

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2005-40

**Efficient Derivative Pricing by
Extended Method of Moments**

P. GAGLIARDINI¹
C. GOURIEROUX²
E. RENAULT³

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ University of St. Gallen and University of Lugano.

² CREST, CEPREMAP (Paris) and University of Toronto.

³ CIRANO-CIREQ (Montreal) and University of North Carolina at Chapel Hill.

Acknowledgements : We thank D. Andrews, X. Chen, R. Davidson, R. Garcia, J. Jackwerth, L. Mancini, A. Melino, M. Stutzer, F. Trojani and E. Zivot for helpful comments.

EFFICIENT DERIVATIVE PRICING BY EXTENDED METHOD OF MOMENTS

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

First version: July 2004
This version: December 2005

*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.

Acknowledgements: We thank D. Andrews, X. Chen, R. Davidson, R. Garcia, J. Jackwerth, L. Mancini, A. Melino, M. Stutzer, F. Trojani and E. Zivot for helpful comments.

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); indeed, GMM assumes that all restrictions are 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, Kernel 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 produits 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 current and lagged values of prices and income. The parameters of interest are the preference parameter γ and the psychological discount rate δ . 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 estimator and to improve its finite sample properties¹. The basic idea is to estimate jointly the structural parameter θ 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 also valid for the other assets, whose payoffs are 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 time-to-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 θ based on the orthogonality conditions (2) and \hat{E}_{t_0} is a (functional) estimator of the conditional expectation. For the application to derivative pricing, the interest is focused on estimation of the conditional moment $c_{t_0}(1, K) = E_{t_0}(a)$ of function $a = M_{t_0,t_0+1}(p_{1,t_0+1} - K)^+$, which is the product of the sdf by the derivative

¹See 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), Smith (2000), Kitamura (2001), Ai, Chen (2003), Kitamura, Tripathi, Ahn (2004), Newey, Smith (2004), Bonnal, Renault (2004).

payoff. Clearly, this problem requires a joint estimation of the parameter θ and of the conditional distribution.

However, the problem becomes much more complicated if we consider a more general pricing formula and if we want to account for the observed prices of assets, which are less actively traded (such as a derivative with given moneyness strike and time-to-maturity), when estimating θ and the conditional distribution. Typically, we can observe the prices of the short term zero-coupon bond:

$$B(t, t + 1) = E_t [M_{t,t+1}(\theta)], \quad \text{for all } t, \quad (5)$$

the prices of the underlying asset:

$$p_{1,t} = E_t [M_{t,t+1}(\theta) p_{1,t+1}], \quad \text{for all } t, \quad (6)$$

and for instance the at-the-money call price at date t_0 , the current date, say:

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

In this situation, the structural parameter θ 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)]. These two types of moment restrictions are difficult to take jointly into account. This explains the approaches, which have been followed in practice (but also in the academic literature):

i) The observations of derivative prices can be neglected, the parameter θ estimated by a standard GMM method based on (5)-(6) and the derivative price of interest approximated by (4). The drawback of this approach is that the risk premium parameters are generally non identifiable from historical dynamics alone, and some of them have to be fixed a priori (frequently to zero). In this respect, the standard CCAPM equilibrium model (1) is not representative of a general pricing formula. By assuming known a priori the number and types of highly traded assets, it implies that the structural parameters γ, δ are identifiable from historical dynamics of the price of underlying asset. In a more general setting, some risk premia parameters can be unidentifiable from these histories. This arises in reduced form pricing models (see Section 5 below for an example with stochastic volatility), as well as in equilibrium models with nonproportional trading costs, which can create a lack of trading on an endogenous set of assets. Moreover, an additional drawback of this approach is that, by discarding information on derivative prices, we incur in an efficiency loss for estimation of other derivative prices.

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 this 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 given derivatives written on this asset for an efficient pricing of other derivatives. Equivalently, 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 kernel 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 conditional moments 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, 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 a 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 in general provide 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 time-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 estimated derivative prices and estimated 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 lack of liquidity on derivative markets. Section 6 concludes. Since the focus of the paper is mainly on the structural interpretations of moment restrictions and on the application to derivative pricing, detailed regularity conditions and proofs of asymptotic results for the XMM estimator are gathered in Appendix 1. The asymptotic results for the information based estimator are sketched in Appendix 2 with the purpose of deriving the limiting distribution and proving kernel nonparametric efficiency. Proofs of technical Lemmas are given in Appendix B and C, which are available at web-site: <http://www.istituti.usilu.net/gagliarp/web/proofsXMM.htm>.

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 strongly stationary. Thus, the conditioning (state) variables X are assumed observable. 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 θ_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 estimator² of θ (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 first step estimation of θ_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 θ_0 have local interpretations. The rates of convergence are nonparametric for both parameter θ_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 θ_0 .

These two cases are said with full- and limited-information, respectively. In Subsection 2.1, we consider a general framework in which the structural parameter θ 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 kernel 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:

$$\begin{aligned} E[g(Y; \theta_0) | X = x] &= 0, \quad \forall x, \\ E[\tilde{g}(Y; \theta_0) | X = x_0] &= 0, \end{aligned} \tag{8}$$

where θ_0 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 an 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 θ_0 satisfies both marginal and conditional (local) restrictions:

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

²Whenever θ is identifiable from the marginal moment restrictions.

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 θ .

- i) If θ 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 θ will converge at a parametric rate.
- ii) If θ is not identifiable from the marginal restrictions only, but is identifiable when both types of restrictions are jointly considered, we can expect different parametric, or non-parametric rates of convergence according to the function (component) of parameter θ , which is considered. This is the general situation.

More precisely, the identification assumptions are the following:

Assumption A.1: *Parameter θ 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 θ is locally identifiable from marginal and conditional moment restrictions, that is, the matrix:*

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

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 θ 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 θ . However, the parameter of primary interest for our purpose 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, we need to discuss the interpretation of parameter of interest β_0 , which corresponds to a set of derivative prices at some given date in the application. More precisely, we have to distinguish between the mapping $x \mapsto E_0 [a(Y; \theta_0) | X = x]$, usually interpreted in terms of predictor, 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 Back, Brown (1992)], whose true value $(\theta'_0, \beta'_0)'$ satisfies the extended set of moment restrictions:

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

Since the dimension of β is equal to the number of moments of interest, that is, the dimension of a , the extended parameter θ^* is also globally or locally identified under Assumptions A.1, A.2. The extended problem always involves restrictions conditional on a given value of the conditioning variable (the restrictions defining β_0), even if θ_0 is defined by means of uniform restrictions only.

ii) Moment estimator

We will now consider moment estimators for $\theta^* = (\theta', \beta')'$ 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 an historical sample average and a kernel estimator of the conditional moment, respectively. More precisely, let us introduce a kernel estimator³ 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, where $X_t = Y_{t-1}$. The kernel density estimator is defined by:

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

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 is used to approximate a conditional moment $E_0(g_2|x_0) = E_0[g_2(Y; \theta)|X = x_0]$ by:

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

Under standard regularity conditions including the bandwidth conditions: $Th_T^d \rightarrow \infty$, $(Th_T^d)^{1/2} h_T^2 \rightarrow 0$ as $T \rightarrow \infty$, 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 (stationary) 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 $\widetilde{E}(\cdot|x_0)$, respectively.

³Another type of nonparametric estimator of the conditional density can be used.

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

$$\widehat{\theta}_T^*(\Omega) = \arg \min_{\theta^* = (\theta', \beta')} \widehat{g}_T(\theta^*)' \Omega \widehat{g}_T(\theta^*),$$

where

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

and Ω is a weighting matrix.

Under standard regularity conditions, the associated (kernel) moment estimator of parameter β 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 Ω [see Appendix 1 for the list of regularity conditions and the derivation of the asymptotic properties of $\widehat{\theta}^*$]. The expression of the asymptotic variance $w^2 V_Z(\Omega)$ for the optimal choice of Ω is provided next.

iii) Kernel nonparametric efficiency bound

Definition 2: The kernel nonparametric 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 Ω . It defines the functional $a \rightarrow \mathcal{B}_Z(x_0, a)$.⁴

The main result of this subsection, provided in Proposition 1 below, is that the kernel nonparametric efficiency bound for β_0 depends on the selected instrument Z only through the local identification content of the corresponding marginal restrictions:

$$E[g_1(Y, X, \theta_0)] = E[Z \cdot g(Y, \theta_0)] = 0.$$

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:

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

⁴Other notions of "nonparametric efficiency" have been introduced in the statistical literature, such as the approach based on linear forms of the parameter of interest [see e.g. Stein (1956), Severini, Tripathi (2001)] or the minimax lower bounds [see e.g. Donoho, Liu (1991), Fan (1993)]. However, these notions focus on the functional dependence of parameter β_0 from x_0 , whereas, in derivative pricing, x_0 is given and the interesting functional dependence is w.r.t. a . This is why we emphasize the mapping $a \rightarrow \mathcal{B}_Z(x_0, a)$. Moreover, these alternative notions of "nonparametric efficiency" are less operational for practical applications such as derivative pricing. They will not be addressed in this paper. Finally, the kernel nonparametric efficiency bound $\mathcal{B}_Z(x_0, a)$ does not depend on the kernel K . The effect of the kernel on the asymptotic distribution of $\widehat{\theta}_T^*$ is summarized in the scale factor w^2 .

The columns of R_Z generate the null space of matrix J_1^Z . Moreover, let us denote by \tilde{R} a $p \times s_Z$ matrix whose columns complete those of R_Z to get a basis of \mathbb{R}^p . Then, the $p \times p$ matrix $R_1 = \begin{pmatrix} \tilde{R} & R_Z \end{pmatrix}$ is non singular, and allows to define a new parametrization:

$$\eta = R_1^{-1}\theta = \left(\eta'_1, \eta'_2\right)' \in \mathbb{R}^{s_Z} \times \mathbb{R}^{p-s_Z}. \quad (11)$$

The vector η_1 defines s_Z linear combinations of structural parameters θ_0 , which are identified from the marginal restrictions, while η_2 corresponds to $p - s_Z$ linear combinations for which the marginal restrictions are not sufficiently 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 η_1 can be estimated at a standard parametric rate, whereas η_2 features a nonparametric rate of convergence induced by the local conditional moment restrictions. Thus, the intuition of the main result below is the following: as far as the kernel nonparametric efficiency bound for β_0 is concerned, parameters η_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 β_0 . This is why the efficiency bound for β_0 depends on the instrument Z only through the information matrix $I_{0,Z}$:

$$I_{0,Z} = f_X(x_0) \begin{pmatrix} E_0 \left(\frac{\partial g_2}{\partial \eta'_2} \right) & 0 \\ E_0 \left(\frac{\partial a}{\partial \eta'_2} \right) & -Id_L \end{pmatrix}' \begin{pmatrix} V_0(g_2) & Cov_0(g_2, a) \\ Cov_0(a, g_2) & V_0(a) \end{pmatrix}^{-1} \begin{pmatrix} E_0 \left(\frac{\partial g_2}{\partial \eta'_2} \right) & 0 \\ E_0 \left(\frac{\partial a}{\partial \eta'_2} \right) & -Id_L \end{pmatrix}, \quad (12)$$

where all moments are conditional on $X = x_0$. The matrix $I_{0,Z}^{-1}$ is similar to a standard GMM efficiency bound for estimation of parameters $\left(\eta'_2, \beta'\right)'$ from moment restrictions based on functions $\left(g'_2, a' - \beta\right)'$, 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 (12). Then, the kernel nonparametric efficiency bound $a \rightarrow \mathcal{B}_Z(x_0, a)$ 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,

$$\begin{aligned} \mathcal{B}_Z(x_0, a) &= \frac{1}{f_X(x_0)} \left\{ V_0 a - \text{Cov}_0(a, g_2)(V_0 g_2)^{-1} \text{Cov}_0(g_2, a) \right. \\ &\quad + \left[E_0 \left(\frac{\partial a}{\partial \theta'} \right) R_Z - \text{Cov}_0(a, g_2)(V_0 g_2)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} \right) R_Z \right] \\ &\quad \left[R'_Z E_0 \left(\frac{\partial g_2'}{\partial \theta} \right) (V_0 g_2)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} \right) R_Z \right]^{-1} \\ &\quad \left. \left[R'_Z E_0 \left(\frac{\partial a'}{\partial \theta} \right) - R'_Z E_0 \left(\frac{\partial g_2'}{\partial \theta} \right) (V_0 g_2)^{-1} \text{Cov}_0(g_2, a) \right] \right\}, \end{aligned}$$

where all moments are conditional on $X = x_0$, and evaluated at the true parameter value θ_0 .

Proof. See Appendix 1. ■

The matrix R_Z is not unique, but is defined up to a post-multiplication by a non-singular matrix. The above kernel nonparametric 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 [see Andrews, Stock (2005) for a general presentation]. 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:

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

They correspond to a finite number of weak moment restrictions⁵ constructed from instrument $Z_T(X) = K \left(\frac{X - x_0}{h_T} \right) / [h_T^d f_X(x_0)]$. This instrument depends on the number of observations, and is "weak" due to localization around $X = x_0$ induced by the kernel⁶; this explains the different rate of convergence of the structural parameters, when this "weak" instrument is used. Since the instruments depend on T , the problem considered above is not a special case of the standard literature on weak instruments or weak IV asymptotics [see e.g. Staiger, Stock (1997), Stock, Wright (2000), Yogo (2004), Andrews, Marmor (2005), Andrews, Stock (2005)]. For instance, the functions of the parameters which are weakly

⁵Thus, our framework differs from the literature introducing a large number of weak instruments [see e.g. Hansen, Hausman, Newey (2004), Hahn, Hausman, Kuersteiner (2004), Andrews, Stock (2005), Newey, Windmeijer (2005)].

⁶More precisely, the corresponding sample moment $\sqrt{T} h_T^d \tilde{E} [g_2(\theta) | x_0]$ in Definition 1 of the estimator is of order $\sqrt{T} h_T^d$ for $\theta \neq \theta_0$, which is lower than \sqrt{T} , as in the standard case of weak instruments [see Assumption C in Stock, Wright (2000)].

(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. The different setting explains also why asymptotically efficient estimation methods are derived below in our framework, whereas "there does not seem to be any dominant estimation method (for the linear model) with weak IV's" [Andrews, Stock (2005), p.20].

2.1.2 Efficiency bound with optimal instruments

i) Optimal instruments

The main lesson of the previous subsection is that the instrument Z only matters for estimation of β_0 through the null space of matrix $J_1^Z = E_0 \left[\frac{\partial g_1}{\partial \theta'}(Y, X; \theta_0) \right]$. Larger this null space is, larger the vector η_2 of structural parameters, which are non-identified from marginal restrictions and must be estimated at a non-parametric rate jointly with β , 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 kernel nonparametric efficiency bound $\mathcal{B}_Z(x_0, a)$ cannot be larger than $\mathcal{B}_W(x_0, a)$.

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

$$E_0 \left[\frac{\partial g}{\partial \theta'}(Y; \theta_0) \mid X = x \right] v = 0, \quad (13)$$

almost surely for the marginal distribution P^X of X . Vectors in N_0 define linear combinations of parameters θ 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 is minimal as soon as it coincides with N_0 . Let us consider in particular:

$$Z = E_0 \left(\frac{\partial g'}{\partial \theta}(Y; \theta_0) \mid X \right) W(X), \quad (14)$$

where $W(X)$ is a positive definite matrix (P^X -almost surely). 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:

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

or:

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

that is, v belongs to N_0 . Therefore, the choice (14) 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 θ 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 β_0 , the set of optimal instruments is larger than the standard one derived by Hansen (1982) and Chamberlain (1987) for efficient estimation of structural parameter θ . 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 β 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 θ .

ii) The identification assumption

Another useful formulation of Assumption A.2* is derived by considering a matrix R of dimension $p \times (p - s)$, say, whose columns constitute a basis of the space N_0 defined in (13). 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, $E_0 \left[\partial g_2 (Y; \theta_0) / \partial \theta' | X = x_0 \right] R$ 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 results above, the matrix R coincides with the matrix R_Z corresponding to the optimal instrument Z in Lemma 1. A new parametrization:

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

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 η_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 (12) 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 nonparametric 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,*

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

where all moments are conditional on $X = x_0$, evaluated at the true parameter value θ_0 , and matrix R is defined in Assumption A.2*.

The efficiency bound of Proposition 2 is not modified by post-multiplying 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 θ 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 θ is full-information identifiable (as in the CCAPM framework). Since θ can be estimated at a parametric rate using the marginal moment restrictions, it can be assumed known for the computation of the kernel nonparametric efficiency bound. This explains why the second term of the decomposition of the efficiency bound involving derivatives with respect to θ vanishes.

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

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

$$E_0(a|x_0) = E_0 \left[a(Y; \theta_0) - \text{Cov}_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} g_2(Y; \theta_0) \mid x_0 \right].$$

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 given 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:*

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

where all moments are conditional on $X = x_0$ and evaluated at θ_0 .

This is the formula in Proposition 2 with $g_2 = \tilde{g}$ and $R = Id$, since no linear combination of parameter θ 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 criterion (KLIC) based approach.

In this section, we develop an information based equivalent of the XMM estimator. The goal is to correct a structural drawback of the XMM approach as introduced in the previous section, namely a 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 existing literature on information based estimation considers a setting with uniform moment restrictions, and assumes the full-information identifiability of parameter θ . 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 discrepancy 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:

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

where $\tilde{\theta}$ is an estimator of θ and \tilde{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

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

which can be written as:

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

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 θ depends on the moment of interest a .

Therefore, it is important to introduce a corrected estimation method, which is both coherent and kernel nonparametrically efficient.

3.2 Information based estimator

The (unconstrained) kernel estimator $\hat{f}(y|x)$ is a consistent estimator of the conditional pdf. However, it is not 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 $\hat{f}(y|x)$, and satisfies the moment restrictions.

In this section, we consider the joint estimator defined by:

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

The objective function includes the two following 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 a Kullback-Leibler information criterion is used for the conditioning value x_0 corresponding to the conditional moment of interest. This second component corresponds to an "empirical likelihood"-type approach applied to the distribution conditional on x_0 . 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 θ without ensuring positivity. Therefore, the objective function is easily concentrated with respect to f^1, \dots, f^T . Next, the information criterion provides a solution $\hat{f}_0(\cdot|x_0)$ satisfying the unit mass and positivity restrictions⁷. In particular, the computation of the estimator only involves the optimization of a concentrated criterion with respect to parameter θ and a Lagrange multiplier of dimension $\dim(g_2)$ [see Appendix 2 for the concentration of the objective function].

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

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

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

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:

$$\begin{aligned} (\hat{f}_0(\cdot|x_0), \hat{\theta}) &= \arg \min_{f, \theta} \int \log \left[\frac{f(y)}{\hat{f}(y|x_0)} \right] f(y) dy, \\ \text{s.t.} \quad &\int f(y) dy = 1, \int f(y) \tilde{g}(y; \theta) dy = 0. \end{aligned} \tag{15}$$

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 θ and the

⁷See e.g. Stutzer (1995), (1996), and Kitamura-Stutzer (1997).

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 θ only, and adopt a smooth empirical likelihood approach. Smith (2004) establishes the results of Kitamura, Tripathi, Ahn (2004) for the Generalized Empirical Likelihood (GEL) methodology⁸. 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 nonparametrically efficient estimator of the moment of interest can also be derived in a two-step approach. Indeed, the structural parameter θ can be estimated consistently and efficiently by means of the uniform restrictions only. This allows to separate the estimation of θ , 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:

$$\begin{aligned} \widehat{f}_0(\cdot|x_0) &= \arg \min_f \int \log[f(y)/\widehat{f}(y|x_0)] f(y) dy, \\ \text{s.t.} \quad &\int f(y) dy = 1, \int f(y) g_2(y; \widehat{\theta}) dy = 0, \end{aligned} \quad (16)$$

and $\widehat{\theta}$ is any estimator of θ 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 implies no 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 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 data. When the parameter is finite dimensional, it is usually estimated by maximum likelihood⁹. Alternatively the

⁸See also Ai-Chen (2003).

⁹This approach can be extended to functional parameters, leading for instance to indirect spline estimation of the state price density [see e.g. Gouriéroux, Monfort (2005)].

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 date, 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 assuming a fixed set of liquid derivatives [see e.g. Duan (1994), 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 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 $r_t = \log(p_t/p_{t-1})$. We assume that the state variable process (Y_t) is a Markov process of order one under the historical probability, including r_t as its first component. The state process is assumed observable for both the investor (for pricing formula) and the econometrician (for estimation by XMM). Then, the price at t of a European call with moneyness strike s and residual maturity one can be written as¹⁰ :

$$c_t(s) = E \left[m(\underline{Y}_{t+1}; \theta) (\exp r_{t+1} - s)^+ \mid Y_t \right], \quad (17)$$

where $m(\underline{Y}_{t+1}; \theta)$ is the path dependent stochastic discount factor. The finite dimensional parameter θ 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 (the current date, say), 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:

¹⁰More precisely, this is the call price in percentage of the underlying asset price. Since $(p_{t+1} - sp_t)^+ = p_t(\exp r_{t+1} - s)^+$, the call or put written on p_{t+1} can also be written on $\exp r_{t+1}$.

$$c_{t_0}(s_k) = E \left[m(\underline{Y}_{t+1}; \theta) (\exp r_{t+1} - s_k)^+ | Y_t = y_{t_0} \right], \quad k = 1, \dots, K. \quad (18)$$

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

$$\begin{aligned} E[m(\underline{Y}_{t+1}; \theta) | Y_t = y_t] &= 1, \quad \forall y_t, \\ E[m(\underline{Y}_{t+1}; \theta) \exp r_{t+1} | Y_t = y_t] &= 1, \quad \forall y_t, \end{aligned} \quad (19)$$

respectively, assuming for simplicity a zero riskfree rate.

The second subset of constraints on θ are uniform with respect to the conditioning value, whereas the conditioning value is fixed in the first subset. The distinction between both 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. This is the case for a derivative with given moneyness strike and given residual maturity, due to periodic issuing in option markets¹¹. For expository purpose, we have considered above the simplest case, where option prices are observed at date t_0 only. The method might be improved by introducing derivative prices observed at several dates different from t_0 , and the extension to this case is straightforward. In particular, the most informative dates are not necessarily dates $t_0 - 1, t_0 - 2, \dots$ close to t_0 , but the dates corresponding to an environment $y_t \sim y_{t_0}$. However, what is important to realize is that it is not possible to specify a moment condition for a derivative, which is valid at *all* dates, since the number and the characteristics of highly traded derivatives, that are, time-to-maturity and moneyness strike, are different on consecutive days, especially due to the effect of the periodic issuing. Thus, moment conditions for observed derivative prices are necessarily local.

Different pricing formulas are derived below depending whether or not the uniform moment restrictions are totally taken into account.

4.2 Derivative pricing with limited-information

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

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

Under limited-information, the only restrictions $E(g_2|y_{t_0})$, which are taken into account, correspond to the same conditioning value. There is a set of $K + 2$ restrictions:

¹¹See the discussion in Aït-Sahalia, Lo (1998) for the evolution of the set of liquid options on S&P.

$$\begin{aligned}
E[m(\underline{Y}_{t+1}; \theta)(\exp r_{t+1} - s_k)^+ - c_{t_0}(s_k)|Y_t = y_{t_0}] &= 0, \quad k = 1, \dots, K, \\
E[m(\underline{Y}_{t+1}; \theta) - 1|Y_t = y_{t_0}] &= 0, \\
E[m(\underline{Y}_{t+1}; \theta) \exp r_{t+1} - 1|Y_t = y_{t_0}] &= 0.
\end{aligned} \tag{20}$$

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.

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 of the conditional (historical) pdf. In particular, the derivative prices are 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). Thus, 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 date of interest t_0 , or at dates t with $y_t \sim y_{t_0}$ [see e.g. Ait-Sahalia and Lo (1998) approach for comparison].

At a first sight, it may seem surprising to get consistent derivative prices, when only a small number of derivatives (3 in the Monte-Carlo study of Section 5) are observed. In fact, our approach can be seen as a method, which creates "artificial" observed derivative prices at well-chosen dates. More precisely, at any date t for which the conditioning variables are close to the current value $y_t \sim y_{t_0}$, say, we introduce the derivatives with the same strike and time-to-maturity, but with a price computed for y_{t_0} (instead of y_t). This approximation has no impact on consistency due to the continuity of the deterministic derivative price formula with respect to the conditioning variable.

Finally, the limited-information method differs from the entropy based approaches 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 (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 (18) and (19) are taken into account. Derivative pricing is 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 θ from underlying asset price dynamics.

i) Full-information identifiability

If parameter θ is identifiable from uniform moment restrictions (19), restrictions (18) can asymptotically be neglected for the estimation of θ . A GMM estimator $\hat{\theta}$ of θ can be computed by using restrictions (19) 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 (20).

ii) Full-information underidentifiability

As seen in Section 2.1, a part of the parameters can be identified from the underlying 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 (18), 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 (13), with moment function g corresponding to restrictions (19). In this case, the estimation has to be performed with the general criterion introduced in Section 3.2.

When θ is full-information underidentified, there exists a multiplicity of values of parameter θ , 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 (19)]. 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

Both approaches use jointly historical information [by means of the kernel estimate of the conditional pdf, and possibly by uniform moment restrictions (19)], and cross-sectional information by moment restrictions (18). Moreover, they are consistent when T tends to infinity with K fixed, whenever θ is identifiable from the whole set of uniform and local moment restrictions.

When θ 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. First, 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 may be an advantage from the financial point of view). Second, it is similar to the standard 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 θ 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 detect an instability

of the risk premium. Thus, we get a misspecification test for the hypothesis of constant θ parameter. For instance, the sdf specification might be deduced from a preference based interpretation with constant preference parameters. The above practice can allow to detect preference parameters depending on state variables, such as wealth level for instance [Ait-Sahalia and Lo (2000), Jackwerth (2000)].

Note, however, that the methodology differs from the usual implied Black-Scholes volatility in one respect. In our framework the varying parameter depends on y_t , whereas this dependence is not explicit in the Black-Scholes practice. As a consequence, if function $y \rightarrow \theta(y)$ is estimated, the pair including the historical pdf $f(y_t | y_{t-1})$ and the sdf $m[y_{t+1}, \theta(y_t)]$ satisfy the arbitrage-free restrictions. To summarize, the approach above can also be used to estimate a new pricing kernel, if the assumption of constant θ parameter is rejected.

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), Ball, Roma (1994), and Das, Sundaram (1999), with a risk premium introduced in the return equation¹². 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 kernel nonparametric efficiency bounds for limited- and mixed-information restrictions are computed in Section 5.3 for the prices of European calls. We discuss how these bounds 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; see Appendix 3 for the verification and discussion of primitive regularity conditions for XMM estimation in this stochastic volatility framework.

5.1 The design

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

$$r_t = \gamma\sigma_t^2 + \sigma_t\varepsilon_t, \quad (21)$$

where (ε_t) is a standard Gaussian white noise, σ_t^2 denotes the volatility, and γ 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 riskfree rate.

The volatility (σ_t^2) is stochastic, with a dynamic independent of the shocks (ε_t) on returns. It follows an autoregressive gamma process (ARG), which is the time discretized Cox-Ingersoll-Ross process [see Gouriéroux, Jasiak (2005)]. The transition distribution of

¹²See Gouriéroux, Sufana (2004) for the multivariate extension based on Wishart autoregressive process.

the stochastic volatility is characterized by the conditional Laplace transform (conditional moment generating function):

$$\Psi_t(u) = E [\exp (-u\sigma_{t+1}^2) | \sigma_t^2] = \exp [-a(u)\sigma_t^2 - b(u)], \quad (22)$$

where: $a(u) = \rho \frac{u}{1+cu}$, $b(u) = \delta \log(1 + cu)$. The positive parameter ρ is the first-order autocorrelation of the variance process (σ_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:

$$Y_t = (r_t, \sigma_t^2)'. \quad (23)$$

Model (21)-(23) is completed by a parametric specification of the stochastic discount factor for period $(t, t + 1)$. The sdf is specified as:

$$M_{t,t+1} = \exp (-\nu_0 - \nu_1\sigma_{t+1}^2 - \nu_2\sigma_t^2 - \nu_3r_{t+1}), \quad (24)$$

where $\nu_0, \nu_1, \nu_2, \nu_3$ are parameters. The exponential affine specification (24) is compatible with no-arbitrage restrictions and provides simple pricing formulas.

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

$$\begin{aligned} & \begin{cases} E_t(M_{t,t+1}) = 1, \\ E_t(M_{t,t+1} \exp r_{t+1}) = 1, \end{cases} \\ \iff & \begin{cases} E_t \exp [-\nu_0 - \nu_1\sigma_{t+1}^2 - \nu_2\sigma_t^2 - \nu_3r_{t+1}] = 1, \\ E_t \exp [-\nu_0 - \nu_1\sigma_{t+1}^2 - \nu_2\sigma_t^2 - (\nu_3 - 1)r_{t+1}] = 1, \end{cases} \\ \iff & \begin{cases} E_t \exp \left[-\nu_0 - \left(\nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2} \right) \sigma_{t+1}^2 - \nu_2\sigma_t^2 \right] = 1, \\ E_t \exp \left[-\nu_0 - \left(\nu_1 + (\nu_3 - 1)\gamma - \frac{(\nu_3 - 1)^2}{2} \right) \sigma_{t+1}^2 - \nu_2\sigma_t^2 \right] = 1, \end{cases} \\ & \text{(by integrating } r_{t+1} \text{ conditional on } \sigma_{t+1}^2) \\ \iff & \begin{cases} \nu_0 + a \left(\nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2} \right) \sigma_t^2 + \nu_2\sigma_t^2 + b \left(\nu_1 + \nu_3\gamma - \frac{\nu_3^2}{2} \right) = 0, \\ \nu_0 + a \left[\nu_1 + (\nu_3 - 1)\gamma - \frac{(\nu_3 - 1)^2}{2} \right] \sigma_t^2 + \nu_2\sigma_t^2 + b \left[\nu_1 + (\nu_3 - 1)\gamma - \frac{(\nu_3 - 1)^2}{2} \right] = 0. \end{cases} \end{aligned} \quad (25)$$

Since the above conditions have to be satisfied for any admissible value of σ_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:

$$\begin{aligned}\nu_0 &= -b \left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2} \right) = -\delta \log [1 + c(\nu_1 + \gamma^2/2 - 1/8)], \\ \nu_2 &= -a \left(\nu_1 + \nu_3 \gamma - \frac{\nu_3^2}{2} \right) = -\rho \frac{\nu_1 + \gamma^2/2 - 1/8}{1 + c(\nu_1 + \gamma^2/2 - 1/8)}.\end{aligned}\quad (26)$$

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 ν_1 is unrestricted. 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 dynamic assumed for stochastic volatility, but is the general case when state variables Y_t follow an affine process. Indeed, in this case, the specification of a parametric exponential affine sdf does not select a unique pricing kernel.

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

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

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:

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

where $E^Q[\cdot]$ denotes expectation w.r.t. the risk-neutral probability and $BS(h, s, \sigma^2)$ denotes the Black-Scholes price of a European call with moneyness strike s , time-to-maturity h , and constant volatility σ^2 . The derivative price is easily computed by Monte-Carlo, since, under the risk-neutral probability, the returns still follow stochastic volatility model (21)-(22) with risk premium parameter $\gamma^* = -1/2$ and ARG volatility parameters (see Appendix B):

$$\rho^* = \frac{\rho}{[1 + c(\nu_1 + \gamma^2/2 - 1/8)]^2}, \quad \delta^* = \delta, \quad c^* = \frac{c}{1 + c(\nu_1 + \gamma^2/2 - 1/8)}.\quad (27)$$

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 volatility, and the pattern of the implied Black-Scholes volatility as function of the log-moneyness strike, respectively. The simulations are performed for the following set of values for the parameters¹³:

$\gamma = 0.5$	$\rho = 0.96$	$\delta = 3.60$	$c = 7.34 \cdot 10^{-7}$
$\nu_0 = 3.82 \cdot 10^{-2}$	$\nu_1 = -1.44 \cdot 10^4$	$\nu_2 = 1.40 \cdot 10^4$	$\nu_3 = 1$

The opposite sign of coefficients ν_1 and ν_2 corresponds to a volatility feedback effect on the sdf [see Bekaert, Wu (2000)]. This effect distinguishes the consequence of a shock on expected volatility, measured by $\nu_1\rho + \nu_2 > 0$, and the shock on the volatility surprise $\nu_1 < 0$.

[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. Moreover, the Black-Scholes implied volatility¹⁴ admits a smile. Since in this model stochastic volatility is independent of shocks on returns, the smile is symmetric as a function of the log-moneyness strike, and thus asymmetric as a function of the moneyness strike.

5.2 The observations and the model

In the next sections, we assume that the observations of the state variables 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 moneyness 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:

$$M_{t,t+1}(\theta) = \exp(-\nu_0 - \nu_1\sigma_{t+1}^2 - \nu_2\sigma_t^2 - \nu_3r_{t+1}),$$

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

ii) Full-information identifiability

¹³Historical parameters ρ , δ and c are such that the ARG volatility process matches the stationary mean, variance and first-order correlation of the discretely sampled CIR process estimated by Andersen, Benzoni and Lund (2002). Risk premium parameter ν_1 is such that the stationary mean of volatility process under the risk-neutral distribution matches the value corresponding to CIR estimates in Bakshi, Cao and Chen (1997). Both estimates refer to S&P 500 index over recent consistent sample periods.

¹⁴We select a time-to-maturity $h = 5$ for the European call and the relevant information at date t is the volatility σ_t^2 , whose value is set equal to the stationary mean $E(\sigma_t^2)$.

Let us now discuss the identifiability of parameter θ 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 θ can be identified, including parameter ν_3 . Therefore, in this model the structural parameter θ 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 ν_0, ν_1, ν_2 . Then, the degree of underidentification from asset dynamics is equal to 1. The null space N_0 defined in equation (13) has dimension 1, and is spanned by (see Appendix 3):

$$R = \begin{pmatrix} -\delta \frac{c}{1+c\nu_1} \\ 1 \\ -\rho \frac{1}{(1+c\nu_1)^2} \\ 0 \end{pmatrix} = \begin{pmatrix} -2.67 \cdot 10^{-6} \\ 1 \\ -0.981 \\ 0 \end{pmatrix}. \quad (28)$$

Component $R'\theta$ and parameters ν_0, ν_1, ν_2 are full-information unidentifiable. Typically, parameters ν_0, ν_1, ν_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 $\nu_1 = 0$, that is, no risk premium on volatility surprise. Under this restriction ν_0, ν_1, ν_2 become full-information identifiable from asset price dynamics, and the estimation problem is greatly simplified. This second approach is often followed in the financial literature, at risk of a misspecification in the identification restriction. Moreover, such an approach can have misleading consequences when considering the confidence interval for derivative prices, which will likely be too narrow.

The first approach is considered in this paper.

iii) Limited-information identifiability and time-to-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 time-to-maturity $h = 1$, only parameters $\nu_{0,t_0} = \nu_0 + \nu_2 \sigma_{t_0}^2$ and ν_1 can be identified from asset dynamics and observed derivative prices. This allows to identify the prices of derivatives with the same time-to-maturity 1, but not to identify the prices of derivatives with larger time-to-maturity. However, parameters ν_0 and ν_2 can be identified separately by means of observed prices of derivatives with time-to-maturity larger than 1. To summarize, when the structural parameter is limited-information underidentified, it can be necessary to use derivative prices with different times-to-maturity to estimate the derivative prices at all maturities. Note, however, that derivative

prices with large time-to-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 (σ_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 macro-finance literature on Consumption CAPM

Stock, Wright (2000) [see also Yogo (2004)] considered an application to asset pricing, in which the sdf is deduced from the optimisation of an expected CRRA utility function. The sdf is:

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

where C_t denotes the consumption¹⁵. In our framework both δ and γ 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 a lack of observations on derivative prices, and concern some additional risk premium parameters.

5.3 Kernel nonparametric efficiency bounds

Two cases are distinguished according to the type of information.

i) Limited-information

The cross-sectional restrictions are:

$$\begin{aligned} E[M_{t_0,t_0+1}(\theta) - 1 | y_{t_0}] &= 0, \\ E[M_{t_0,t_0+1}(\theta) \exp r_{t_0+1} - 1 | y_{t_0}] &= 0, \\ E[M_{t_0,t_0+1}(\theta) (\exp r_{t_0+1} - s)^+ - c_{t_0}(s) | y_{t_0}] &= 0, \quad s = 0.98, 1, 1.02. \end{aligned} \quad (29)$$

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

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

The identifiable parameters are $\nu_{0,t_0} = \nu_0 + \nu_2 \sigma_{t_0}^2$, ν_1 , ν_3 , but are sufficient to identify the conditional moments of interest, which have the same time-to-maturity. We provide in Figure 3 the kernel 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, with the current factor y_{t_0} corresponding to a

¹⁵In this framework, the price of the consumption good is $q_t = 1$, and the returns y_t have to be interpreted as real returns.

variance $\sigma_{t_0}^2$ equal to the stationary expectation $E(\sigma_t^2)$. This efficiency bound gives the statistical accuracy for the estimation of a call price based only on the informational content of no-arbitrage restrictions at date t_0 .

[Insert Figure 3: Kernel 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 with the standardization $\sqrt{w^2/Th_T^2 f_X(x_0)} =$

2. We adopt this standardization to illustrate the pattern of the kernel nonparametric efficiency bound as a function of the moneyness strike. For the relevant sample size T and bandwidth h_T , this pattern is the same, but the widths of the bands are much narrower (see the next section). For expository purpose, we consider symmetric confidence bands, which do not account for the positivity of derivative prices. These bands have to be truncated at zero to satisfy the positivity restriction, which gives asymmetric bands. However, in practice, when the bands are narrow, the truncation effect is negligible and arises only for large strikes. The width of the confidence interval for derivative price $E(a(s)|y_{t_0})$ depends on moneyness strike s . The interval is generally wider for almost at-the-money (ATM) options, whereas it is narrower when the derivative is deep in-the-money (ITM), or deep out-of-the-money (OTM). Indeed, for moneyness strikes approaching zero or infinity, the kernel nonparametric efficiency bound goes to zero, since the derivative price has to be equal to the underlying asset price or equal to zero, respectively, by no-arbitrage. Finally, 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 time-to-maturity, let us consider the kernel nonparametric efficiency bound for a European call with time-to-maturity $h = 40$ days. At date t_0 , the prices of three derivatives with the same time-to-maturity $h = 40$ and strikes $s = 0.9, 1, 1.1$, respectively, are assumed to be observed. In this case, the whole parameter θ is limited-information identifiable. The efficiency bound is displayed in Figure 4 below¹⁶.

[Insert Figure 4: Kernel nonparam. efficiency bound, limited information, time-to-maturity 40]

The confidence interval admits similar pattern, but is generally larger, compared to time-to-maturity $h = 1$.

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. Finally, lower and upper bounds for European call prices can be derived from some observed call prices with the same time-to-maturity. They correspond to the smallest and largest elements in the set of functions, which are positive, decreasing, convex, smaller than 1 and compatible with these prices. Due to the lack of liquidity, the number of observed derivative prices with the same time-to-maturity is generally small, and the difference between the upper and lower bounds

¹⁶The efficiency bound is computed according to Corollary 2 using similar moment restrictions as in (29). In particular, moment restrictions at horizon $h = 40$ are used for bond, underlying asset, and derivatives.

very large (for instance, it can be equal to 100% if this number is less or equal to 1). As for positivity discussed earlier, the unconstrained confidence bands are rather narrow in practice, and these lower and upper bounds constraints can be neglected.

ii) Mixed limited- and full-information

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 (28). The kernel nonparametric efficiency bound for a European call with time-to-maturity $h = 1$ is displayed in Figure 5.

[Insert Figure 5: Kernel nonparametric efficiency bound, mixed-information]

The confidence interval is very close to that obtained in the limited-information framework (see Figure 3). In this example, the additional information, which is contained in the no-arbitrage restrictions for bond and underlying asset at dates different from t_0 , is not relevant for estimation of derivative prices at t_0 . Similarly, in Figure 6 we display the mixed-information kernel nonparametric efficiency bound for a call option with time-to-maturity $h = 40$, when the price of three derivatives with the same time-to-maturity and strikes $s = 0.9, 1, 1.1$, respectively, are observed.

[Insert Figure 6: Kernel nonparam. efficiency bound, mixed-information, time-to-maturity 40]

Also in this case, the confidence band with mixed-information is very close to the one with limited-information (see Figure 4).

The situation is different in Figure 7, where we display limited- and mixed-information kernel nonparametric efficiency bounds for a European call option with time-to-maturity $h = 40$, when only two derivative prices are observed, namely for moneyness strike $s = 1$ and 1.1, and time-to-maturity $h = 40$.

[Insert Figure 7: Kernel nonparametric efficiency bound, mixed- and limited-information]

The mixed-information confidence band is narrower than the limited information one for ITM calls. Compared to Figures 4 and 6, the fact that the derivative price for moneyness strike 0.9 is not observed implies a widening of the confidence band in this moneyness region. This effect is less pronounced for mixed-information, due to the contribution of uniform no-arbitrage restrictions. In general, we expect that the uniform restrictions on bond and underlying asset are relevant to reduce the confidence bands in moneyness regions where few derivative prices are observed.

Finally, in Figure 8 we display the limited- and mixed-information kernel nonparametric efficiency bound for a European call with time-to-maturity $h = 40$, when the prices of three derivatives with time-to-maturity 20 and strikes $s = 0.9, 1, 1.1$, respectively, are observed.

[Insert Figure 8: Kernel nonparametric efficiency bound, time-to-maturity 40 and 20]

The maturity of the observed derivatives does not correspond with the maturity of interest. This explains why the kernel nonparametric efficiency bound is much larger compared with Figures 4 and 6, and, in particular, it is different from zero for all moneyness strikes¹⁷. Thus, observed derivative prices at the maturity of interest have a large informational content for the estimation of other derivative prices. Finally, note that in Figure 8 limited- and mixed-information confidence bands differ.

5.4 Monte-Carlo

In this section, we report the results of a Monte-Carlo experiment 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 time-to-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 $S = 2500$ paths of return-volatility process (r_t, σ_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, σ_t^2) follows the same stochastic volatility process in direct and in reversed time. The information based estimator of structural parameter θ_0 and of European call prices $E[a(h, s)|y_{t_0}]$ at date t_0 for time-to-maturity $h = 2$, and different moneyness 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 $h_{r,T} = 0.0031$ and $h_{\sigma,T} = 0.0014$, respectively¹⁸. Finally, the selected sample size $T = 250$ corresponds to about 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, median, 95% confidence interval, as well as the 5% and 95% quantiles of the estimated European call prices for time-to-maturity $h = 2$ and different values of the moneyness strike between $s = 0.96$ and $s = 1.04$.

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

For comparison, we also report for each moneyness strike the corresponding true derivative price, and the 95% asymptotic confidence interval based on the kernel nonparametric efficiency bound, computed according to Section 5.3. As seen previously in Figures 5 and 6,

¹⁷In this case, the bounds based on smallest and largest decreasing convex functions is degenerate and equal to 100%.

¹⁸The bandwidths are selected in order to get an appropriate smoothing of the joint pdf of (r_t, σ_t) at sample size $T = 250$.

the width of the 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 0.5% of the true price for strike $s = 0.97$, whereas it amounts to about 4% for strike $s = 0.99$ and more than 300% for strike $s = 1.03$. Finally, the truncation effect due to positivity of call prices arises only for deep OTM strikes $s = 1.03$ and $s = 1.04$, whereas it is negligible for all other strikes.

Let us first consider the finite sample bias of estimated derivative prices. This bias is positive for call options close to the money ($s = 0.99$ and $s = 1.01$), whereas prices of deep ITM ($s = 0.97$) and deep OTM ($s = 1.03$) calls are generally underestimated. The corresponding relative pricing errors are less than 0.1% for ITM strikes, whereas they are about 1% for strike $s = 1.01$, and about 20% for the deep OTM strike $s = 1.03$.

Let us now consider the finite sample accuracy of the estimated derivative prices. The 95% confidence intervals and the 95% - 5% interquantile ranges of estimated call prices feature patterns across strikes similar to the patterns of the kernel nonparametric efficiency bound, but they are somewhat wider (for strikes close to the money). For instance, the 95% - 5% interquantile range is about 5% of the median (or mean) call price for strike $s = 0.99$, and about 400% of the mean for the deep OTM strike $s = 1.03$ ¹⁹. In particular, these bounds are much larger than those typically reported in the literature based on fully parametric specifications. A narrow parametric bound 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 kernel 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 9 the histograms of estimated derivative prices for different strikes.

[Figure 9: Histograms of estimated derivative prices]

These finite sample distributions feature some non-Gaussian patterns. The non-Gaussian patterns are more pronounced for deep ITM and OTM strikes, for which right skewed and fat tailed distributions are observed, whereas for strikes not far away from the money the distributions are closer to normality²⁰.

ii) Structural parameter

Although the focus of this paper is on estimation of conditional moments corresponding to derivative prices, it is interesting to consider also the results for the estimator of structural parameter θ . In Table 2, we display the mean, median, standard deviation and 95% - 5% interquantile range of estimator $\hat{\theta}$.

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

The estimators of the different components are rather biased for sample size $T = 250$. For parameters ν_0 , ν_1 and ν_2 , the bias is such that average estimates are larger, in absolute

¹⁹For strikes $s = 0.97$ and $s = 1.03$, the interquantile range is highly skewed, with a median very close to the lower bound (see also Figure 9 below).

²⁰Note however the different scales of the horizontal axis in the four panels.

value, than true parameters. The converse is true for median estimates. The estimators feature large standard deviation and rather wide interquantile ranges. These results are confirmed by the histograms of the estimates, which are displayed in Figure 10.

[Figure 10: Histograms of estimated structural parameters]

The finite sample distributions of parameters ν_0, ν_1, ν_2 are highly non-normal, in particular skewed to the right for ν_0 and ν_2 (resp. to the left for ν_1), with fat tails. The distribution of parameter ν_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, that are the parametric rate $T^{1/2}$ for ν_3 , and the nonparametric rate $(Th_{r,T}h_{\sigma,T})^{1/2}$ for ν_0, ν_1, ν_2 , respectively. These different rates of convergence are a consequence of market incompleteness, which cause parameters ν_0, ν_1, ν_2 related to the risk premium for stochastic volatility to be full-information non-identifiable.

6 Concluding remarks

The aim of this paper is 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. The difficulty in applying the GMM in such a framework is due to two fundamental market characteristics, that are, lack of liquidity and incompleteness. On the one hand, the lack of liquidity of some assets, such as a derivative with given moneyness strike and time-to-maturity, implies that the corresponding no-arbitrage moment restrictions are local, that is, valid for a given value of the conditioning variable only, instead of uniform, as for liquid assets. On the other hand, the market incompleteness implies that risk premium parameters are not necessarily identifiable from historical data on the prices of liquid underlying assets, and that some of them can only be deduced from less frequent cross-sectional observations on derivative prices.

These difficulties of the standard GMM are solved by the Extended Method of Moments (XMM), which explains how to appropriately account for both uniform and local moment restrictions from liquid and less liquid assets, when the risk premia parameters are possibly full-information unidentifiable. The XMM approach allows for efficient estimation of derivative prices, and for consistent estimation of risk premia, even if the number of observed derivative prices is small.

Lack of liquidity and market incompleteness show up in some non-standard properties of the XMM approach. The inclusion of both local and uniform moment restrictions implies different rates of convergence for the different risk premia parameters. In particular, these rates of convergence are nonparametric, when identification is ensured only by the infrequent cross-sectional observations on derivative prices. Moreover, the confidence bands for estimated derivative prices are much wider than usually reported in both the theoretical and applied literature.

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., and A., Lo (1998): "Nonparametric Estimation of State Price Densities Implicit in Financial Asset Prices", *Journal of Finance*, 53, 499-547.
- Aït-Sahalia, Y., and A., Lo (2000): "Nonparametric Risk Management and Implied Risk Aversion", *Journal of Econometrics*, 94, 9-51.
- 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.
- Andersen, T., Benzoni, L., and J., Lund (2002): "An Empirical Investigation of Continuous-Time Models for Equity Returns", *Journal of Finance*, 57, 1239-1284.
- Andrews, D. (1991): "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables", *Journal of Multivariate Analysis*, 38, 187-203.
- Andrews, D. (1994): "Empirical Process Methods in Econometrics", in *Handbook of Econometrics*, Vol. 4, R., Engle and D., McFadden, Eds, Amsterdam, North Holland, 2247-2294.
- Andrews, D., and V., Marmer (2005): "Exactly Distribution Free Inference in Instrumental Variables Regressions with Possibly Weak Instruments", Cowles Foundation DP.
- Andrews, D., and J., Stock (2005): "Inference with Weak Instruments", Yale University DP.
- 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.
- Bakshi, G., Cao, C., and Z., Chen (1997): "Empirical Performance of Alternative Option Pricing Models", *Journal of Finance*, 52, 2003-2049.
- 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.
- Bekaert, G., and G., Wu (2000): "Asymmetric Volatility and Risk in Equity Markets", *Review of Financial Studies*, 13, 1-42.
- Bickel, P., Klaassen, A., Ritov, Y., and J., Wellner (1993): *Efficient and Adaptive Estimation for Semi-Parametric Models*. The John Hopkins University Press, Baltimore, MD.
- Billingsley, P. (1965): *Ergodic Theory and Information*. John Wiley, New York.

- Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley, New York.
- 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.
- Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Springer, New-York.
- Bosq, D., and J.-P., Lecoutre (1987): *Théorie de l'estimation fonctionnelle*. Economica. Paris.
- 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, Princeton, NJ.
- Cressie, N., and T., Read (1984): "Multinomial Goodness of Fit Tests", *Journal of the Royal Statistical Society*, 8, 46, 440-464.
- Das, S., and R., Sundaram (1999): "Of Smiles and Smirks: A Term Structure Perspective", *Journal of Financial and Quantitative Analysis*, 34, 211-239.
- 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.
- Donoho, D., and R., Liu, (1991): "Geometrizing Rate of Convergence II", *Annals of Statistics*, 19, 633-667.
- Duan, J. (1994): "Maximum Likelihood Estimation Using Price Data of the Derivative Contract", *Mathematical Finance*, 4, 155-167.
- 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 (2005): "Autoregressive Gamma Process", forthcoming *Journal of Forecasting*.
- Gourieroux, C., and A., Monfort (2005): "Pricing with Splines", forthcoming *Annales d'Economie et de Statistique*.
- Gourieroux, C., and R., Sufana (2004): "Derivative Pricing with Multivariate Stochastic Volatility", HEC Montreal DP.
- Hahn, J., Hausman, J., and G., Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations", *Econometrics Journal*, 7, 272-306.
- Hansen, B. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays", *Econometric Theory*, 12, 347-359.

Hansen, C., Hausman, J., and W., Newey (2004): "Many Instruments, Weak Instruments, and Microeconomic Practice", Working Paper, MIT.

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.

Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability*, 12, 141-153.

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. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns", *Review of Financial Studies*, 13, 433-451.

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.

Newey, W., and F., Windmeijer (2005): "Many Weak Moment Asymptotics for Generalized Empirical Likelihood Estimators", Working Paper, MIT.

- Owen, A. (1991): "Empirical Likelihood for Linear Models", *Annals of Statistics*, 19, 1725-1747.
- Owen, A. (2001): *Empirical Likelihood*. Chapman and Hall, New-York.
- Pollard, D. (1990): *Empirical Processes: Theory and Applications*. CBMS Conference Series in Probability and Statistics, Vol. 2, Hayward, CA.
- Potscher, B., and I., Prucha (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes", *Econometrica*, 57, 675-683.
- Qin, J., and L., Lawless (1994): "Empirical Likelihood and General Estimating Equations", *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.
- Staiger, D., and J., Stock (1997): "Instrumental Variables with Weak Instruments", *Econometrica*, 65, 557-586.
- 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., 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.
- Tenreiro, C. (1995): "Estimation fonctionnelle: Applications aux tests d'adéquation et de paramètre constant", Ph.D. Thesis, University of Coimbra.
- Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak", *Review of Economic Studies*, 86, 797-810.

Table 1

Maturity: h=2, Sample size: T=250					
Derivative price $\beta = E(a(s) y_{t_0}) (\times 100)$					
Strike $s = 0.96$			Strike $s = 0.97$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
4.000	3.999 – 4.002		3.002	2.997 – 3.007	
Mean	95% confidence interval		Mean	95% confidence interval	
4.000	3.999 – 4.001		3.001	2.997 – 3.006	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
4.000	4.000	4.001	3.000	2.999	3.006
Strike $s = 0.99$			Strike $s = 1.01$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
1.125	1.103 – 1.146		0.128	0.107 – 0.149	
Mean	95% confidence interval		Mean	95% confidence interval	
1.126	1.092 – 1.159		0.129	0.097 – 0.162	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
1.125	1.099	1.153	0.129	0.104	0.155
Strike $s = 1.03$			Strike $s = 1.04$		
True price	Nonparametric Bound		True price	Nonparametric Bound	
0.003	–0.003 – 0.008		0.000	–0.002 – 0.002	
Mean	95% confidence interval		Mean	95% confidence interval	
0.002	–0.003 – 0.007		0.000	–0.001 – 0.001	
Median	0.05-quant.	0.95-quant.	Median	0.05-quant.	0.95-quant.
0.001	0.000	0.008	0.000	0.000	0.001

Table 2

Maturity: h=2, Sample size: T=250					
Parameter θ					
	True value	Mean	Median	Stand. dev.	5% / 95% quant.
ν_0	0.038	0.072	0.031	0.117	–0.014 / 0.293
$\nu_1 (\times 10^{-4})$	–1.440	–1.829	–0.930	2.611	–6.546 / 0.353
$\nu_2 (\times 10^{-4})$	1.397	1.767	0.872	2.522	–0.355 / 6.331
$\nu_3 (\times 10^{-2})$	0.010	0.008	0.006	0.099	–0.154 / 0.173

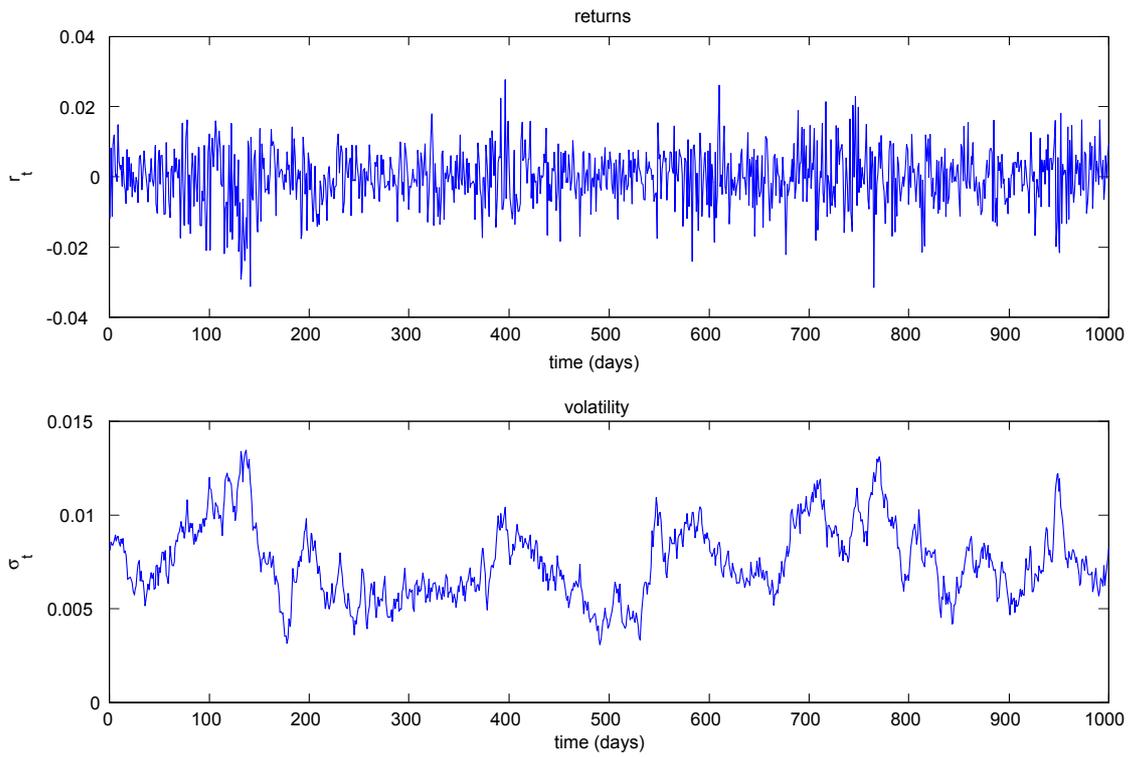


Figure 1: Simulated series of return r_t (upper Panel) and volatility series σ_t (lower Panel) for the ARG stochastic volatility process.

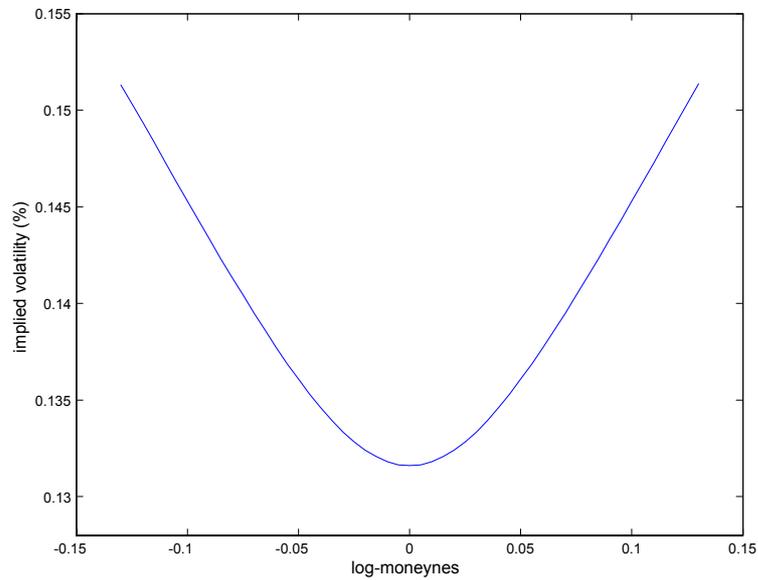


Figure 2: Implied Black-Scholes volatility (annualized %) as a function of the log-moneyness strike $\log s$ for a European call with time-to-maturity $h = 5$.

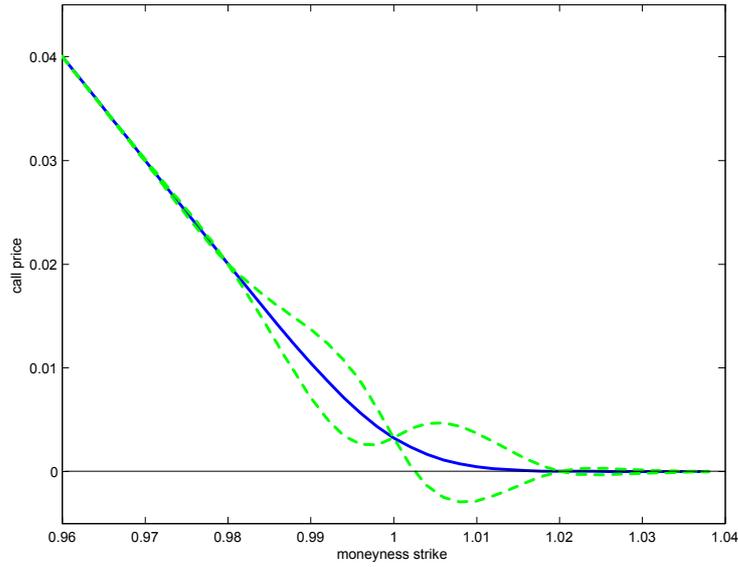


Figure 3: Limited information kernel nonparametric efficiency bound for a European call with time-to-maturity 1. Derivative prices for moneyness strikes 0.98, 1, 1.02 and time-to-maturity 1 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed lines to 95% symmetric pointwise confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

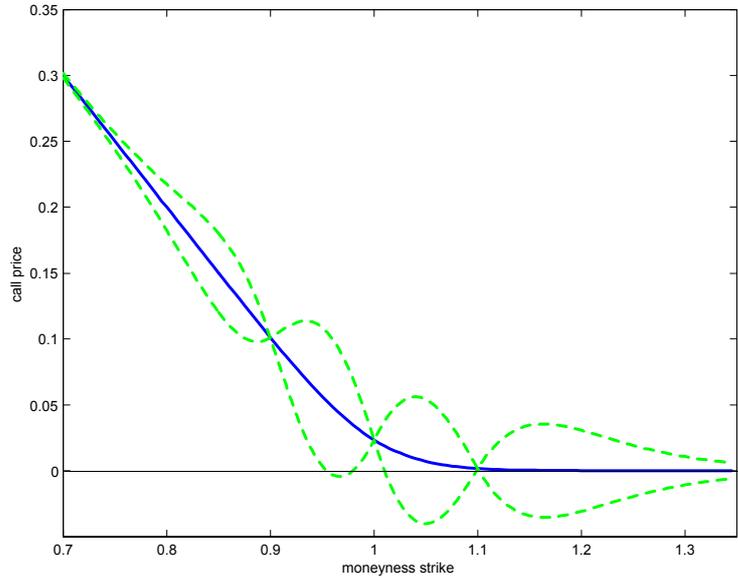


Figure 4: Limited information kernel nonparametric efficiency bound for a European call with time-to-maturity 40. Derivative prices for moneyness strikes 0.9, 1, 1.1 and time-to-maturity 40 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed lines to pointwise 95% symmetric confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

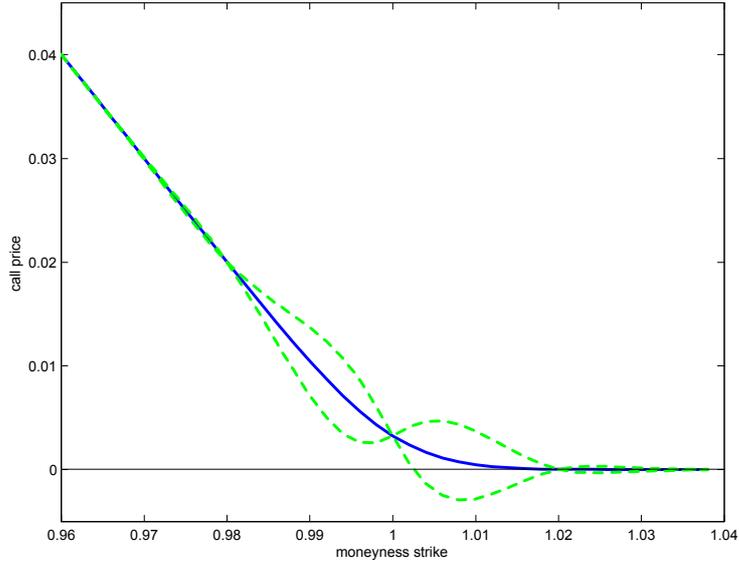


Figure 5: Mixed information kernel nonparametric efficiency bound for a European call with time-to-maturity 1. Derivative prices for moneyness strikes 0.98, 1, 1.02 and time-to-maturity 1 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed lines to pointwise 95% symmetric confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

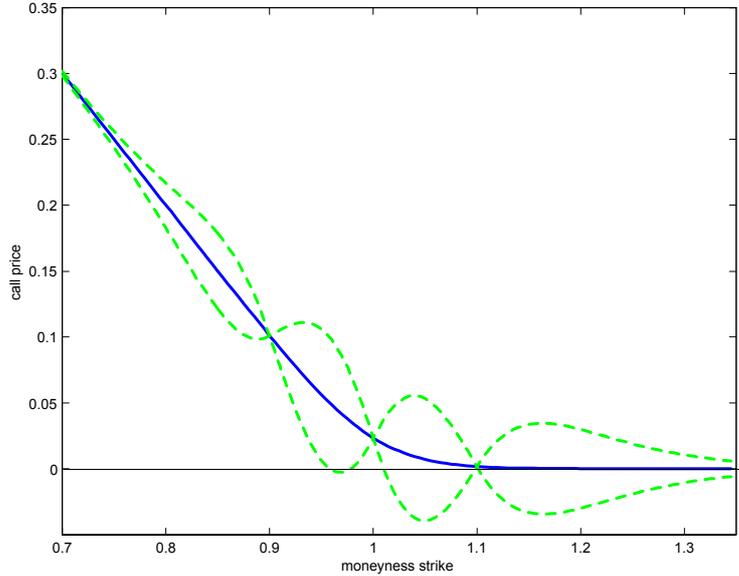


Figure 6: Mixed information kernel nonparametric efficiency bound for a European call with time-to-maturity 40. Derivative prices for moneyness strikes 0.9, 1, 1.1 and time-to-maturity 40 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed lines to pointwise 95% symmetric confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

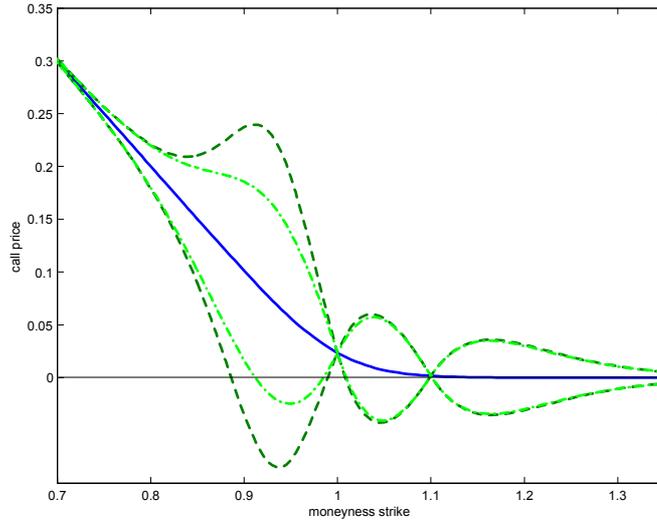


Figure 7: Limited information (dashed line) and mixed information (dash-dotted line) kernel nonparametric efficiency bound for a European call with time-to-maturity 40. Derivative prices for moneyness strikes 1, 1.1 and time-to-maturity 40 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed and dash-dotted lines to pointwise 95% symmetric confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

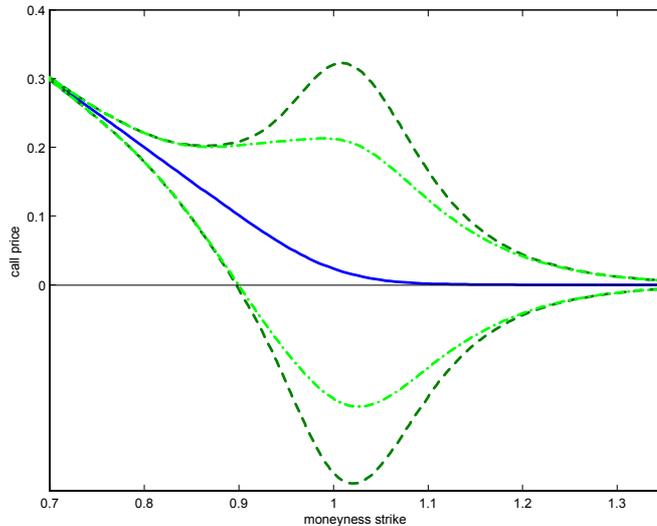


Figure 8: Limited information (dashed line) and mixed information (dash-dotted line) kernel nonparametric efficiency bound for a European call with time-to-maturity 40. Derivative prices for moneyness strikes 0.9, 1, 1.1 and time-to-maturity 20 are observed. The solid line corresponds to the price $E(a(s)|y_{t_0})$, the dashed and dash-dotted lines to pointwise 95% confidence intervals $E(a(s)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, s)^{1/2}$.

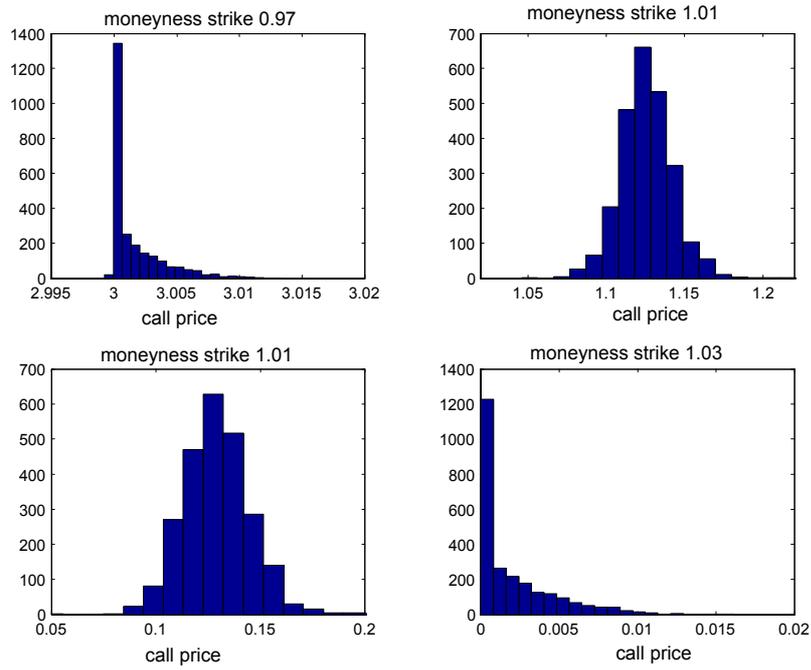


Figure 9: Information based XMM estimator: histograms of estimated derivative prices.

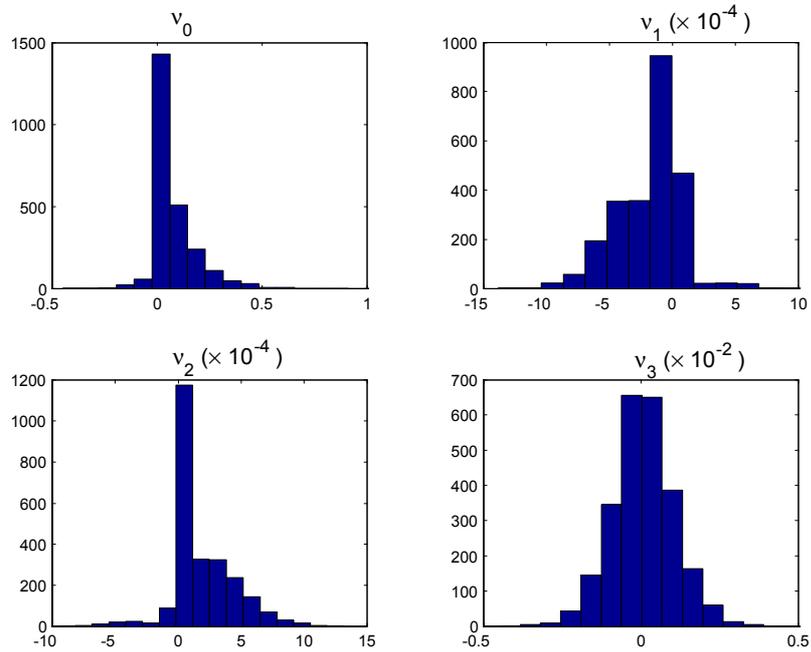


Figure 10: Information based XMM estimator: histograms of estimated structural parameters.

APPENDIX 1

Asymptotic properties of XMM estimator

In this Appendix, we derive the asymptotic properties of the (kernel) moment estimator $\widehat{\theta}_T^*$ for given weighting matrix Ω and instruments Z . Let $\Theta \subset \mathbb{R}^p$, $B \subset \mathbb{R}^L$ be compact sets. The kernel moment estimator is defined by:

$$\widehat{\theta}_T^* = \arg \min_{\theta^* = (\theta', \beta')' \in \Theta \times B} Q_T(\theta^*) = \widehat{g}_T(\theta^*)' \Omega \widehat{g}_T(\theta^*),$$

where

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

and $g_1(Y, X; \theta) = Z \cdot g(y; \theta)$, $g_2 = (g', \widetilde{g}')$. Due to the different rates of convergence of the empirical moments in $\widehat{g}_T(\theta^*)$, it is not possible to use the standard approach for GMM framework to derive the asymptotic properties of $\widehat{\theta}_T^*$. For instance, to prove consistency, we cannot rely on a.s. uniform convergence of criterion Q_T to some limit deterministic criterion. Indeed, after dividing Q_T by T , the part of the criterion involving conditional moment restrictions is asymptotically negligible. Therefore, the limit criterion would only take into account marginal moment restrictions, and could not allow for identifying parameter θ in the full-information underidentified case. To prove consistency, we follow an alternative approach relying on empirical process methods [see Stock, Wright (2000) for a similar approach].

Let us introduce the vector of standardized theoretical moments:

$$m_T(\theta^*) = \left(\sqrt{T} E_0 [g_1(Y_t, X_t; \theta)]', \sqrt{Th_T^d} E_0 [g_2(Y_t; \theta) | x_0]', \sqrt{Th_T^d} E_0 [a(Y_t; \theta) - \beta | x_0]' \right)',$$

and define the associated empirical process:

$$\Psi_T(\theta) = \widehat{g}_T(\theta^*) - m_T(\theta^*) \equiv T^{-1/2} \sum_{t=1}^T g_{t,T}(\theta), \quad \theta \in \Theta.$$

Indeed, due to the linearity of \widehat{g}_T w.r.t. β , the empirical process Ψ_T depends on parameter θ , but not on parameter β .¹

In this Appendix we will use the following notation. $L^2(F_Y)$ denotes the Hilbert space of real-valued functions, which are square integrable w.r.t. the distribution F_Y of r.v. Y , and $\|\cdot\|_{L^2(F_Y)}$ is the corresponding L^2 -norm. Linear space $L^p(\mathcal{X})$, $p > 0$, of p -integrable functions w.r.t. Lebesgue measure λ on set \mathcal{X} is defined similarly. For matrix A , $\|A\|$ denotes matrix norm $\|A\| = [Tr(AA')]^{1/2}$. In particular, when A is a vector, $\|A\|$ is the standard Euclidean norm $\|A\| = (A'A)^{1/2}$. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and vector $y \in \mathbb{R}^d$, we set $|\alpha| := \sum_{i=1}^d \alpha_i$, $y^\alpha := y_1^{\alpha_1} \cdot \dots \cdot y_d^{\alpha_d}$, and $\partial^{|\alpha|} f / \partial y^\alpha := \partial^{|\alpha|} f / \partial y_1^{\alpha_1} \dots \partial y_d^{\alpha_d}$. Symbol \implies denotes weak convergence

¹Compared to the standard definition of empirical process, triangular array $g_{t,T}(\theta)$ is not zero-mean, because of the bias term in the nonparametric component. However, it is shown in the proof of Lemma A.1 (see Appendix B), that this bias is asymptotically negligible under Assumption A.16 below, and process $\Psi_T(\theta)$ is asymptotically equivalent to a zero-mean process.

in the space of bounded real functions on Θ , equipped with the uniform metric [see e.g. Andrews (1994)]. Symbol $\|f\|_\infty$ denotes the sup-norm $\|f\|_\infty = \sup_{y \in \mathcal{Y}} \|f(y)\|$ of a function f defined on space \mathcal{Y} . We denote by $C^m(\mathcal{Y})$ the space of functions f on \mathcal{Y} , which are continuously differentiable up to order $m \in \mathbb{N}$, and $\|D^m f\| := \sum_{|\alpha|=m} \|\partial^{|\alpha|} f / \partial y^\alpha\|_\infty$. Finally, we denote by g_2^* the function $g_2^* = (g_2', a') = (g', \tilde{g}', a')$ [see equations (8) and (9) in the text].

A.1.1 Regularity assumptions

Let us introduce the following set of regularity conditions in addition to Assumptions A.1 and A.2 in the text:

Assumption A.3: *The parameter sets $\Theta \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^L$ are compact and the true parameter $\theta_0^* = (\theta_0', \beta_0')$ is in the interior of $\Theta \times B$.*

Assumption A.4: *The process $\left\{ (X_t', Y_t')' : t \in \mathbb{N} \right\}$ on $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^d$ is strictly stationary and geometric strong mixing.*

Assumption A.5: *The function $g_2^*(\cdot; \theta)$ is in $L^2(F_Y)$, for any $\theta \in \Theta$, where F_Y is the stationary cdf of Y_t . There exists a basis of functions $\{\psi_j : j \in \mathbb{N}\}$ in $L^2(F_Y)$, such that $\|\psi_j\|_{L^2(F_Y)} = 1$, $j \in \mathbb{N}$, and:*

$$g_2^*(y; \theta) = \sum_{j=1}^{\infty} c_j(\theta) \psi_j(y), \quad y \in \mathcal{Y},$$

for any $\theta \in \Theta$, where $\{c_j(\theta) : j \in \mathbb{N}\}$ is a sequence of coefficient vectors. Moreover, there exist $r > 2$ and a sequence $\{\lambda_j > 0 : j \in \mathbb{N}\}$, such that $\sum_{j=1}^{\infty} \lambda_j < \infty$, and:

$$\sum_{j=1}^{\infty} \lambda_j \left(E_0 \left[\|Z_t \psi_j(Y_t)\|^r \right]^{2/r} + E_0 \left[\psi_j(Y_t)^2 | X_t = x_0 \right] \right) < \infty,$$

$$\lim_{J \rightarrow \infty} \sup_{\theta \in \Theta} \sum_{j=J}^{\infty} \frac{1}{\lambda_j} \|c_j(\theta)\|^2 = 0.$$

Assumption A.6: *The matrices:*

$$S_0 = \lim_{T \rightarrow \infty} V_0 \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T g_1(Y_t, X_t; \theta_0) \right], \quad \Sigma_0 = V_0 [g_2^*(Y_t; \theta_0) | X_t = x_0],$$

exist and are positive definite.

Assumption A.7: *The stationary density f of X_t is in class $C^m(\mathcal{X})$ for some $m \in \mathbb{N}$, $m \geq 2$, is such that $\|f\|_\infty < \infty$ and $\|D^m f\|_\infty < \infty$.*

Assumption A.8: *For $l > 0$, the stationary density $f_{t,t+l} \equiv f_l$ of (X_t, X_{t+l}) is such that:*

$$\sup_{l > 0} \|f_l\|_\infty < \infty.$$

Moreover, for $t_1 < t_2 < t_3 < t_4$, the stationary density f_{t_1, t_2, t_3, t_4} of $(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})$ is such that:

$$\sup_{t_1 < t_2 < t_3 < t_4} \|f_{t_1, t_2, t_3, t_4}\|_\infty < \infty.$$

Assumption A.9: For any $\theta \in \Theta$, the function:

$$x \mapsto \varphi(x; \theta) = E[g_2^*(Y_t; \theta) | X_t = x] f(x),$$

is in class $C^m(\mathcal{X})$, such that:

$$\sup_{\theta \in \Theta} \|D^m \varphi(\cdot; \theta)\|_\infty < \infty.$$

Assumption A.10: For any $\theta, \tau \in \Theta$, the functions:

$$E[g_2^*(Y_t; \theta) | X_t = \cdot] f(\cdot) \quad , \quad E[g_2^*(Y_t; \theta) g_2^*(Y_t; \tau)' | X_t = \cdot] f(\cdot),$$

are continuous at $x = x_0$.

Assumption A.11: The instrument Z is given by $Z = H(X)$, where the function H defined on \mathcal{X} is continuous at $x = x_0$.

Assumption A.12: The mapping:

$$x \mapsto E \left[\sup_{\theta \in \Theta} \|g_2^*(Y_t; \theta)\|^2 | X_t = x \right] f(x),$$

is bounded. Moreover, there exists $\beta > 2$ such that:

$$E \left[\sup_{\theta \in \Theta} \|g_2^*(Y_t; \theta)\|^\beta \right] < \infty.$$

Assumption A.13: For any $\theta, \tau \in \Theta$:

$$\begin{aligned} \|E[g_2^*(Y_t; \theta) | X_t = \cdot] f(\cdot)\|_\infty &< \infty \\ \sup_{l > 0} \left\| E \left[g_2^*(Y_t; \theta) g_2^*(Y_{t-l}; \tau)' | X_t = \cdot, X_{t-l} = \cdot \right] f_l(\cdot, \cdot) \right\|_\infty &< \infty. \end{aligned}$$

Assumption A.14: The moment function:

$$\theta \mapsto \left(E[g_1(Y_t, X_t; \theta)]', E[g_2^*(Y_t; \theta) | X_t = x_0]' \right)',$$

is continuous on Θ .

Assumption A.15: The weighting matrix Ω is positive definite.

Assumption A.16: The bandwidth h_T is such that:

$$T h_T^{d+2m} \rightarrow 0, \quad \text{as } T \rightarrow \infty,$$

and there exists $\alpha < 1/2 - 1/\beta$, where β is defined in Assumption A.12, such that:

$$T^\alpha h_T^d / \log T \rightarrow \infty, \quad \text{as } T \rightarrow \infty.$$

Assumption A.17: The kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Parzen kernel of order m , that is,
i) $K \geq 0$ and $\int_{\mathbb{R}^d} K(u) du = 1$,
ii) K is bounded, $\lim_{\|u\| \rightarrow \infty} \|u\|^d K(u) = 0$ and $w^2 := \int_{\mathbb{R}^d} K(u)^2 du < \infty$,
iii) $\int_{\mathbb{R}^d} u^\alpha K(u)^2 du = 0$ for any $\alpha \in \mathbb{N}^d$ such that $|\alpha| < m$, and:

$$\int_{\mathbb{R}^d} K(u) \|u\|^m dv < \infty.$$

Assumption A.18: The function:

$$x \mapsto \varphi_j(x) = E [\psi_j(Y_t)^2 | X_t = x] f(x),$$

is in class $C^2(\mathcal{X})$, for any $j \in \mathbb{N}$, such that:

$$\sup_{j \in \mathbb{N}} \|D^2 \varphi_j\|_\infty < \infty.$$

Assumption A.19: The following inequalities hold:

$$\begin{aligned} \sup_{j \in \mathbb{N}} \|E [\psi_j(Y_t) | X_t = \cdot] f(\cdot)\|_\infty &< \infty, \\ \sup_{j \in \mathbb{N}} \sup_{l > 0} \|E [\psi_j(Y_t) \psi_j(Y_{t-l}) | X_t = \cdot, X_{t-l} = \cdot] f_l(\cdot, \cdot)\|_\infty &< \infty, \\ \sup_{j \in \mathbb{N}} \|E [|\psi_j(Y_t)|^r | X_t = \cdot] f(\cdot)\|_\infty &< \infty, \end{aligned}$$

where r is defined in Assumption A.5.

Assumption A.20: For any $\theta \in \Theta$:

$$E [\|g_1(Y_t, X_t; \theta)\|^4] < \infty, \quad E [\|g_2^*(Y_t; \theta)\|^4] < \infty.$$

Assumption A.21: For any $\theta \in \Theta$,

$$\begin{aligned} \sup_{t_1 \leq t_2 \leq t_3 \leq t_4} \|E [&\|g_2^*(Y_{t_1}; \theta)\| \|g_2^*(Y_{t_2}; \theta)\| \|g_2^*(Y_{t_3}; \theta)\| \|g_2^*(Y_{t_4}; \theta)\| \\ &| X_{t_1} = \cdot, X_{t_2} = \cdot, X_{t_3} = \cdot, X_{t_4} = \cdot] f_{t_1, t_2, t_3, t_4}(\cdot, \cdot, \cdot, \cdot)\|_\infty &< \infty. \end{aligned}$$

Assumption A.22: Function $g_2^*(y; \theta)$ is twice continuously differentiable w.r.t. (y, θ) .

Assumption A.23: There exist $\delta, \gamma > 1$ and $\tau > 2$, such that:

$$\begin{aligned} E \left[\left\| \frac{\partial g_1}{\partial \theta'} (Y_t, X_t; \theta_0) \right\|^\tau \right] < \infty, \quad E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial g_1}{\partial \theta'} (Y_t, X_t; \theta) \right\|^\delta \right] < \infty, \\ E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 g_1}{\partial \theta_i \partial \theta_j} (Y_t, X_t; \theta) \right\|^\gamma \right] < \infty, \quad i, j = 1, \dots, p. \end{aligned}$$

Assumption A.24: The mapping:

$$x \mapsto E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial g_2^*}{\partial \theta'} (Y_t; \theta) \right\|^2 | X_t = x \right] f(x),$$

is bounded. Moreover,

$$E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial g_2^*}{\partial \theta} (Y_t; \theta) \right\|^\beta \right] < \infty.$$

for $\beta > 2$ defined in Assumption A.12.

Assumption A.25: *Functions:*

$$\begin{aligned} \theta &\longmapsto \left(E \left[\frac{\partial g_1'}{\partial \theta} (Y_t, X_t; \theta) \right], E \left[\frac{\partial g_2^*'}{\partial \theta} (Y_t; \theta) | X_t = x_0 \right] \right), \\ \theta &\longmapsto E \left[\frac{\partial^2 g_1}{\partial \theta_i \partial \theta_j} (Y_t, X_t; \theta) \right], \quad i, j = 1, \dots, p, \end{aligned}$$

are continuous on Θ .

Assumption A.5 is needed to prove stochastic equicontinuity of process Ψ_T along the lines of Andrews (1991) [see the proof of Lemma A.1 in Appendix B]. Note that standard results for stochastic equicontinuity [e.g. Hansen (1996)] do not apply here, since the kernel component in Ψ_T does not allow for uniformly bounded moments of order larger than two for functions $g_{t,T}(\theta)$. Let us now discuss the bandwidth conditions in Assumption A.16. The condition $Th_T^{d+2m} \rightarrow 0$ is the standard assumption for a negligible asymptotic bias. Condition $T^\alpha h_T^d / \log T \rightarrow \infty$, for $\alpha < 1/2 - 1/\beta$, is stronger than the standard condition $Th_T^d \rightarrow \infty$; it is used to prove the consistency of kernel regression estimator $\tilde{E}[g_2(Y; \theta) | x_0]$, uniformly in $\theta \in \Theta$ (see Lemma B.1 in Appendix B). Such a stronger bandwidth condition is also necessary to ensure negligible second-order term in the asymptotic expansion of the kernel moment estimator. Indeed, in the full-information underidentified case, some linear combinations of parameter θ_0 are estimated at a nonparametric rate $1/\sqrt{Th_T^d}$, whereas other linear combinations are estimated at a parametric rate $1/\sqrt{T}$. Thus, we need to ensure that the second-order term with smallest rate of convergence is negligible w.r.t. the first-order term with largest rate of convergence:

$$\left(1/\sqrt{Th_T^d} \right)^2 = o(1/\sqrt{T}) \iff T^{1/2} h_T^d \rightarrow \infty.$$

This condition is satisfied under Assumption A.16. Finally, the bandwidth condition in Assumption A.16 can be satisfied when $d < 2m(\beta - 2)/(\beta + 2)$. In particular, $m = 2$ is sufficient when $d < 4$, if $\beta > 14$.

A.1.2 Consistency

To study the asymptotic properties of the kernel moment estimator we have to derive the asymptotic distribution of the empirical process Ψ_T . This asymptotic distribution is given in Lemma A.1 below, which is proved in Appendix B. The proof uses consistency and asymptotic normality of kernel estimators [e.g. Bosq (1998), Bosq, Lecoutre (1987)], the Liapunov CLT [Billingsley (1965)], results on kernel M-estimators [Tenreiro (1995)], weak convergence of empirical processes [Pollard (1990)], and a proof of stochastic equicontinuity similar to Andrews (1991).

Lemma A.1: *Under Assumptions A.1-A.25:*

$$\Psi_T \implies \Psi,$$

where $\Psi(\theta)$, $\theta \in \Theta$, denotes the Gaussian stochastic process defined on Θ with covariance function $V_0(\theta, \tau) = E[\Psi(\theta)\Psi(\tau)']$ given by:

$$V_0(\theta, \tau) = \begin{pmatrix} S_0(\theta, \tau) & 0 \\ 0 & \frac{w^2}{f(x_0)}\Sigma_0(\theta, \tau) \end{pmatrix},$$

for $\theta, \tau \in \Theta$, where,

$$S_0(\theta, \tau) = \sum_{k=-\infty}^{\infty} \text{Cov}[g_1(Y_t, X_t; \theta), g_1(Y_{t-k}, X_{t-k}; \tau)],$$

$$\Sigma_0(\theta, \tau) = \text{Cov}[g_2^*(Y_t; \theta), g_2^*(Y_t; \tau) | X_t = x_0].$$

In particular $\Psi_T(\theta_0) \xrightarrow{d} N(0, V_0(\theta_0, \theta_0))$ where,

$$V_0(\theta_0, \theta_0) = \begin{pmatrix} S_0 & 0 \\ 0 & w^2\Sigma_0/f(x_0) \end{pmatrix}, \quad (\text{A.1})$$

and matrices S_0, Σ_0 are defined in Assumption A.6.

Block diagonal elements of matrix $V_0(\theta_0, \theta_0)$ are the standard asymptotic variance-covariance matrices of sample average, and kernel regression estimators, respectively. Lemma A.1 implies that marginal and conditional moment restrictions are asymptotically independent, and that the convergence is uniform w.r.t. θ .

We have the following proposition:

Proposition A.2: Under Assumptions A.1-A.25, the (kernel) moment estimator $\widehat{\theta}_T^*$ is consistent:

$$\left\| \widehat{\theta}_T^* - \theta_0^* \right\| \xrightarrow{p} 0, \quad \text{as } T \rightarrow \infty.$$

Proof: Write the criterion as:

$$Q_T(\theta^*) = [\Psi_T(\theta) + m_T(\theta^*)]' \Omega [\Psi_T(\theta) + m_T(\theta^*)], \quad \theta^* \in \Theta \times B.$$

For any $\varepsilon > 0$, we have:

$$\begin{aligned} P \left[\left\| \widehat{\theta}_T^* - \theta_0^* \right\| \geq \varepsilon \right] &\leq P \left[\inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} Q_T(\theta^*) \leq Q_T(\theta_0^*) \right] \\ &\leq P \left[\inf_{\theta \in \Theta} \Psi_T(\theta)' \Omega \Psi_T(\theta) + \inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} 2m_T(\theta^*)' \Omega \Psi_T(\theta) - Q_T(\theta_0^*) \leq \right. \\ &\quad \left. - \inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} m_T(\theta^*)' \Omega m_T(\theta^*) \right]. \end{aligned} \quad (\text{A.2})$$

Let us compare the asymptotic orders of the different terms. From Lemma A.1 and Continuous Mapping Theorem [CMT, Billingsley (1968)], we have:

$$\inf_{\theta \in \Theta} \Psi_T(\theta)' \Omega \Psi_T(\theta) = O_p(1), \quad (\text{A.3})$$

$$\inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} m_T(\theta^*)' \Omega \Psi_T(\theta) = O_p(\sqrt{T}). \quad (\text{A.4})$$

Similarly, since $\Psi_T(\theta_0) = O_p(1)$ and $m_T(\theta_0^*) = 0$, we deduce:

$$Q_T(\theta_0^*) = O_p(1).$$

Let $\lambda > 0$ be the smallest eigenvalue of Ω (Assumption A.15). We get:

$$\begin{aligned} & m_T(\theta^*)' \Omega m_T(\theta^*) \\ \geq & T\lambda \left(\|E_0[g_1(Y_t, X_t; \theta)]\|^2 + h_T^d \|E_0[g_2(Y_t; \theta) | X_t = x_0]\|^2 + h_T^d \|E_0[a(Y_t; \theta) | X_t = x_0] - \beta\|^2 \right), \\ \geq & Th_T^d \lambda \left(\|E_0[g_1(Y_t, X_t; \theta)]\|^2 + \|E_0[g_2(Y_t; \theta) | X_t = x_0]\|^2 + \|E_0[a(Y_t; \theta) | X_t = x_0] - \beta\|^2 \right), \end{aligned}$$

for T large, and any $\theta^* \in \Theta \times B$. From continuity of moment functions (Assumption A.14), compactness of $\Theta \times B$ (Assumption A.3) and global identification (Assumption A.1), we have:

$$\inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} m_T(\theta^*)' \Omega m_T(\theta^*) \geq CT h_T^d, \quad (\text{A.5})$$

for some constant $C > 0$. Thus from (A.2)-(A.5) we get:

$$P \left[\left\| \widehat{\theta}_T^* - \theta_0^* \right\| \geq \varepsilon \right] \leq P \left[Z_T \leq -CT h_T^d \right],$$

where Z_T is a random variable of order $O_p(\sqrt{T})$. From bandwidth Assumption A.16, we have $\sqrt{T} = o(Th_T^d)$, and we deduce:

$$P \left[\left\| \widehat{\theta}_T^* - \theta_0^* \right\| \geq \varepsilon \right] \rightarrow 0, \text{ as } T \rightarrow \infty.$$

■

A.1.3 Rate of convergence

In this section, we derive the rate of convergence of $\widehat{\theta}_T^*$.

Lemma A.3: *Under Assumptions A.1-A.25:*

$$\left\| \widehat{\theta}_T^* - \theta_0^* \right\| = O_p \left(1/\sqrt{Th_T^d} \right).$$

Proof: We follow the approach in the proof of Lemma A1 in Stock, Wright (2000). Since $\widehat{\theta}_T^*$ is the minimizer of Q_T we have:

$$Q_T(\widehat{\theta}_T^*) - Q_T(\theta_0^*) = \left[\Psi_T(\widehat{\theta}_T) + m_T(\widehat{\theta}_T^*) \right]' \Omega \left[\Psi_T(\widehat{\theta}_T) + m_T(\widehat{\theta}_T^*) \right] - \Psi_T(\theta_0)' \Omega \Psi_T(\theta_0) \leq 0,$$

that is,

$$m_T(\widehat{\theta}_T^*)' \Omega m_T(\widehat{\theta}_T^*) + 2m_T(\widehat{\theta}_T^*)' \Omega \Psi_T(\widehat{\theta}_T) + d_{1,T} \leq 0,$$

where $d_{1,T} = \Psi_T(\widehat{\theta}_T)' \Omega \Psi_T(\widehat{\theta}_T) - \Psi_T(\theta_0)' \Omega \Psi_T(\theta_0)$. Using:

$$\begin{aligned} m_T(\widehat{\theta}_T^*)' \Omega m_T(\widehat{\theta}_T^*) & \geq \lambda \left\| m_T(\widehat{\theta}_T^*) \right\|^2, \\ m_T(\widehat{\theta}_T^*)' \Omega \Psi_T(\widehat{\theta}_T) & \geq - \left\| m_T(\widehat{\theta}_T^*) \right\| \left\| \Omega \Psi_T(\widehat{\theta}_T) \right\|, \end{aligned}$$

we deduce:

$$\left\| m_T(\widehat{\theta}_T^*) \right\|^2 - 2d_{2,T} \left\| m_T(\widehat{\theta}_T^*) \right\| + d_{3,T} \leq 0, \quad (\text{A.6})$$

where:

$$d_{2,T} = \left\| \Omega \Psi_T(\widehat{\theta}_T) \right\| / \lambda \quad \text{and} \quad d_{3,T} = d_{1,T} / \lambda = \left[\Psi_T(\widehat{\theta}_T)' \Omega \Psi_T(\widehat{\theta}_T) - \Psi_T(\theta_0)' \Omega \Psi_T(\theta_0) \right] / \lambda.$$

Inequality (A.6) implies:

$$\left\| m_T(\widehat{\theta}_T^*) \right\| \leq d_{2,T} + (d_{2,T}^2 - d_{3,T})^{1/2}.$$

Let us now derive the order of the RHS. From Lemma A.1 and CMT we have:

$$\begin{aligned} d_{2,T} &\leq \sup_{\theta \in \Theta} \left\| \Omega \Psi_T(\theta) \right\| / \lambda = O_p(1), \\ |d_{3,T}| &\leq 2 \sup_{\theta \in \Theta} \left| \Psi_T(\theta)' \Omega \Psi_T(\theta) \right| / \lambda = O_p(1). \end{aligned}$$

We get:

$$\left\| m_T(\widehat{\theta}_T^*) \right\| = O_p(1).$$

Define:

$$G(\theta^*) = \left(E_0 [g_1(Y_t, X_t; \theta)]', E_0 [g_2(Y_t; \theta) | x_0]', E_0 [a(Y_t; \theta) - \beta | x_0]' \right),$$

for $\theta^* \in \Theta \times B$. Since $\left\| m_T(\theta^*) \right\|^2 \geq Th_T^d \left\| G(\theta^*) \right\|^2$, $\theta^* \in \Theta \times B$, we deduce:

$$\left\| G(\widehat{\theta}_T^*) \right\| = O_p \left(1 / \sqrt{Th_T^d} \right).$$

By the mean-value theorem we can write:

$$\left\| \frac{\partial G}{\partial \theta^{*'}}(\widetilde{\theta}_T^*) \left(\widehat{\theta}_T^* - \theta_0^* \right) \right\| = O_p \left(1 / \sqrt{Th_T^d} \right),$$

where $\widetilde{\theta}_T^*$ is between $\widehat{\theta}_T^*$ and θ_0^* . Since $\widehat{\theta}_T^*$ converges to θ_0^* by Proposition A.2, and $\partial G / \partial \theta^{*'}(\theta^*)$ is continuous by Assumption A.25, we have:

$$\frac{\partial G}{\partial \theta^{*'}}(\widetilde{\theta}_T^*) \xrightarrow{p} \frac{\partial G}{\partial \theta^{*'}}(\theta_0^*),$$

where $\partial G / \partial \theta^{*'}(\theta_0^*)$ has full rank, by local identification condition A.2. Thus we conclude:

$$\left\| \widehat{\theta}_T^* - \theta_0^* \right\| = O_p \left(1 / \sqrt{Th_T^d} \right).$$

■

The rate of convergence of the components of $\widehat{\theta}_T^*$ is in general the nonparametric rate $1 / \sqrt{Th_T^d}$, due to the full-information unidentified directions. However, it will be seen below that there may exist linear combinations of θ_0^* which are estimated at a parametric rate $1 / \sqrt{T}$. These linear combinations correspond to the full-information identified directions.

A.1.4 Asymptotic normality

In this section, we derive the asymptotic distribution of the kernel moment estimator $\widehat{\theta}_T^*$. We have to distinguish between the linear combinations of θ_0 , which are full-information identifiable, and the linear combinations of θ_0 , which are full-information non-identifiable. Indeed, the former linear combinations feature a parametric rate of convergence $1/\sqrt{T}$, whereas a nonparametric rate $1/\sqrt{Th_T^d}$ is expected for the latter linear combinations.

Let $\eta = R_1^{-1}\theta = (\eta_1', \eta_2') \in \mathbb{R}^{s_Z} \times \mathbb{R}^{p-s_Z}$, where $R_1 = (\widetilde{R}, R_Z)$, be the new parametrization introduced in equation (11) of the text. Parameter η_1 defines the s_Z directions of full-information identification, whereas η_2 gives the $p - s_Z$ full-information underidentified directions. Furthermore, let us introduce the matrix:

$$J_0 = \begin{pmatrix} E_0\left(\frac{\partial g_1}{\partial \theta'}\right) \widetilde{R} & 0 & 0 \\ 0 & E_0\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_Z & 0 \\ 0 & E_0\left(\frac{\partial a_0}{\partial \theta'}|x_0\right) R_Z & -Id_L \end{pmatrix} = \begin{pmatrix} E_0\left(\frac{\partial g_1}{\partial \eta_1}\right) & 0 & 0 \\ 0 & E_0\left(\frac{\partial g_2}{\partial \eta_2}|x_0\right) & 0 \\ 0 & E_0\left(\frac{\partial a}{\partial \eta_2}|x_0\right) & -Id_L \end{pmatrix}. \quad (\text{A.7})$$

Under Assumption A.2, matrix J_0 has full column-rank. Matrix J_0 is the asymptotic matrix of derivatives of standardized moment conditions. Indeed let us introduce the standardization matrix:

$$D_T = \begin{pmatrix} \sqrt{T}Id_{s_Z} & 0 & 0 \\ 0 & \sqrt{Th_T^d}Id_{p-s_Z} & 0 \\ 0 & 0 & \sqrt{Th_T^d}Id_L \end{pmatrix},$$

and define the matrix:

$$R_T = \begin{pmatrix} \widetilde{R} & R_Z & 0 \\ 0 & 0 & Id_L \end{pmatrix} D_T^{-1} = \begin{pmatrix} T^{-1/2}\widetilde{R} & (Th_T^d)^{-1/2}R_Z & 0 \\ 0 & 0 & (Th_T^d)^{-1/2}Id_L \end{pmatrix}.$$

Then, $R_T^{-1}(\theta', \beta')' = (\sqrt{T}\eta_1', \sqrt{Th_T^d}\eta_2', \sqrt{Th_T^d}\beta')'$, and we have the following Lemma A.4, proved in Appendix B using the ULLN and the CLT for mixing processes in Potscher, Prucha (1989), and Herrndorf (1984), respectively.

Lemma A.4: *Let $\widetilde{\theta}_T^*$ be such that $\|\widetilde{\theta}_T^* - \theta_0^*\| = O_p(1/\sqrt{Th_T^d})$. Then, under Assumptions A.1-A.25, we have:*

$$p \lim \frac{\partial \widehat{g}_T}{\partial \theta^{*'}}(\widetilde{\theta}_T^*) R_T = J_0.$$

In particular, matrix J_0 is block diagonal w.r.t. parameters η_1 and $(\eta_2', \beta')'$ reflecting the different rates of convergence, $1/\sqrt{T}$ and $1/\sqrt{Th_T^d}$, respectively.

The joint asymptotic distribution of $(\sqrt{T}(\widehat{\eta}_{1,T} - \eta_{1,0})', \sqrt{Th_T^d}(\widehat{\eta}_{2,T} - \eta_{2,0})', \sqrt{Th_T^d}(\widehat{\beta}_T - \beta_0)')'$ is provided in the next proposition.

Proposition A.5: *Under Assumptions A.1-A.25, kernel moment estimator $\widehat{\theta}_T^*$ is asymptotically*

normal:

$$\begin{pmatrix} \sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0}) \\ \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0}) \\ \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0) \end{pmatrix} \xrightarrow{d} N\left(0, (J_0' \Omega J_0)^{-1} J_0' \Omega V_0(\theta_0, \theta_0) \Omega J_0 (J_0' \Omega J_0)^{-1}\right).$$

Proof: The first-order condition for kernel moment estimator $\hat{\theta}_T^*$ is:

$$\frac{\partial \hat{g}_T'}{\partial \theta^*}(\hat{\theta}_T^*) \Omega \hat{g}_T(\hat{\theta}_T^*) = 0.$$

By a mean-value expansion we can write:

$$\frac{\partial \hat{g}_T'}{\partial \theta^*}(\hat{\theta}_T^*) \Omega \hat{g}_T(\theta_0^*) + \frac{\partial \hat{g}_T'}{\partial \theta^*}(\hat{\theta}_T^*) \Omega \frac{\partial \hat{g}_T}{\partial \theta^{*'}}(\hat{\theta}_T^*) (\hat{\theta}_T^* - \theta_0^*) = 0,$$

where $\tilde{\theta}_T^*$ is between $\hat{\theta}_T^*$ and θ_0^* . By multiplying this first-order condition by invertible matrix R_T' , we get:

$$R_T' \frac{\partial \hat{g}_T'}{\partial \theta^*}(\hat{\theta}_T^*) \Omega \hat{g}_T(\theta_0^*) + R_T' \frac{\partial \hat{g}_T'}{\partial \theta^*}(\hat{\theta}_T^*) \Omega \frac{\partial \hat{g}_T}{\partial \theta^{*'}}(\hat{\theta}_T^*) R_T \begin{pmatrix} \sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0}) \\ \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0}) \\ \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0) \end{pmatrix} = 0.$$

Let us define:

$$\hat{J}_T = \frac{\partial \hat{g}_T}{\partial \theta^{*'}}(\hat{\theta}_T^*) R_T, \quad \tilde{J}_T = \frac{\partial \hat{g}_T}{\partial \theta^{*'}}(\tilde{\theta}_T^*) R_T.$$

From Lemma A.3 and A.4, we have:

$$p \lim \hat{J}_T = p \lim \tilde{J}_T = J_0. \quad (\text{A.8})$$

Thus, $\tilde{J}_T \Omega \tilde{J}_T$ is non-singular with probability approaching 1, and we can write:

$$\begin{pmatrix} \sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0}) \\ \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0}) \\ \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0) \end{pmatrix} = -(\tilde{J}_T \Omega \tilde{J}_T)^{-1} \tilde{J}_T \Omega g_T(\theta_0^*).$$

Since $m_T(\theta_0^*) = 0$, we get:

$$\begin{pmatrix} \sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0}) \\ \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0}) \\ \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0) \end{pmatrix} = -(\tilde{J}_T \Omega \tilde{J}_T)^{-1} \tilde{J}_T \Omega \Psi_T(\theta_0).$$

The result follows from Lemma A.1 and (A.8). ■

A.1.5 Optimal weighting matrix

From Proposition A.5, the optimal weighting matrix Ω_0 , which minimizes the asymptotic variance

of $\left(\sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0})', \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0})', \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0)'\right)'$, is:

$$\Omega_0 = V_0(\theta_0, \theta_0)^{-1},$$

and the corresponding variance of the efficient estimator is:

$$\left(J_0' V_0(\theta_0, \theta_0)^{-1} J_0 \right)^{-1}.$$

Since $V_0(\theta_0, \theta_0)$ and J_0 are block diagonal w.r.t. η_1 and (η_2', β') , the variance of the efficient estimator of β is the lower right $L \times L$ block of matrix:

$$\frac{w^2}{f(x_0)} \left[\begin{pmatrix} E_0 \left(\frac{\partial g_2}{\partial \eta_2} | x_0 \right) & 0 \\ E_0 \left(\frac{\partial a}{\partial \eta_2} | x_0 \right) & -Id_L \end{pmatrix}' V_0 \left(\begin{pmatrix} g_2 \\ a \end{pmatrix} | x_0 \right)^{-1} \begin{pmatrix} E_0 \left(\frac{\partial g_2}{\partial \eta_2} | x_0 \right) & 0 \\ E_0 \left(\frac{\partial a}{\partial \eta_2} | x_0 \right) & -Id_L \end{pmatrix} \right]^{-1}.$$

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 kernel nonparametric efficiency of the information based estimator (Proposition 3).

A.2.1 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:

$$\begin{aligned} \mathcal{L} &= \frac{1}{T} \sum_{t=1}^T \int \frac{[\widehat{f}(y|x_t) - f^t(y)]^2}{\widehat{f}(y|x_t)} dy + h_T^d \int \log [f^0(y)/\widehat{f}(y|x_0)] f^0(y) dy \\ &\quad - 2 \frac{1}{T} \sum_{t=1}^T \mu_t \left(\int f^t(y) dy - 1 \right) - h_T^d \mu_0 \left(\int f^0(y) dy - 1 \right) \\ &\quad - 2 \frac{1}{T} \sum_{t=1}^T \lambda_t' \int g(y; \theta) f^t(y) dy - h_T^d \lambda_0' \int g_2(y; \theta) f^0(y) dy. \end{aligned}$$

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:

$$f^t(y) = \widehat{f}(y|x_t) + \mu_t \widehat{f}(y|x_t) + \lambda_t' g(y; \theta) \widehat{f}(y|x_t), \quad t = 1, \dots, T, \quad (\text{A.9})$$

$$f^0(y) = \widehat{f}(y|x_0) \exp \left(\lambda_0' g_2(y; \theta) + \mu_0 - 1 \right). \quad (\text{A.10})$$

The Lagrange multipliers are deduced from the constraints. From (A.9), we get:

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

and:

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

Similarly, from (A.10) we deduce the value of Lagrange multiplier μ_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.9) and (A.10), $\mu_0, \lambda_t, \mu_t, t = 1, \dots, T$ can be eliminated to get the concentrated functional parameters:

$$\begin{aligned} f^t(y; \theta) &= \widehat{f}(y|x_t) - \widetilde{E}(g(\theta)|x_t)' \widetilde{V}(g(\theta)|x_t)^{-1} \left[g(y; \theta) - \widetilde{E}(g(\theta)|x_t) \right] \widehat{f}(y|x_t), \\ & \quad t = 1, \dots, T, \\ f^0(y; \theta, \lambda_0) &= \frac{\exp \lambda'_0 g_2(y; \theta)}{\widetilde{E}(\exp \lambda'_0 g_2(\theta)|x_0)} \widehat{f}(y|x_0), \end{aligned} \quad (\text{A.11})$$

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 $\widehat{\theta}_T$ is solution of the saddle point problem [see Kitamura-Stutzer (1997) in a marginal framework]:

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

where

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

and the conditional density estimators are:

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

A.2.2 Asymptotic expansions

i) Asymptotic expansion of the concentrated objective function

Since the conditional moment restrictions are satisfied asymptotically, we have $\widehat{\lambda}_{0,T} \xrightarrow{p} 0$, 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 derive the expansion w.r.t. λ_0 . We have:

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

Therefore, we can asymptotically concentrate w.r.t. λ_0 :

$$\lambda_0 \simeq -\widetilde{V} (g_2(\theta) | x_0)^{-1} \widetilde{E} (g_2(\theta) | x_0), \quad (\text{A.12})$$

and the asymptotic expansion of the concentrated objective function becomes:

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

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

$$\begin{aligned} \widetilde{E} (g(\theta) | x_t) & \simeq \widetilde{E} (g(\theta_0) | x_t) + E_0 \left(\frac{\partial g}{\partial \theta'} (\theta_0) | x_t \right) (\theta - \theta_0), \\ \widetilde{V} (g(\theta) | x_t) & \simeq V_0 (g(\theta_0) | x_t), \end{aligned}$$

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

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

where functions g, g_2 are evaluated at θ_0 .

ii) Asymptotic expansion of $\widehat{\theta}_T$

In order to derive the asymptotic expansion of $\widehat{\theta}_T$, we have to carefully distinguish between the directions of θ converging at a parametric rate and those converging 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:

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

The asymptotic expansion of $\hat{\eta}_{1,T}$ is 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:

$$\begin{aligned} \sqrt{T} (\hat{\eta}_{1,T} - \eta_1^0) & \simeq - \left[\frac{1}{T} \sum_{t=1}^T \tilde{R}' E_0 \left(\frac{\partial g'}{\partial \theta} \mid x_t \right) V_0(g|x_t)^{-1} E_0 \left(\frac{\partial g}{\partial \theta'} \mid x_t \right) \tilde{R} \right]^{-1} \\ & \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{R}' E_0 \left(\frac{\partial g'}{\partial \theta} \mid x_t \right) V_0(g|x_t)^{-1} \int g(y; \theta_0) \hat{f}(y|x_t) dy \\ & \simeq - \left(\tilde{R}' E \left[E_0 \left(\frac{\partial g'}{\partial \theta} \mid x_t \right) V_0(g|x_t)^{-1} E_0 \left(\frac{\partial g}{\partial \theta'} \mid x_t \right) \right] \tilde{R} \right)^{-1} \\ & \cdot \sqrt{T} \int \int \tilde{R}' E_0 \left(\frac{\partial g'}{\partial \theta} \mid x \right) V_0(g|x)^{-1} g(y; \theta_0) \hat{f}(y, x) dx dy. \end{aligned}$$

Thus $\hat{\eta}_{1,T}$ converges at a parametric rate.

The asymptotic expansion of $\hat{\eta}_{2,T}$ can be deduced from the maximization of the second component of $\mathcal{L}_c(\eta)$. Estimator $\hat{\eta}_{2,T}$ converges at a nonparametric rate, and terms involving $(\hat{\eta}_{1,T} - \eta_1^0)$ can be neglected. We get:

$$\begin{aligned} \sqrt{Th_T^d} (\hat{\eta}_{2,T} - \eta_2^0) & \simeq - \left[R' E_0 \left(\frac{\partial g_2'}{\partial \theta} \mid x_0 \right) V_0(g_2|x_0)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} \mid x_0 \right) R \right]^{-1} \\ & \cdot R' E_0 \left(\frac{\partial g_2'}{\partial \theta} \mid x_0 \right) V_0(g_2|x_0)^{-1} \sqrt{Th_T^d} \int g_2(y; \theta_0) \hat{f}(y|x_0) dy. \end{aligned} \tag{A.13}$$

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

Let us consider the expansion of $f^0(y; \theta, \lambda_0)$ in (A.11) 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.12). Thus, we get:

$$\begin{aligned} &\widehat{f}_0(y|x_0) \\ &= f_0(y; \widehat{\theta}_T, \widehat{\lambda}_{0,T}) \\ &\simeq \widehat{f}(y|x_0) - \widetilde{E}(g_2(\widehat{\theta}_T)|x_0)' \widetilde{V}(g_2(\widehat{\theta}_T)|x_0)^{-1} \left(g_2(y; \widehat{\theta}_T) - \widetilde{E}(g_2(\widehat{\theta}_T)|x_0) \right) \widehat{f}(y|x_0) \\ &\simeq \widehat{f}(y|x_0) - \widetilde{E}(g_2(\widehat{\theta}_T)|x_0)' V_0(g_2|x_0)^{-1} g_2(y; \theta_0) f(y|x_0). \end{aligned} \tag{A.14}$$

Moreover,

$$\begin{aligned} \widetilde{E}(g_2(\widehat{\theta}_T)|x_0) &\simeq \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy + E_0 \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) (\widehat{\theta}_T - \theta_0) \\ &\simeq \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy + E_0 \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R (\widehat{\eta}_{2,T} - \eta_2^0) \\ &\quad (\text{since the contribution of } \widehat{\eta}_{1,T} - \eta_1^0 \text{ is asymptotically negligible}) \\ &= (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy, \end{aligned}$$

from (A.13), 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 substituting in (A.14), we get:

$$\widehat{f}_0(y|x_0) \simeq \widehat{f}(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V_0(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy. \tag{A.15}$$

iv) Asymptotic expansion of the moment of interest

We have:

$$\begin{aligned}
\widehat{E}(a|x_0) &= \int a(y; \widehat{\theta}_T) \widehat{f}_0(y|x_0) dy \\
&\simeq \int a(y; \theta_0) f(y|x_0) dy + \int \frac{\partial a}{\partial \theta'}(y; \theta_0) f(y|x_0) dy (\widehat{\theta}_T - \theta_0) \\
&\quad + \int a(y; \theta_0) [\widehat{f}_0(y|x_0) - f(y|x_0)] dy \\
&\simeq E(a|x_0) + E_0 \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R (\widehat{\eta}_{2,T} - \eta_2^0) \\
&\quad + \int a(y; \theta_0) \left\{ \widehat{f}(y|x_0) - f(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V_0 (g_2|x_0)^{-1} \right. \\
&\quad \left. (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \right\} dy \quad [\text{from (A.15)}] \\
&= E(a|x_0) - E_0 \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R \left[R' E_0 \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]^{-1} \\
&\quad \cdot R' E_0 \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \quad [\text{from (A.13)}] \\
&\quad + \int a(y; \theta_0) [\widehat{f}(y|x_0) - f(y|x_0)] dy \\
&\quad - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy.
\end{aligned}$$

Thus, we get:

$$\begin{aligned}
&\widehat{E}(a|x_0) - E(a|x_0) \\
&\simeq \int a(y; \theta_0) \delta \widehat{f}(y|x_0) dy - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy \\
&\quad - \left[E_0 \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R - Cov_0(a, g_2|x_0) V_0 (g_2|x_0)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right] \\
&\quad \cdot \left[R' E_0 \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} E_0 \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R \right]^{-1} \\
&\quad \cdot R' E_0 \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V_0 (g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy, \tag{A.16}
\end{aligned}$$

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

A.2.3 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.16), 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, from the asymptotic normality of integrals of kernel estimators, 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:

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

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

Regularity conditions in the stochastic volatility model

In this Appendix, we discuss conditions that ensure the regularity assumptions for XMM estimator (see Appendix 1.1) in the stochastic volatility model.

A.3.1 Identification

Let us first consider the identifiability of structural parameter θ (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:

$$E_0 \left(\left(\begin{array}{c} 1 \\ \exp r_{t+1} \end{array} \right) \frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) | y_t \right) v = 0, \quad \forall y_t. \quad (\text{A.17})$$

We know that θ_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 ε is small and v satisfies (A.17), 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 ε . 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 (26), the parameters θ , 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 γ 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:

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

and matrix R is given by:

$$R = \begin{pmatrix} -\delta\frac{c}{1+c\nu_1} \\ 1 \\ -\rho\frac{1}{(1+c\nu_1)^2} \\ 0 \end{pmatrix}. \quad (\text{A.18})$$

ii) Verification 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:

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

We have:

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

From (27), we have:

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

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

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

Thus, we get:

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

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

A.3.2 Stationary distribution

Let us consider process $\{Y_t = (r_t, \sigma_t^2) : t \in \mathbb{Z}\}$, where r_t and σ_t^2 are defined in equations (21), (22) of the text. Markov process Y_t is exponential affine:

$$\begin{aligned} E \left[e^{-z' Y_{t+1}} \mid \underline{Y}_t \right] &= E \left[e^{-ur_{t+1} - v\sigma_{t+1}^2} \mid \underline{Y}_t \right] = E \left[e^{-(\gamma u + v)\sigma_{t+1}^2} E \left[e^{-u\sigma_{t+1}\varepsilon_{t+1}} \mid (\sigma_t^2), \underline{Y}_t \mid \underline{Y}_t \right] \right] \\ &= E \left[e^{-(\gamma u + v - \frac{1}{2}u^2)\sigma_{t+1}^2} \mid \sigma_t^2 \right] = \exp \left[-a \left(\gamma u + v - \frac{1}{2}u^2 \right) \sigma_t^2 - b \left(\gamma u + v - \frac{1}{2}u^2 \right) \right] \\ &\equiv \exp \left[-A(z)' Y_t - B(z) \right], \end{aligned}$$

where:

$$A(z) = \left(0, a \left(\gamma u + v - \frac{1}{2}u^2 \right) \right)', \quad B(z) = b \left(\gamma u + v - \frac{1}{2}u^2 \right),$$

for $z = (u, v)' \in \mathbb{C}^2$ such that $\text{Re}(\gamma u + v - \frac{1}{2}u^2) > -1/c$, and functions a and b are defined in (22).

i) Strict stationarity and geometric strong mixing

From Proposition 2 of Gouriou, Jasiak (2005), the ARG process (σ_t^2) is stationary if $0 \leq \rho < 1$, with marginal invariant distribution such that $[(1 - \rho)/c] \sigma_t^2 \sim \gamma(\delta)$, where $\gamma(\delta)$ denotes the gamma distribution with parameter δ . Thus, when $\rho < 1$, process (Y_t) admits the marginal invariant distribution:

$$f(y) = \frac{1}{\sigma} \phi \left(\frac{r - \gamma \sigma^2}{\sigma} \right) \frac{[(1 - \rho)/c]^\delta}{\Gamma(\delta)} e^{\frac{1-\rho}{c} \sigma^2} (\sigma^2)^{\delta-1}, \quad y = (r, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ = \mathcal{Y}. \quad (\text{A.19})$$

To prove that (Y_t) is geometrically strong mixing, we use Proposition 4.2 of Darolles, Gouriou, Jasiak (2005), and verify the condition:

$$\lim_{h \rightarrow \infty} \frac{\partial A}{\partial z'}(0)^h = 0. \quad (\text{A.20})$$

We have:

$$\frac{\partial A}{\partial z'}(0) = \begin{pmatrix} 0 & 0 \\ \gamma \rho & \rho \end{pmatrix}.$$

Condition (A.20) is satisfied if $\rho < 1$. Thus, with $X_t = Y_{t-1}$, we conclude that Assumption A.4 is satisfied if $0 \leq \rho < 1$.

ii) Smoothness of the marginal distribution

Stationary distribution f in (A.19) is in $C^\infty(\mathcal{Y})$. Moreover, we have:

$$f(y) \leq C_1 e^{\frac{1-\rho}{c}\sigma^2} (\sigma^2)^{\delta-3/2}, \quad y \in \mathcal{Y},$$

for a constant $C_1 > 0$. Thus, $\|f\|_\infty < \infty$ if, and only if, $\delta \geq 3/2$. Moreover, we have the following Lemma A.6, proved in Appendix B.

Lemma A.6 : $\|D^m f\| < \infty$ if, and only if, $\delta \geq 3/2 + m$.

Thus, Assumption A.7 is satisfied if $\delta \geq 3/2 + m$. For instance, for $m = 2$, we get $\delta \geq 7/2$.

A.3.3 Existence of moments

The moment function $g_2^*(y; \theta)$, which corresponds to the estimating constraints (17)-(19), is given by:

$$g_2^*(y_t; \theta) = e^{-\theta_0 - \theta_1 \sigma_{t+1}^2 - \theta_2 \sigma_t^2 - \theta_3 r_{t+1}} \begin{pmatrix} 1 \\ e^{r_{t+1}} \\ (e^{r_{t+1}} - s_1)^+ \\ \vdots \\ (e^{r_{t+1}} - s_K)^+ \\ (e^{r_{t+1}} - s)^+ \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ c_{t_0}(s_1, h) \\ \vdots \\ c_{t_0}(s_K, h) \\ c_{t_0}(s, h) \end{pmatrix}, \quad y_t = (r_{t+1}, \sigma_{t+1}^2, \sigma_t^2)'$$

The relevant state and conditioning variables are $Y_t = (r_{t+1}, \sigma_{t+1}^2, \sigma_t^2)$, and $X_t = (r_t, \sigma_t^2)$, respectively.

The following Lemma A.7, proved in Appendix B, provides a condition for $g_2^*(\cdot; \theta) \in L^2(F_Y)$ [see Assumption A.5].

Lemma A.7: *The function $g_2^*(\cdot; \theta)$ is in $L^2(F_Y)$ if, and only if:*

$$\theta \in \Gamma = \left\{ (\theta_0, \theta_1, \theta_2, \theta_3)' \in \mathbb{R}^4 \mid \theta_2 > -1/2c, \theta_1 > -1/2c - \gamma\theta_3 + \theta_3^2 + (1 + \gamma - 2\theta_3)^+ \right\}.$$

Thus, the condition $g_2^*(\cdot; \theta) \in L^2(F_Y)$ is satisfied, whenever the risk premia parameters θ_1 and θ_2 for stochastic volatility are above some thresholds. In particular, the lower bound for θ_1 depends on θ_3 . For the values of c and γ in Section 5, and intermediate values of θ_3 , the lower bounds for θ_1 and θ_2 are of the order -10^7 .