices in an Exchange Economy", *Econometrica*, 4, 1429-1445.

Newey, W., and R., Smith (2004) : "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-255.

Owen, A. (1991) : "Empirical Likelihood for Linear Models", *The Annals of Statistics*, 19, 1725-1747.

Owen, A. (2001): *Empirical Likelihood*, New-York, Chapman and Hall.

Qin, J., and L., Lawless (1994): "Empirical Likelihood and General Estimating Equations", *The Annals of Statistics*, 22, 300-325.

Rubinstein, M. (1994): "Implied Binomial Trees", *Journal of Finance*, 49, 771-818.

Severini, T., and G., Tripathi (2001) : "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models", *Journal of Econometrics*, 102, 23-66.

Smith, R., (1997): "Alternative Semiparametric Likelihood Approaches to Generalized Method of Moments Estimation", *Economic Journal*, 107, 503-519.

Smith, R., (2000): "Empirical Likelihood Estimation and Inference", in "Applications of Differential Geometry to Econometrics", P. Marriot and M., Salmon, Eds, Cambridge University Press.

Smith, R., (2004): "Local GEL Estimation with Conditional Moment Restrictions", Working Paper.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation", in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1, 187-195.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification", *Econometrica*, 68, 1055-1096.

Stutzer, M. (1995) : "A Bayesian Approach to Diagnosis of Asset Pricing Models", *Journal of Econometrics*, 68, 367-397.

Stutzer, M. (1996) : "A Simple Nonparametric Approach to Derivative Security Valuation", *Journal of Finance*, 51, 1633-1652.

**Table 1**

Maturity: H=2, Sample size: T=250					
Derivative price $\beta = E(a(s) y_{t_0}) (\times 100)$					
Strike $s = 0.95$			Strike $s = 0.96$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
5.007	4.990 – 5.026		4.029	4.014 – 4.043	
Mean	95% confidence interval		Mean	95% confidence interval	
5.029	4.904 – 5.154		4.046	3.944 – 4.148	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
5.035	4.921	5.121	4.047	3.958	4.125
Strike $s = 0.97$			Strike $s = 0.99$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
3.079	3.070 – 3.088		1.419	1.413 – 1.424	
Mean	95% confidence interval		Mean	95% confidence interval	
3.089	3.029 – 3.150		1.413	1.376 – 1.451	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
3.089	3.037	3.138	1.413	1.382	1.443
Strike $s = 1.01$			Strike $s = 1.03$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
0.424	0.419 – 0.429		0.089	0.084 – 0.095	
Mean	95% confidence interval		Mean	95% confidence interval	
0.426	0.392 – 0.459		0.087	0.048 – 0.125	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
0.426	0.395	0.452	0.087	0.053	0.119
Strike $s = 1.04$			Strike $s = 1.05$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
0.037	0.031 – 0.044		0.015	0.010 – 0.022	
Mean	95% confidence interval		Mean	95% confidence interval	
0.035	0 – 0.077		0.014	0 – 0.045	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
0.031	0.031	0.074	0.009	0.000	0.044

**Table 2**

Maturity: H=2, Sample size: T=250					
Parameter $\theta$					
	True	Mean	Median	Stand. dev.	5% / 95% quant.
$\nu_0$ ( $\times 10^4$ )	-3.37	-10.0	-5.08	19.5	-38.8 / 0.36
$\nu_1$	10	4.78	6.57	16.4	-17.6 / 16.0
$\nu_2$	-8.498	-12.4	-8.63	16.9	-45.6 / 1.38
$\nu_3$	1	-0.30	0.10	5.58	-10.5 / 9.05

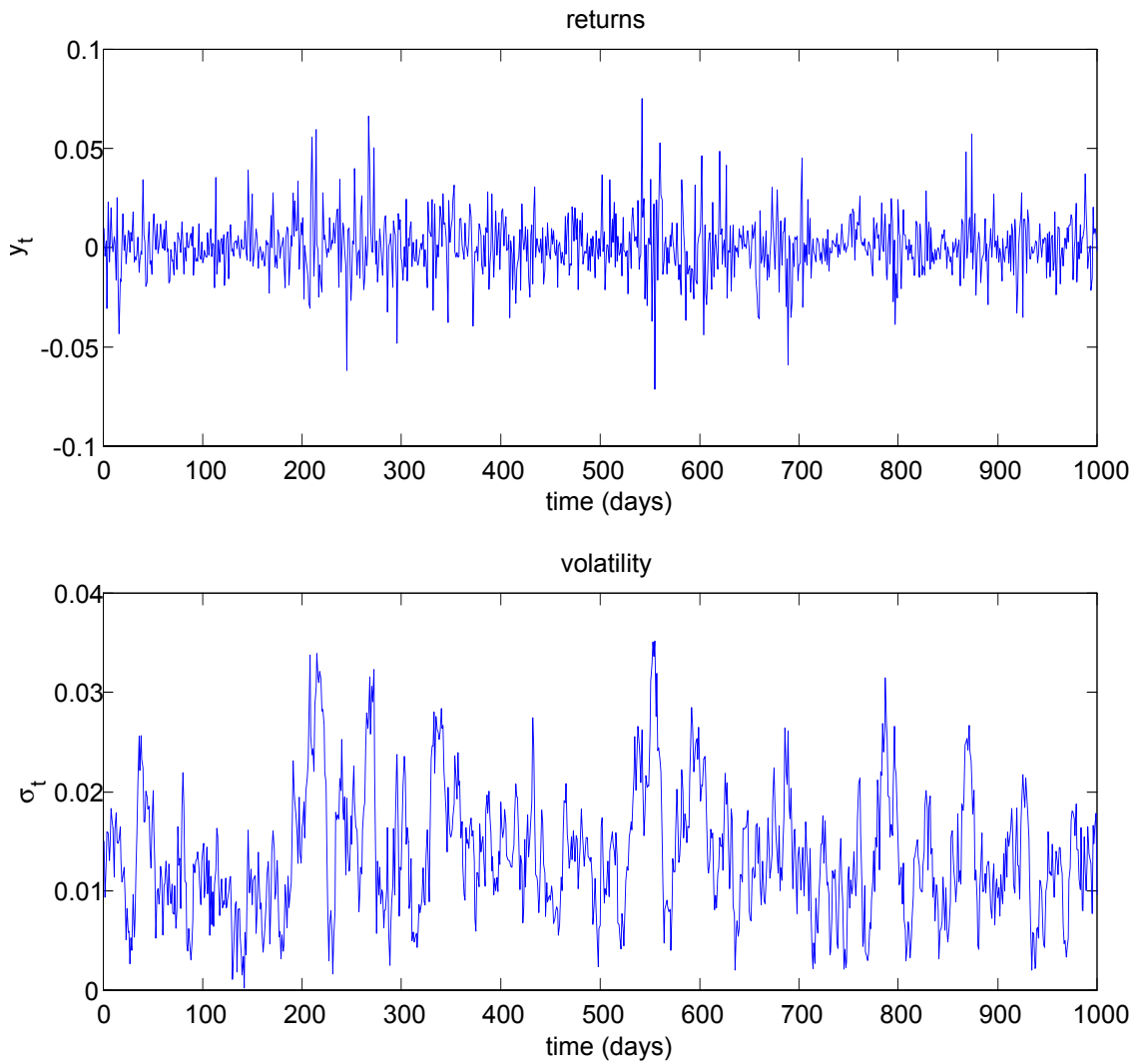


Figure 1: Simulated series of return  $r_t$  (upper Panel) and volatility series  $\sigma_t$  (lower Panel) for the ARG stochastic volatility process.

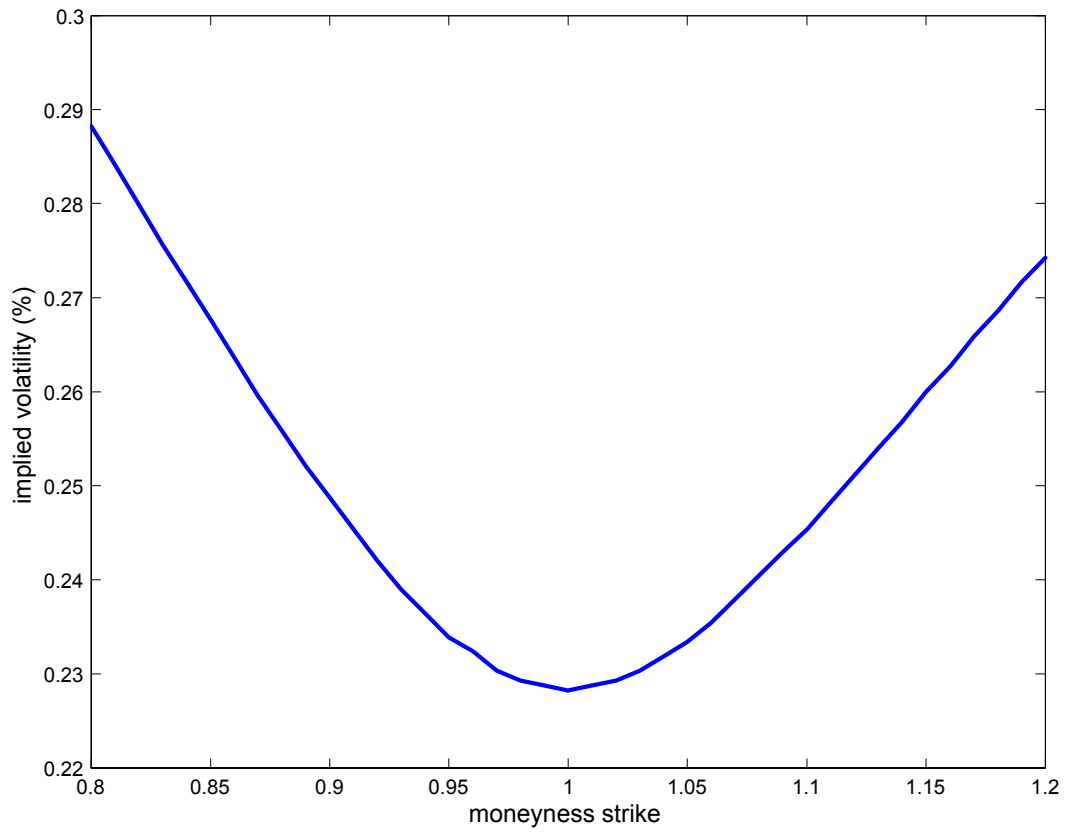


Figure 2: Implied Black-Scholes volatility (annualized %) as a function of the moneyness strike  $k$  for a European call with residual maturity  $h = 1$ .



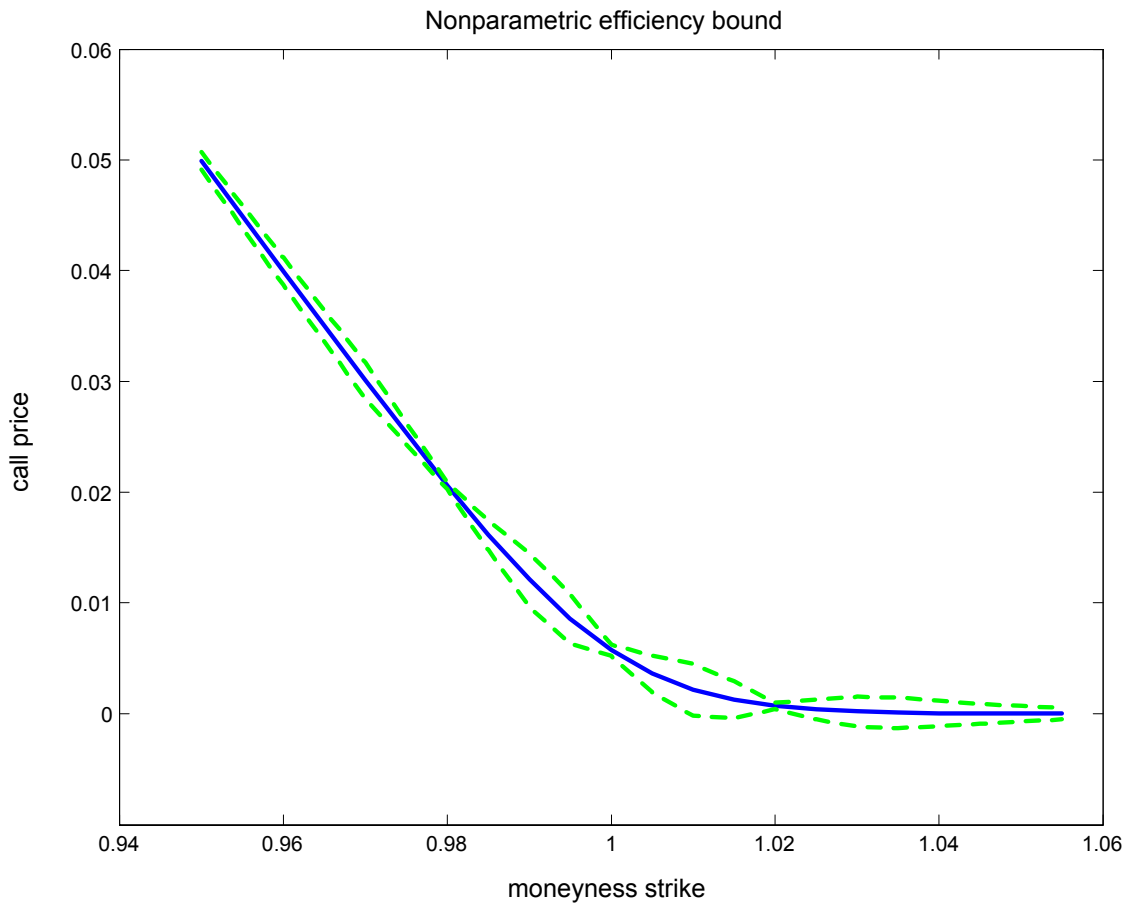


Figure 3: Nonparametric efficiency bound (limited information) for a European call with maturity 1. The solid line corresponds to the price  $E(a(s)|y_{t_0})$ , the dashed lines to pointwise confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ .

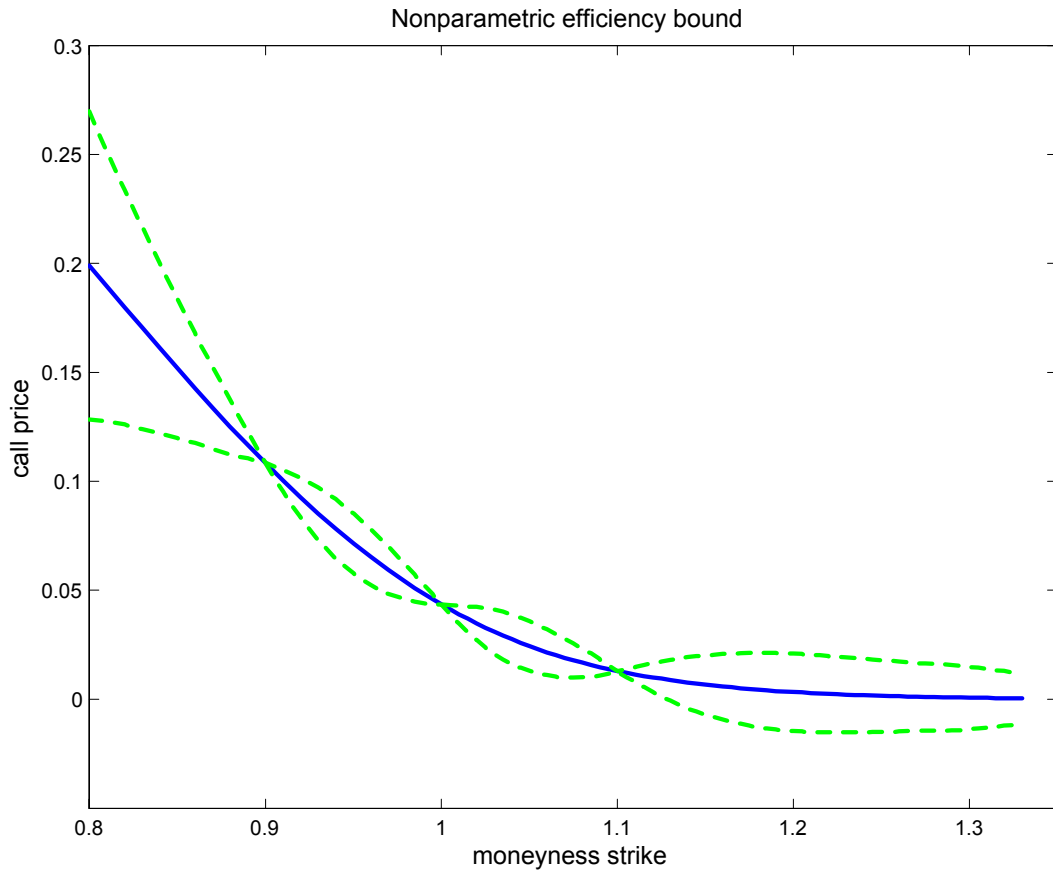


Figure 4: Nonparametric efficiency bound (limited information) for a European call with maturity 60. The solid line corresponds to the price  $E(a(s)|y_{t_0})$ , the dashed lines to pointwise confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ .

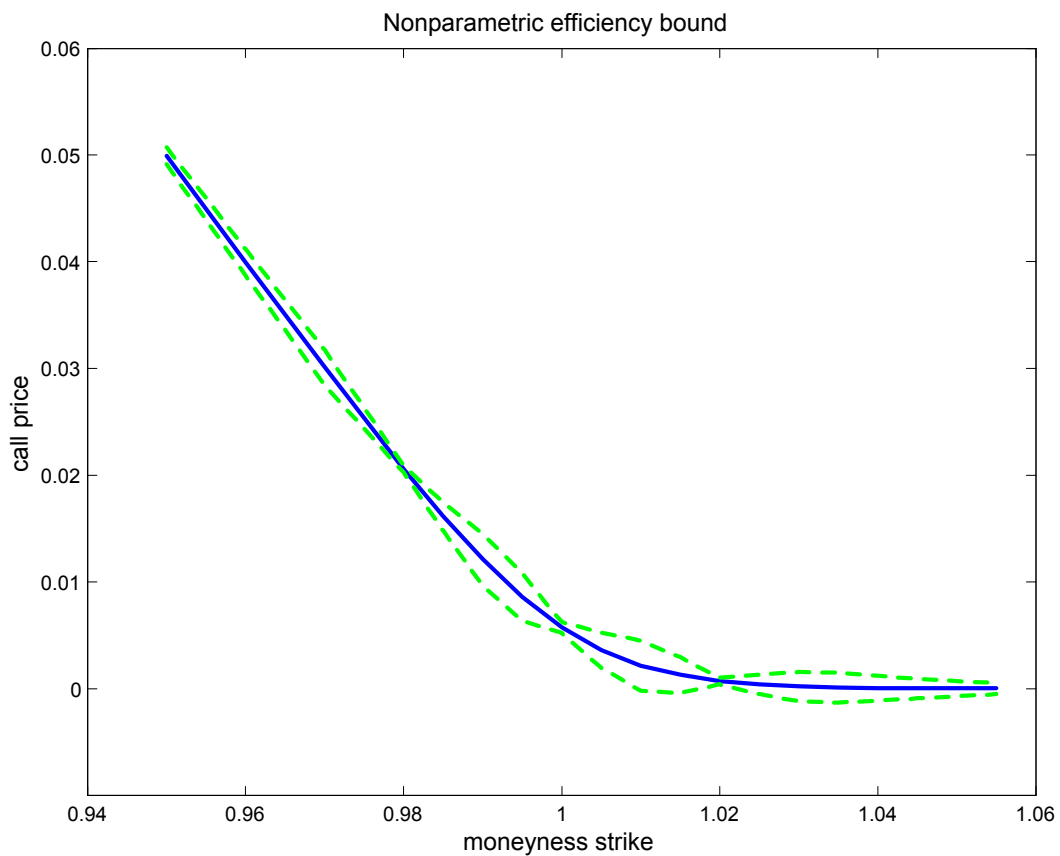


Figure 5: Nonparametric efficiency bound (mixed information) for a European call with maturity 1. The solid line corresponds to the price  $E(a(s)|y_{t_0})$ , the dashed lines to pointwise confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ .

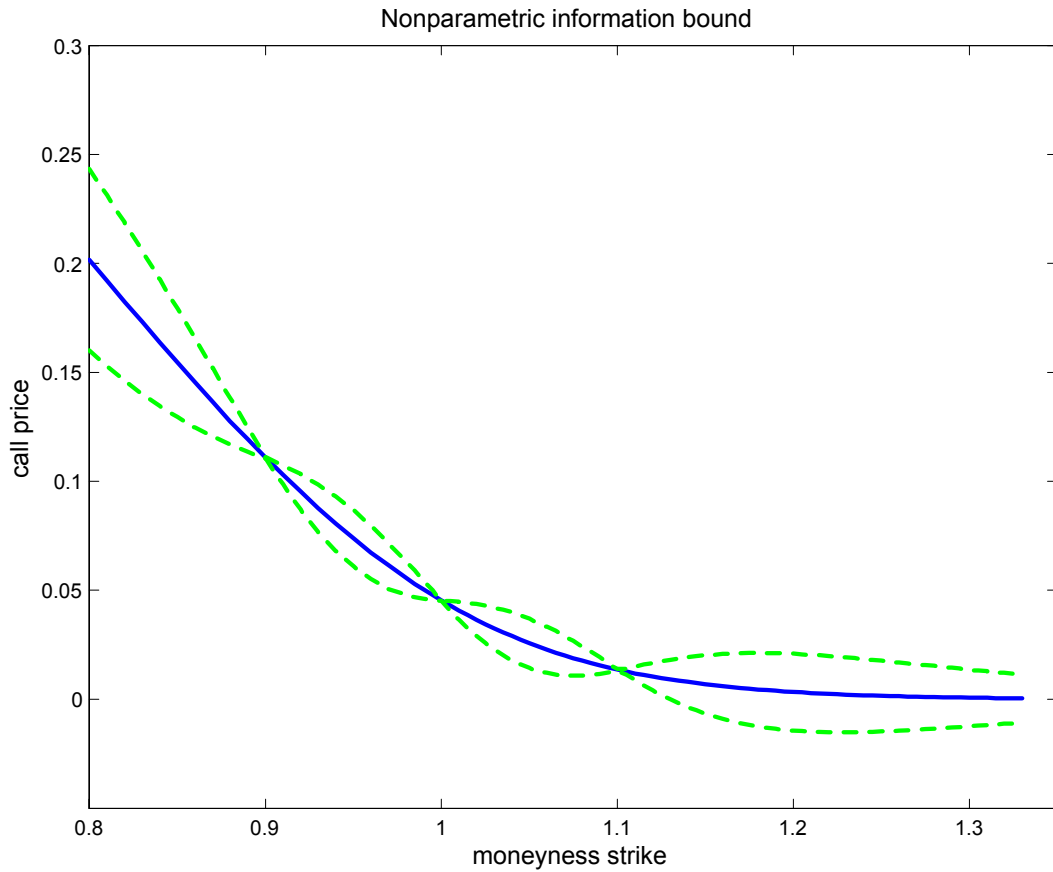


Figure 6: Nonparametric efficiency bound (mixed information) for a European call with maturity 60. The solid line corresponds to the price  $E(a(s)|y_{t_0})$ , the dashed lines to pointwise confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ .

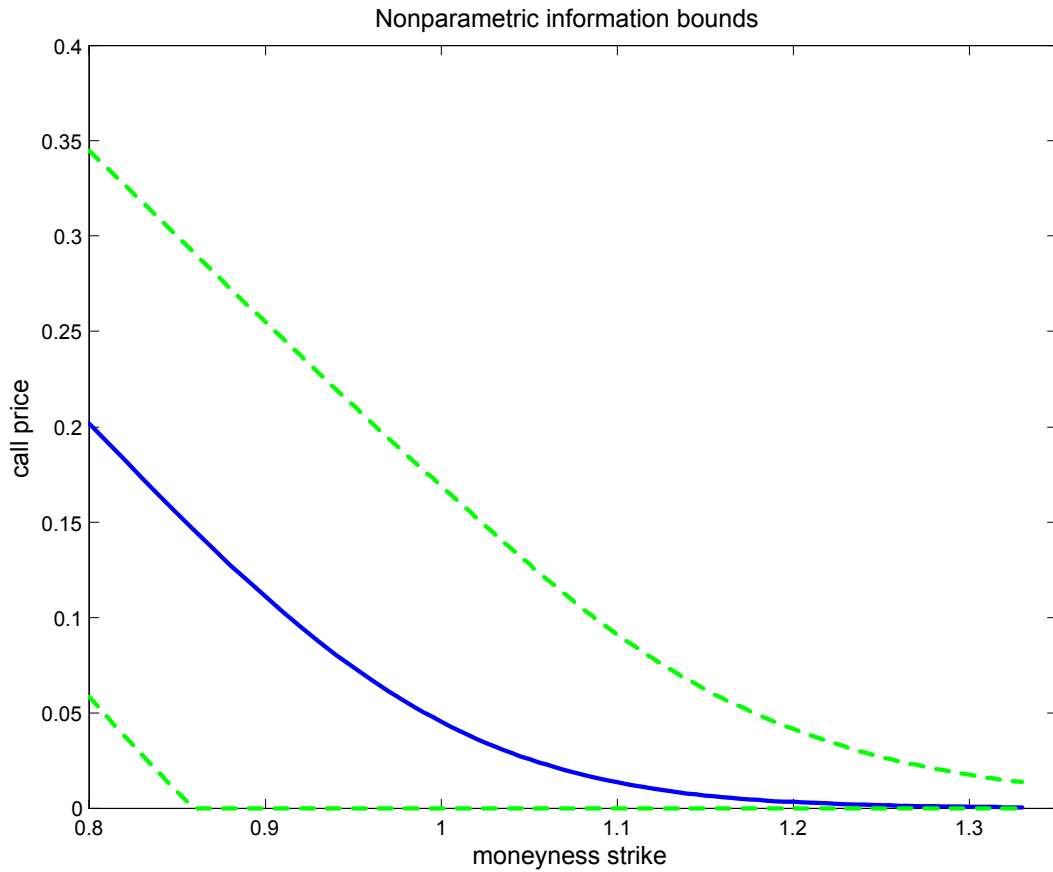


Figure 7: Nonparametric efficiency bound (mixed information) for a European call with maturity 60 when prices of calls with maturity 20 are observed. The solid line corresponds to the price  $E(a(s)|y_{t_0})$ , the dashed lines to point-wise confidence intervals  $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$ .

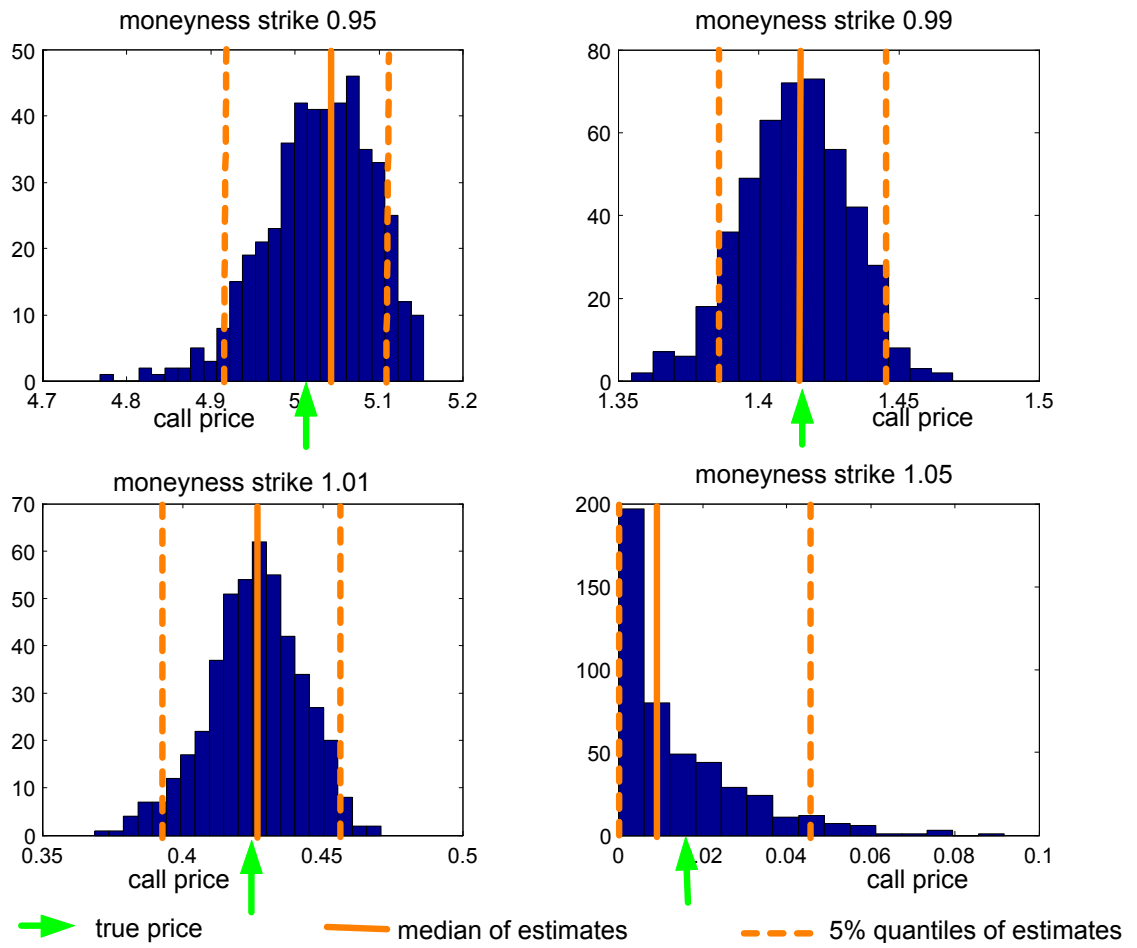


Figure 8: Information based XMM estimator: histograms of estimated derivative prices.

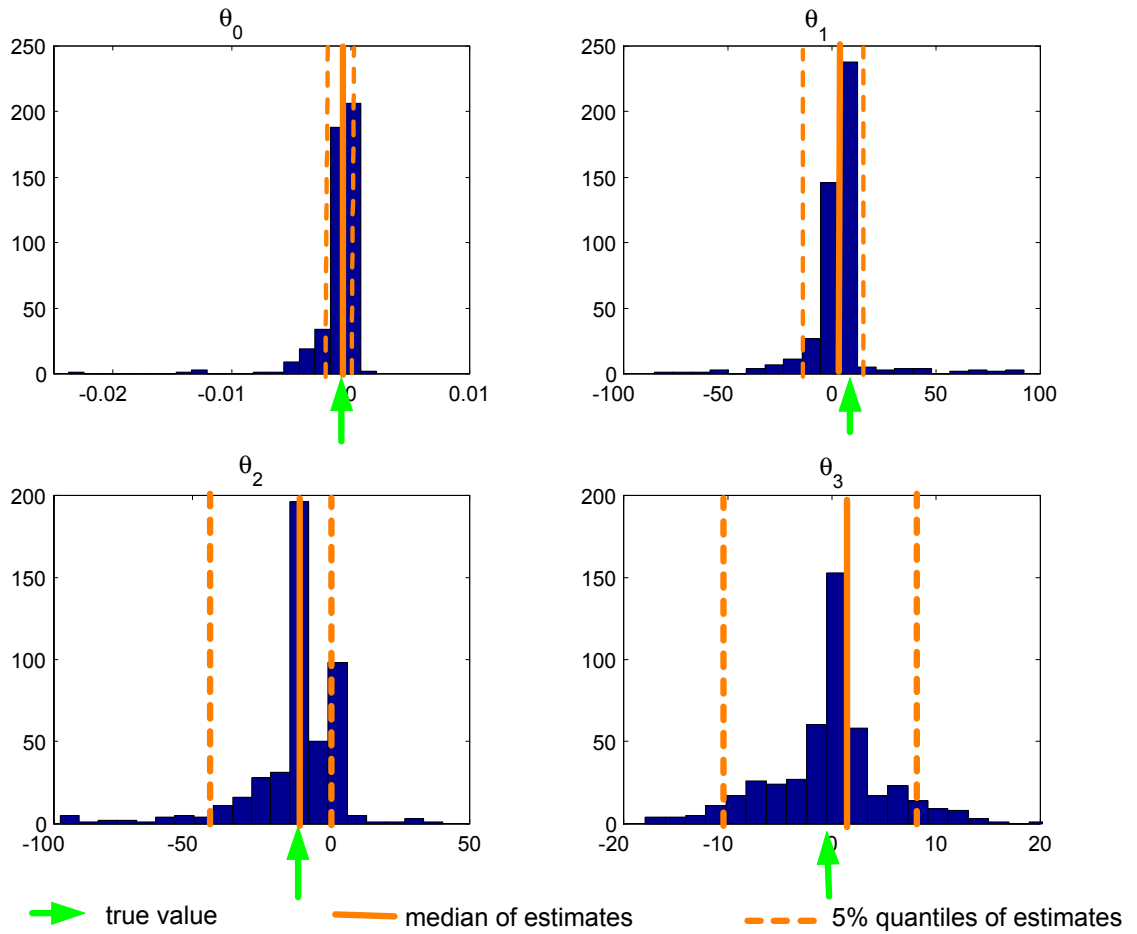


Figure 9: Information based XMM estimator: histograms of estimated structural parameters.