

CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass



<http://www.cass.city.ac.uk/cea/index.html>

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Efficient Derivative Pricing by Extended Methods of Moments

Patrick Gagliardini, Christian Gourieroux and Eric Renault

CEA@Cass Working Paper Series

WP-CEA-02-2008

EFFICIENT DERIVATIVE PRICING BY EXTENDED METHOD OF MOMENTS

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

First version: July 2004

This version: May 2007

*University of Lugano and Swiss Finance Institute.

†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, J. Fan, R. Garcia, J. Jackwerth, O. Linton, L. Mancini, A. Melino, C. Ortelli, O. Scaillet, M. Stutzer, F. Trojani, A. Vedolin, B. Werker, E. Zivot and anonymous referees for helpful comments. The first author gratefully acknowledges financial support of the Swiss National Science Foundation through the NCCR FINRISK network. The second author gratefully acknowledges financial support of NSERC Canada and of the chair AXA: "Large Risks in Insurance".

Efficient Derivative Pricing by Extended Method of Moments

Abstract

This paper extends GMM and information theoretic estimation to settings where the conditional moment restrictions are either uniform (i.e. valid for any value of the conditioning variable), or local (i.e. valid for a particular value of the conditioning variable only). The parameter of interest can be either a structural parameter, or a local conditional moment. This is the framework for option pricing based on both historical data on the underlying asset and cross-sectional data on derivative assets, as a consequence of the rather small traded volumes on derivatives. We derive the asymptotic properties of the estimators and a kernel efficiency bound. The asymptotic behavior is not standard since the speed of convergence depends on the type of parameter considered. The results are applied to the derivative pricing problem using a factor model that is parametric in the stochastic discount factor and nonparametric in the conditional distribution of the observed factors. The extended method of moments is compared with the cross-sectional calibration approach used on the market for pricing S&P 500 options.

Keywords: Derivative Pricing, Partial Observability, Generalized Method of Moments, Information Theoretic Estimation, Weak Instrument, Kernel Nonparametric Efficiency.

JEL number: C13, C14, G12.

1 Introduction

The Generalized Method of Moments (GMM) has been introduced by Hansen (1982) and Hansen, Singleton (1982) to estimate parameters defined by Euler conditions. For example, 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, information theoretic approaches, such as empirical likelihood, minimum chi-square, or exponential tilting, have been proposed to simplify the derivation of GMM estimators and to improve their finite sample properties.¹ The basic idea is to estimate jointly the structural parameter θ and the nuisance infinite dimensional parameter under the moment restrictions.

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) [e.g., Harrison, Kreps (1979) and Hansen, Richard (1987)]. 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 relative price² at date t_0 of a European call written on p_1 , with moneyness strike k

¹See e.g. Owen (1991), Qin, Lawless (1994), Hansen, Heaton, Yaron (1996), Kitamura, Stutzer (1997), Imbens, Spady, Johnson (1998), Kitamura (2001), Ai, Chen (2003), Kitamura, Tripathi, Ahn (2004), Newey, Smith (2004), Antoine, Bonnal, Renault (2006).

²That is, the ratio between the price of the option and the price of the underlying asset.

and time-to-maturity 1 is:

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

A simple estimator of this relative option price is given by:

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

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. Hence, for derivative pricing the interest is focused on estimation of conditional moment $c_{t_0}(1, k) = E_{t_0}(a)$ of function $a = M_{t_0, t_0+1}(p_{1, t_0+1}/p_{1, t_0} - k)^+$, which is the product of the sdf by the derivative payoff, given the conditioning variable observed at t_0 . This problem requires a joint estimation of the parameter θ and the conditional distribution at t_0 .

The above GMM setting presupposes that prices of all tradable assets are observed at regular dates. The problem becomes more complicated if we want to account for assets that are less actively traded. For instance, it is really rare to observe an option again, which happens to have the same time-to-maturity as well as the same moneyness. As explained in Section 2, such a partial observability of derivative prices is inherent in option markets. As an example, let us assume for a moment that we 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 only:

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

In this case, the structural parameter θ is subject to two types of moment restrictions, which are satisfied either for multiple environments [uniform moment restrictions (5) and (6)], or only for a given one [local moment restriction (7)]. The standard GMM methodology does not apply in this case.

The aim of this paper is to use jointly regularly observed prices of the underlying asset, and partially observed prices of derivatives written on this asset, to efficiently estimate other derivative prices. In other words, we extend the standard GMM approach to estimate conditional moments under two types of conditional moment restrictions, which are either uniform or local with respect to the conditioning variable. In Section 2, we describe the problem of partial observability of derivative prices and review the main methodologies for empirical derivative pricing either proposed in the academic literature, or used in the trading rooms. The main features of the data are illustrated on S&P 500 option prices. The standard academic literature has neglected the problem of partial observability. This feature is taken into account by practitioners when they try to reconstitute daily the pricing function by pure cross-sectional approaches. However, this latter method is not compatible with the absence of arbitrage opportunities and provides erratic results for the daily estimated structural parameters. The estimation method proposed in our paper takes into account both the historical data on the underlying asset price and cross-sectional observations of derivative prices. It is consistent with the absence of arbitrage opportunities, provides more stable results, and confidence intervals are available. In Section 3, we study the set of moment estimators for both structural parameters and conditional moments of interest. The uniform and local moment restrictions 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 that 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 kernel moment method, called basic extended method of moments (XMM), which minimizes the asymptotic mean squared error of the estimated conditional moment of interest (or its asymptotic variance in a given class of kernel moment estimators). However, in general the basic XMM does not provide an estimated model consistent with the absence of arbitrage opportunities. Section 4 introduces an information based approach to estimate jointly the structural parameter and the conditional distribution. The information based XMM has the same asymptotic properties as the basic XMM, and is compatible with the absence of arbitrage opportunities. Section 5 considers the implementation for the derivative pricing problem. We discuss the patterns of the asymptotic confidence intervals as functions of time-to-maturity and moneyness strike. We compare

XMM with the standard cross-sectional calibration approach on a data set of S&P 500 options. Section 6 concludes. The detailed regularity conditions and the proofs of the asymptotic results for the basic XMM estimator are gathered in Appendix 1. The asymptotic results for the information based XMM estimator are sketched in Appendix 2. Proofs of technical Lemmas are given in Appendices B and C, which are available at the web-site: <http://www.istituti.usilu.net/gagliarp/proofsXMM.htm>.

2 Data and practice for option pricing

2.1 Option data

Three types of option price datasets can be distinguished, namely quotes, trading prices without taking into account the traded volume and the trading intensity, and trading prices for liquid options only. The approaches relying on absence of arbitrage opportunities are meaningful for the third type of dataset. Let us consider the options written on the S&P 500 index, which are traded at the Chicago Board Options Exchange (CBOE) and are often considered as highly traded.³ Our dataset consists of daily closing prices of OptionMetrics for call and put contracts in 2005. We use the same dataset in the application in Section 5. In order to ensure a minimal liquidity, the option market is organized by CBOE with periodic issuing of new option contracts [Hull (2005), p. 187]. The admissible times-to-maturity at issuing are 1-month, 2-month, 3-month, 6-month, 9-month, 12-month, ... The issuing dates during the year for these different options are provided in Table 1.

[Insert Table 1: Expiration months of S&P 500 options]

For instance, new 12-month options can be issued every three months, when the previous ones become 9-month options. Table 1 and Figure 1 show the cycle of times-to-maturity for the options which can be quoted on the market [see Schwartz (1987), Figure 1, and Pan (2002), Figure 2, for a similar discussion].

[Insert Figure 1: Times-to-maturity of quoted S&P 500 options in 2005]

³The question of illiquid derivatives is much more pronounced for options on corporate stocks, for credit derivatives or weather derivatives.

The options are issued for a limited number of moneyness strikes around the value of the index at the issuing date. For the options with large time-to-maturity at issuing, the moneyness during the life of the option can become far from the current value of the index. Additional options can be issued at intermediate dates to correct this feature.

A significant number of options are not actively traded daily, not traded at all, or even not quoted. For the discussion below and the empirical application, we consider the options with a daily traded volume of more than 4000 contracts. Each contract corresponds to 100 options, and 4000 contracts to a value between 5 millions \$ and 7 millions \$ on average. These amounts are rather small, when compared to the daily traded volume of a portfolio mimicking the index, such as the SPDR, which is about 120 millions \$.⁴

The times-to-maturity and moneyness strikes of the actively traded options vary in time, due to the issuing cycle and the trading activity. This feature is illustrated in Figures 2 and 3.

[Insert Figure 2: Times-to-maturity of actively traded S&P 500 options in June 2005]

[Insert Figure 3: Moneyness strikes of actively traded S&P 500 options in June 2005]

Figure 2 provides the times-to-maturity, whereas Figure 3 displays the moneyness strikes, for the actively traded options at the different trading days of June 2005. The deterministic nonstationary feature observed in Figure 2 is due to the periodic issuing. The random variability of the times-to-maturity and moneyness strikes is due to the trading activity.

Let us now consider a given day, June 1st 2005. The prices of actively traded options are displayed in Figure 4.

[Insert Figure 4: Prices of actively traded S&P 500 options in June 1st 2005]

There are 15 options⁵ with a daily traded volume larger than 4000 contracts for the 4 liquid times-to-maturity. We observe the two major reasons for option trading, which are arbitrage and protection (insurance). Close to at-the-money strikes, both puts

⁴On 03/15/2007.

⁵This number is equal to 11 and to 19 for the following two days.

and calls are actively traded, which allows for put-call parity strategies. In the region $k < 1$ (resp. $k > 1$), we have essentially traded put contracts (resp. call contracts).

These data and the discussion above show that

- (i) An option with a given time-to-maturity and moneyness strike is never actively traded every trading day (partial observability of option prices).
- (ii) The daily number of actively traded options and their characteristics vary in time. A part of this dynamics is due to the nonstationary deterministic periodic issuing. Another part is due to the random trading activity.

We assume in this paper that this random selection can be considered as exogenous.⁶

2.2 Calibration based on option data

Let us now review the main approaches proposed in the literature to calibrate (estimate) the parameters of an option pricing model.

i) Parametric analysis of time series of option prices

Let us assume a parametric sdf $M_{t,t+1}(\theta) = m(Y_{t+1}; \theta)$, where Y_t are latent factors with a parametric transition density $f(y_t|y_{t-1}; \theta)$. This parametric model can be used to express the relative option prices as deterministic functions of parameter θ and factor value y_t . For instance, the relative price of a European call option with time-to-maturity h and moneyness k is $c_t(h, k) = c(h, k; \theta, Y_t)$, say. These pricing equations are the basis for an estimation approach relying on n option prices observed at all dates, denoted $c_t(h_j, k_j)$, $j = 1, \dots, n$, $t = 1, \dots, T$ [see e.g. Duan (1994)].⁷ If the number of observed option prices n is equal to the number of factors, $\dim Y = n$, the deterministic relations

$$c_t(h_j, k_j) = c(h_j, k_j; \theta, Y_t), \quad j = 1, \dots, n, \quad t = 1, \dots, T, \quad (8)$$

can be inverted to get the unobserved factor values Y_t as functions of parameter θ and observed option prices. By applying the Jacobian formula, it is possible to derive the

⁶This assumption is implicit in all methodologies described in Section 2.2 and its relaxation is out of the scope of the present paper.

⁷The method extends easily to account for a time series of observations on the underlying asset by noting that the underlying asset is an option with zero moneyness strike.

likelihood function based on observed option prices and to estimate θ by maximum likelihood. In a similar vein, Pan (2002) uses the restrictions (8) to define an implied state GMM estimator. When the number of latent factors is larger than the number n of option prices considered in (8), a simulation based estimation procedure can be applied [e.g., Chernov and Ghysels (2000)].

These approaches have two drawbacks. First, it is assumed that a number of options with given characteristics ⁸ is actively traded at all dates, which is questionable in the light of the discussion in Section 2.1. ⁹ An exception is Pan (2002), who proves consistency and asymptotic normality of the implied state GMM estimator with time varying h_j, k_j . However, this comes at the cost of introducing unrealistic assumptions on the process of the characteristics of traded options. Second, the method does not work if the number of observed option prices is strictly larger than the number of factors. In such a case, there is no solution to the system of pricing equations (8) applied to all observed data. This can arise since, as seen above, the number of actively traded options varies in time.

ii) Semi-parametric analysis of time series of option prices

Let us still assume a parametric sdf $M_{t,t+1}(\theta) = m(Y_{t+1}; \theta)$, where Y_t is an observable factor, and available relative call prices $c_t(h_j, k_j) = E_t \left[M_{t,t+h_j}(\theta) (p_{t+h_j}/p_t - k_j)^+ \right]$, $j = 1, \dots, n$, $t = 1, \dots, T$. Parameter θ can be estimated by GMM based on the conditional moment restrictions associated with the pricing conditions [see e.g. Buraschi and Jackwerth (2001), Brennan, Liu and Xia (2005)]. For option pricing, the drawbacks of this approach are twofold. First, it assumes that a number of options with given characteristics are actively traded at all dates. Second, the method provides the estimated structural parameter θ , but gives no information on option prices themselves, which require also the estimation of the nonparametric transition.

iii) Cross-sectional approaches

⁸The time invariance of h_j, k_j is needed to get stationary sequences of option prices and apply the standard asymptotic theory for ML estimation.

⁹The problem of partial observability has been circumvented by building artificial series of near-the-money short options; for instance a time series can be constructed by considering the actively traded options with the shortest time-to-maturity and the moneyness strike closest to 1 [e.g., Chernov and Ghysels (2000)]. A related practice considers average prices of near-the-money short options [Santa-Clara, Yan (2004)]. It has also been suggested to draw at random one actively traded option per day [Eraker (2004)]. However, these practices imply measurement errors, biased results and spurious arbitrage opportunities.

Cross-sectional approaches can be used with parametric or nonparametric models.

The parametric cross-sectional approach is largely used in the trading rooms as follows. Let us consider the system of pricing equations (8) written for a given date t_0 and unobservable factor Y_{t_0} , that is, $c_{t_0}(h_j, k_j) = c(h_j, k_j; \theta, Y_{t_0})$, $j = 1, \dots, n$, with $n \geq \dim Y + \dim \theta$. At date t_0 , the parameter and factor values can be calibrated by minimizing $\sum_{j=1}^n [c_{t_0}(h_j, k_j) - c(h_j, k_j; \theta, Y_{t_0})]^2$ with respect to both θ and Y_{t_0} .¹⁰ This provides calibrated values $\hat{\theta}_{t_0}$ and \hat{Y}_{t_0} . Contrary to the previous methods, this approach can be applied if the number of options n_{t_0} and their characteristics h_{j,t_0}, k_{j,t_0} vary in time, which explains its success among practitioners.¹¹ Its drawbacks are threefold. First, the daily estimates $\hat{\theta}_{t_0}$ of the structural parameters are often very erratic in time, which is not compatible with the structural interpretation of parameter θ used to derive the option pricing formula $c(h, k; \theta, Y_t)$. Second, the approximated option prices $\hat{c}_{t_0}(h_j, k_j) = c(h_j, k_j; \hat{\theta}_{t_0}, \hat{Y}_{t_0})$ are not compatible with the absence of arbitrage opportunities and observed prices. Third, the method provides no confidence intervals for estimated parameters and option prices.

The cross-sectional approach can be extended to a nonparametric framework, where $c_{t_0}(h, k) = c(h, k; t_0)$ and c is a smooth function [see e.g. Ait-Sahalia, Lo (1998), (2000) and Jackwerth (2000)]. The method is consistent if the number of observed option prices n_{t_0} at date t_0 is large ($n_{t_0} \rightarrow \infty$) and if the moneyness strikes (and the times-to-maturity) are "continuously" distributed. Both assumptions are clearly not satisfied from the discussion in Section 2.1.

iv) Mixed time series and cross-sectional approaches

Finally, the maximum entropy approach combines time series information on the underlying asset price with cross-sectional observations on option prices [see e.g. Jackwerth, Rubinstein (1996), Buchen, Kelly (1996), Stutzer (1996)]. For instance, in the approach followed by Stutzer (1996), the historical transition is estimated nonparametrically from the time series of underlying asset prices. Then, at a given day t_0 , the risk-neutral transition at time-to-maturity h , say, is derived by looking for the distribution that is the closest to the historical one, subject to the pricing restrictions corresponding to time-to-maturity h . The main drawback of this approach is to

¹⁰An alternative approach with similar properties is the calibration based on implied Black-Scholes volatilities.

¹¹The parametric cross-sectional approach is also widely used by academic researchers, e.g. Bakshi, Cao and Chen (1997).

calibrate independently for the different times-to-maturity. First, this diminishes the number of actively traded options which can be used at each step. Second, this is not compatible with the absence of dynamic arbitrage opportunities. Moreover, there is no financial justification for choosing the historical distribution as a benchmark, that is, for minimizing the risk premium. The introduction of a parametric pricing kernel provides the required degree of freedom for risk correction.

3 Extended Method of Moments

The aim of this section is to introduce the basic XMM method in a general setting.

3.1 Parameters and constraints

Let us consider a general semi-parametric framework with both uniform and local conditional moment restrictions:

$$E[g(Y; \theta_0) | X = x] = 0, \quad \forall x, \quad (9)$$

$$E[\tilde{g}(Y; \theta_0) | X = x_0] = 0, \quad (10)$$

where $\theta_0 \in \mathbb{R}^p$ is an unknown structural parameter and x_0 is a given value of X . In the derivative pricing application (Section 4.3), local moment conditions (10) are associated with the option prices, and x_0 is the value of the state variables at current date t_0 . As in the standard GMM approach, in a first step we replace the uniform conditional restrictions by a set of marginal restrictions, by introducing a finite number of instrumental variables. Then, in a second step, we look for optimal instruments.

Let us introduce a matrix of instruments $Z = H(X)$ with dimensions $q_1 \times \dim(g)$, and let function g_1 define the corresponding marginal restrictions: $E[Z \cdot g(Y; \theta_0)] =: E[g_1(Y, X; \theta_0)] = 0$. Therefore, structural parameter θ_0 satisfies both marginal and conditional (local) restrictions:

$$E[g_1(Y, X; \theta_0)] = 0, \quad E[g_2(Y; \theta_0) | X = x_0] = 0, \quad (11)$$

where $g_2 = (\tilde{g}, g)'$ is obtained by gathering all conditional restrictions for environment x_0 . The parameter of primary interest for the application to derivative pricing

is the conditional moment:

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

where a is a given function. For the interpretation of parameter β_0 , we have to distinguish between the mapping $x \mapsto E[a(Y; \theta_0) | X = x]$, which is the prediction function, and its value at a given point x_0 , that is $\beta_0 = E[a(Y; \theta_0) | X = x_0]$, which can be considered as a standard parameter. In other words, it is important to distinguish the ex-ante option price, that is, before observing the value of factor X , and the ex-post option price, once X has been observed. The latter interpretation is the relevant one for the derivative pricing application and is used for developing the estimation approach. Thus, the total parameter to be estimated is the extended vector $\theta^* = (\theta', \beta')'$ [see Back, Brown (1992)], whose true value $(\theta'_0, \beta'_0)'$ satisfies the extended set of moment restrictions:

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

3.2 Kernel moment estimators

For expository purpose, we assume that variables X and Y have identical dimension d . Let us introduce a kernel estimator of the conditional density $f(y|x_0)$:

$$\hat{f}(y|x_0) = \frac{1}{h_T^d} \sum_{t=1}^T K\left(\frac{y_t - y}{h_T}\right) K\left(\frac{x_t - x_0}{h_T}\right) / \sum_{t=1}^T K\left(\frac{x_t - x_0}{h_T}\right), \quad (12)$$

where K is the d -dimensional kernel and h_T the bandwidth. The kernel density estimator is used to estimate conditional moment $E[g_2(Y; \theta) | X = x_0]$ by

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

Definition 1 A kernel moment estimator $\hat{\theta}_T^* = (\hat{\theta}_T', \hat{\beta}_T')'$ of parameter $\theta_0^* = (\theta_0', \beta_0')'$ is defined by:

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

where

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

\widehat{E} and $\widehat{E}[\cdot | x_0]$ denote an historical sample average and a kernel estimator of the conditional moment, respectively, and Ω is a positive definite weighting matrix.

The asymptotic properties of kernel moment estimators are derived in Appendix 1 using empirical process methods. Let us introduce a linear change of parameters from θ to $\eta = (\eta'_1, \eta'_2)'$, say, such that the subset of parameters η_1 are locally identified by the marginal moment conditions in (11), whereas the subset of parameters η_2 are identifiable only if both sets of conditions in (11) are used (see Appendix 1 for a complete discussion of identification in this framework). The estimator of parameter η_1 converges at a parametric rate, whereas the rate of convergence is nonparametric for parameters η_2 and β as stated in Proposition 1.

Proposition 1 *Under Assumptions A.1-A.25 in Appendix 1, the kernel moment estimator $\widehat{\theta}_T^*$ is consistent and asymptotically normal:*

$$\begin{pmatrix} \sqrt{T} (\widehat{\eta}_{1,T} - \eta_{1,0}) \\ \sqrt{Th_T^d} (\widehat{\eta}_{2,T} - \eta_{2,0}) \\ \sqrt{Th_T^d} (\widehat{\beta}_T - \beta_0) \end{pmatrix} \xrightarrow{d} N \left(B_\infty, \left(J'_0 \Omega J_0 \right)^{-1} J'_0 \Omega V_0 \Omega J_0 \left(J'_0 \Omega J_0 \right)^{-1} \right),$$

where $(\eta'_{1,0}, \eta'_{2,0})'$ is the true value of the transformed structural parameter, the bias is $B_\infty = -\sqrt{\bar{c}} \left(J'_0 \Omega J_0 \right)^{-1} J'_0 \Omega b_0$, $\bar{c} := \lim Th_T^{d+2m} \in [0, \infty)$ from Assumption A.17, and matrices J_0 , V_0 , and vector b_0 are given in Appendix 1 and depend on the kernel.

3.3 Weak instruments

The different rates of convergence of the components of kernel moment estimators are related to the use of weak instruments [see Andrews, Stock (2006) for a survey]. The marginal moment restrictions in (11) are obtained by introducing standard instruments satisfying the usual conditions. The local moment restrictions in (11) corresponding to a given value x_0 of the conditioning variable can be approximately written as $E [g_2(Y; \theta_0) | X = x_0] \simeq E \left[\frac{1}{f_X(x_0)h_T^d} K \left(\frac{X-x_0}{h_T} \right) g_2(Y; \theta_0) \right]$, where f_X denotes the unconditional pdf of variable X . They correspond to a finite number of

weak moment restrictions constructed from instrument $Z_T = K\left(\frac{X-x_0}{h_T}\right) / [h_T^d f_X(x_0)]$. This instrument depends on the number of observations through the bandwidth, and is weak in the sense of Stock and Wright (2000) due to the localization around $X = x_0$ induced by the kernel. Since the instruments depend on T , the problem considered above is not a special case of the standard literature on weak IV asymptotics [e.g., Staiger, Stock (1997), Stock, Wright (2000)]. For instance, the functions of the parameters which are weakly (resp. strongly) identified are not known a priori, and the rates of convergence of the estimators differ from the rates of convergence obtained in the applications considered earlier in the literature.

3.4 The basic XMM estimator

Let us now focus on the parameter of interest β_0 . To simplify the exposition, we assume that β_0 is one-dimensional, and that the kernel K is a given product kernel $K(u) = \prod_{i=1}^d \kappa(u_i)$ of order $m \geq 2$. It is proved in Appendix 1 that for a bandwidth sequence $h_T = cT^{-\frac{1}{2m+d}}$, where $c > 0$ is a constant, the asymptotic Mean Squared Error (MSE) M_T of $\hat{\beta}_T$ achieves the optimal d -dimensional nonparametric rate and is such that $T^{\frac{2m}{2m+d}} M_T \rightarrow M(\Omega, Z, c)$ as $T \rightarrow \infty$, where $M(\Omega, Z, c) > 0$ is a constant depending on weighting matrix Ω , instrument Z and bandwidth constant c .

Definition 2 *The kernel nonparametric efficiency bound $\mathcal{M}(x_0, a)$ for estimating $\beta_0 = E(a|x_0)$ is the smallest possible value of $M(\Omega, Z, c)$ corresponding to the optimal choice of weighting matrix Ω , instrument Z and bandwidth constant c .*

The efficient kernel moment estimator is called XMM estimator. The optimal weighting matrix Ω^* and the optimal bandwidth constant c^* are derived in Appendix 1. In Proposition 2 below, we focus on the set of optimal instruments and the kernel non-parametric efficiency bound. The idea of the proof (see Appendix 1) is based on a linear change of parameter from θ to $\eta^* = (\eta_1^*, \eta_2^*)$, where η_1^* are the linear combinations of θ that are locally identified by uniform moment restrictions (9), while parameter η_2^* is locally identifiable only if both local and uniform moment restrictions (9) and (10) are jointly considered. Linear combinations η_1^* and η_2^* are said full-information identifiable, and full-information underidentified, respectively. An optimal instrument is such that the corresponding marginal moment restrictions identify all full-information identifiable parameters.

Proposition 2 (i) *Any instrument:*

$$Z^* = E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X), \quad (13)$$

where $W(X)$ is a p.d. matrix, P -a.s., is optimal.

(ii) *The kernel non-parametric efficiency bound $\mathcal{M}(a, x_0)$ is given by the lower-right element of matrix*

$$\left(J_0^{*'} \left(\frac{1}{c^{*d}} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{*2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_0^* \right)^{-1}, \quad (14)$$

where $w^2 = \int_{\mathbb{R}^d} K(u)^2 du$, $w_m := \int_{\mathbb{R}} u^m \kappa(u) du$, optimal bandwidth constant c^* is given in Appendix 1,

$$J_0^* = \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R^* & 0 \\ E \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R^* & -1 \end{pmatrix}, \quad R^* = \partial \theta / \partial \eta_2^{*'}, \quad \Sigma_0 = V \left(\begin{pmatrix} g_2 \\ a \end{pmatrix} \middle| x_0 \right),$$

and

$$b(x) = \frac{1}{m!} \frac{1}{f_X(x)} \begin{pmatrix} \Delta^m (E(g_2|x) f_X(x)) - E(g_2|x) \Delta^m f_X(x) \\ \Delta^m (E(a|x) f_X(x)) - E(a|x) \Delta^m f_X(x) \end{pmatrix},$$

with $\Delta^m := \sum_{i=1}^d \partial^m / \partial x_i^m$ and all functions evaluated at θ_0 .

Since we focus on estimation of local moment β_0 , the set of optimal instruments is larger than the standard set of instruments for efficient estimation of a structural parameter θ identified by (9). While in the standard framework $W(X) = (V[g(Y, \theta_0)|X])^{-1}$ is the efficient weighting of the conditionally heteroskedastic moment conditions [e.g., Chamberlain (1987)], any choice of a positive definite matrix $W(X)$ is valid, when β_0 is the parameter of interest. The set of optimal instruments is independent of the selected kernel. Moreover, the matrix in (14) has the form of a GMM efficiency bound for estimating parameters η_2^*, β from orthogonality function $(g_2', a - \beta)'$. Since the moment restrictions are conditional on $X = x_0$, the variance of the orthogonality conditions is replaced by the MSE matrix $\frac{1}{c^{*d}} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{*2m} w_m^2 b(x_0) b(x_0)'$ of the kernel regression estimator (up to a scale factor), and the expectations in J_0^* are conditional on $X = x_0$.

To eliminate the asymptotic bias in Proposition 1, we can restrict our analysis to bandwidth sequences such that $Th_T^{2m+d} = o(1)$ (i.e. $\bar{c} = 0$) and introduce an appropriate efficiency notion in this restricted class of estimators. For a weighting matrix Ω which is diagonal w.r.t. η_1 and (η_2, β) ,¹² the asymptotic variance of $\sqrt{Th_T^d} (\hat{\beta}_T - \beta_0)$ is of the form $w^2 V(Z, \Omega)$, where $V(Z, \Omega)$ is independent of the kernel K .

Definition 3 *The bias free kernel non-parametric efficiency bound $\mathcal{B}(a, x_0)$ is the smallest asymptotic variance $V(Z, \Omega)$ corresponding to the optimal choice of the weighting matrix Ω and of the instrument Z .*

This notion of efficiency is independent of the selected bandwidth sequence h_T and kernel K . The set of optimal instruments is the same as before and is given by (13).

Corollary 3 *The bias free kernel non-parametric efficiency bound $\mathcal{B}(a, x_0)$ is*

$$\begin{aligned} \mathcal{B}(x_0, a) &= \frac{1}{f_X(x_0)} \{V(a) - \text{Cov}(a, g_2)V(g_2)^{-1}\text{Cov}(g_2, a) \\ &\quad + \left[E\left(\frac{\partial a}{\partial \theta'}\right) R^* - \text{Cov}(a, g_2)V(g_2)^{-1}E\left(\frac{\partial g_2}{\partial \theta'}\right) R^* \right] \left[R^{*\prime} E\left(\frac{\partial g_2}{\partial \theta}\right) V(g_2)^{-1}E\left(\frac{\partial g_2}{\partial \theta'}\right) R^* \right]^{-1} \\ &\quad \left[R^{*\prime} E\left(\frac{\partial a}{\partial \theta}\right) - R^{*\prime} E\left(\frac{\partial g_2}{\partial \theta}\right) V(g_2)^{-1}\text{Cov}(g_2, a) \right] \}, \end{aligned}$$

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

When the structural parameter θ itself is full-information identifiable, the bias free kernel nonparametric efficiency bound becomes:

$$\mathcal{B}(x_0, a) = \frac{1}{f_X(x_0)} \{V(a|x_0) - \text{Cov}(a, g_2|x_0)V(g_2|x_0)^{-1}\text{Cov}(g_2, a|x_0)\}. \quad (15)$$

Since the conditional moment of interest is also equal to $E(a|x_0) = E[a(Y; \theta_0) - \text{Cov}(a, g_2|x_0)V(g_2|x_0)^{-1}g_2(Y; \theta_0) | x_0]$, the bound (15) is simply 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 [see also Brown and Newey (1998)], and extended to a conditional framework by Antoine,

¹²It is sufficient to consider this class, since the optimal weighting matrix satisfies this condition (see Appendix 1).

Bonnal and Renault (2006). In the general case, the formula of $\mathcal{B}(x_0, a)$ in Corollary 3 distinguishes the gain in information coming from the conditional moment restrictions and the cost due to full-information underidentification of θ . In the rest of the paper, we assume that the bandwidth is selected, such that the asymptotic bias can be neglected, and we focus on the efficiency notion in Definition 3.

4 Information based XMM estimator

4.1 Lack of coherency of the basic XMM estimator

It is expected that an appropriate estimation approach for the conditional moment (derivative price) $E(a|x_0) = E[a(Y; \theta_0)|X = x_0]$ provides an estimator of the type:

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

where $\tilde{\theta}$ is an estimator of θ and \tilde{f} is an estimator of the conditional density. This condition is not satisfied by the basic XMM estimator. Therefore, in the application to derivative pricing, the basic XMM may yield an estimated state price density with negative values. This is not compatible with the absence of arbitrage opportunity. This feature will be corrected by introducing an information based XMM estimator.

4.2 Information based XMM estimator

The (unconstrained) kernel estimator $\hat{f}(y|x)$ is a consistent estimator of the conditional pdf, but does not take into account the parameterized moment restrictions. The kernel density estimator can be improved by looking for the pdf that is the closest to $\hat{f}(y|x)$, and satisfies the moment restrictions. We consider the joint estimator:

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

The objective function includes two components: a chi-square distance is used for the optimization with respect to the conditional distributions associated with the sample values of the conditioning variable, whereas 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 . The two types of constraints are taken into account: 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 $\hat{f}^*(\cdot|x_0)$ satisfying the unit mass and positivity restrictions.¹³ In particular, the computation of the estimator only requires 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 XMM estimator of the conditional moment of interest is defined by:

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

The asymptotic properties of this estimator are derived in Appendix 2.

Proposition 4 *When the bandwidth is such that $Th_T^{d+2m} = o(1)$, the estimator $\hat{E}^*(a|x_0)$ is consistent, converges at rate $\sqrt{Th_T^d}$, is asymptotically normal and kernel nonparametrically efficient:*

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

for all a .

4.3 Application to derivative pricing

The information based XMM estimator can be applied to derivative pricing once the pricing model and the uniform and local moment conditions have been made explicit. We consider a semi-parametric pricing model in which the historical transition is

¹³See e.g. Stutzer (1996) and Kitamura, Stutzer (1997).

functional and the risk premia parameter involved in the sdf is finite dimensional. We assume a finite number n of derivative prices observed at a given date t_0 (the current day, say) and T serial observations of the return of the underlying asset corresponding to the current and previous days $t = t_0 - T + 1, \dots, t_0$. The data set is doubly indexed by n and T . The asymptotics is with respect to T , $T \rightarrow \infty$, with n fixed. 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 (which applies a pricing formula) and the econometrician (which applies information based XMM). Then, the relative price at t of a European call with moneyness strike k and time-to-maturity 1 can be written as:

$$c_t(k, 1) = E [m(Y_{t+1}, Y_t; \theta)(\exp r_{t+1} - k)^+ | Y_t], \quad (16)$$

where $m(Y_{t+1}, Y_t; \theta)$ is the stochastic discount factor, which depends on the future and current value of the state variables. The finite dimensional parameter θ characterizes the risk premia, whereas the historical conditional distribution of Y_{t+1} given Y_t is left unspecified. Then, the moment restrictions are twofold. Some local constraints concern the observed derivative prices at t_0 , and are given by:

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

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

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

respectively, assuming for simplicity a zero riskfree rate. Further, let us assume that our interest is in the price at date t_0 of a European call with time-to-maturity 1 and moneyness strike k . Its relative price is equal to the conditional moment:

$$E(a(k) | y_{t_0}) := E[m(Y_{t+1}, Y_t; \theta)(\exp r_{t+1} - k)^+ | Y_t = y_{t_0}]. \quad (19)$$

There is a total set of $n + 2$ local moment restrictions $E(g_2|y_{t_0}) = 0$ given by:

$$E[m(Y_{t+1}, Y_t; \theta)(\exp r_{t+1} - k_j)^+ - c_{t_0}(k_j, 1)|Y_t = y_{t_0}] = 0, \quad j = 1, \dots, n,$$

$$E[m(Y_{t+1}, Y_t; \theta) - 1|Y_t = y_{t_0}] = 0, \tag{20}$$

$$E[m(Y_{t+1}, Y_t; \theta) \exp r_{t+1} - 1|Y_t = y_{t_0}] = 0.$$

The application of information based XMM to this setting has the following advantages:

- (i) It is compatible with the absence of arbitrage opportunities;
- (ii) It uses both historical and cross-sectional data;
- (iii) It can be applied independently at several consecutive dates, and in this setting the number and characteristics of the options considered are allowed to vary with t_0 ;
- (iv) When it is applied independently at several consecutive dates $t_0, t_0 + 1, \dots$, the successive estimators of the risk premium parameter θ are less erratic than in a pure cross-sectional approach due to the common use of historical data;
- (v) If the semi-parametric pricing model is well-specified, the estimated parameters and option prices are consistent whenever T tends to infinity, even if the number of options actively traded at t_0 is small. It may seem surprising to get consistent estimation of derivative prices, when only a small number of derivatives are included in the estimation procedure. In fact, the (information based) XMM approach implicitly creates additional "artificial" observed derivative prices by using the deterministic pricing formula, which is continuous with respect to the conditioning variable. More precisely, at any date $t < t_0$ for which the conditioning variables are close to the current value $y_t \sim y_{t_0}$, artificial observations are created by applying the pricing formula to derivatives with the same strikes $k_j, j = 1, \dots, n$ and times-to-maturity, but computed for y_{t_0} instead of y_t .

5 XMM estimation in practice

We will now illustrate the theoretical results presented in the previous sections and discuss the practical implementation of the information based XMM on real data. For the first purpose, we consider a true dynamics for the asset and derivatives prices that is a discrete time version of the stochastic volatility model of Heston (1993), Ball, Roma (1994) with a risk premium introduced in the return equation. From this DGP, we deduce the parametric expressions of the sdf and of the option prices (see Section 5.1). In Section 5.2, we focus on some theoretical properties of the information based XMM approach corresponding to a semi-parametric model with the above parametric specification of the sdf and an unconstrained transition density for the state variables. We discuss the identifiability of the parameters and provide the patterns of the kernel nonparametric efficiency bounds. In Section 5.3, we consider data on S&P 500 options for the trading days in June 2005. We apply the standard pure cross-sectional calibration method [see Section 2.2 iii)] using the parametric expression of option prices implied by the parametric stochastic volatility model. We also apply the information based XMM estimator using the parametric sdf.

5.1 The data generating process

In the DGP, there is 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, 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 and Jasiak (2006), Darolles, Gouriéroux and Jasiak (2006)]. The transition distribution of the stochastic volatility is defined through the conditional Laplace transform (conditional moment generating

function):

$$E \left[\exp(-u\sigma_{t+1}^2) \mid \underline{\sigma}_t^2 \right] = \exp[-a(u)\sigma_t^2 - b(u)], \quad (22)$$

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 the no-arbitrage restrictions and provides simple pricing formulas.

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

$$E_t(M_{t,t+1}) = 1, \quad E_t(M_{t,t+1} \exp r_{t+1}) = 1.$$

We have the following proposition, proved in Appendix 3.

Proposition 5 *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 with a riskfree asset and a 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 [Gouriéroux and Monfort (2007)].

The relative price at t of a European call with moneyness strike k and time-to-maturity h is given by:

$$c_t(k, h) = E_t \left(M_{t,t+1} \dots M_{t+h-1,t+h} [\exp(r_{t+1} + \dots + r_{t+h}) - k]^+ \right). \quad (25)$$

The option price can be written in terms of the Black-Scholes price and integrated volatility $\sigma_{t+1}^2(h) = (\sigma_{t+1}^2 + \dots + \sigma_{t+h}^2) / h$. We get:

$$c_t(k, h) = E_t^Q [BS(h, k, \sigma_{t+1}^2(h))], \quad (26)$$

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

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

5.2 Theoretical properties of the information based XMM estimator

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 . We consider a semi-parametric model in which the sdf is compatible with the DGP above:

$$M_{t,t+1}(\theta) = \exp(-\nu_0 - \nu_1 \sigma_{t+1}^2 - \nu_2 \sigma_t^2 - \nu_3 r_{t+1}), \quad (28)$$

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

i) Full-information identifiability

Let us now discuss the identifiability of parameter θ from the uniform conditional

moment 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 5, three independent linear combinations of parameter θ are full-information identifiable, including parameter ν_3 , which implies $\dim \eta_1^* = 3$ and $\dim \eta_2^* = 1$ (see Section 3.4). The Jacobian matrix $R^* = \partial\theta/\partial\eta_2^{*'} w.r.t.$ the full-information underidentified directions is (see Appendix 3):

$$R^* = \left(-\delta \frac{c}{1+c(\nu_1+\gamma^2/2-1/8)}, \quad 1, \quad -\rho \frac{1}{[1+c(\nu_1+\gamma^2/2-1/8)]^2}, \quad 0 \right)'. \quad (29)$$

These directions are not fixed a priori, but depend on the parameter values. Parameters ν_0, ν_1, ν_2 are identified only if the cross-sectional restrictions from observed derivative prices are taken into account, and will be estimated at nonparametric rates. Conversely, the estimator of ν_3 features a parametric rate of convergence.

ii) Kernel nonparametric efficiency bound

Let us now derive the kernel efficiency bounds for derivative prices corresponding to the stochastic volatility DGP with parameters given by $\gamma = 0.5$, $\rho = 0.96$, $\delta = 3.60$, $c = 7.34 \cdot 10^{-7}$, $\nu_0 = 3.82 \cdot 10^{-2}$, $\nu_1 = -1.44 \cdot 10^4$, $\nu_2 = 1.40 \cdot 10^4$ and $\nu_3 = 1$. Historical parameters ρ , δ and c are such that the ARG volatility process matches the stationary mean, variance and first-order autocorrelation 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. 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 a shock on the volatility surprise $\nu_1 < 0$.

Let us first assume that, at the current date t_0 , we observe the price of $n = 3$ actively traded call options with moneyness strikes $k_1 = 0.9$, $k_2 = 1$, $k_3 = 1.1$, respectively, and time-to-maturity $h = 40$ days. The kernel nonparametric efficiency bound to estimate call option prices at time-to-maturity $h = 40$ is displayed in Figure

5 for a conditioning value y_{t_0} corresponding to a return r_{t_0} equal to 0 and a variance $\sigma_{t_0}^2$ equal to the historical mean (see Corollary 3).

[Insert Figure 5: Kernel nonparametric efficiency bound, time-to-maturity 40]

The solid line corresponds to the call price $E(a(k)|y_{t_0})$, the dashed lines to confidence intervals $E(a(k)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, k)^{1/2}$, as a function of moneyness k , 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 width of the confidence band is much narrower.¹⁴ 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, where the bands are narrow, the truncation effect is negligible and arises only for large strikes. The width of the confidence interval for derivative price $E(a(k)|y_{t_0})$ depends on moneyness strike k . The interval is wider for almost at-the-money options, whereas it is narrower when the derivative is deep in-the-money, or deep out-of-the-money. 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.¹⁵ Finally, the width of the interval is zero when k corresponds to the moneyness strikes of the observed calls.

In Figure 6, we display the kernel nonparametric efficiency bound for estimating a European call with time-to-maturity $h = 40$, when the prices of three derivatives with time-to-maturity 20 and moneyness strikes $k = 0.9, 1, 1.1$, respectively, are observed.

[Insert Figure 6: Kernel nonparametric efficiency bound, time-to-maturity 40.
Observed derivative prices at time-to-maturity 20]

The time-to-maturity of the observed derivatives is not equal to the time-to-maturity of interest. This explains why the kernel nonparametric efficiency bound is much larger compared with Figure 5 and, in particular, it is different from zero for all

¹⁴For the sample size, the bandwidth and the kernel used in the application (Section 5.3), we have $\sqrt{w^2/Th_T^2} f_X(x_0) \simeq .007$ and the bands are about 300 times narrower than the bands in Figure 5.

¹⁵However, the relative accuracy can be poor in these moneyness regions.

moneyiness strikes.

5.3 Application to S&P 500 options

Let us now compare the pure cross-sectional calibration and the information based XMM approaches on S&P 500 options in June 2005 (see Section 2.1 for a description of the dataset). For this purpose, we adopt a parametric risk-neutral specification for the calibration approach and a parametric sdf for the XMM approach, both compatible with the model described in Section 5.1. To account for a non-zero riskfree rate, factor r_t is now the underlying asset return in excess of the riskfree rate $r_{f,t}$, parameter ν_0 is replaced by $\nu_0 + r_{f,t}$ in the sdf, and k by the discounted moneyiness strike $B(t, t+h)k$. The term structure of riskfree interest rates is assumed exogenous, and is estimated at t_0 by a cubic spline interpolation of market yields at available maturities. It is important to remark that we do not assume the complete parametric specification (21), (22), (24), (27). The calibration approach assumes only the parametric risk-neutral ARG dynamics (27), but leaves the sdf unspecified. The XMM approach assumes only the parametric sdf (24), but leaves the historical transition unspecified. Both approaches require the definition of the volatility factor σ_t . To follow the usual practice, the factor value will be estimated daily in the calibration approach. For the XMM approach, the volatility factor has to be observable; we will consider the standard RiskMetrics volatility.

i) Cross-sectional calibration

From equations (26) and (27), relative option prices are function of the current value of volatility σ_t and of the three parameters $\theta = (c^*, \delta^*, \rho^*)$ characterizing the risk-neutral distribution of stochastic volatility: $c_t(h, k) = c(h, k; \theta, \sigma_t)$. Function c is computed by Fourier Transform methods as in Carr and Madan (1999). The calibrated parameter $\hat{\theta}_{t_0}$ and volatility $\hat{\sigma}_{t_0}$ for the trading days of June 2005 are displayed in Table 2.

[Insert Table 2: Calibrated parameters (cross-sectional approach) for S&P 500 options in June 2005]

We also report the Root Mean Squared Errors $RMSE_{t_0} = \left\{ \frac{1}{n_{t_0}} \sum_{j=1}^{n_{t_0}} \left[c \left(h_j, k_j; \hat{\theta}_{t_0}, \hat{\sigma}_{t_0}^2 \right) - c_{t_0}(h_j, k_j) \right]^2 \right\}^{1/2}$ as goodness of fit measure. The calibrated parameters $\hat{\delta}_{t_0}^*$, $\hat{\rho}_{t_0}^*$ and

$\hat{c}_{t_0}^*$, the calibrated volatility $\hat{\sigma}_{t_0}$, and the goodness of fit measure are highly varying with the day. Likely, the variability of the goodness of fit is due to a large extent to the small and time-varying number of observations in the calibration.

ii) Information based XMM estimation

Let us now consider the semi-parametric setting defined in Section 5.2. Excess return r_t is computed as the logarithmic return from daily closing prices of the S&P 500 index minus the daily yield of a 1-month US Treasury bond. In the XMM approach, the factors are assumed observable by both the investor and the econometrician. We assume that the observed volatility σ_t^2 is obtained by the exponential smoothing filter $\sigma_t^2 = (1 - \lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$, with $\lambda = .94$, as in RiskMetrics. For each trading day t_0 of June 2005, we estimate by information based XMM the structural parameter $\theta = (\nu_0, \nu_1, \nu_2, \nu_3)'$ and 5 option prices $c_{t_0}(k, h)$ at constant time-to-maturity $h = 20$ and moneyness strikes $k = .96, .98, 1, 1.02, 1.04$, corresponding to puts for $k < 1$, and to calls for $k \geq 1$. The estimator is defined as in Section 4.3, using the current and previous $T = 1000$ historical observations on the state variables, and the derivative prices at t_0 of S&P 500 options with daily trading volume larger than 4000 contracts (see Section 2.1). To simplify, we treat the index as a tradable asset.¹⁶ The details on the implementation of the estimator are given in Appendix 2.4. The estimation results are displayed in Table 3.

[Insert Table 3: Estimated structural parameters and option prices (information based XMM approach) for S&P 500 options in June 2005]

A direct comparison of the estimated structural parameters given in Tables 2 and 3 is difficult, even if the complete parametric model of Section 5.1 is well-specified. On the one hand, the calibration method identifies three functions of the parameters, which do not allow for deducing the values of the four risk premia parameters ν [see the expressions in (27) and Proposition 5]. On the other hand, the XMM approach estimates the risk premia parameters and the unrestricted transition density. Thus, it is possible to deduce XMM estimates of the risk-neutral parameters δ^*, ρ^*, c^* involved

¹⁶Since the index is computed without dividends, a more precise approach consists in correcting for dividends the no-arbitrage restriction on the index, without correcting the restrictions for the derivatives. This correction is often neglected in practice and is of minor importance compared to partial observability of option prices, which is the focus of our analysis.

in the calibration. Instead of performing this academic exercise, which assumes that the parametric model is well-specified, we focus directly on a comparison of estimated option prices. However, even if the parameters are difficult to compare, the XMM estimated parameters use the historical information and are much more stable over time. Finally, from the computational point of view, the calibration method is rather time consuming since it requires the inversion of the Fourier Transform for all option prices, for each evaluation of the criterion function and its partial derivatives w.r.t. the parameters, and for each step of the optimization algorithm. On the contrary, the complete XMM estimation on a given day requires about one minute.

iii) Comparison of estimated option prices

In Figure 7 we display the estimated relative prices of S&P 500 options in June 1st and June 2nd 2005, as a function of discounted moneyness strike and time-to-maturity.

[Insert Figure 7: Estimated call and put price functions for S&P 500 options in June 1st and June 2nd, 2005]

The upper two Panels refer to June 1st 2005, the lower two to June 2nd 2005. In the right Panels, the solid lines correspond to XMM estimates, and in the left Panels to calibrated estimates. Circles correspond to observed prices of highly traded options. On June 1st, there are four times-to-maturity with highly traded options. They correspond to 12, 57, 77, and 209 days, respectively. For the long time-to-maturity 209, the traded option is one call (resp., one put for $h = 57$ and two calls for $h = 77$). The other highly traded options correspond to time-to-maturity 12. In some cases, put and call with identical moneyness strike can be highly traded, and can immediately reveal that the prices of these liquid options are not compatible with the no-arbitrage assumptions. As usual in the literature [see e.g. Aït-Sahalia and Lo (1998)], when this problem arises, a single type of option is selected for each moneyness strike, that is a put if $k < 1$, a call, otherwise (unless arbitrage opportunities are introduced, in which case no option is selected). After this operation, we get 7 highly traded options for time-to-maturity 12. Similarly, we get 8 highly traded options for 3 times-to-maturity in June 2nd 2005. Both calibration and XMM methodologies can be used in practice for pricing any type of put and call, in particular derivatives with no highly traded times-to-maturity or no highly traded moneyness strikes. For

this reason, Figure 7 also includes (dashed lines) the times-to-maturity 120 days for June 1st, and 119 days for June 2nd 2005, which are not traded. As usual, the reported option prices correspond to puts if $k < 1$, to calls, otherwise. The calibration method assumes a parametric risk-neutral model, whereas the XMM risk-neutral model is semiparametric. This explains why the XMM approach is more flexible. For instance, by construction the XMM prices coincide with the market prices for highly liquid traded options, whereas there is some discrepancy with the calibration method, and this discrepancy can be large. This feature is not due to the special parametric risk-neutral model which has been selected. It will also be observed with more complicated parametric models, or with the nonparametric approach of Aït-Sahalia and Lo (1998). By construction, the risk-neutral stochastic volatility model, which underlies the calibration method, provides smooth symmetric option pricing functions, with similar types of curvature when the time-to-maturity increases. The more flexible XMM approach provides skewed option pricing functions, which means that the method is naturally able to capture the so-called leverage effect. Moreover, the curvature and the leverage effects can be adjusted as a function of time-to-maturity in a less constrained way compared to fully parametric models. Similarly, the XMM method will be also more flexible w.r.t. changes of the benchmark daily traded volume used to select the options, for instance passing from 4000 to 3000 traded contracts.

Let us now consider the dynamics of the option pricing function. In Figure 8 we display the time series of implied volatilities at fixed time-to-maturity $h = 20$ and moneyness strikes $k = .96, .98, 1, 1.02, 1.04, 1.06$ for all trading days of June 2005.

[Insert Figure 8: Time series of implied volatilities for S&P 500 options in June 2005]

Circles correspond to XMM implied volatilities and squares to calibrated implied volatilities. As a result of the use of the historical information on the underlying asset, XMM implied volatilities are more stable over time, especially for moneyness strikes close to at-the-money. Since most of the highly traded options are for moneyness strikes close to the money, the calibration approach is rather sensitive to partial observability of option prices with extreme strikes.

The average levels of the two time series of implied volatilities in Figure 8 are different. In particular, for $k \leq 1$, XMM implied volatilities are typically larger than

calibrated ones, whereas the opposite holds for $k > 1.02$. This is because the more flexible XMM approach captures the smirk in the implied volatility curve, while the calibrated model features a smile or a flat pattern. For the at-the-money strike $k = 1$, the two implied volatility time series are close in the trading days around Friday June 17th 2005, when time-to-maturity $h = 20$ is highly traded. Finally, we observe some evidence of the week-end effect in XMM implied volatilities for $k = .98$, while this effect is hidden by noise for the calibrated implied volatilities.

6 Concluding remarks

There exist significant differences between estimated S&P 500 option prices reported in the literature for similar periods. As noted by Singleton (2006), p. 404, "the differences may be partly attributable to the different estimation strategies", which can use cross-sectional or time series approaches. Likely, they are also largely due to the treatment of partial observability of option prices. The aim of our paper was to address this missing data challenge. We have noted that the partial observability of an asset price can be represented by a local moment condition. This leads to extend the standard GMM approach to account for both uniform and local conditional moment restrictions. We have developed the complete asymptotic theory for the basic and information based XMM estimators. The XMM approach to derivative pricing is compatible with the absence of arbitrage opportunities, uses both historical and cross-sectional data, provides estimates which are stable over time, gives an accurate estimator even if the number of observed option prices is rather small, and is numerically tractable.

For the derivative pricing problem, the XMM approach is easily extended when the option prices are selected on a small number of consecutive days ($t_0, t_0 - 1, t_0 - 2$, say), possibly with decreasing weights, as is usually done in the trading rooms for pure cross-sectional methods. The XMM method cannot include however all prices of actively traded options, due to the nonstationarity of the issuing procedure and the selectivity effects. The modelling and analysis of activity on derivative markets is left for long term future research.

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 Returns", *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.
- 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, eds. Engle, R., and D., McFadden, Amsterdam, North Holland, 2247-2294.
- Andrews, D., and J., Stock (2006): "Inference with Weak Instruments", in *Advances in Economics and Econometrics. Theory and Application*, 9th World Congress of the Econometric Society, eds. Blundell, R., Newey, W., and T., Persson, Cambridge University Press.
- Antoine, B., Bonnal, H., and E., Renault (2007): "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood", *Journal of Econometrics*, 138, 461-487.
- 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.
- 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.
- Bekaert, G., and G., Wu (2000): "Asymmetric Volatility and Risk in Equity Markets", *Review of Financial Studies*, 13, 1-42.

- Billingsley, P. (1965): *Ergodic Theory and Information*. John Wiley, New York.
- Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley, New York.
- Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Springer, New-York.
- Brennan, M., Liu, X., and Y., Xia (2005): "Option Pricing Kernels and the ICAPM", Working Paper.
- Brown, B., and W., Newey (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.
- Buraschi, A., and J., Jackwerth (2001): "The Price of a Smile: Hedging and Spanning in Option Markets", *Review of Financial Studies*, 14, 495-527.
- Carr, P., and D., Madan (1999): "Option Pricing and the Fast Fourier Transform", *Journal of Computational Finance*, 2, 61-73.
- Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.
- Chernov, M., and E., Ghysels (2000): "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation", *Journal of Financial Economics*, 56, 407-458.
- Darolles, S., Gouriéroux, C., and J., Jasiak (2006): "Structural Laplace Transform and Compound Autoregressive Models", *Journal of Time Series Analysis*, 27, 477-504.
- Duan, J. (1994): "Maximum Likelihood Estimation Using Price Data of the Derivative Contract", *Mathematical Finance*, 4, 155-167.
- Eraker, B. (2004): "Does Option Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices", *Journal of Finance*, 59, 1367-1403.
- Gouriéroux, C., and J., Jasiak (2006): "Autoregressive Gamma Process", *Journal of Forecasting*, 25, 129-152.
- Gouriéroux, C., and A., Monfort (2007): "Econometric Specification of Stochastic Discount Factor Models", *Journal of Econometrics*, 136, 509-530.

Hansen, B. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays", *Econometric Theory*, 12, 347-359.

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 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.

Harrison, J., and D., Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381-408.

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.

Hull, J. (2005): *Options, Futures and Other Derivatives*, 6.th Edition, Prentice Hall.

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

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

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

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

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

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

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

Pan, J. (2002): "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study", *Journal of Financial Economics*, 63-50.

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.

Santa-Clara, P., and S., Yan (2004): "Jump and Volatility Risk and Risk Premia: A New Model and Lessons from S&P 500 Options", NBER Working Paper.

Schwartz, E. (1987): "The Stochastic Behaviour of Commodity Prices: Implications for Valuation and Hedging", *Journal of Finance*, 52, 923-973.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Singleton, K., (2006): *Empirical Dynamic Asset Pricing*, Princeton University Press.

Staiger, D., and J., Stock (1997): "Instrumental Variables with Weak Instruments", *Econometrica*, 65, 557-586.

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

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 Paris VI.

Figure 1: Times-to-maturity of quoted S&P 500 options in 2005.

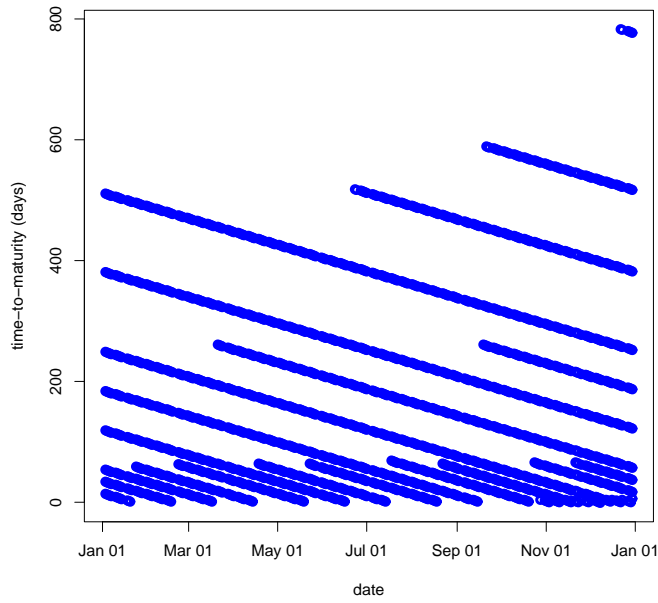
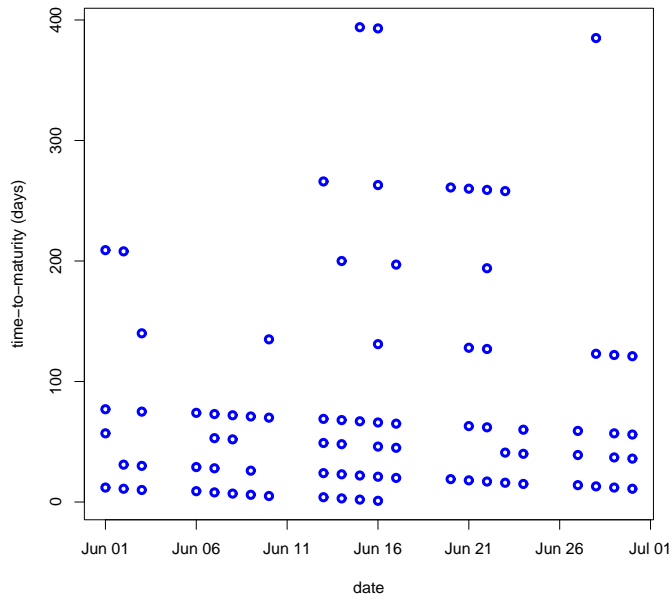
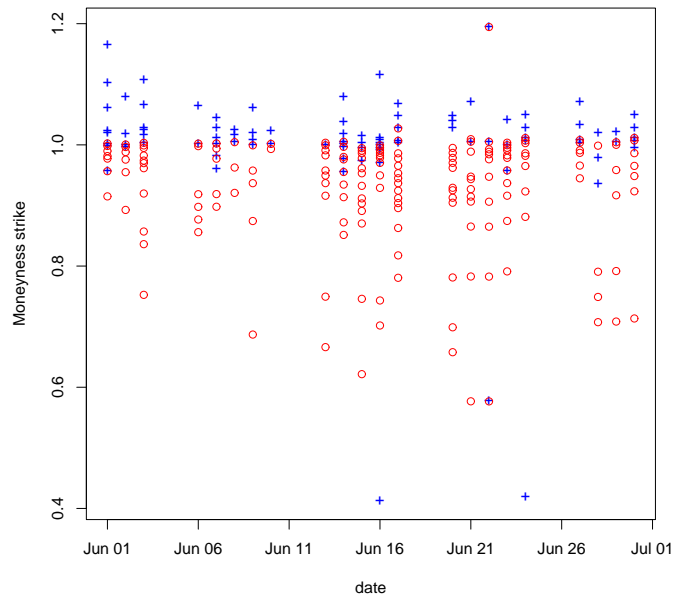


Figure 2: Times-to-maturity of actively traded S&P 500 options in June 2005.



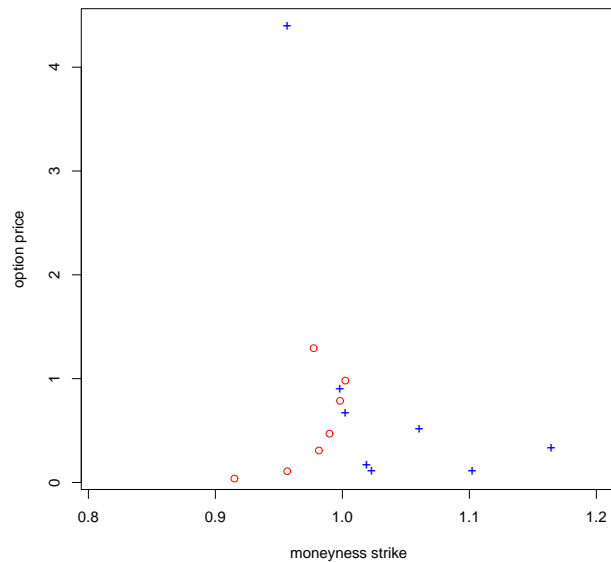
Traded options on the S&P 500 index with daily volume larger than 4000 contracts are considered.

Figure 3: Moneyness strikes of actively traded S&P 500 options in June 2005.



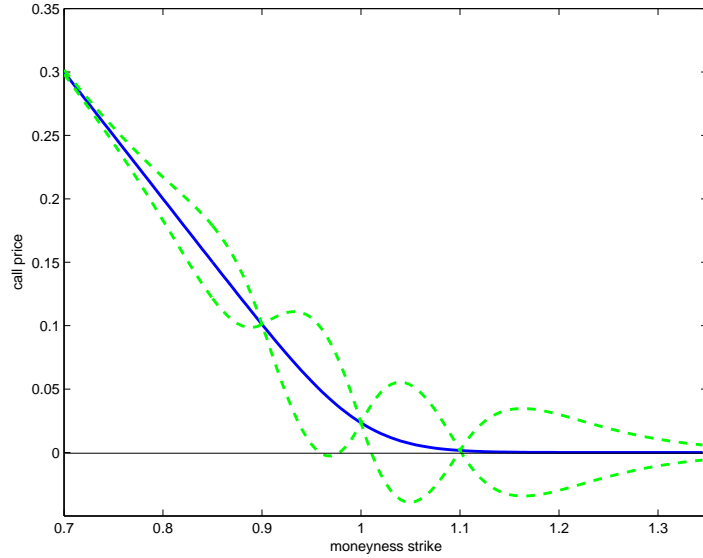
Traded options on the S&P 500 index with daily volume larger than 4000 contracts are considered. Crosses correspond to call contracts, and circles to put contracts.

Figure 4: Prices of actively traded S&P 500 options in June 1st 2005.



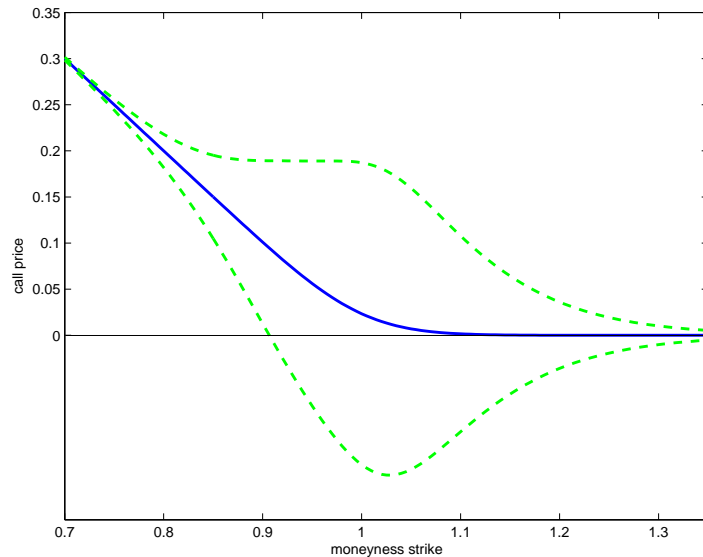
Traded options on the S&P 500 index with daily volume larger than 4000 contracts are considered. Relative option prices are reported in percentage. Crosses correspond to call contracts, circles to put contracts, respectively.

Figure 5: Kernel nonparametric efficiency bound, time-to-maturity 40.



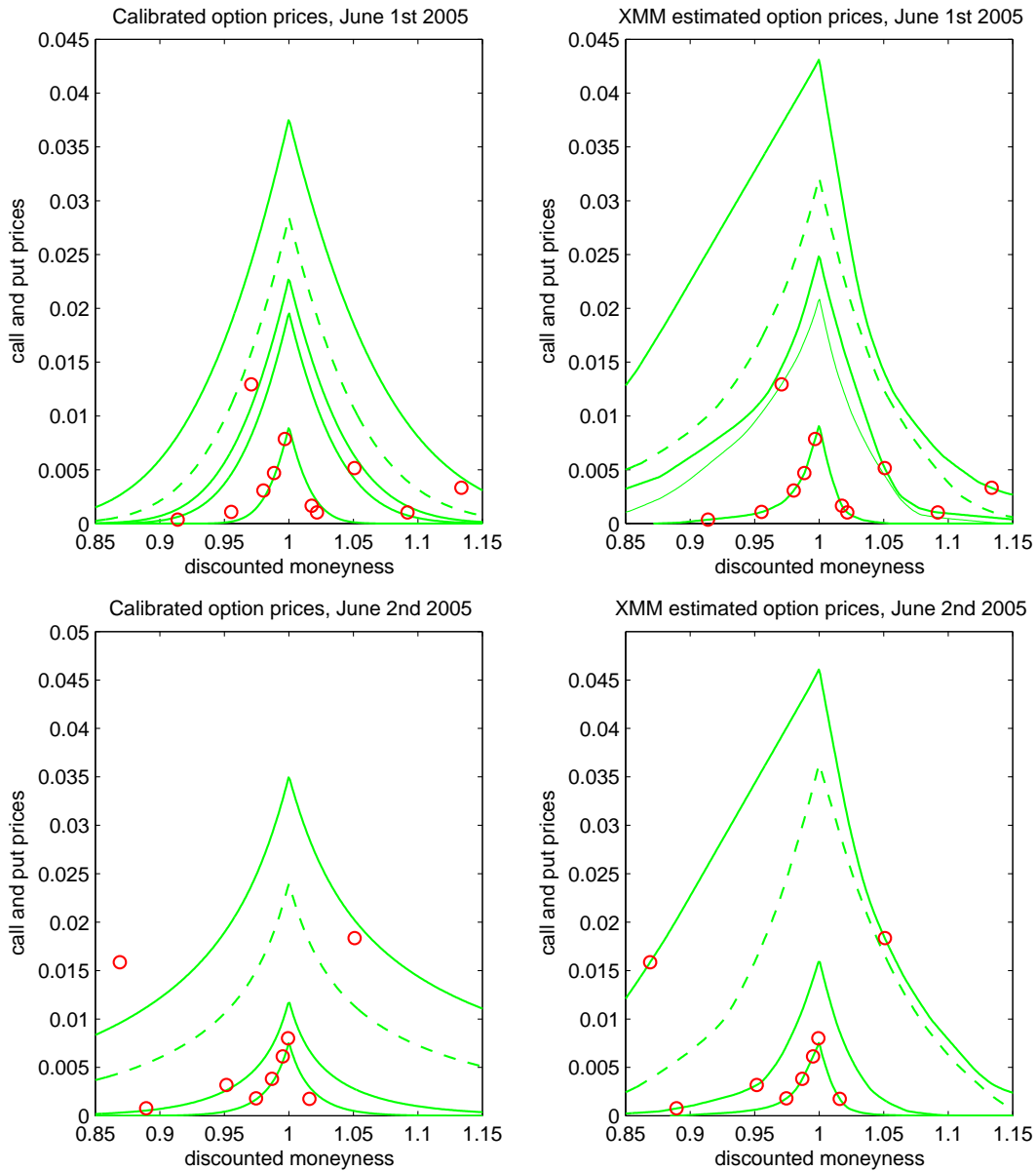
Derivative prices for moneyness strikes 0.9, 1, 1.1 and time-to-maturity 40 are observed. The solid line corresponds to the relative price $E(a(k)|y_{t_0})$, the dashed lines to pointwise 95% symmetric confidence intervals $E(a(k)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, k)^{1/2}$, using the standardization $\sqrt{w^2/Th_T^2 f_X(x_0)} = 2$.

Figure 6: Kernel nonparametric efficiency bound, time-to-maturity 40. Observed derivative prices at time-to-maturity 20.



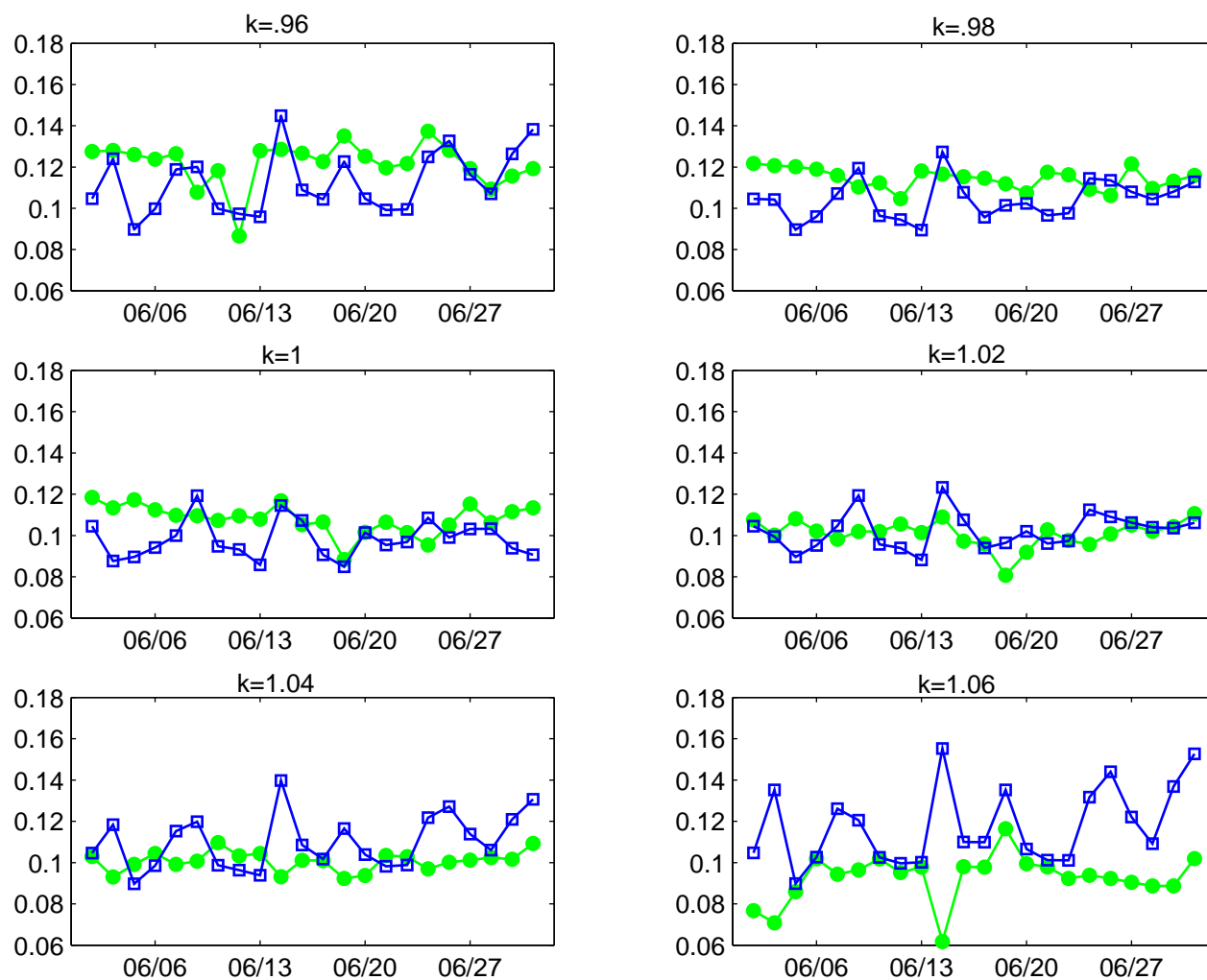
Derivative prices for moneyness strikes 0.9, 1, 1.1 and time-to-maturity 20 are observed. The solid line corresponds to the relative price $E(a(k)|y_{t_0})$ at time-to-maturity 40, the dashed lines to pointwise 95% confidence intervals $E(a(k)|y_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(y_{t_0}, k)^{1/2}$, using $\sqrt{w^2/Th_T^2 f_X(x_0)} = 2$.

Figure 7: Estimated call and put price function for S&P 500 options at June 1st and June 2nd, 2005.



In the upper right Panel, the solid lines correspond to estimated relative option prices as a function of discounted moneyness strike $B(t, t+h)k$ for the highly traded times-to-maturity $h = 12, 57, 77, 209$ at June 1st 2005, obtained by information based XMM. The dashed line corresponds to XMM estimated prices for the non-traded time-to-maturity $h = 120$. The price curves correspond to puts if $B(t, t+h)k < 1$, to calls otherwise. In the upper left Panel, the solid and dashed lines are the price curves obtained by the parametric pricing model of Section 5.1 with the calibrated parameters in Table 2 for times-to-maturity $h = 12, 57, 77, 209$, and $h = 120$, respectively. In both Panels, circles correspond to observed S&P 500 option prices with daily trading volume larger than 4000 contracts. The two lower Panels correspond to June 2nd, 2005 with highly traded times-to-maturity $h = 11, 31, 208$, and non-traded time-to-maturity $h = 119$.

Figure 8: Time series of implied volatilities of S&P 500 options in June 2005.



In Panels 1-6, annualized implied volatilities at time-to-maturity $h = 20$ and moneyness strike $k = .96$, $k = .98$, $k = 1$, $k = 1.02$, $k = 1.04$ and $k = 1.06$, respectively, are displayed for each trading day in June 2005. Circles are implied volatilities computed from option prices estimated by the information based XMM approach, squares are implied volatilities from the cross-sectional calibration approach. The ticks on the horizontal axis correspond to Mondays.

Table 2: Calibrated parameters (cross-sectional approach) for S&P 500 options in June 2005.

Day t_0	Calibrated parameters				Goodness of fit $RMSE_{t_0}$
	$\hat{\delta}_{t_0}^*$	$\hat{\rho}_{t_0}^*$	$\hat{c}_{t_0}^* (\times 10^{-7})$	$\hat{\sigma}_{t_0}$	
1.6.05	24.7303	0.47588	9.59115	0.00347	0.00161
2.6.05	0.05400	0.99994	94.2208	0.00670	0.00219
3.6.05	87.7199	0.99999	0.09906	0.00480	0.00114
6.6.05	1.68952	0.99625	6.57284	0.00519	0.00194
7.6.05	0.41250	0.99998	41.2201	0.00563	0.00071
8.6.05	10.8766	0.81153	12.5623	0.00102	0.00061
9.6.05	2.26166	0.99106	6.58317	0.00501	0.00191
10.6.05	3.15278	0.96637	6.79999	0.00499	0.00032
13.6.05	0.21085	0.99909	8.84000	0.00555	0.00035
14.6.05	0.02125	0.99999	101.489	0.00874	0.00267
15.6.05	7.37374	0.88993	10.5002	0.00238	0.00055
16.6.05	0.61316	0.99909	18.7014	0.00510	0.00103
17.6.05	0.13482	0.98520	197.089	0.00521	0.00107
20.6.05	3.60121	0.96010	7.00061	0.00567	0.00071
21.6.05	3.61166	0.96181	7.00976	0.00493	0.00244
22.6.05	4.97660	0.91445	8.50086	0.00495	0.00179
23.6.05	0.43675	0.99909	39.2002	0.00630	0.00047
24.6.05	0.09603	0.99999	109.058	0.00709	0.00158
27.6.05	0.44010	0.99909	23.0004	0.00624	0.00163
28.6.05	4.05385	0.99342	7.28798	0.00394	0.00080
29.6.05	0.10119	0.99999	88.8541	0.00664	0.00167
30.6.05	0.04738	0.99999	277.505	0.00754	0.00194

Calibrated parameter $\hat{\theta}_{t_0}$, volatility $\hat{\sigma}_{t_0}$ and goodness of fit measure $RMSE_{t_0}$ for each trading day t_0 of June 2005. Calibration is performed using a Fourier Transform approach to compute option prices. At each day t_0 , the sample consists of the derivative prices at t_0 of S&P 500 options with daily volume larger than 4000 contracts.

Table 3: Estimated structural parameters and option prices (information based XMM approach) for S&P 500 options in June 2005.

Day t_0	Structural parameters				Option prices ($\times 10^{-2}$)				
	$\hat{\nu}_0$ ($\times 10^{-6}$)	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}_3$	$k = .96$	$k = .98$	$k = 1$	$k = 1.02$	$k = 1.04$
1.6.05	1.163	-.126	.165	.860	.201	.527	1.227	.537	.144
2.6.05	1.152	-.127	.165	.860	.203	.518	1.169	.466	.097
3.6.05	1.152	-.127	.165	.860	.192	.513	1.213	.543	.125
6.6.05	1.136	-.127	.166	.860	.181	.502	1.157	.486	.152
7.6.05	1.125	-.126	.165	.860	.195	.475	1.128	.449	.125
8.6.05	1.127	-.127	.165	.860	.106	.427	1.125	.484	.133
9.6.05	1.114	-.125	.163	.860	.153	.444	1.100	.483	.180
10.6.05	1.104	-.125	.164	.860	.038	.378	1.125	.518	.146
13.6.05	1.091	-.125	.163	.860	.203	.494	1.107	.478	.152
14.6.05	1.081	-.124	.163	.860	.206	.481	1.208	.550	.097
15.6.05	1.071	-.124	.163	.860	.196	.471	1.079	.440	.134
16.6.05	1.062	-.123	.162	.860	.175	.464	1.093	.426	.134
17.6.05	1.054	-.123	.162	.860	.243	.440	.889	.292	.094
20.6.05	1.047	-.123	.162	.860	.188	.402	1.035	.391	.100
21.6.05	1.041	-.123	.162	.860	.159	.488	1.090	.492	.147
22.6.05	1.037	-.122	.161	.860	.170	.478	1.033	.443	.143
23.6.05	1.056	-.123	.162	.860	.257	.417	.967	.425	.115
24.6.05	1.063	-.122	.161	.860	.204	.392	1.077	.472	.130
27.6.05	1.057	-.122	.161	.860	.158	.524	1.188	.514	.136
28.6.05	1.058	-.121	.161	.860	.111	.417	1.082	.488	.143
29.6.05	1.053	-.126	.165	.860	.139	.448	1.142	.509	.138
30.6.05	1.060	-.128	.167	.860	.156	.472	1.162	.571	.179

Estimated structural parameter $\hat{\theta}$ and relative option prices $\hat{c}_{t_0}(k, h)$ at time-to-maturity $h = 20$ for each trading day t_0 of June 2005. Option prices correspond to puts for $k < 1$, and to calls for $k \geq 1$. Estimation is performed using information based XMM. At each day t_0 , the sample consists of the current and previous $T = 1000$ observations on the state variables, and the derivative prices at t_0 of S&P 500 options with daily volume larger than 4000 contracts.

APPENDIX 1

Asymptotic properties of the basic XMM estimator

In this Appendix, we first derive (Sections A.1.1-A.1.5) the asymptotic properties of the kernel moment estimator $\widehat{\theta}_T^*$ for given weighting matrix Ω and instruments Z (proof of Proposition 1). Then, in Section A.1.6, we discuss the identification of the structural parameter from the conditional moment restrictions. Finally, in Sections A.1.7 and A.1.8, we derive the optimal instruments and the kernel nonparametric efficiency bound (proofs of Proposition 2 and Corollary 3).

Let $\Theta \subset \mathbb{R}^p$, $B \subset \mathbb{R}^L$ be compact sets. The kernel moment estimator is defined by:

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

where

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

$g_1(Y, X; \theta) = Z \cdot g(Y; \theta)$, $g_2 = (g', \widetilde{g})'$ and a is a given function of dimension L . 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 the 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 local conditional moment restrictions is asymptotically negligible. Therefore, the limit criterion would only take into account marginal moment restrictions, and could not identify parameter θ_0 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:

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

and define the associated empirical process:

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

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 the Frobenius norm $\|A\| = [Tr(AA')]^{1/2}$. In particular, when A is a vector, $\|A\|$ is the standard Euclidean norm $\|A\| =$

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. The zero-mean process is $\Psi_T(\theta) - E[\Psi_T(\theta)]$. In order to avoid to introduce further notation, we refer to $\Psi_T(\theta)$ as the empirical process.

$(A' A)^{1/2}$. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and vector $y \in \mathbb{R}^d$, we set $|\alpha| := \sum_{i=1}^d \alpha_i$, $y^\alpha := y_1^{\alpha_1} \dots y_d^{\alpha_d}$, 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 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 continuous function f defined on set \mathcal{Y} . We denote by $C^m(\mathcal{Y})$ the space of functions f on \mathcal{Y} that are continuously differentiable up to order $m \in \mathbb{N}$, $\|D^m f\| := \sum_{|\alpha|=m} \|\partial^{|\alpha|} f / \partial y^\alpha\|_\infty$, and $\Delta^m f := \sum_{i=1}^d \partial^m f / \partial y_i^m$. We denote by g_2^* the function defined by $g_2^*(y; \theta) = (g_2(y; \theta)', a(y; \theta)')$. Finally, all functions of θ are evaluated at θ_0 , when the argument is not explicit.

A.1.1 Regularity assumptions

Let us introduce the following set of regularity conditions:

Assumption A.1: *The structural parameter $\theta_0 \in \mathbb{R}^p$ is globally identified, that is,*

$$\left(E[g_1(Y, X; \theta)]', E[g_2(Y; \theta) | X = x_0]' \right)' = 0 \implies \theta = \theta_0,$$

for any $\theta \in \Theta$.

Assumption A.2: *The structural parameter θ_0 is locally identified, that is, the matrix*

$$\begin{pmatrix} E \left[\frac{\partial g_1}{\partial \theta'}(Y, X; \theta_0) \right] \\ E \left[\frac{\partial g_2}{\partial \theta'}(Y; \theta_0) | X = x_0 \right] \end{pmatrix}$$

has full column-rank.

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 $\Theta \times B$, where $\beta_0 = E[a(Y; \theta_0) | X = x_0]$.*

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 geometrically strongly 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:

$$\sum_{j=1}^{\infty} \lambda_j \left(E[\|Z_t \psi_j(Y_t)\|^r]^{2/r} + E[\psi_j(Y_t)^2 | X_t = x_0] \right) < \infty, \quad \limsup_{J \rightarrow \infty} \sup_{\theta \in \Theta} \sum_{j=J}^{\infty} \frac{1}{\lambda_j} \|c_j(\theta)\|^2 = 0.$$

Assumption A.6: *The matrices:*

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

exist and are positive definite.

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

Assumption A.8: For $t_1 < t_2$, the stationary density f_{t_1, t_2} of (X_{t_1}, X_{t_2}) is such that $\sup_{t_1 < t_2} \|f_{t_1, t_2}\|_\infty < \infty$. 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 $x \mapsto \varphi(x; \theta) = E[g_2^*(Y_t; \theta) | X_t = x] f_X(x)$ is in class $C^m(\mathcal{X})$, such that $\sup_{\theta \in \Theta} \|D^m \varphi(\cdot; \theta)\|_\infty < \infty$ and $\partial^{|\alpha|} \varphi / \partial x^\alpha$ is uniformly continuous on $\mathcal{X} \times \Theta$ for any $\alpha \in \mathbb{N}^d$ with $|\alpha| = m$.

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

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

are continuous at $x = x_0$.

Assumption A.11: The instrument Z is given by $Z = H(X)$, where H is a matrix function of dimension $q_1 \times \dim(g)$ defined on \mathcal{X} and is continuous at $x = x_0$.

Assumption A.12: The mapping $x \mapsto E[\sup_{\theta \in \Theta} \|g_2^*(Y_t; \theta)\|^2 | X_t = x] f_X(x)$ is bounded. Moreover, there exists $\delta > 2$ such that:

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

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

$$\begin{aligned} \|E[g_2^*(Y_t; \theta) | X_t = \cdot] f_X(\cdot)\|_\infty &< \infty \\ \sup_{t_1 < t_2} \left\| E\left[g_2^*(Y_{t_1}; \theta) g_2^*(Y_{t_2}; \tau)' | X_{t_1} = \cdot, X_{t_2} = \cdot\right] f_{t_1, t_2}(\cdot, \cdot)\right\|_\infty &< \infty. \end{aligned}$$

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 kernel $K(u) = \prod_{i=1}^d \kappa(u_i)$, $u \in \mathbb{R}^d$, is a Parzen product kernel of order m ,

that is,

i) $\int_{\mathbb{R}^d} K(u) du = 1$,

ii) K is bounded, $\lim_{\|u\| \rightarrow \infty} \|u\|^d K(u) = 0$, $\int_{\mathbb{R}^d} |K(u)| du < \infty$ and $w^2 := \int_{\mathbb{R}^d} K(u)^2 du < \infty$,

iii) $\int_{\mathbb{R}} u^l \kappa(u) du = 0$ for any $l \in \mathbb{N}$ such that $l < m$, and:

$$\int_{\mathbb{R}} \kappa(u) |u|^m du < \infty.$$

Assumption A.17: The bandwidth h_T is such that $Th_T^{d+2m} \rightarrow \bar{c} \in [0, \infty)$, as $T \rightarrow \infty$, and there exists $\alpha < 1/2 - 1/\delta$, where δ is defined in Assumption A.12, such that $T^\alpha h_T^d / \log T \rightarrow \infty$, as $T \rightarrow \infty$.

Assumption A.18: The function $x \mapsto \varphi_j(x) = E[\psi_j(Y_t)^2 | X_t = x] f_X(x)$ 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:

$$\begin{aligned} \sup_{j \in \mathbb{N}} \|E[\psi_j(Y_t) | X_t = \cdot] f_X(\cdot)\|_\infty &< \infty, \\ \sup_{j \in \mathbb{N}} \sup_{t_1 < t_2} \|E[\psi_j(Y_{t_1}) \psi_j(Y_{t_2}) | X_{t_1} = \cdot, X_{t_2} = \cdot] f_{t_1, t_2}(\cdot, \cdot)\|_\infty &< \infty, \\ \sup_{j \in \mathbb{N}} \|E[|\psi_j(Y_t)|^r | X_t = \cdot] f_X(\cdot)\|_\infty &< \infty, \end{aligned}$$

where r is defined in Assumption A.5.

Assumption A.20: For any $\theta \in \Theta$: $E[\|g_1(Y_t, X_t; \theta)\|^4] < \infty$, $E[\|g_2^*(Y_t; \theta)\|^4] < \infty$.

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 $\gamma_1, \gamma_2 > 1$ and $\tau > 2$, such that:

$$\begin{aligned} E\left[\left\|\frac{\partial g_1}{\partial \theta}(Y_t, X_t; \theta_0)\right\|^\tau\right] < \infty, \quad E\left[\sup_{\theta \in \Theta} \left\|\frac{\partial g_1}{\partial \theta}(Y_t, X_t; \theta)\right\|^{\gamma_1}\right] < \infty, \\ E\left[\sup_{\theta \in \Theta} \left\|\frac{\partial^2 g_1}{\partial \theta_i \partial \theta_j}(Y_t, X_t; \theta)\right\|^{\gamma_2}\right] < \infty, \quad i, j = 1, \dots, p. \end{aligned}$$

Assumption A.24: The mapping:

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

is bounded. Moreover,

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

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

Assumption A.25: Functions:

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

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.17. The condition $Th_T^{d+2m} \rightarrow \bar{c} \in [0, \infty)$ is standard in nonparametric regression analysis. When $\bar{c} > 0$, the bandwidth features the optimal d -dimensional rate of convergence, whereas when $\bar{c} = 0$ the asymptotic bias is negligible. Condition $T^\alpha h_T^d / \log T \rightarrow \infty$, for $\alpha < 1/2 - 1/\delta$, is stronger than the standard condition $Th_T^d \rightarrow \infty$; it is used to prove the consistency of kernel regression estimator $\widehat{E}[g_2(Y; \theta) | x_0]$, 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 terms 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.17. Finally, the bandwidth condition in Assumption A.17 can be satisfied when $d < 2m(\delta - 2)/(\delta + 2)$. In particular, $m = 2$ is sufficient when $d < 4$, if $\delta > 14$.

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

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

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

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

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

and continuous function b is given by

$$b(\theta) = \frac{\sqrt{\lim Th_T^{d+2m}}}{m!} w_m \frac{1}{f_X(x_0)} \begin{pmatrix} 0 \\ \Delta^m \varphi(x_0; \theta) - \frac{\varphi(x_0; \theta)}{f_X(x_0)} \Delta^m f_X(x_0) \end{pmatrix}, \quad \theta \in \Theta,$$

with $\varphi(x; \theta) := E[g_2^*(Y_t; \theta) | X_t = x] f_X(x)$, $w_m := \int_{\mathbb{R}} u^m \kappa(u) du$. In particular $\Psi_T(\theta_0) \xrightarrow{d} N(\sqrt{\bar{c}} b_0, V_0)$ where

$$b_0 := \frac{w_m}{m!} \frac{1}{f_X(x_0)} \begin{pmatrix} 0 \\ \Delta^m \varphi(x_0; \theta) - \frac{\varphi(x_0; \theta)}{f_X(x_0)} \Delta^m f_X(x_0) \end{pmatrix}, \quad V_0 := V_0(\theta_0, \theta_0) = \begin{pmatrix} S_0 & 0 \\ 0 & w^2 \Sigma_0 / f_X(x_0) \end{pmatrix}, \quad (\text{A.1})$$

matrices S_0, Σ_0 are defined in Assumption A.6, and $\bar{c} := \lim Th_T^{d+2m}$.

The block diagonal elements of matrix V_0 in (A.1) are the standard asymptotic variance-covariance matrices of sample average, and kernel regression estimators, respectively. The bias function $b(\theta)$ is zero for the marginal moments, and is equal to the kernel regression bias for the conditional moments. 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:

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

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

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

where Z_T is a random variable of order $O_p(\sqrt{T})$. From bandwidth Assumption A.17, we have $\sqrt{T} = o(Th_T^d)$, and we deduce $P \left[\left\| \widehat{\theta}_T^* - \theta_0^* \right\| \geq \varepsilon \right] \rightarrow 0$, as $T \rightarrow \infty$. \blacksquare

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:

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

that is,

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

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

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

we deduce:

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

where:

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

Inequality (A.6) implies:

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

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

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

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

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

for $\theta^* \in \Theta \times B$. Since $\|m_T(\theta^*)\|^2 \geq Th_T^d \|G(\theta^*)\|^2$, $\theta^* \in \Theta \times B$, we deduce: $\|G(\widehat{\theta}_T^*)\| = O_p\left(1/\sqrt{Th_T^d}\right)$.

By the mean-value theorem we can write ²:

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

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

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 $1/\sqrt{T}$.

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 that are locally identifiable from the marginal restriction $E[g_1(Y, X; \theta)] = 0$ in (11), and the linear combinations of θ that are identifiable only if both set of restrictions in (11) are taken into account. 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 us consider the matrix $J_1^Z = E\left[\partial g_1(Y, X; \theta_0)/\partial \theta'\right]$, whose rank is denoted by s_Z , $s_Z \leq p$. The upper index Z emphasizes that this matrix depends on instrument Z . The null space of matrix J_1^Z , denoted $\ker J_1^Z$, is characterized by a $p \times (p - s_Z)$ matrix R_Z such that:

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

The columns of R_Z generate $\ker J_1^Z$. Moreover, let us denote by \widetilde{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 $R_1 = \begin{pmatrix} \widetilde{R} & R_Z \end{pmatrix}$ is non singular, and allows us to define a new parametrization:

$$\eta = R_1^{-1}\theta = \left(\eta_1', \eta_2'\right)' \in \mathbb{R}^{s_Z} \times \mathbb{R}^{p-s_Z}.$$

Parameter η_1 defines the s_Z directions that are locally identifiable from the marginal restrictions $E[g_1(Y, X; \theta)] = 0$ alone, whereas η_2 defines the remaining $p - s_Z$ directions that are not. Furthermore, let us introduce the matrix:

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

²More precisely, the mean-value theorem is applied separately for any component of function G , and the intermediary point $\widehat{\theta}_T^*$ can differ across components.

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:

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

Then, $R_T^{-1}(\theta', \beta')' = \left(\sqrt{T}\eta_1', \sqrt{Th_T^d}\eta_2', \sqrt{Th_T^d}\beta' \right)'$, 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 $\tilde{\theta}_T^*$ be such that $\|\tilde{\theta}_T^* - \theta_0^*\| = O_p\left(1/\sqrt{Th_T^d}\right)$. Then, under Assumptions A.1-A.25, we have:*

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

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 $\left(\sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0})', \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0})', \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0)' \right)'$ is provided in the next proposition.

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

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

where $B_\infty = -\sqrt{\lim Th_T^{2m+d}} \left(J_0' \Omega J_0 \right)^{-1} J_0' \Omega b_0$.

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

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

By a mean-value expansion we can write:

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

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

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

Let us define:

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

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

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

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

$$\left(\sqrt{T} (\widehat{\eta}_{1,T} - \eta_{1,0})', \sqrt{T h_T^d} (\widehat{\eta}_{2,T} - \eta_{2,0})', \sqrt{T h_T^d} (\widehat{\beta}_T - \beta_0)' \right)' = - \left(\widetilde{J}_T \Omega \widetilde{J}_T \right)^{-1} \widetilde{J}_T \Omega g_T (\theta_0^*).$$

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

$$\left(\sqrt{T} (\widehat{\eta}_{1,T} - \eta_{1,0})', \sqrt{T h_T^d} (\widehat{\eta}_{2,T} - \eta_{2,0})', \sqrt{T h_T^d} (\widehat{\beta}_T - \beta_0)' \right)' = - \left(\widetilde{J}_T \Omega \widetilde{J}_T \right)^{-1} \widetilde{J}_T \Omega \Psi_T (\theta_0).$$

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

A.1.5 Optimal weighting matrix

When the bandwidth is such that $h_T = cT^{-1/(2m+d)}$, for some constant $c > 0$, from Proposition A.5 the asymptotic MSE of $\left(\sqrt{T} (\widehat{\eta}_{1,T} - \eta_{1,0})', \sqrt{T h_T^d} (\widehat{\eta}_{2,T} - \eta_{2,0})', \sqrt{T h_T^d} (\widehat{\beta}_T - \beta_0)' \right)'$ is $\left(J_0' \Omega J_0 \right)^{-1} J_0' \Omega M_0 \Omega J_0 \left(J_0' \Omega J_0 \right)^{-1}$, where $M_0 := V_0 + c^{2m+d} b_0 b_0'$. The optimal weighting matrix is $\Omega_0 = M_0^{-1}$, and the corresponding best MSE is $\left(J_0' M_0^{-1} J_0 \right)^{-1}$. Since M_0 and J_0 are block diagonal w.r.t. η_1 and $(\eta_2', \beta)'$, and $T h_T^d = c^d T^{\frac{2m}{2m+d}}$, the asymptotic MSE M_T of the efficient estimator of β is such that $M_T = T^{-\frac{2m}{2m+d}} \min_{\Omega} M(\Omega, Z, c) =: T^{-\frac{2m}{2m+d}} M(Z, c)$, where:

$$M(Z, c) = e' \left(J_{0,Z}' \left(\frac{w^2}{c^d f_X(x_0)} \Sigma_0 + c^{2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_{0,Z} \right)^{-1} e, \quad (\text{A.9})$$

$e = \begin{pmatrix} 0 \\ Id_L \end{pmatrix}'$, $b(x) = \frac{1}{m!} \left(\frac{\Delta^m \varphi(x; \theta_0)}{f_X(x)} - \frac{\varphi(x; \theta_0)}{f_X(x)^2} \Delta^m f_X(x) \right)$ and

$$J_{0,Z} =: \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_Z & 0 \\ E \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R_Z & -Id_L \end{pmatrix} = \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \eta_2} | x_0 \right) & 0 \\ E \left(\frac{\partial a}{\partial \eta_2} | x_0 \right) & -Id_L \end{pmatrix}.$$

A.1.6 Identification from conditional moment restrictions

In this Section, we discuss the identification of parameter θ from the conditional moment restrictions (9) and (10) in the text.

Assumption I: *The structural parameter θ_0 is locally identifiable from the conditional moment restrictions:*

$$\begin{cases} E \left[\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right] v = 0, P^X\text{-a.s.} \\ E \left[\frac{\partial \tilde{g}}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] v = 0 \end{cases} \implies v = 0, \text{ for any } v \in \mathbb{R}^p.$$

To derive the optimal instruments and the kernel nonparametric efficiency bound, it is important to distinguish between: (i) the linear combinations of θ that are locally identified when the sole uniform conditional moment restrictions $E[g(Y; \theta) | X = x] = 0, \forall x$, in (9) are considered; and (ii) the linear combinations of θ that are locally identified only if both local and uniform conditional moment restrictions in (9) and (10) are taken into account. These linear combinations are said full-information identifiable, and full-information unidentifiable, respectively. More specifically, we introduce the linear subspace

$$N^* = \left\{ v \in \mathbb{R}^p : E \left[\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right] v = 0, P^X\text{-a.s.} \right\},$$

and denote by $p - s^*$ its dimension, $s^* \leq p$. Further, let R^* be a $p \times (p - s^*)$ matrix, whose columns constitute a basis of the space N^* , and let \tilde{R}^* be a $p \times s^*$ matrix such that $R_1^* = (\tilde{R}^*, R^*)$ is non-singular. Then,

$$\eta^* = (R_1^*)^{-1} \theta = (\eta_1^*, \eta_2^*)' \in \mathbb{R}^{s^*} \times \mathbb{R}^{p-s^*} \quad (\text{A.10})$$

defines a new parameterization. Parameter η_1^* defines the full-information identifiable directions, while parameter η_2^* defines the full-information unidentifiable directions.

Assumption I can be rewritten in terms of matrix $R^* = \partial \theta / \partial \eta_2^{*'}$, that is, the Jacobian matrix w.r.t. the full-information unidentifiable directions. Indeed, any vector $v \in \mathbb{R}^p$ satisfying $E \left[\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right] v = 0, P^X\text{-a.s.}$, can be written as $v = R^* c$, where $c \in \mathbb{R}^{p-s^*}$. Then, Assumption I is equivalent to $E \left[\partial \tilde{g} (Y; \theta_0) / \partial \theta' | X = x_0 \right] R^* c = 0 \implies c = 0$, that is,

$$E \left[\frac{\partial \tilde{g}}{\partial \theta'} (Y; \theta_0) | X = x_0 \right] R^* \text{ has full column rank.} \quad (\text{A.11})$$

A.1.7 Optimal instruments

Let $Z^* = E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X)$, where $W(X)$ is a positive definite matrix P^X -a.s.. We assume that

Assumption IV: *The matrix $W(X)$ is such that Assumptions A.3-A.25 are satisfied with instrument Z^* .*

Proof of Proposition 2: (i) Let us first prove that instrument Z^* is admissible, that is, satisfies Assumptions A.1 and A.2. We have, for any vector $v \in \mathbb{R}^p$:

$$\begin{aligned} E \left[\frac{\partial g_1}{\partial \theta'} (Y, X; \theta_0) \right] v = 0 &\iff E \left[E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X) E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) v \right] = 0 \\ &\iff v' E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X) E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) v = 0, \quad P^X\text{-a.s.}, \\ &\iff E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) v = 0, \quad P^X\text{-a.s.} \end{aligned} \quad (\text{A.12})$$

Thus, Assumption A.2 follows from Assumption I. Then, Assumption A.1 is also satisfied if Θ is taken small enough, which is sufficient for the validity of the asymptotic results.

Let us now prove that instrument Z^* is optimal. In the expression of $M(Z, c)$ in (A.9), instrument Z affects matrix $J_{0,Z}$ only. Moreover, from the definition of R_Z in A.1.4, matrix $J_{0,Z}$ depends on instrument Z only through $\ker J_1^Z$, where $J_1^Z = E \left[\partial g_1(Y, X; \theta_0) / \partial \theta' \right]$. In particular, the larger the null space $\ker J_1^Z$, the larger the vector η_2 of structural parameters that are unidentifiable from marginal restrictions and must be estimated at a non-parametric rate jointly with β . In other words, if Z and W are two alternative admissible sets of instruments such that $\ker J_1^Z \subset \ker J_1^W$, then $M(Z, c) \leq M(W, c)$. Since $N^* \subset \ker J_1^Z$ for any admissible instrument, it follows that Z^* is an optimal instrument if we can show that $N^* = \ker J_1^{Z^*}$. The latter equality follows from (A.12).

Since $N^* = \ker J_1^{Z^*}$, matrix R^* coincides with matrix R_{Z^*} for any optimal instrument Z^* . From (A.9), the concentrated criterion for (any) optimal instrument Z^* becomes

$$M(c) = e' \left(J_0^{*'} \left(\frac{1}{c^d} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_0^* \right)^{-1} e.$$

(ii) Let us now derive the optimal bandwidth sequence and the kernel non-parametric efficiency bound. To simplify, in the rest of the proof we assume $L = 1$. Then, function $M(c)$ is scalar, and the first-order condition for minimization of $M(c)$ w.r.t. c is given by

$$\frac{dM(c)}{dc} = e' \left[\left(J_0^{*'} \tilde{\Sigma}^{-1} J_0^* \right)^{-1} J_0^{*'} \tilde{\Sigma}^{-1} \left(-\frac{d}{c^{d+1}} \frac{w^2}{f_X(x_0)} \Sigma_0 + 2mc^{2m-1} w_m^2 b(x_0) b(x_0)' \right) \tilde{\Sigma}^{-1} J_0^* \left(J_0^{*'} \tilde{\Sigma}^{-1} J_0^* \right)^{-1} \right] e = 0$$

with $\tilde{\Sigma} := \frac{1}{c^d} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{2m} w_m^2 b(x_0) b(x_0)'$ and $e = (0', 1)' \in \mathbb{R}^{p-s_z} \times \mathbb{R}$. The solution c^* is $c^* = \xi^{1/(2m+d)}$ where ξ satisfies the equation

$$\xi = \frac{w^2 d}{2m w_m^2 f_X(x_0)} \frac{e' \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} \Sigma_0 A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e}{e' \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} b(x_0) b(x_0)' A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e},$$

where $A(\xi) := \frac{w^2}{f_X(x_0)} \Sigma_0 + \xi b(x_0) b(x_0)'$. The optimal bandwidth sequence is $h_T = c^* T^{-1/(2m+d)}$. The kernel nonparametric efficiency bound is $M(c^*)$. ■

A.1.8 Bias free kernel non-parametric efficiency bound (proof of Corollary 3)

Let us now assume that the bandwidth is such that $Th_T^{2m+d} = o(1)$. Then $B_\infty = 0$ in Proposition A.5, and the kernel moment estimator is asymptotically unbiased. The proof that $Z^* = E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X)$ is an optimal instrument is similar to the proof of Proposition 2 (i), replacing $M(Z, c)$ in (A.9) with $V(Z) = \frac{w^2}{f_X(x_0)} e' \left(J_{0,Z}' \Sigma_0^{-1} J_{0,Z} \right)^{-1} e$, which is the asymptotic variance of $\hat{\beta}_T$. Thus, the bias free kernel nonparametric efficiency bound is $\mathcal{B}(a, x_0) = \frac{w^2}{f_X(x_0)} e' \left(J_0^{*'} \Sigma_0^{-1} J_0^* \right)^{-1} e$. Corollary 3 follows from the formula of the inverse of a block matrix.

APPENDIX 2

Information based XMM estimator

In this Appendix we first derive the asymptotic expansion of the objective function and of the estimators (Sections A.2.1-A.2.2), in order to prove in Section A.2.3 the asymptotic kernel nonparametric efficiency of the information based XMM estimator (Proposition 4). Then, in Section A.2.4 we discuss the practical implementation of the information based XMM estimator.

A.2.1 Concentration with respect to functional parameters

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 $f^t, t = 1, \dots, T$, and f^0 are:

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

which yield:

$$f^t(y) = \widehat{f}(y|x_t) + \mu_t \widehat{f}(y|x_t) + \lambda'_t g(y; \theta) \widehat{f}(y|x_t), \quad t = 1, \dots, T, \quad (\text{A.13})$$

$$f^0(y) = \widehat{f}(y|x_0) \exp (\lambda'_0 g_2(y; \theta) + \mu_0 - 1). \quad (\text{A.14})$$

The Lagrange multipliers are deduced from the constraints. From (A.13), we get:

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

and:

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

$t = 1, \dots, T$. Similarly, from (A.14) 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.13) and (A.14), $\mu_0, \lambda_t, \mu_t, t = 1, \dots, T$ can be eliminated to get the concentrated functional parameters:

$$\begin{aligned} f^t(y; \theta) &= \widehat{f}(y|x_t) - \widehat{E}(g(\theta)|x_t)' \widehat{V}(g(\theta)|x_t)^{-1} \left[g(y; \theta) - \widehat{E}(g(\theta)|x_t) \right] \widehat{f}(y|x_t), \quad t = 1, \dots, T, \\ f^0(y; \theta, \lambda_0) &= \frac{\exp(\lambda_0' g_2(y; \theta))}{\widehat{E}[\exp(\lambda_0' g_2(\theta)) | x_0]} \widehat{f}(y|x_0), \end{aligned} \quad (\text{A.15})$$

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

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

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

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

and the conditional density estimators are:

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

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:

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

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

$$\lambda_0 \simeq -\widehat{V}(g_2(\theta) | x_0)^{-1} \widehat{E}(g_2(\theta) | x_0), \quad (\text{A.16})$$

and the asymptotic expansion of the concentrated objective function becomes:

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

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

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

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

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

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. To this goal, we consider the new parameterization in (A.10). To simplify the notation, we omit the star in parameter η . Then, we have:

$$E\left(\frac{\partial g}{\partial \theta'} | x_t\right) (\theta - \theta_0) = E\left(\frac{\partial g}{\partial \theta'} | x_t\right) \widetilde{R}^* (\eta_1 - \eta_{1,0}).$$

We get:

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

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

$$\begin{aligned} \sqrt{T} (\widehat{\eta}_{1,T} - \eta_{1,0}) &\simeq - \left[\frac{1}{T} \sum_{t=1}^T \widetilde{R}^{*'} E\left(\frac{\partial g'}{\partial \theta} | x_t\right) V(g|x_t)^{-1} E\left(\frac{\partial g}{\partial \theta'} | x_t\right) \widetilde{R}^* \right]^{-1} \\ &\quad \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^T \widetilde{R}^{*'} E\left(\frac{\partial g'}{\partial \theta} | x_t\right) V(g|x_t)^{-1} \int g(y; \theta_0) \widehat{f}(y|x_t) dy \\ &\simeq - \left(\widetilde{R}^{*'} E \left[E\left(\frac{\partial g'}{\partial \theta} | x_t\right) V(g|x_t)^{-1} E\left(\frac{\partial g}{\partial \theta'} | x_t\right) \right] \widetilde{R}^* \right)^{-1} \\ &\quad \cdot \sqrt{T} \int \int \widetilde{R}^{*'} E\left(\frac{\partial g'}{\partial \theta} | x\right) V(g|x)^{-1} g(y; \theta_0) \widehat{f}(y, x) dx dy. \end{aligned}$$

Thus $\widehat{\eta}_{1,T}$ converges at a parametric rate. The bias terms induced by the kernel density estimator can be neglected asymptotically, since $Th_T^{d+2m} = o(1)$.

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

$$\begin{aligned} \sqrt{Th_T^d}(\widehat{\eta}_{2,T} - \eta_{2,0}) &\simeq - \left[R^{*'} E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2 | x_0)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R^* \right]^{-1} \\ &\quad \cdot R^{*'} E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2 | x_0)^{-1} \sqrt{Th_T^d} \int g_2(y; \theta_0) \widehat{f}(y | x_0) dy. \end{aligned} \quad (\text{A.17})$$

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

Let us consider the expansion of $f^0(y; \theta, \lambda_0)$ in (A.15) around $\lambda_0 = 0$. We have:

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

from (A.16). Thus, we get:

$$\begin{aligned} \widehat{f}^*(y | x_0) &= f_0(y; \widehat{\theta}_T, \widehat{\lambda}_{0,T}) \\ &\simeq \widehat{f}(y | x_0) - \widehat{E}(g_2(\widehat{\theta}_T) | x_0)' \widehat{V}(g_2(\widehat{\theta}_T) | x_0)^{-1} \left(g_2(y; \widehat{\theta}_T) - \widehat{E}(g_2(\widehat{\theta}_T) | x_0) \right) \widehat{f}(y | x_0) \\ &\simeq \widehat{f}(y | x_0) - \widehat{E}(g_2(\widehat{\theta}_T) | x_0)' V(g_2 | x_0)^{-1} g_2(y; \theta_0) f(y | x_0). \end{aligned} \quad (\text{A.18})$$

Moreover,

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

from (A.17), where

$$M = E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R^* \left[R^{*'} E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2 | x_0)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R^* \right]^{-1} R^{*'} E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2 | x_0)^{-1},$$

is an orthogonal projector for the inner product defined by $V(g_2 | x_0)^{-1}$. After substituting in (A.18), we get:

$$\widehat{f}^*(y | x_0) \simeq \widehat{f}(y | x_0) - f(y | x_0) g_2(y; \theta_0)' V(g_2 | x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y | x_0) dy. \quad (\text{A.19})$$

iv) Asymptotic expansion of the moment of interest

We have:

$$\begin{aligned}
\widehat{E}^*(a|x_0) &= \int a(y; \widehat{\theta}_T) \widehat{f}^*(y|x_0) dy \\
&\simeq \int a(y; \theta_0) f(y|x_0) dy + \int \frac{\partial a}{\partial \theta'}(y; \theta_0) f(y|x_0) dy (\widehat{\theta}_T - \theta_0) + \int a(y; \theta_0) [\widehat{f}^*(y|x_0) - f(y|x_0)] dy \\
&\simeq E(a|x_0) + E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R^* (\widehat{\eta}_{2,T} - \eta_{2,0}) \\
&\quad + \int a(y; \theta_0) \left\{ \widehat{f}(y|x_0) - f(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V(g_2|x_0)^{-1} \right. \\
&\quad \left. (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \right\} dy \quad [\text{from (A.19)}] \\
&= E(a|x_0) - E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R^* \left[R^{*'} E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R^* \right]^{-1} \\
&\quad \cdot R^{*'} E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \quad [\text{from (A.17)}] \\
&\quad + \int a(y; \theta_0) [\widehat{f}(y|x_0) - f(y|x_0)] dy - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy.
\end{aligned}$$

Thus, we get:

$$\begin{aligned}
\widehat{E}^*(a|x_0) - E(a|x_0) &\simeq \int a(y; \theta_0) \delta \widehat{f}(y|x_0) dy - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy \\
&\quad - \left[E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R^* - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R^* \right] \\
&\quad \cdot \left[R^{*'} E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R^* \right]^{-1} R^{*'} E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy,
\end{aligned} \tag{A.20}$$

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.20), 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:

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

where the asymptotic variance is such that:

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

Since $W(x_0)/f_X(x_0)$ corresponds to the bias free kernel nonparametric efficiency bound $\mathcal{B}(x_0, a)$ (see Corollary 3), the kernel nonparametric efficiency of the information based XMM estimator is proved.

A.2.4 Implementation

In the application of Section 5.3, the estimator is computed as follows. We first get an estimator of θ with a two-step approach:

- (i) In the first step, we minimize the criterion using fixed weighting matrices

$$\tilde{\theta} = \arg \min_{\theta} \frac{1}{T} \sum_{t=1}^T \widehat{E}(g(\theta)|x_t)' \widehat{V}(g(\bar{\theta})|x_t)^{-1} \widehat{E}(g(\theta)|x_t) + \frac{1}{2} h_T^d \widehat{E}(g_2(\theta)|x_0)' \widehat{V}(g_2(\bar{\theta})|x_0)^{-1} \widehat{E}(g_2(\theta)|x_0),$$

where $\bar{\theta}$ is a given parameter value.

- (ii) In the second step, we use $\tilde{\theta}$ to estimate the weighting matrices

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{T} \sum_{t=1}^T \widehat{E}(g(\theta)|x_t)' \widehat{V}(g(\tilde{\theta})|x_t)^{-1} \widehat{E}(g(\theta)|x_t) + \frac{1}{2} h_T^d \widehat{E}(g_2(\theta)|x_0)' \widehat{V}(g_2(\tilde{\theta})|x_0)^{-1} \widehat{E}(g_2(\theta)|x_0).$$

Then, we get $\hat{\lambda} = \lambda_0(\hat{\theta})$ by solving

$$\hat{\lambda} = \arg \min_{\lambda} \log \widehat{E} \left[\exp \left(\lambda' g_2(\hat{\theta}) \right) | x_0 \right].$$

The kernel estimators \widehat{E} and \widehat{V} are computed using a bivariate Gaussian product kernel, and the bandwidths for r_t and σ_t are selected by the standard Silverman (1986) rule of thumb.

APPENDIX 3

Stochastic volatility model

In this Appendix, we prove Proposition 5 and discuss parameter identification in the stochastic volatility model of Section 5.1 (see Appendix B on the web for further discussion of the XMM regularity assumptions of Appendix 1).

A.3.1 Proof of Proposition 5

The no-arbitrage restrictions are:

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

Since the above conditions have to be satisfied for any admissible value of σ_t^2 , we get:

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

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 conclude:

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

A.3.2 Identification

Let us first consider the identifiability of structural parameter θ (Assumption I in Appendix 1) and provide the expression of the Jacobian matrix R^* w.r.t. the directions of full-information underidentification.

i) Computation of matrix R^*

The null space N^* associated with the uniform restrictions is the linear space of vectors $v \in \mathbb{R}^4$ such that (see A.1.6):

$$E \left(\begin{pmatrix} 1 \\ \exp r_{t+1} \end{pmatrix} \frac{\partial M_{t,t+1}}{\partial \theta'} (\theta_0) \mid y_t \right) v = 0, \quad \forall y_t. \quad (\text{A.21})$$

We know that θ_0 satisfies the no-arbitrage restrictions:

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

We deduce that any $\theta = \theta_0 + v\varepsilon$, where ε is small and v satisfies (A.21), is also such that:

$$E \left(M_{t,t+1}(\theta) \begin{pmatrix} 1 \\ \exp r_{t+1} \end{pmatrix} \middle| y_t \right) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \forall y_t,$$

at first order in ε . Therefore, the vectors in N^* are the directions $d\theta = \theta - \theta_0$ of infinitesimal parameter changes that are compatible with no-arbitrage. From Proposition 5 and the definitions of functions a, b [see equation (22)], the parameters θ that are compatible with no-arbitrage are characterized by the nonlinear restrictions:

$$\nu_0 = -b(\nu_1 + \gamma^2/2 - 1/8), \quad \nu_2 = -a(\nu_1 + \gamma^2/2 - 1/8), \quad \nu_3 = \gamma + 1/2.$$

Thus, the tangent set is spanned by the vector:

$$v = \begin{pmatrix} d\nu_0/d\nu_1 \\ d\nu_1/d\nu_1 \\ d\nu_2/d\nu_1 \\ d\nu_3/d\nu_1 \end{pmatrix} = \begin{pmatrix} -db(\nu_1 + \gamma^2/2 - 1/8)/d\nu_1 \\ 1 \\ -da(\nu_1 + \gamma^2/2 - 1/8)/d\nu_1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\delta \frac{c}{1+c\lambda_1} \\ 1 \\ -\rho \frac{1}{(1+c\lambda_1)^2} \\ 0 \end{pmatrix},$$

where $\lambda_1 := \nu_1 + \gamma^2/2 - 1/8$. We deduce that matrix R^* is given by:

$$R^* = \begin{pmatrix} -\delta \frac{c}{1+c\lambda_1}, 1, -\rho \frac{1}{(1+c\lambda_1)^2}, 0 \end{pmatrix}'. \quad (\text{A.22})$$

ii) Checking Assumption I

Let us now verify that Assumption I is satisfied when the conditional restrictions include the observed price of a European call. From (A.11), we have to prove that $E \left(\frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) (\exp r_{t+1} - k)^+ \middle| y_t \right) R^* \neq 0$, $\forall k > 0$. We have:

$$\begin{aligned} & E \left(\frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) (\exp r_{t+1} - k)^+ \middle| y_t \right) R^* \\ &= -E \left(M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ (1, \sigma_{t+1}^2, \sigma_t^2, r_{t+1}) R^* \middle| y_t \right) \\ &= \left[\frac{\delta c}{1+c\lambda_1} + \frac{\rho}{(1+c\lambda_1)^2} \sigma_t^2 \right] E \left(M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ \middle| y_t \right) - E \left(M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ \sigma_{t+1}^2 \middle| y_t \right). \end{aligned}$$

From (27), we have:

$$\frac{\delta c}{1+c\lambda_1} + \frac{\rho}{(1+c\lambda_1)^2} \sigma_t^2 = \rho^* \sigma_t^2 + \delta^* c^* = E_t^Q [\sigma_{t+1}^2],$$

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

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

Thus, we get:

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

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