

EFFICIENT DERIVATIVE PRICING BY THE EXTENDED METHOD OF MOMENTS

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

First version: July 2004

This version: October 2008

*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, F. Corsi, R. Davidson, J. Fan, R. Garcia, J. Jackwerth, J. Jasiak, O. Linton, L. Mancini, A. Melino, C. Ortelli, O. Scaillet, M. Stutzer, F. Trojani, H. Uhlig, 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/Risk Foundation: "Large Risks in Insurance".

Efficient Derivative Pricing by the Extended Method of Moments

Abstract

In this paper we introduce the Extended Method of Moments (XMM) estimator. This estimator accommodates a more general set of moment restrictions than the standard Generalized Method of Moments (GMM) estimator. More specifically, the XMM differs from the GMM in that it can handle not only uniform conditional moment restrictions (i.e. valid for any value of the conditioning variable), but also local conditional moment restrictions valid for a given fixed value of the conditioning variable. The local conditional moment restrictions are of particular relevance in derivative pricing for reconstructing the pricing operator at a given day, by using the information in a cross-section of observed traded derivative prices and a time series of underlying asset returns. The estimated derivative prices are consistent for large time series dimension, but fixed number of cross-sectionally observed derivative prices. The asymptotic properties of the XMM estimator are non-standard, since the combination of uniform and local conditional moment restrictions induces different rates of convergence (parametric and nonparametric) for the parameters.

Keywords: Derivative Pricing, Trading Activity, 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) was introduced by Hansen (1982) and Hansen, Singleton (1982) to estimate a structural parameter θ identified by Euler conditions:

$$p_{i,t} = E_t [M_{t,t+1}(\theta)p_{i,t+1}], \quad i = 1, \dots, n, \quad \forall t, \quad (1.1)$$

where $p_{i,t}$, $i = 1, \dots, n$, are the observed prices of n financial assets, E_t denotes the expectation conditional on the available information at date t , and $M_{t,t+1}(\theta)$ is the stochastic discount factor. Model (1.1) is semi-parametric. The GMM estimates parameter θ regardless of the conditional distribution of the state variables. This conditional distribution however becomes relevant when the Euler conditions (1.1) are used for pricing derivative assets. Indeed, when the derivative payoff is written on $p_{i,t+1}$ and its current price is not observed on the market, the derivative pricing requires the joint estimation of parameter θ and the conditional distribution of the state variables.

The Extended Method of Moments (XMM) estimator extends the standard GMM to accommodate a more general set of moment restrictions. The standard GMM is based on uniform conditional moment restrictions such as (1.1), that are valid for any value of the conditioning variables. The XMM can handle not only the uniform moment restrictions, but also local moment restrictions that are valid for a given value of the conditioning variables only. This leads to a new field of application to derivative pricing, as the XMM can be used for reconstructing the pricing operator on a given day, by using the information in a cross-section of observed traded derivative prices and a time series of underlying asset returns. To illustrate the principle of XMM, let us consider the S&P 500 index and its derivatives. Suppose an investor at date t_0 is interested in estimating the price $c_{t_0}(h, k)$ of a call option with time-to-maturity h and moneyness strike k that is currently not (actively) traded on the market. She has data on a time series of T daily returns of the S&P 500 index, as well as on a small cross-section of current option prices $c_{t_0}(h_j, k_j)$, $j = 1, \dots, n$, of n highly traded derivatives. The XMM approach provides the estimated prices $\hat{c}_{t_0}(h, k)$ for different values of moneyness strike k and time-to-maturity h , that interpolate the observed prices of highly traded derivatives and satisfy the hypothesis of absence of arbitrage opportunities. These estimated prices are consistent for a large number of dates T , but a fixed, even small,

number of observed derivative prices n .

To highlight the specificity of XMM with respect to GMM, we present in Section 2 an application to the S&P 500 index and its derivatives. First we show that the trading activity on the index option market is rather weak and the daily number of reliable derivative prices is small. Then we explain why the time series observations on the underlying index induce uniform moment restrictions, whereas the observed cross-sectional derivative prices correspond to the local moment restrictions. The XMM estimator minimizes the discrepancy of the historical transition density from a kernel density estimator, subject to both types of moment restrictions. The estimation criterion includes a Kullback-Leibler Information Criterion associated with the local moment restrictions to ensure the compatibility of the estimated derivative prices with the absence of arbitrage opportunities. The comparison of the XMM estimator and the standard calibration estimator for S&P 500 options data shows clearly that XMM outperforms the traditional approach.

The theoretical properties of the XMM estimator are presented in Section 3 in a general framework. We discuss the parameter identification under both uniform and local moment restrictions and derive the efficiency bound. We prove that the XMM estimator is consistent for a fixed number n of cross-sectional observations associated with the local restrictions and a large number T ($T \rightarrow \infty$) of time series observations associated with uniform restrictions. Moreover, the XMM estimator is asymptotically normal and semi-parametrically efficient. Section 4 concludes. The proofs of the theoretical results are gathered in Appendix A. Proofs of technical Lemmas are given in Appendix B on the web-site <http://www.istituti.usilu.net/gagliarp/proofsXMM.htm>.

2 The XMM applied to derivative pricing

2.1 The activity on the index option market

In order to maintain a minimum activity, the Chicago Board Options Exchange (CBOE) enhances the market of options on the S&P 500 index by periodically issuing 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, etc, up to a maximal time-to-maturity. For instance, new 12-month options are issued every three months, when the options from the previous issuing attain the time-to-maturity of 9-month. This induces a cycle in the times-to-maturity of quoted options [see e.g. Schwartz (1987), Figure 1, and Pan (2002), Figure 2]. For any admissible time-to-maturity, the options are issued for a limited number of strikes around the value of the underlying asset at the issuing date.

This strategy restricts the number of options available on the market. Among these, only a few options are traded on a daily basis. This phenomenon is illustrated in Section 2.6, where we select the call and put options with a daily traded volume of more than 4000 contracts in June 2005. Since each contract corresponds to 100 options, 4000 contracts are worth between 5 millions USD and 7 millions USD, on average. The daily number of the highly traded options varies between a minimum of 7 in June 10, 2005 and a maximum of 31 in June 16, 2005. The corresponding times-to-maturity and moneyness strikes also vary in time. For instance in June 10, 2005 the actively traded times-to-maturity are 5, 70 and 135 days, while in June 16, 2005 the actively traded times-to-maturity are 1, 21, 46, 66, 131, 263 and 393 days. In brief, the number of highly traded derivatives on a given day is rather small. Moreover, the number of options and the moneyness strikes and times-to-maturity vary from one day to another due to the issuing cycle and the trading activity.

2.2 Calibration based on current derivative prices

The underlying asset features a regular trading activity, and the associated returns can be observed and are expected to be stationary. In contrast, the prices of a given derivative on two consecutive dates are not always observable. Therefore the associated returns cannot be computed. Even if these option returns were available, they would be nonstationary due to the issuing cycle discussed above. To circumvent the difficulty in modelling the trading activity and its potential nonstationary effects on prices, the standard methodology consists in calibrating daily the pricing operator ¹. More precisely, let us assume that at date t_0 the option prices $c_{t_0}(h_j, k_j)$, $j = 1, \dots, n$, are observed, and that a parametric model

¹Or, in calibrating daily the Black-Scholes implied volatility surface.

for the risk-neutral dynamics of the relevant state variables implies a parameterized pricing formula $c_{t_0}(h, k; \theta)$ for the option prices at date t_0 . The unknown parameter θ is usually estimated daily by minimizing the least-squares criterion ² [see e.g. Bakshi et al. (1997)]:

$$\hat{\theta}_{t_0} = \arg \min_{\theta} \sum_{j=1}^n [c_{t_0}(h_j, k_j) - c_{t_0}(h_j, k_j; \theta)]^2. \quad (2.1)$$

This practice has the following drawbacks: First, the estimated parameters $\hat{\theta}_{t_0}$ are generally erratic over time. Second, the approximated prices $\hat{c}_{t_0}(h_j, k_j) = c_{t_0}(h_j, k_j; \hat{\theta}_{t_0})$ are not compatible with the absence of arbitrage opportunities, since the approximated option prices $\hat{c}_{t_0}(h_j, k_j)$ differ from the observed prices $c_{t_0}(h_j, k_j)$ for highly traded options. Third, even if the estimates can be shown to be consistent when n is large and the options characteristics (h_j, k_j) are well distributed [see e.g. Ait-Sahalia and Lo (1998)], in practice n is small and the estimates are not very accurate.

2.3 Semi-parametric pricing model

In general, the underlying asset is much more actively traded than each of its derivatives. For instance, the daily traded volume of a portfolio mimicking the S&P 500 index, such as the SPDR, is about one hundred millions USD. It is possible to improve the above calibration approach by considering jointly the time series of observations on the underlying asset and the cross-sectional data on its derivatives. For this purpose the specification cannot be reduced to the risk-neutral dynamics only, but has to define in a coherent way the historical and risk-neutral dynamics of the variables of interest. Let us consider a discrete time model. We denote by r_t the logarithmic return of the underlying asset between dates $t - 1$ and t . We assume that the information available to the investors at date t is generated by the random vector X_t of state variables with dimension d , including the return r_t as the first component, and that X_t is also observable by the econometrician. The process (X_t) on $\mathcal{X} \subset \mathbb{R}^d$ is supposed to be strictly stationary and Markov under the historical probability with transition density $f(x_t|x_{t-1})$.

²A penalized Least-Squares (LS) criterion [Jackwerth (2000)], a local LS criterion under shape constraints [Ait-Sahalia, Duarte (2003)], a LS criterion under convolution constraints [Bondarenko (2003)], or an entropy criterion [Buchen, Kelly (1996), Stutzer (1996), Jackwerth, Rubinstein (1996)], in a nonparametric setting.

We consider a semi-parametric pricing model defined by the historical dynamics and a stochastic discount factor (sdf) [Hansen, Richard (1987)]. The historical dynamics of process (X_t) is left unconstrained. We adopt a parametric specification for the sdf $M_{t,t+1}(\theta) = m(X_{t+1}; \theta)$, where $\theta \in \mathbb{R}^p$ is an unknown vector of risk premia parameters. This model implies a semi-parametric specification for the risk-neutral distribution. For instance, the relative price at time t of a European call with moneyness strike k and time-to-maturity h written on the underlying asset is given by:

$$c_t(h, k) = E [M_{t,t+h}(\theta)(\exp R_{t,h} - k)^+ | X_t],$$

where $M_{t,t+h}(\theta) = m(X_{t+1}; \theta) \cdots m(X_{t+h}; \theta)$ is the sdf between t and $t+h$, and $R_{t,h} = r_{t+1} + \cdots + r_{t+h}$ is the return of the underlying asset in this period. The option price depends on both the finite-dimensional sdf parameter θ and the functional historical parameter f .

We are interested in estimating the pricing operator at a given date t_0 , that is, the mapping that associates any European call option payoff $\varphi_{t_0}(h, k) = (\exp R_{t_0,h} - k)^+$ with its price $c_{t_0}(h, k)$ at time t_0 , for any time-to-maturity h and any moneyness strike k . The data consists of a finite number n of derivative prices $c_{t_0}(h_j, k_j)$, $j = 1, \dots, n$, observed at date t_0 , and T serial observations of the state variables X_t corresponding to the current and previous days $t = t_0 - T + 1, \dots, t_0$. The no-arbitrage assumption implies the moment restrictions for the observed asset prices.

2.4 The moment restrictions

The moment restrictions are twofold. The constraints concerning the observed derivative prices at t_0 are given by:

$$c_{t_0}(h_j, k_j) = E [M_{t,t+h_j}(\theta)(\exp R_{t,h_j} - k_j)^+ | X_t = x_{t_0}], \quad j = 1, \dots, n. \quad (2.2)$$

The constraints concerning the riskfree asset and the underlying asset are:

$$\begin{cases} E[M_{t,t+1}(\theta) | X_t = x] = B(t, t+1), & \forall x \in \mathcal{X}, \\ E[M_{t,t+1}(\theta) \exp r_{t+1} | X_t = x] = 1, & \forall x \in \mathcal{X}, \end{cases} \quad (2.3)$$

respectively, where $B(t, t+1)$ denotes the price at time t of the short-term riskfree bond. The conditional moment restrictions (2.2) are local, since they hold for a single value of

the conditioning variable only, namely the value x_{t_0} of the state variable at time t_0 . This is because we consider only observations of the derivative prices $c_{t_0}(h_j, k_j)$ at date t_0 . Conversely, the prices of the underlying asset and the riskfree bond are observed for all trading days. Therefore the conditional moment restrictions (2.3) hold for all values of the state variables. They are called the uniform moment restrictions. The distinction between the uniform and local moment restrictions is a consequence of the differences between the trading activities of the underlying asset and its derivatives. Technically, it is the essential feature of the XMM that distinguishes this method from its predecessor GMM.

There are $n + 2$ local moment restrictions at date t_0 given by:

$$\left\{ \begin{array}{l} E[M_{t,t+h_j}(\theta)(\exp R_{t,h_j} - k_j)^+ - c_{t_0}(h_j, k_j)|X_t = x_{t_0}] = 0, \quad j = 1, \dots, n, \\ E[M_{t,t+1}(\theta) - B(t_0, t_0 + 1)|X_t = x_{t_0}] = 0, \\ E[M_{t,t+1}(\theta) \exp r_{t+1} - 1|X_t = x_{t_0}] = 0. \end{array} \right. \quad (2.4)$$

Let us denote the local moment restrictions (2.2) as:

$$E[\tilde{g}(Y; \theta)|X = x_0] = 0, \quad (2.5)$$

where $Y_t = (X_{t+1}, \dots, X_{t+\bar{h}})'$ is the \tilde{d} -dimensional vector of relevant future values of the state variables, \bar{h} is the largest time-to-maturity of interest, $x_0 \equiv x_{t_0}$, and the time index is suppressed. Similarly, the uniform moment restrictions (2.3) are written as:

$$E[g(Y; \theta)|X = x] = 0, \quad \forall x \in \mathcal{X}. \quad (2.6)$$

Then, the whole set of local moment restrictions (2.4) corresponds to:

$$E[g_2(Y; \theta)|X = x_0] = 0, \quad (2.7)$$

where $g_2 = (g', \tilde{g}')$. Since we are interested in estimating the pricing operator at a given date t_0 , the value x_0 is considered as a given constant.

We assume that the sdf parameter θ is identified from the two sets of conditional moment restrictions (2.5) and (2.6). In general, some linear combinations η_2^* , say, of the components of θ are unidentifiable from the uniform moment restrictions (2.6) on the risk-free asset and the underlying asset, only. The identification of these linear combinations

requires local moment restrictions (2.5) on the cross-sectional derivative prices at t_0 . Intuitively, some of these linear combinations η_2^* correspond to genuine risk-premia parameters. This point is illustrated in Section 3.2 on a stochastic volatility model used in the application.

2.5 The XMM estimator

The XMM estimator presented in this section is related to the recent literature on the information based GMM [e.g., Kitamura, Stutzer (1997), Imbens, Spady, Johnson (1998)]. It provides estimators of both the sdf parameter θ and the historical transition density $f(y|x)$. By using the parameterized sdf, the information based estimator of the historical transition density defines the estimated state price density for pricing derivatives.

The XMM approach involves a consistent nonparametric estimator of the historical transition density $f(y|x)$, such as the kernel density estimator:

$$\hat{f}(y|x) = \frac{1}{h_T^{\bar{d}}} \sum_{t=1}^T \tilde{K} \left(\frac{y_t - y}{h_T} \right) K \left(\frac{x_t - x}{h_T} \right) / \sum_{t=1}^T K \left(\frac{x_t - x}{h_T} \right), \quad (2.8)$$

where K (resp. \tilde{K}) is the d -dimensional (resp. \bar{d} -dimensional) kernel, h_T is the bandwidth and (x_t, y_t) , $t = 1, \dots, T$, are the historical sample data.³ Next, this kernel density estimator is improved by selecting the conditional pdf that is the closest to $\hat{f}(y|x)$, and satisfies the moment restrictions, as defined below.

Definition 1. *The XMM estimator $(\hat{f}^*(\cdot|x_0), \hat{f}^*(\cdot|x_1), \dots, \hat{f}^*(\cdot|x_T), \hat{\theta})$ consists of the functions f_0, f_1, \dots, f_T defined on $\mathcal{Y} \subset \mathbb{R}^{\bar{d}}$, and the parameter value θ , that minimize the objective function:*

$$L_T = \frac{1}{T} \sum_{t=1}^T \int \frac{[\hat{f}(y|x_t) - f_t(y)]^2}{\hat{f}(y|x_t)} dy + h_T^{\bar{d}} \int \log \left[\frac{f_0(y)}{\hat{f}(y|x_0)} \right] f_0(y) dy,$$

subject to the constraints:

$$\begin{aligned} \int f_t(y) dy &= 1, \quad t = 1, \dots, T, & \int f_0(y) dy &= 1, \\ \int g(y; \theta) f_t(y) dy &= 0, \quad t = 1, \dots, T, & \int g_2(y; \theta) f_0(y) dy &= 0. \end{aligned} \quad (2.9)$$

³For expository purpose, the dates previous to t_0 , at which data on (X, Y) are available, have been re-indexed as $t = 1, \dots, T$ and accordingly the asymptotics in T correspond to a long history before t_0 .

The objective function L_T has two components. The first component involves the chi-square distance between the density f_t and the kernel density estimator $\hat{f}(\cdot|x_t)$ at any sample point x_t , $t = 1, \dots, T$. The second component corresponds to the Kullback-Leibler information criterion (KLIC) between the density f_0 and the kernel estimator $\hat{f}(\cdot|x_0)$ at the given value x_0 . In addition to the unit mass restrictions for the density functions, the constraints include the uniform moment restrictions (2.6) written for all sample points, and the whole set of local moment restrictions (2.7) at x_0 . The combination of two types of discrepancy measures is motivated by computational and financial reasons. The chi-square criterion evaluated at the sample points allows for closed form solutions $f_1(\theta), \dots, f_T(\theta)$ for a given θ (see Appendix A.2.1). Therefore, the objective function can be easily concentrated with respect to functions f_1, \dots, f_T , which reduces the dimension of the optimization problem. The KLIC criterion evaluated at x_0 ensures that the minimizer f_0 satisfies the positivity restriction [see e.g. Stutzer (1996) and Kitamura, Stutzer (1997)]. The positivity of the associated state price density at t_0 guarantees the absence of arbitrage opportunities in the estimated derivative prices. The estimator of $\hat{\theta}$ minimizes the concentrated objective function:

$$\mathcal{L}_T^c(\theta) = \frac{1}{T} \sum_{t=1}^T \hat{E}(g(\theta)|x_t)' \hat{V}(g(\theta)|x_t)^{-1} \hat{E}(g(\theta)|x_t) - h_T^d \log \hat{E}\left(\exp\left(\lambda(\theta)' g_2(\theta)\right) | x_0\right), \quad (2.10)$$

where the Lagrange multiplier $\lambda(\theta) \in \mathbb{R}^{n+2}$ is such that:

$$\hat{E}\left[g_2(\theta) \exp\left(\lambda(\theta)' g_2(\theta)\right) | x_0\right] = 0, \quad (2.11)$$

for all θ , and $\hat{E}(g(\theta)|x_t)$ and $\hat{V}(g(\theta)|x_t)$ denote the expectation and variance of $g(Y; \theta)$, respectively, w.r.t. the kernel estimator $\hat{f}(y|x_t)$. The first part of the concentrated objective function (2.10) is reminiscent from the conditional version of the continuously updated GMM [Ai, Chen (2003), Antoine, Bonnal, Renault (2007)]. The estimator of $f(y|x_0)$ is given by:

$$\hat{f}^*(y|x_0) = \frac{\exp\left(\lambda(\hat{\theta})' g_2(y; \hat{\theta})\right)}{\hat{E}\left[\exp\left(\lambda(\hat{\theta})' g_2(\hat{\theta})\right) | x_0\right]} \hat{f}(y|x_0), \quad y \in \mathcal{Y}. \quad (2.12)$$

This conditional density is used to estimate the pricing operator at time t_0 .

Definition 2. *The XMM estimator of the derivative price $c_{t_0}(h, k)$ is:*

$$\hat{c}_{t_0}(h, k) = \int M_{t_0, t_0+h}(\hat{\theta}) (\exp R_{t_0, h} - k)^+ \hat{f}^*(y|x_0) dy, \quad (2.13)$$

for any time-to-maturity $h \leq \bar{h}$ and any moneyness strike k . The estimator of the pricing operator density at time t_0 up to time-to-maturity \bar{h} is $M_{t_0, t_0+\bar{h}}(\hat{\theta}) \hat{f}^*(y|x_0)$.

The constraints (2.9) imply that the estimator $\hat{c}_{t_0}(h, k)$ is equal to the observed option price $c_{t_0}(h_j, k_j)$ when $h = h_j$ and $k = k_j$, $j = 1, \dots, n$. The large sample properties of estimators $\hat{\theta}$ and $\hat{c}_{t_0}(h, k)$ in Definitions 1 and 2 are examined in Section 3. These estimators are consistent and asymptotically normal for large samples T of the time series of underlying asset returns, but a fixed number n of observed derivative prices at t_0 . The linear combinations of θ that are identifiable from the uniform moment restrictions (2.6) on the riskfree asset and the underlying asset only, are estimated at the standard parametric rate \sqrt{T} . Any other direction η_2^* in the parameter space (see Section 2.4) and the derivative prices as well are estimated at the rate $\sqrt{Th_T^d}$ corresponding to nonparametric estimation of conditional expectations given $X = x_0$. The estimators of derivative prices are asymptotically efficient.⁴

2.6 Application to S&P 500 options

In this section we compare the XMM estimation and a traditional cross-sectional calibration approach using the data on the S&P 500 options in June 2005 with daily trading volume larger than 4000 contracts (see Section 2.1).

i) Cross-sectional calibration

The cross-sectional calibration is based on a parametric stochastic volatility model for the risk-neutral distribution Q . We assume that under Q the S&P 500 return is such that:

$$r_t = r_{f,t} - \frac{1}{2}\sigma_t^2 + \sigma_t \varepsilon_t^*, \quad (2.14)$$

⁴Asymptotically equivalent estimators are obtained by replacing the integral w.r.t. the kernel density estimator in (2.13) by the discrete sum involved in a kernel regression estimator. These latter estimators involve a smoothing w.r.t. X only, and are used in the application in Section 2.6.

where $r_{f,t}$ is the riskfree rate between $t-1$ and t , (ε_t^*) is a standard Gaussian white noise and σ_t^2 denotes the volatility. The volatility (σ_t^2) is stochastic, independent of shocks (ε_t^*) on returns and follows an Autoregressive Gamma (ARG) process [see Gouriéroux and Jasiak (2006), Darolles, Gouriéroux and Jasiak (2006)]. This model is a time-discretized Heston model [Heston (1993)]. The risk-neutral transition distribution of the stochastic volatility is defined from its conditional Laplace transform:

$$E^Q \left[\exp(-u\sigma_{t+1}^2) \mid \underline{\sigma}_t^2 \right] = \exp[-a^*(u)\sigma_t^2 - b^*(u)], \quad u \geq 0, \quad (2.15)$$

where $E^Q[\cdot]$ denotes the expectation under Q , $a^*(u) = \rho^* \frac{u}{1+c^*u}$ and $b^*(u) = \delta^* \log(1+c^*u)$. Parameter ρ^* , $\rho^* > 0$, is the risk-neutral first-order autocorrelation of volatility process (σ_t^2) ; parameter δ^* , $\delta^* \geq 0$, describes its (conditional) risk-neutral over-/under-dispersion; parameter c^* , $c^* > 0$, is a scale parameter.

The relative option prices are function of the current value of volatility σ_t^2 and of the three parameters $\theta = (c^*, \delta^*, \rho^*)$, that is, $c_t(h, k) = c(h, k; \theta, \sigma_t^2)$, say. Function c is computed by Fourier Transform methods as in Carr and Madan (1999).⁵ The volatility σ_t^2 is calibrated daily jointly with θ by minimizing the mean square errors as in (2.1). The calibrated parameter $\hat{\theta}_{t_0}$ and volatility $\hat{\sigma}_{t_0}$ for the first ten trading days of June 2005 are displayed in Table 1.

[Insert Table 1 : Calibrated parameters (cross-sectional approach)
for the S&P 500 options in June, 2005]

We also report the Root Mean Squared Errors $RMSE_{t_0} = \left\{ \frac{1}{n_{t_0}} \sum_{j=1}^{n_{t_0}} \left[c(h_j, k_j; \hat{\theta}_{t_0}, \hat{\sigma}_{t_0}^2) - c_{t_0}(h_j, k_j) \right]^2 \right\}^{1/2}$ as goodness of fit measure. We observe that 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 vary in time and are quite erratic. The variation of the goodness of fit is due to a large extent to the small and time-varying number of derivative prices used in the calibration.

ii) XMM estimation

⁵Equations (5)-(6) in Carr, Madan (1999) are used to compute the option price as a function of time-to-maturity h and discounted moneyness $B(t, t+h)k$. The term structure of riskfree bond prices is assumed exogenous and is estimated at date t_0 by cubic spline interpolation of market yields at available maturities.

In the XMM estimation we consider the bivariate vector of state variables:

$$X_t = (\tilde{r}_t, \sigma_t^2)', \quad (2.16)$$

where $\tilde{r}_t := r_t - r_{f,t}$ is the daily logarithmic return from closing prices of the S&P 500 index in excess of the riskfree rate, and σ_t^2 is an observable volatility factor. More specifically, σ_t^2 is the one-scale realized volatility computed from 30-minute S&P 500 returns [e.g., Andersen et al. (2003)].⁶ The parameterized sdf is exponential affine:

$$M_{t,t+1}(\theta) = e^{-r_{f,t+1}} \exp(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 \tilde{r}_{t+1}), \quad (2.17)$$

where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$. This specification of the sdf is justified by the fact that the corresponding semi-parametric model for the risk-neutral distribution nests the parametric model used for cross-sectional calibration in Section i) above. More precisely, the risk-neutral stochastic volatility model (2.14)-(2.15) can be derived from an historical stochastic volatility model of the same type and the exponential affine sdf (2.17) with appropriate restrictions on parameter θ (see Appendix A.3).

For each trading day t_0 of June 2005, we estimate by XMM the sdf parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$ and 5 option prices $c_{t_0}(h, k)$ at a constant time-to-maturity $h = 20$ days and moneyness strikes $k = .96, .98, 1, 1.02, 1.04$. The options are puts for $k < 1$, and calls for $k \geq 1$. The estimator is defined as in Section 2.5, using the current and previous $T = 1000$ daily historical observations on the state variables, and the derivative prices of the actively traded S&P 500 options at t_0 . We use a product Gaussian kernel and select two bandwidths for the state variables according to the standard rule of thumb [Silverman (1986)]. The estimation results for the first ten trading days of June 2005 are displayed in Table 2.

[Insert Table 2 : Estimated sdf parameters and option prices (XMM approach)
for the S&P 500 options in June, 2005]

A direct comparison of the estimated structural parameters given in Tables 1 and 2 is difficult, since the parameters in Table 1 concern the risk-neutral dynamics whereas those in Table 2 concern the sdf. For this reason, we focus in Sections iii) and iv) below on a direct comparison of estimated option prices. Nevertheless, the following two remarks are

⁶A similar one time scale realized volatility measure is the underlying state variable for variance swaps.

interesting: First, the XMM estimated parameters are much more stable over time than the calibrated parameters, since they use the same historical information on the underlying asset. Second, 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, at each evaluation of the criterion function and its partial derivatives w.r.t. the parameters, and at each step of the optimization algorithm. On the contrary, the full XMM estimation takes about one minute on a standard computer.

iii) Static comparison of estimated option prices

In Figure 1 we display the estimated relative prices of S&P 500 options on June, 1 and June 2, 2005, as a function of the discounted moneyness strike $\tilde{k} = B(t, t + h)k$ and time-to-maturity h .

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

The results for June, 1 (resp. June, 2) are displayed in the top (resp. bottom) panels. In the panels displayed on the right hand side, the solid lines correspond to XMM estimates, and in the panels on the left hand side to calibrated estimates. Circles correspond to the observed prices of highly traded options. On June, 1, there are four highly traded times-to-maturity, which are 12, 57, 77, and 209-day, respectively. For the longest time-to-maturity $h = 209$, there is only one actively traded call option (resp., one put for $h = 57$ and two calls for $h = 77$). All remaining highly traded options correspond to time-to-maturity 12. We find that both put and call with identical moneyness strike and time-to-maturity can be highly traded, which reveals that these prices are not compatible with the no-arbitrage assumptions. In the literature [see e.g. Aït-Sahalia and Lo (1998)], this problem is solved by selecting the put if $\tilde{k} < 1$, and the call, otherwise. After applying this procedure, we end up with 11 highly traded options in June, 1, and 8 highly traded options for the 3 times-to-maturity in June, 2, 2005. Both calibration and XMM can also be used for pricing puts and calls that do not correspond to highly traded times-to-maturity and highly traded moneyness strikes. To show this, Figure 1 includes (dashed lines) the times-to-maturity 120 for June, 1, 2005, and 119 for June, 2, 2005, which are not traded. The calibration

method assumes a parametric risk-neutral model, whereas the XMM risk-neutral model is semiparametric. Any specification error in the fully parametric pricing model is detrimental for the calibration approach, while the XMM approach is more flexible. An advantage of XMM is that, by construction, the estimated derivative prices coincide with the market prices for highly traded options, while the calibrated prices differ from the observed prices, and these discrepancies can be large. This is not due to the specific parametric risk-neutral model that was used. It would also occur if a more complicated parametric model, or the nonparametric approach proposed in Aït-Sahalia and Lo (1998), was used. By construction, the risk-neutral stochastic volatility model that underlies the calibration method produces smooth symmetric option pricing functions w.r.t. the log-moneyness [see Gouriéroux, Jasiak (2001), Chapter 13.1.5], with similar types of curvature when the time-to-maturity increases. In contrast, the XMM produces skewed option pricing functions, which means that the method captures the leverage effect and its term structure.

iv) Dynamic comparison of estimated option prices

Let us now consider the dynamics of the option pricing function. In Figure 2 we display the time series of implied volatilities at a fixed time-to-maturity $h = 20$ and moneyness strikes $k = .96, .98, 1, 1.02, 1.04, 1.06$ for all trading days in June, 2005.

[Insert Figure 2: Time series of implied volatilities for S&P 500 options in June, 2005]

The XMM implied volatilities are indicated by circles and the calibrated implied volatilities are marked by squares. The XMM implied volatilities are more stable over time because of the use of the historical information on the underlying asset. Since most of the highly traded options have moneyness strikes close to at-the-money, the calibration approach is rather sensitive to unfrequently observed option prices with extreme strikes.

The sample means of the two time series of implied volatilities in Figure 2 are different. In particular, for $k = .96, .98, 1$, the XMM implied volatilities are on average larger than the calibrated ones, and smaller for $k = 1.04$ and $k = 1.06$. The reason is that the XMM approach captures the smirk, i.e. the skewness, in the implied volatility curve, while the calibrated model can reproduce either a smile, or a flat pattern only. Finally, we observe

some evidence of the week-end effect in the XMM implied volatilities for $k = .98, 1, 1.02$ while this effect is missing in the calibrated implied volatilities.

2.7 Information loss and precautionary risk management

In the above implementation of XMM, we considered a single cross-section of observed derivative prices only, date-by-date. This procedure is in line with current practice. The approach can be easily extended by increasing the finite number of cross-sections to, for instance, the last 10 trading days,⁷ without a significant effect on the asymptotic properties of the estimators. An interesting question is whether, instead of the XMM estimator, one could use a GMM estimator that exploits the observed option prices of all trading days in the sample. In this subsection we argue that this is possible, but the large sample properties of such a GMM estimator are unknown. In particular, the convergence rate of the estimator of the sdf parameter θ is hard to assess, and it can be difficult to derive the confidence intervals for option prices.

Let us first focus on the parameter θ and recall that the standard GMM requires a time-invariant number of uniform moment restrictions. A time-invariant set of uniform conditions can be formulated as:

$$E [\tilde{g}(Y_t; \theta) | X_t] = 0, \quad \forall t, \quad (2.18)$$

where $\tilde{g}(Y_t; \theta) = \text{vec}(\tilde{G}_t(\theta))$ and $\tilde{G}_t(\theta) = [\tilde{g}_{h,k,t}(\theta)]$ is a matrix with elements $\tilde{g}_{h,k,t}(\theta) = [M_{t,t+h}(\theta) (\exp R_{t,h} - k)^+ - c_t(h, k)] Z_{h,k,t}$ indexed by the time-to-maturity h and the moneyness strike k . The activity indicator $Z_{h,k,t}$ for derivative (h, k) at date t is such that $Z_{h,k,t} = 1$, if this derivative is traded at date t , and $= 0$, otherwise. The vector $Z_t = \text{vec}(Z_{h,k,t})$ of activity indicators is included in the state variables vector X_t . The vector $\tilde{g}(Y_t; \theta)$ ideally contains the moment restrictions for all the options (h, k) that can be traded somewhen on the market. The time-to-maturity h is an integer between 1 and a maximal time-to-maturity H (e.g., $H = 500$, that is, 2 years). The number of potential moneyness strikes k is also large. Indeed we have $k = K/p_t$, where the available strikes K at a given date t are about hundred equally-spaced values around the current underlying

⁷Or the 10 past trading days with the values of the state variable closest to x_0 .

asset price p_t .⁸ Both p_t and K are multiple of the price tick. As p_t moves over time, a large number of potential moneyness strikes k are generated. Even when this number is finite, we get a very large number of conditional moment restrictions in (2.18) which are weakly informative. Indeed, at any given date, most of these conditional moment restrictions concern non-traded derivatives and thus amount to trivial equalities.

An additional feature of the estimation problem is the nonstationary behavior of the activity indicators Z_t . This nonstationarity is induced by the issuing cycle and the seasoning effect in times-to-maturity, the dynamics in admissible moneyness strikes, and the trading activity. Therefore the large sample properties of the GMM estimator $\tilde{\theta}$, say, derived from (2.18) are likely nonstandard. The issuing and seasoning effect in times-to-maturity alone is not expected to significantly impact the root- T convergence of the GMM estimator $\tilde{\theta}$, since a given time-to-maturity can be traded recurrently according to a deterministic cyclical pattern. However, the root- T convergence of the estimator $\tilde{\theta}$ requires a recurrence property for the trading activity such that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Z_{h,k,t} = \pi(h, k) > 0$, P -a.s., for a relevant set of options (h, k) . This set of options (h, k) has to allow for the identification of θ . This recurrence assumption is rather strong. Its theoretical justification requires a structural model of the trading activity. In practice, the frequencies $\frac{1}{T} \sum_{t=1}^T Z_{h,k,t}$ are close to zero for most options (h, k) and it is difficult to test empirically the validity of the recurrence hypothesis.⁹

The weak informational content of the conditional moment restrictions (2.18) gives some insight on the convergence rates of the GMM estimator $\tilde{\theta}$. Even if at least one derivative price is observed at each date, we cannot always expect to recover the standard parametric rate of convergence for all components of $\tilde{\theta}$. To see this, let us suppose that a single derivative price $c_t(h_t, k_t)$ is observed at each date t , and that a single component η_2^* of θ is unidentifiable from the uniform moment restrictions of the underlying asset and the risk-free asset. From standard GMM theory it follows that the rate of convergence of the GMM

⁸The exact number of admissible strikes depends on the time-to-maturity.

⁹A similar issue arises in the estimation of the term structure of zero-coupon bonds. In this application however the concern for the recurrence property is less pronounced, since the zero-coupon bonds market is organized according to a single characteristic, namely time-to-maturity.

estimator of η_2^* is \sqrt{T} , if $(k_t, h_t) = (k, h)$ is time-invariant, which is quite unrealistic. In contrast in XMM, where the current derivative price only is used, Proposition 1 in Section 3.3 and Corollary 6 in Section 3.5 show that the rate of convergence of the XMM estimator of η_2^* is the nonparametric rate $\sqrt{Th_T^d}$. Intuitively, the information on η_2^* contained in the observed derivative price at t is a function, say, of moneyness strike k_t , time-to-maturity h_t and state x_t . Therefore, the information on η_2^* contained in the entire time series of observed option prices is $\sum_{t=1}^T i(k_t, h_t, x_t)$. The parametric rate is reached, if this quantity is of order T as $T \rightarrow \infty$. In general, intermediate convergence rates between the parametric rate and the nonparametric one ¹⁰ are expected. The nonparametric convergence rate for the GMM estimator may occur, for instance, when the structural parameter η_2^* is a risk premium for extreme risk. The point is that there is no way to get reliable information about extreme risks that may lead to an improved convergence rate, since the trading frequencies of derivatives with very deep in- or out-of-the-money strikes are typically quite low.

Let us now consider the option prices. An estimator of the pricing operator at t_0 can be obtained by plugging into equations (2.12) and (2.13) the GMM estimator $\tilde{\theta}$ from (2.18). If the recurrence property is satisfied, and the convergence rate of the GMM estimator $\tilde{\theta}$ is larger than the nonparametric rate $\sqrt{Th_T^d}$, the results of Section 3 can be used to show that the estimation of θ is irrelevant for the estimation of the option prices. The estimated option prices are asymptotically normal and the confidence bands are derived from the results in Section 3 as if parameter θ were known. However, if the recurrence property is not satisfied and some components of $\tilde{\theta}$ converge at the nonparametric rate $\sqrt{Th_T^d}$, then the estimation of θ cannot be neglected even asymptotically. This complicates the derivation of the large sample distribution and confidence bands of the estimated option prices. In particular, neglecting the effect of the estimation of θ when θ has a slow rate of convergence, leads to spuriously narrow confidence bands for option prices and underestimation of risk.

In conclusion, the advantage of focusing on a finite number of cross-sections of derivative prices is that the XMM estimator has a known rate of convergence and an explicit asymptotic (normal) distribution. In particular, the XMM provides prudential confidence intervals [see Section 3.4 iii)] for option prices, in line with precautionary risk management.

¹⁰Once kernel localization is applied at each date.

3 Theoretical properties of XMM

Let us now examine the theoretical properties of the XMM estimator introduced in Section 2 for derivative pricing. First we discuss the parameter of interest and the moment restrictions. Next we derive the identification conditions and explain how they differ from the standard GMM conditions. Under the identification conditions, we derive the optimal instruments and the kernel nonparametric efficiency bound. Finally, we prove the consistency and kernel nonparametric efficiency of the XMM estimator.

3.1 The parameter of interest

Let us consider a semi-parametric estimation of the conditional moments:

$$E_0 [a(Y; \theta_0) | X = x_0], \quad (3.1)$$

subject to the uniform and local conditional moment restrictions:

$$E_0 [g(Y; \theta_0) | X = x] = 0, \quad \forall x \in \mathcal{X}, \quad P_0\text{-a.s.}, \quad (3.2)$$

$$E_0 [\tilde{g}(Y; \theta_0) | X = x_0] = 0, \quad (3.3)$$

where θ_0 is the true parameter value and $E_0[\cdot]$ denotes the expectation under the true DGP P_0 . The unknown parameter value θ_0 is in set $\Theta \subset \mathbb{R}^p$, variables (X, Y) are in $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^{\tilde{d}}$, and x_0 is a given value in \mathcal{X} . The moment functions g and \tilde{g} are given vector-valued functions on $\mathcal{Y} \times \Theta$. Function a on $\mathcal{Y} \times \Theta$ has dimension L . In the derivative pricing application in Section 2, vector $a(Y; \theta)$ is the product of sdf $M_{t,t+h}(\theta)$ and payoff $(\exp R_{t,h} - k)^+$ on the L derivatives of interest. Vector functions g and \tilde{g} define the uniform moment restrictions on the riskfree asset and underlying asset, and the local moment restrictions on observed derivative prices [see (2.5) and (2.6)].

The conditional moment of interest (3.1) can be viewed as an additional parameter β_0 in $B \subset \mathbb{R}^L$ identified by the local conditional moment restrictions:

$$E_0 [a(Y; \theta_0) - \beta_0 | X = x_0] = 0. \quad (3.4)$$

Thus, all parameters to be estimated are in vector $\theta_0^* = (\theta_0', \beta_0')$ that satisfies the extended set of uniform and local moment restrictions (3.2), (3.3), (3.4) [see Back, Brown (1992) for

a setting with unconditional moment restrictions]. In this interpretation, β_0 is the parameter of interest while θ_0 is a nuisance parameter.

3.2 Identification

Let us now consider the identification of extended parameter θ_0^* . From moment restriction (3.4), it follows that β_0 is identified if θ_0 is identified. Thus, we can restrict the analysis to the identification of θ_0 .

i) The identification assumption

Assumption 1: *The true value of the parameter θ_0 is globally identified:*

$$\left\{ \begin{array}{l} E_0 [g(Y; \theta)|X = x] = 0, \forall x \in \mathcal{X}, P_0\text{-a.s.} \\ E_0 [\tilde{g}(Y; \theta)|X = x_0] = 0 \end{array} \right. , \theta \in \Theta \Rightarrow \theta = \theta_0.$$

Assumption 2: *The true value of the parameter θ_0 is locally identified:*

$$\left\{ \begin{array}{l} E_0 \left[\frac{\partial g}{\partial \theta'}(Y; \theta_0)|X = x \right] \alpha = 0, \forall x \in \mathcal{X}, P_0\text{-a.s.} \\ E_0 \left[\frac{\partial \tilde{g}}{\partial \theta'}(Y; \theta_0)|X = x_0 \right] \alpha = 0 \end{array} \right. , \alpha \in \mathbb{R}^p \Rightarrow \alpha = 0.$$

We need to distinguish the linear transformations of θ_0 , $\alpha'\theta_0$, say, where $\alpha \in \mathbb{R}^p$, that are identifiable from the uniform conditional moment restrictions (3.2) alone, from the linear transformations of θ_0 that are identifiable only from both uniform and local conditional moment restrictions (3.2) and (3.3). The former transformations are called full-information identifiable, while the latter ones are called full-information unidentifiable. Let us consider the linear space:

$$\mathcal{J}^* = \left\{ \alpha \in \mathbb{R}^p : E_0 \left[\frac{\partial g}{\partial \theta'}(Y; \theta_0)|X = x \right] \alpha = 0, \forall x \in \mathcal{X}, P_0\text{-a.s.} \right\},$$

of dimension $\dim \mathcal{J}^* = s^* \leq p$, say. The full-information identified transformations are $\alpha'\theta_0$ with $\alpha \in (\mathcal{J}^*)^\perp$, while the full-information unidentifiable transformations are $\alpha'\theta_0$ with $\alpha \in \mathbb{R}^p \setminus (\mathcal{J}^*)^\perp$. There exist parameterizations of the moment functions such that $p - s^*$ components of the parameter vector are full-information identifiable, and s^*

components are full-information unidentifiable. Indeed, let us consider a linear change of parameter:

$$\eta^* = \begin{pmatrix} \eta_1^* \\ \eta_2^* \end{pmatrix} = \begin{pmatrix} R_1' \theta \\ R_2' \theta \end{pmatrix}, \quad (3.5)$$

where $R^* = [R_1, R_2]$ is an orthogonal (p, p) matrix, R_1 is a $(p, p - s^*)$ matrix whose columns span $(\mathcal{J}^*)^\perp$, and R_2 is a (p, s^*) matrix whose columns span \mathcal{J}^* . The matrices R_1 and R_2 depend on the DGP P_0 . The subvectors $\eta_1^* \in \mathbb{R}^{p-s^*}$ and $\eta_2^* \in \mathbb{R}^{s^*}$ are full-information identifiable, and full-information underidentified, respectively. In particular, under Assumption 2 the matrix $E_0 \left[\frac{\partial \tilde{g}}{\partial \theta'}(Y; \theta_0) | X = x_0 \right] R_2$ has full column rank. Thus, a necessary order condition for local identification is that the number of local moment restrictions is larger than or equal to s^* , i.e. the dimension of the linear space \mathcal{J}^* characterizing the full-information unidentifiable parameters.

The standard GMM considers uniform moment restrictions only, and assumes full-information identification for the full vector θ_0 . To illustrate the difference with our setting, let us consider the parametric stochastic volatility model that is compatible with the parametric risk-neutral specification of Section 2.6 i) and the semi-parametric model of Section 2.6 ii) [see Appendix A.3.1]. The riskfree rate is set equal to zero for expository purpose. The historical distribution is such that:

$$r_t = \gamma \sigma_t^2 + \sigma_t \varepsilon_t, \quad (3.6)$$

where γ is a parameter, $\varepsilon_t \sim IIN(0, 1)$ and the stochastic volatility (σ_t^2) follows an ARG process with parameters ρ , δ and c . The risk-neutral distribution is such that:

$$r_t = \gamma^* \sigma_t^{*2} + \sigma_t \varepsilon_t^*, \quad (3.7)$$

where $\varepsilon_t^* \sim IIN(0, 1)$ and (σ_t^{*2}) follows an ARG process with parameters ρ^* , δ^* and c^* . Finally, the sdf is given by:

$$M_{t,t+1}(\theta) = \exp \left(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 r_{t+1} \right). \quad (3.8)$$

The parameters are subject to restrictions that ensure the coherency of the model [see (A.24) in Appendix A.3.1] and the absence of arbitrage opportunities (see Lemma A.6). The parameter θ_0 is identifiable when the DGP P_0 is compatible with this parametric stochastic

volatility model and the semi-parametric specification of Section 2.6 ii) is used for identification (see Appendix A.3.2 for the proof). However, the dimension of the subspace of vectors $\alpha \in \mathbb{R}^4$ associated with the full-information identifiable transformations $\alpha'\theta_0$ is only 3 (i.e., $s^* = 1$). Thus, the uniform moment restrictions (3.2) on the riskfree asset and the underlying asset are insufficient to identify a unique sdf in the parametric family. A full-information unidentifiable transformation $\eta_2^* \in \mathbb{R}$ is associated with vector $R_2 = \left(-\delta \frac{c}{1+c\lambda_2}, 1, -\rho \frac{1}{(1+c\lambda_2)^2}, 0\right)'$, where $\lambda_2 = \theta_2 + \gamma^2/2 - 1/8$. Since $\frac{\partial \theta}{\partial \eta_2^*} = R_2$, the sdf parameters θ_1 , θ_2 and θ_3 associated with time discounting and stochastic volatility are identifiable only if the local moment restrictions (3.3) from the observed derivative prices are also taken into account, while the sdf parameter θ_4 associated with the underlying asset returns is full-information identifiable.

ii) Admissible instrumental variables

It is also useful to discuss a weaker notion of identification, based on a given matrix of instruments $Z = H(X)$. The uniform conditional moment restrictions (3.2) imply the unconditional moment restrictions $E_0 [Z \cdot g(Y; \theta_0)] =: E_0 [g_1(X, Y; \theta_0)] = 0$. Let us denote the whole set of local conditional moment restrictions at x_0 as $E_0 [g_2(Y; \theta_0) | X = x_0] = 0$, where $g_2 = (g', \tilde{g}')$. Thus, parameter $\theta_0^* = (\theta_0', \beta_0)'$ satisfies the moment restrictions:

$$E_0 [g_1(X, Y; \theta_0)] = 0, \quad (3.9)$$

$$E_0 [g_2(Y; \theta_0) | X = x_0] = 0, \quad (3.10)$$

$$E_0 [a(Y; \theta_0) - \beta_0 | X = x_0] = 0. \quad (3.11)$$

Definition 3. *The instrument Z is admissible if the true value of the parameter θ_0 is globally identified by the moment restrictions (3.9)-(3.10) and locally identified by their differential counterparts.*

We introduce the linear change of parameter:

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} R'_{1Z}\theta \\ R'_{2Z}\theta \end{pmatrix}, \quad (3.12)$$

where $R = [R_{1,Z}, R_{2,Z}]$ is an orthogonal (p, p) matrix, and $R_{2,Z}$ is a (p, s_Z) matrix whose columns span the null space $\mathcal{J}_Z = \text{Ker } E_0 \left[\frac{\partial g_1}{\partial \theta'}(X, Y; \theta_0) \right]$. The subvector of parameters η_1

is locally identified by the unconditional moment restrictions (3.9), whereas the subvector of parameters η_2 is identifiable only from both sets of restrictions (3.9) and (3.10).

3.3 Kernel moment estimators

Let Z be a given admissible instrument. Let us introduce a GMM-type estimator of θ_0^* that is obtained by minimizing a quadratic form of sample counterparts of moments (3.9)-(3.11). The kernel density estimator $\hat{f}(y|x)$ in (2.8) is used to estimate the conditional moments in (3.10) and (3.11):

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

and similarly for $E_0[a(Y; \theta) - \beta|X = x_0]$.

Definition 4. A kernel moment estimator $\hat{\theta}_T^* = (\hat{\theta}_T', \hat{\beta}_T')$ of parameter $\theta_0^* = (\theta_0', \beta_0')$ is defined by:

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

where

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

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

The empirical moments in $\hat{g}_T(\theta^*)$ have different rates of convergence, that are parametric and nonparametric. To derive the asymptotic properties of the kernel moment estimator $\hat{\theta}_T^*$, we prove the weak convergence of a suitable empirical process derived from $\hat{g}_T(\theta^*)$ (Lemma A.1 in Appendix A.1.2). The definition of the empirical process is in the spirit of Stock, Wright (2000), but the choice of weak instruments is different.¹¹ The various rates

¹¹The local moment restrictions (3.10) can be approximately written as $E_0[g_2(Y; \theta_0) | X = x_0] \simeq E_0[Z_T g_2(Y; \theta_0)] = 0$, where $Z_T = K\left(\frac{X - x_0}{h_T}\right) / [h_T^d f_X(x_0)]$ and f_X denotes the unconditional pdf of X . The “instrument” Z_T is weak in the sense of Stock, Wright (2000). However, it depends on T , and the rates of convergence of the estimators differ from the rates of convergence in Stock, Wright (2000).

of convergence in $\widehat{g}_T(\theta^*)$ can be disentangled by using the transformed parameters $(\eta'_1, \eta'_2)'$ and β defined in (3.12) (see Appendix A.1.4).

Proposition 1. *Under Assumptions A.1-A.25 in Appendix A.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 $B_\infty = -\sqrt{\bar{c}} \left(J'_0 \Omega J_0 \right)^{-1} J'_0 \Omega b_0$, with $\bar{c} := \lim Th_T^{d+2m} \in [0, \infty)$ and $m \geq 2$ the order of differentiability of the pdf f_X of X , and matrices J_0 , V_0 , and vector b_0 are given in (A.2) and (A.4) in Appendix A.1, and depend on kernel K .

3.4 Kernel nonparametric efficiency bound

The class of kernel moment estimators helps us define a convenient notion of efficiency for estimating $\beta_0 = E_0(a|x_0) := E_0[a(Y; \theta_0)|X = x_0]$. We first consider a scalar parameter β_0 and derive the optimal weighting matrix, admissible instruments and bandwidth. The kernel moment estimator can be asymptotically biased. If we restrict ourselves to asymptotically unbiased kernel moment estimators for practical purposes, a bias-free kernel nonparametric efficiency bound can be derived for scalar or vector parameter β_0 . We assume that kernel K is a product kernel $K(u) = \prod_{i=1}^d \kappa(u_i)$ of order m .

i) Efficiency bound with optimal rate of convergence

Let us consider a bandwidth sequence $h_T = cT^{-\frac{1}{2m+d}}$, where $c > 0$ is a constant. From Proposition 1, it follows that estimator $\widehat{\beta}_T$ achieves the optimal d -dimensional nonparametric rate of convergence $T^{-\frac{m}{2m+d}}$, and its asymptotic Mean Square Error (MSE) constant $M(\Omega, Z, c, a) > 0$ depends on the weighting matrix Ω , instrument Z , bandwidth constant c and moment function a .

Definition 5. *The kernel nonparametric efficiency bound $\mathcal{M}(x_0, a)$ for estimating $\beta_0 = E_0(a|x_0)$ is the smallest possible value of $M(\Omega, Z, c, a)$ corresponding to the optimal choice of weighting matrix Ω , admissible instrument Z and bandwidth constant c .*

Proposition 2. *Let Assumptions 1, 2 and A.1-A.26 in Appendix A.1 hold. (i) There exist an optimal weighting matrix $\Omega^*(a)$ and an optimal bandwidth constant $c^*(a)$. They are given in (A.9) and (A.10) in Appendix A.1.6. (ii) Any instrument:*

$$Z^* = E_0 \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X), \quad (3.13)$$

where $W(X)$ is a positive definite matrix, P_0 -a.s., is optimal, independent of a .

It is easily verified that $\mathcal{J}_{Z^*} = \mathcal{J}^*$ for any instrument Z^* in (3.13). Thus, an admissible instrument Z is optimal if and only if the corresponding unconditional moment restrictions (3.9) identify all full-information identifiable parameters. Since we focus on the estimation of local conditional moment β_0 , the set of optimal instruments is larger than the standard set of instruments for efficient estimation of a structural parameter θ_0 identified by (3.2). While in the standard framework $W(X) = V_0 [g(Y; \theta_0) | X]^{-1}$ is the efficient weighting matrix for conditionally heteroskedastic moment restrictions [e.g., Chamberlain (1987)], any choice of a positive definite matrix $W(X)$ is asymptotically equivalent for estimating β_0 . Moreover, the set of optimal instruments is independent of the selected kernel.

The expression of the kernel nonparametric efficiency bound is easily formulated in terms of the transformed parameters $(\eta_1^*, \eta_2^*)'$ defined in (3.5).

Proposition 3. *Under Assumptions 1, 2 and A.1-A.26 in Appendix A.1, the kernel nonparametric efficiency bound $\mathcal{M}(a, x_0)$ is given by the lower-right element of the matrix*

$$\left(J_0^{*'} \left(\frac{1}{c^{*d}} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{*2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_0^* \right)^{-1}, \quad (3.14)$$

where $w^2 = \int_{\mathbb{R}^d} K(u)^2 du$, $w_m = \int_{\mathbb{R}} v^m \kappa(v) dv$, $c^* = c^*(a)$ is the optimal bandwidth constant given in (A.9) in Appendix A.1.6,

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

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 .

The matrix in (3.14) resembles the GMM efficiency bound for estimating parameters $(\eta_2^*, \beta)'$ from orthogonality conditions based on function $(g_2', a - \beta)'$, with η_1^* known. Since moment restrictions (3.10)-(3.11) are conditional on $X = x_0$, the variance of the orthogonality conditions is replaced by the asymptotic 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 at x_0 (up to a scale factor), and the expectations in the Jacobian matrix J_0^* are conditional on $X = x_0$. In particular, the efficiency bound depends on the likelihood of observing the conditioning variable close to x_0 by means of $f_X(x_0)$. The estimation of the full-information identified parameter η_1^* is irrelevant for the efficiency bound of β since the estimation of η_1^* achieves a faster parametric rate of convergence.

ii) Bias-free kernel nonparametric efficiency bound

Asymptotically unbiased estimators require a bandwidth sequence such that $\bar{c} = \lim Th_T^{2m+d} = 0$. When $\bar{c} = 0$ in Proposition 1, the asymptotic variance of $\sqrt{Th_T^d} (\hat{\beta}_T - \beta_0)$ can be written as $w^2 V(Z, \Omega, a)$, where $V(Z, \Omega, a)$ is independent of the selected kernel K and bandwidth h_T .

Definition 6. *The bias-free kernel nonparametric efficiency bound $\mathcal{B}(a, x_0)$ is the smallest asymptotic variance $V(Z, \Omega, a)$ corresponding to the optimal choice of the weighting matrix Ω and of the instrument Z .*

Corollary 4. (i) *There exists an optimal choice of the weighting matrix Ω and of the instruments Z . The optimal instruments are given in (3.13) and the optimal weighting matrix in Appendix A.1.7. (ii) The bias-free kernel non-parametric efficiency bound $\mathcal{B}(x_0, a)$ is:*

$$\begin{aligned} \mathcal{B}(x_0, a) &= \frac{1}{f_X(x_0)} \left\{ V(a) - \text{Cov}(a, g_2) V(g_2)^{-1} \text{Cov}(g_2, a) \right. \\ &\quad + \left[E \left(\frac{\partial a}{\partial \theta'} \right) R_2 - \text{Cov}(a, g_2) V(g_2)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} \right) R_2 \right] \left[R_2' E \left(\frac{\partial g_2'}{\partial \theta} \right) V(g_2)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} \right) R_2 \right]^{-1} \\ &\quad \left. \left[R_2' E \left(\frac{\partial a'}{\partial \theta} \right) - R_2' E \left(\frac{\partial g_2'}{\partial \theta} \right) V(g_2)^{-1} \text{Cov}(g_2, a) \right] \right\}, \end{aligned}$$

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

Since the expression of $\mathcal{B}(x_0, a)$ is a quadratic function of a , this formula holds for vector moment functions a , as well. When the parameter θ_0 itself is full-information iden-

tifiable, the bias-free kernel nonparametric efficiency bound becomes:

$$\mathcal{B}(x_0, a) = \frac{1}{f_X(x_0)} \{V(a|x_0) - \text{Cov}(a, g_2|x_0)V(g_2|x_0)^{-1} \text{Cov}(g_2, a|x_0)\}. \quad (3.15)$$

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 (3.15) is simply the variance-covariance matrix of the residual term in the affine regression of a on g_2 performed conditional on x_0 . 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, Bonnal and Renault (2007). In the general case, the efficiency bound $\mathcal{B}(x_0, a)$ balances the gain in information from the local conditional moment restrictions and the efficiency cost due to full-information underidentification of θ_0 .

iii) Illustration with S&P 500 options

Let us derive the bias-free kernel nonparametric efficiency bounds for derivative prices estimation when the DGP P_0 is the parametric stochastic volatility model (3.6)-(3.8) with parameters given by $\gamma = 0.360$, $\rho = 0.960$, $\delta = 1.047$, $c = 3.65 \cdot 10^{-6}$, $\theta_1 = .456 \cdot 10^{-6}$, $\theta_2 = -0.059$, $\theta_3 = 0.114$ and $\theta_4 = 0.860$. The ARG parameters ρ , δ , c are set to match the stationary mean, variance and first-order autocorrelation of the realized volatility σ_t^2 of the S&P 500 index in the period from June, 1, 2001 to Mai, 31, 2005. The risk premia parameters θ_2 and θ_4 for stochastic volatility and underlying asset return, respectively, correspond to the XMM estimates obtained in Section 2.6 ii) for the S&P 500 options in June, 1, 2005 (see Table 2). Parameters γ , θ_1 , θ_3 are then fixed by the no-arbitrage restrictions (A.24). At a current date t_0 , the prices of $n = 11$ actively traded call options are observed, with the same times-to-maturity and moneyness strikes as the S&P 500 put and call options with daily traded volume larger than 4000 contracts in June, 1, 2005 [see Section 2.6 iii)]. The current values of the state variables are the return and the realized volatility of the S&P 500 index on June, 1, 2005.

Let us first consider the time-to-maturity $h = 12$ days. The bias-free kernel nonparametric efficiency bound on the call option prices is displayed in Figure 3.

[Insert Figure 3: Bias-free kernel nonparametric efficiency bound, time-to-maturity 12-day]

The dashed line represents the theoretical call prices $E(a(k)|x_{t_0})$ computed under the DGP P_0 , and the dashed lines represents 95% confidence intervals $E(a(k)|x_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(x_{t_0}, k)^{1/2}$, as a function of moneyness k . The circles indicate the theoretical prices of the observed derivatives at time-to-maturity 12-day. In order to better visualize the pattern of the kernel nonparametric efficiency bound as a function of the moneyness strike, $\sqrt{w^2/Th_T^2}$ is set ten times larger than the value implied by the sample size and the bandwidths used in the empirical application.¹² The width of the confidence interval for derivative price $E(a(k)|x_{t_0})$ depends on moneyness strike k . This width is zero when k corresponds to the moneyness strikes of the observed calls. The width of the confidence interval is close to zero 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 option price has to be equal to the underlying asset price or equal to zero, respectively, by the no-arbitrage condition.¹³ Finally, the confidence intervals in Figure 3 are rather narrow, especially for moneyness strikes $k < 1$. Since the call option price as a function of k is monotonically decreasing and convex, the 7 observed option prices at time-to-maturity $h = 12$ are very informative to estimate the prices at other moneyness strikes.

Let us now consider the time-to-maturity $h = 77$ days. The bias-free kernel nonparametric efficiency bound is displayed in Figure 4.

[Insert Figure 4 : Bias-free kernel nonparametric efficiency bound, time-to-maturity 77-day]

The same standardization for $\sqrt{w^2/Th_T^2}$ as above is used. The width of the confidence intervals is overall larger than in Figure 3. This is due to the different informational content of the set of observed option prices with different times-to-maturity of interest.

3.5 Asymptotic normality and efficiency of the XMM estimator

When the optimal instruments and optimal weighting matrix are used, the kernel moment estimator of β_0 in Definition 4 is kernel nonparametrically efficient. However, in appli-

¹²For expository purpose we consider symmetric confidence bands. These bands have to be truncated at zero to account for the positivity of derivative prices and get asymmetric bands. However, with the correct standardization, the truncation effect is negligible and arises only for large strikes.

¹³However, the relative accuracy can be poor in these moneyness regions.

cation to derivative pricing this estimator does not ensure a positive estimated state price density. This is a consequence of the quadratic GMM-type nature of the kernel moment estimators. The positivity of the estimated state price density is achieved by considering the information based XMM estimator defined in Section 2.5. In the general setting, the XMM estimator of β_0 is:

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

where $\hat{\theta}$ is defined in (2.10)-(2.11) and $\hat{f}^*(y|x_0)$ is defined in (2.12).

The large sample properties of the XMM estimator of β_0 are given in Proposition 5 below, for a bandwidth sequence that eliminates the asymptotic bias (see Appendix A.2).

Proposition 5. *Suppose the bandwidth is such that $\bar{c} = \lim Th_T^{d+2m} = 0$. Then the XMM estimator $\hat{E}^*(a|x_0)$ is consistent, converges at rate $\sqrt{Th_T^d}$, is asymptotically normal and bias-free kernel nonparametrically efficient:*

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

The XMM estimator of β_0 is asymptotically equivalent to the best kernel moment estimator of β_0 with optimal instruments and weighting matrix. The asymptotic distribution of the XMM estimator of θ_0 is given for the transformed parameter $(\eta_1^{*'}, \eta_2^{*'})'$ in (3.5).

Corollary 6. *Suppose that the bandwidth is such that $\bar{c} = \lim Th_T^{d+2m} = 0$. The XMM estimators $\hat{\eta}_1^*$ and $\hat{\eta}_2^*$ are asymptotically equivalent to the kernel moment estimators with optimal weighting matrix and instrument Z as in (3.13) with $W(X) = V_0 [g(Y; \theta_0)|X]^{-1}$.*

In particular, the rates of convergence for the full-information identifiable parameter η_1^* and the full-information unidentifiable parameter η_2^* are \sqrt{T} and $\sqrt{Th_T^d}$, respectively. The different rates of convergence of full-information identifiable and full-information unidentifiable parameters are reflected in the empirical results on the S&P 500 option data displayed in Table 2 in Section 2.6. If the DGP P_0 is the stochastic volatility model (3.6)-(3.8), parameter θ_4 is full-information identifiable [see Section 3.2 (ii)] and its estimates are very stable over time, whereas the estimates of the full-information unidentifiable parameters θ_1 , θ_2 , θ_3 feature a larger time variability. Proposition 5 and Corollary 6 extend the first-order

asymptotic equivalence between information based and quadratic GMM estimators [see Kitamura, Tripathi, Ahn (2004), Kitamura (2007)] to a setting including local conditional moment restrictions and allowing for full-information unidentifiable parameters.

4 Concluding remarks

The literature on joint estimation of historical and risk-neutral parameters is generally based on either Maximum Likelihood (ML), or GMM, type of methods. A part of this literature relies on uniform moment restrictions from a time series of spot prices only, and implicitly assumes that the risk premia parameters are full-information identified [e.g., Bansal, Viswanathan (1993), Hansen, Jagannathan (1997), Stock, Wright (2000)]. This hypothesis is not valid when some risk premia parameters can be identified only from option data. Another part of the literature exploits time series of both spot and a few options prices [e.g., Duan (1994), Chernov, Ghysels (2000), Pan (2002), Eraker (2004)]. However the activity on derivative markets is rather weak, and these approaches typically rely either on artificial option series which are approximately near-the-money and at short time-to-maturity, or on ad-hoc assumptions on time-varying options characteristics.

In this paper we introduce a new XMM estimator of derivative prices using jointly a time series of spot returns and a cross-section of derivative prices. We argue that these two types of data imply different types of conditional moment restrictions, that are either uniform, or local. First, the XMM approach allows for consistent estimation of the sdf parameters θ even if they are full-information unidentifiable. Second, the XMM estimator of the pricing operator at a given date is consistent for a fixed number of cross-sectionally observed derivative prices. These results are due to both the parametric sdf and the deterministic relationships between derivative prices that hold in a no-arbitrage pricing model with a finite number of state variables. The application to the S&P 500 options shows that the new XMM-based calibration approach outperforms the traditional cross-sectional calibration approach while being easy to implement. In particular, the XMM estimated option prices are compatible with the observed option prices for highly traded derivatives, and are more stable over time.

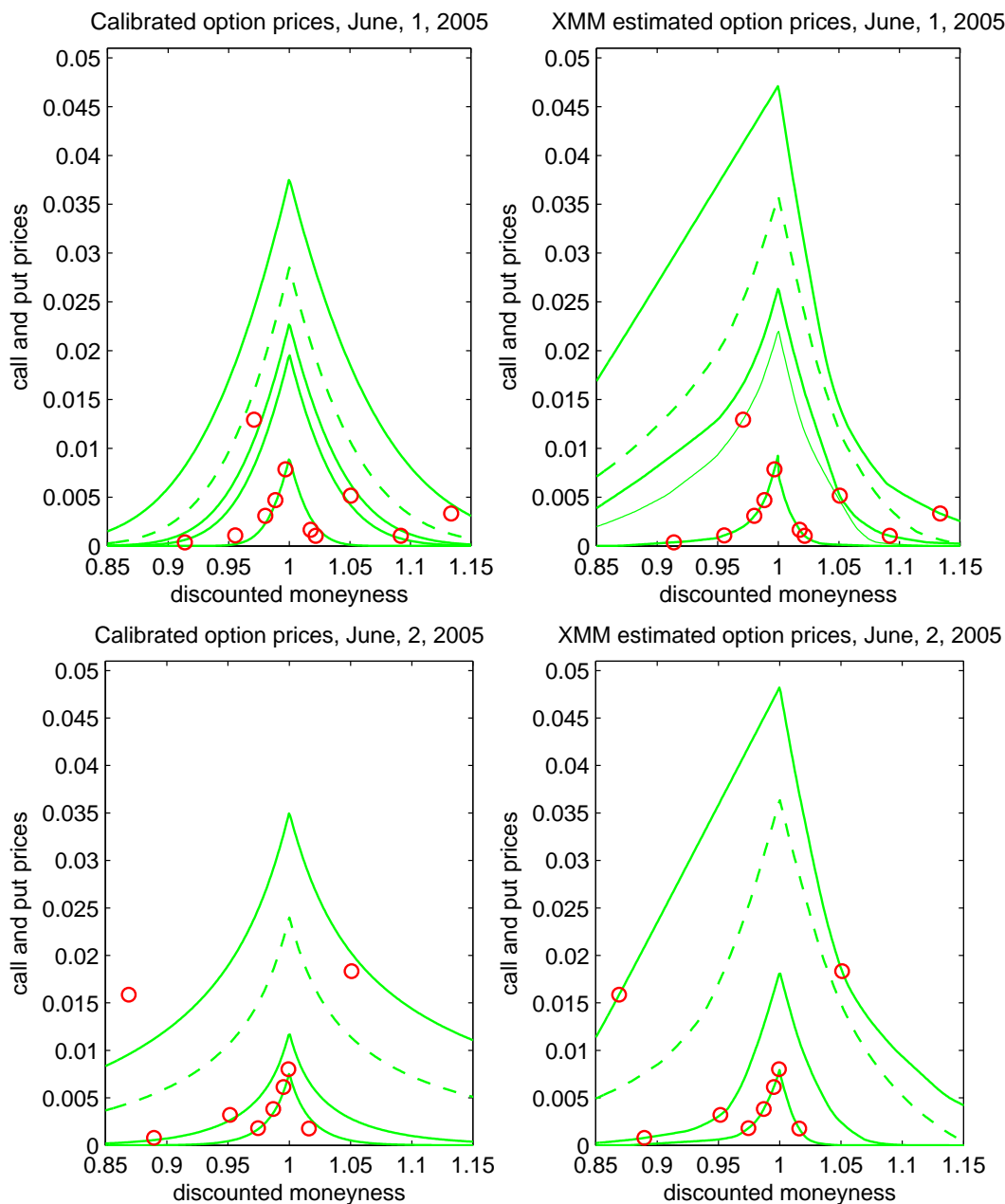
References

- [1] Ai, C., and X., Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", *Econometrica*, 71, 1795-1843.
- [2] Aït-Sahalia, Y., and J., Duarte (2003): "Nonparametric Option Pricing under Shape Restrictions", *Journal of Econometrics*, 116, 9-47.
- [3] Aït-Sahalia, Y., and A., Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Returns", *Journal of Finance*, 53, 499-547.
- [4] Andersen, T., Bollerslev, T., Diebold, F., and P., Labys (2003): "Modeling and Forecasting Realized Volatility", *Econometrica*, 71, 529-626.
- [5] Andrews, D. (1991): "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables", *Journal of Multivariate Analysis*, 38, 187-203.
- [6] Andrews, D. (1994): "Empirical Process Methods in Econometrics", in *Handbook of Econometrics*, Vol. 4, eds. Engle, R., and D., McFadden, North Holland, 2247-2294.
- [7] 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.
- [8] Back, K., and D., Brown (1992): "GMM, Maximum Likelihood and Nonparametric Efficiency", *Economics Letters*, 39, 23-28.
- [9] Back, K., and D., Brown (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.
- [10] Bakshi, G., Cao, C., and Z., Chen (1997): "Empirical Performance of Alternative Option Pricing Models", *Journal of Finance*, 52, 2003-2049.
- [11] Bansal, R., and S., Viswanathan (1993): "No Arbitrage and Arbitrage Pricing: A New Approach", *Journal of Finance*, 48, 1231-1362.
- [12] Billingsley, P. (1965): *Ergodic Theory and Information*. John Wiley, New York.
- [13] Bondarenko, O. (2003): "Estimation of Risk-neutral Densities using Positive Convolution Approximation", *Journal of Econometrics*, 116, 85-112.
- [14] Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Springer, New-York.
- [15] Brown, B., and W., Newey (1998): "Efficient Semiparametric Estimation of Expectations", *Econometrica*, 66, 453-464.
- [16] 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.

- [17] Carr, P., and D., Madan (1999): "Option Pricing and the Fast Fourier Transform", *Journal of Computational Finance*, 2, 61-73.
- [18] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.
- [19] 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.
- [20] Darolles, S., Gouriéroux, C., and J., Jasiak (2006): "Structural Laplace Transform and Compound Autoregressive Models", *Journal of Time Series Analysis*, 27, 477-504.
- [21] Duan, J. (1994): "Maximum Likelihood Estimation Using Price Data of the Derivative Contract", *Mathematical Finance*, 4, 155-167.
- [22] Eraker, B. (2004): "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices", *Journal of Finance*, 59, 1367-1403.
- [23] Gouriéroux, C., and J., Jasiak (2001): *Financial Econometrics: Problems, Models and Methods*, Cambridge University Press.
- [24] Gouriéroux, C., and J., Jasiak (2006): "Autoregressive Gamma Process", *Journal of Forecasting*, 25, 129-152.
- [25] Gouriéroux, C., and A., Monfort (2007): "Econometric Specification of Stochastic Discount Factor Models", *Journal of Econometrics*, 136, 509-530.
- [26] Hansen, B. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays", *Econometric Theory*, 12, 347-359.
- [27] Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.
- [28] Hansen, L., and R., Jagannathan (1997): "Assessing Specification Errors in Stochastic Discount Factor Models", *Journal of Finance*, 52, 557-590.
- [29] 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.
- [30] Hansen, L., and K., Singleton (1982): "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286.
- [31] Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability*, 12, 141-153.
- [32] 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.
- [33] Hull, J. (2005): *Options, Futures and Other Derivatives*, 6.th Edition, Prentice Hall.

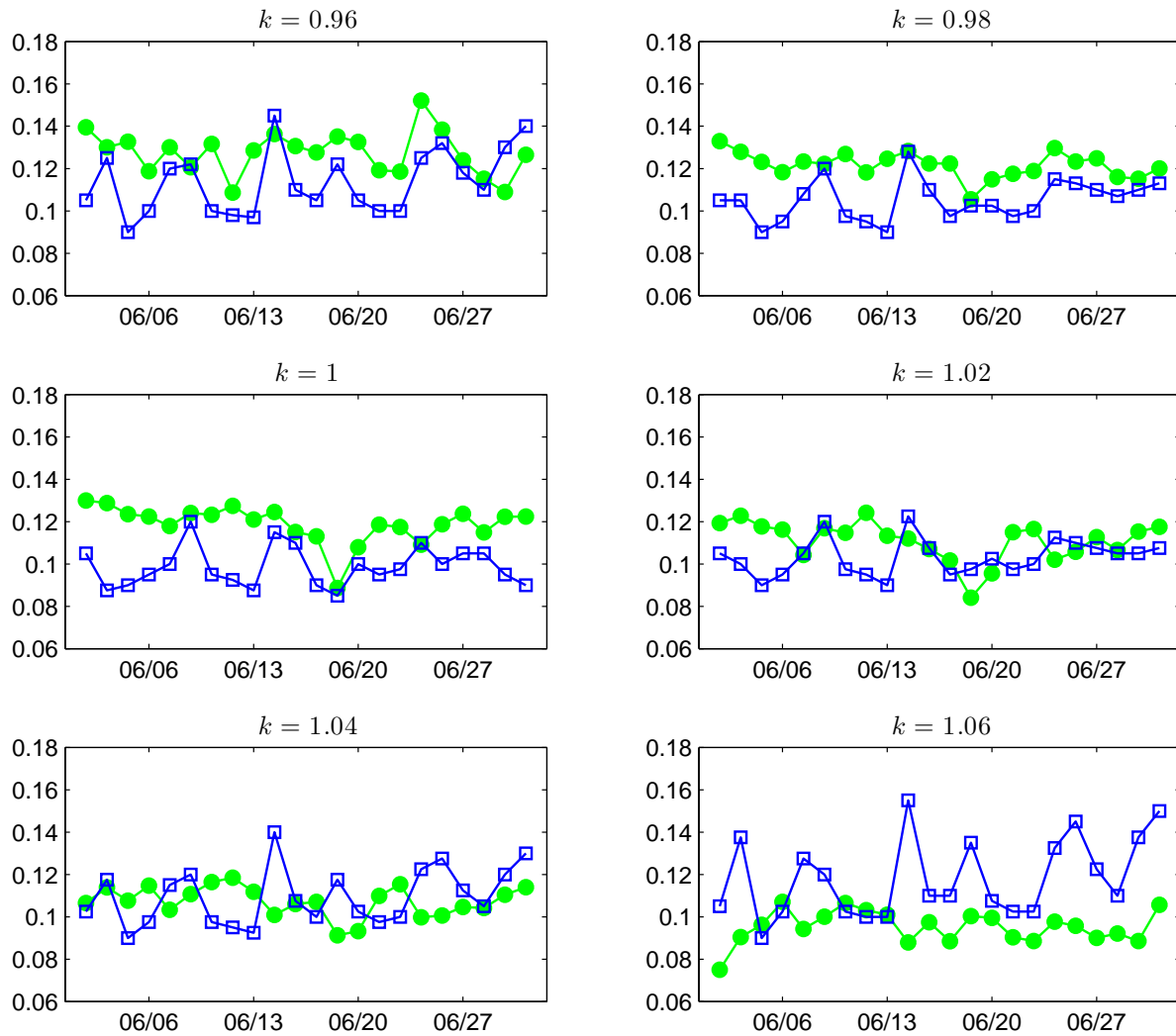
- [34] Hull, J., and A., White (1987): "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance*, 42, 281-300.
- [35] Imbens, G., Spady, R., and P., Johnson (1998): "Information Theoretic Approaches to Inference in Moment Conditions Models", *Econometrica*, 66, 333-357.
- [36] Jackwerth, J. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns", *Review of Financial Studies*, 13, 433-451.
- [37] Jackwerth, J., and M., Rubinstein (1996): "Recovering Probability Distributions from Option Prices", *Journal of Finance*, 51, 1611-1631.
- [38] Kitamura, Y. (2007): "Empirical Likelihood Methods in Econometrics: Theory and Practice" in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 3, eds. Blundell, R., Newey, W., and T., Persson, Econometric Society Monograph Series, Cambridge University Press, 174-237.
- [39] Kitamura, Y., and M., Stutzer (1997): "An Information Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.
- [40] Kitamura, Y., Tripathi, G., and H., Ahn (2004): "Empirical Likelihood Based Inference in Conditional Moment Restriction Models", *Econometrica*, 72, 1667-1714.
- [41] Pan, J. (2002): "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study", *Journal of Financial Economics*, 63-50.
- [42] Pollard, D. (1990): *Empirical Processes: Theory and Applications*. CBMS Conference Series in Probability and Statistics, Vol. 2, Hayward, CA.
- [43] Potscher, B., and I., Prucha (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes", *Econometrica*, 57, 675-683.
- [44] Schwartz, E. (1987): "The Stochastic Behaviour of Commodity Prices: Implications for Valuation and Hedging", *Journal of Finance*, 52, 923-973.
- [45] Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- [46] Stock, J., and J., Wright (2000): "GMM with Weak Identification", *Econometrica*, 68, 1055-1096.
- [47] Stutzer, M. (1996): "A Simple Nonparametric Approach to Derivative Security Valuation", *Journal of Finance*, 51, 1633-1652.
- [48] Tenreiro, C. (1995): "Estimation fonctionnelle: Applications aux tests d'adéquation et de paramètre constant", Ph.D. Thesis, University of Paris VI.

Figure 1: Estimated call and put prices for S&P 500 options at June, 1, and June, 2, 2005.



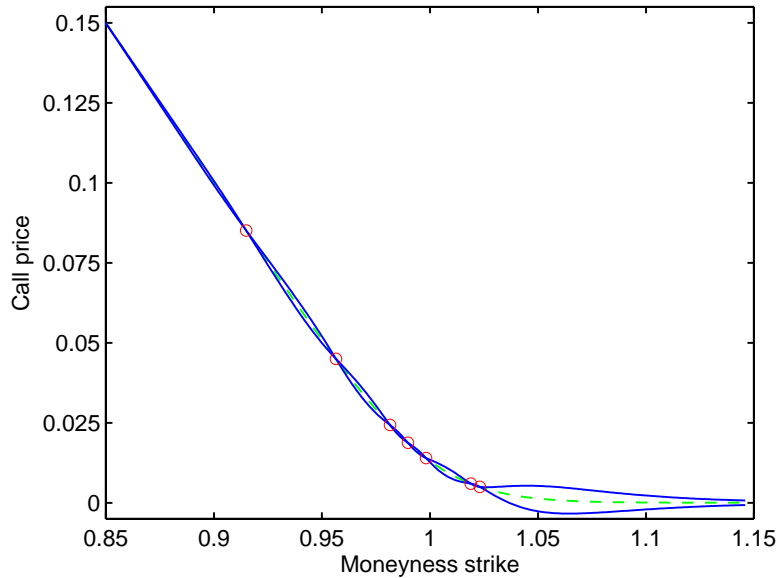
In the upper right Panel, the solid lines correspond to estimated relative option prices as a function of discounted moneyness strike $B(t, t+h)k$ for the highly traded times-to-maturity $h = 12, 57, 77, 209$ at June, 1, 2005, obtained by XMM. The dashed line corresponds to XMM estimated prices for the 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 (2.14)-(2.15) with the calibrated parameters in Table 1 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, 2, 2005 with highly traded times-to-maturity $h = 11, 31, 208$, and non-traded time-to-maturity $h = 119$.

Figure 2: Time series of implied volatilities of S&P 500 options in June, 2005.



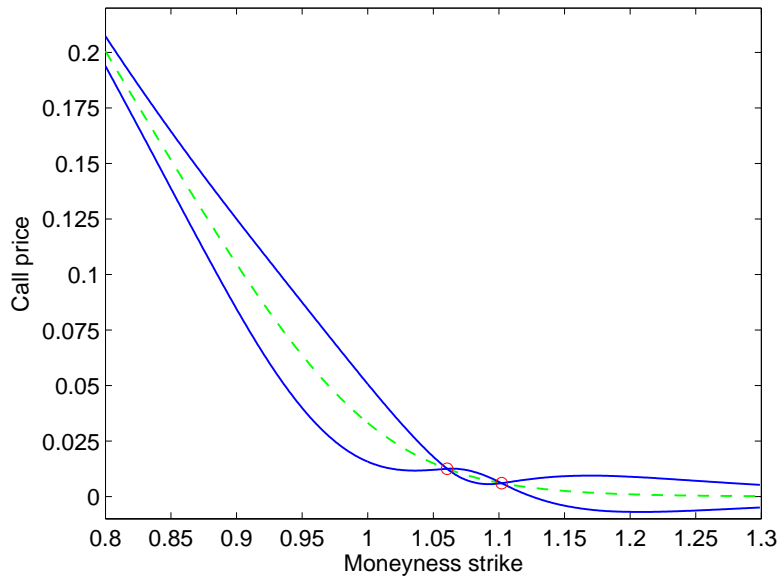
In Panels 1-6, annualized implied volatilities at time-to-maturity $h = 20$ and moneyness strike $k = .96$, $k = .98$, $k = 1$, $k = 1.02$, $k = 1.04$ and $k = 1.06$, respectively, are displayed for each trading day in June 2005. Circles are implied volatilities computed from option prices estimated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

Figure 3: Bias-free kernel nonparametric efficiency bound at June, 1, 2005, time-to-maturity 12.



The dashed line corresponds to the relative call price $E(a(k)|x_{t_0})$ at time-to-maturity $h = 12$, the solid lines to pointwise 95% symmetric confidence intervals $E(a(k)|x_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(x_{t_0}, k)^{1/2}$. The value of $\sqrt{w^2/Th_T^2}$ is set 10 times larger than in the empirical application.

Figure 4: Bias-free kernel nonparametric efficiency bound at June, 1, 2005, time-to-maturity 77.



The dashed line corresponds to the relative call price $E(a(k)|x_{t_0})$ at time-to-maturity $h = 77$, the solid lines to pointwise 95% symmetric confidence intervals $E(a(k)|x_{t_0}) \pm 1.96 \frac{w}{\sqrt{Th_T^2}} \mathcal{B}(x_{t_0}, k)^{1/2}$. The value of $\sqrt{w^2/Th_T^2}$ is set 10 times larger than in the empirical application.

Table 1: Calibrated parameters (cross-sectional approach) for the 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

Calibrated parameter $\hat{\theta}_{t_0}$, volatility $\hat{\sigma}_{t_0}$ and goodness of fit measure $RMSE_{t_0}$ for the first ten trading days t_0 of June 2005. The 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 2: Estimated sdf parameters and option prices (XMM approach) for the S&P 500 options in June, 2005.

Day t_0	Sdf parameters				Option prices ($\times 10^{-2}$)				
	$\hat{\theta}_1 (\times 10^{-6})$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$k = .96$	$k = .98$	$k = 1$	$k = 1.02$	$k = 1.04$
1.6.05	.910	-.059	.121	.860	.269	.631	1.356	.651	.162
2.6.05	.897	-.057	.118	.860	.215	.584	1.342	.686	.204
3.6.05	.900	-.054	.115	.860	.229	.539	1.282	.637	.169
6.6.05	.892	-.056	.118	.860	.155	.496	1.270	.622	.210
7.6.05	.888	-.056	.117	.860	.218	.550	1.227	.514	.147
8.6.05	.891	-.063	.124	.860	.164	.531	1.288	.628	.186
9.6.05	.890	-.063	.124	.860	.223	.574	1.280	.607	.220
10.6.05	.878	-.063	.125	.860	.110	.495	1.326	.701	.233
13.6.05	.864	-.063	.124	.860	.206	.553	1.253	.593	.193
14.6.05	.851	-.063	.125	.860	.250	.588	1.295	.580	.133

Estimated sdf parameter $\hat{\theta}$ and relative option prices $\hat{c}_{t_0}(h, k)$ at time-to-maturity $h = 20$ for the first ten trading days t_0 of June 2005. The option prices correspond to puts for $k < 1$, and to calls for $k \geq 1$. The estimation is performed using 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 kernel moment estimators

The asymptotic properties of the estimators are derived from an appropriate empirical process associated with the empirical moments:

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

The empirical process is:

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

where:

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

Due to the linearity of \widehat{g}_T and m_T w.r.t. β , the empirical process Ψ_T is function of parameter θ , but not of parameter β . Moreover, the empirical process Ψ_T depends on instrument Z by means of function g_1 . Note also that triangular array $g_{t,T}(\theta)$ is not zero-mean, because of the bias term in the nonparametric component.

We use the following notation. Symbol \implies denotes weak convergence in the space of bounded real functions on Θ , equipped with the uniform metric [see e.g. Andrews (1994)]. The Frobenius norm of matrix A is $\|A\| = \left[\text{Tr}(AA') \right]^{1/2}$. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and vector $x \in \mathbb{R}^d$, we set $|\alpha| := \sum_{i=1}^d \alpha_i$, $x^\alpha := x_1^{\alpha_1} \cdot \dots \cdot x_d^{\alpha_d}$, and $\partial^{|\alpha|} f / \partial x^\alpha := \partial^{|\alpha|} f / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$. Symbol $\|f\|_\infty$ denotes the sup-norm $\|f\|_\infty = \sup_{x \in \mathcal{X}} \|f(x)\|$ of a continuous function f defined on set \mathcal{X} . We denote by $C^m(\mathcal{X})$ the space of functions f on \mathcal{X} that are continuously differentiable up to order $m \in \mathbb{N}$, $\|D^m f\|_\infty := \sum_{|\alpha|=m} \|\partial^{|\alpha|} f / \partial x^\alpha\|_\infty$, and $\Delta^m f := \sum_{i=1}^d \partial^m f / \partial x_i^m$. Furthermore, $L^2(F_Y)$ 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. We denote by g_2^* the function defined by $g_2^*(y; \theta) = \left(g_2(y; \theta)', a(y; \theta)' \right)'$. Finally, all functions of θ are evaluated at θ_0 , when the argument is not explicit, and the expectation $E[\cdot]$ is w.r.t. the DGP P_0 .

A.1.1 Regularity assumptions

Let us introduce the following set of regularity conditions:

Assumption A.1: *The instrument Z is given by $Z = H(X)$, where H is a matrix function defined on \mathcal{X} and is continuous at $x = x_0$.*

Assumption A.2: *The true value of the parameter $\theta_0 \in \mathbb{R}^p$ is globally identified with instrument Z , that is,*

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

Assumption A.3: The true value of the parameter θ_0 is locally identified with instrument Z , that is, the matrix

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

has full column-rank.

Assumption A.4: 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 $\beta_0 = E[a(Y; \theta_0) \mid X = x_0]$.

Assumption A.5: The process $\left\{ (X_t', Y_t')' : t \in \mathbb{N} \right\}$ on $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^{\bar{d}}$ is strictly stationary and geometrically strongly mixing.

Assumption A.6: 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:

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

Assumption A.7: The matrices:

$$S_0 = V[g_1(X_t, Y_t; \theta_0)], \quad \Sigma_0 = V[g_2^*(Y_t; \theta_0) \mid X_t = x_0],$$

exist and are positive definite.

Assumption A.8: The stationary density f_X of X_t is in class $C^m(\mathcal{X})$ for some $m \in \mathbb{N}$, $m \geq 2$, and is such that $\|f_X\|_{\infty} < \infty$ and $\|D^m f_X\|_{\infty} < \infty$.

Assumption A.9: 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.10: For any $\theta \in \Theta$, the function $x \mapsto \varphi(x; \theta) = E[g_2^*(Y_t; \theta) \mid X_t = x] f_X(x)$ 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.11: For any $\theta, \tau \in \Theta$, the functions:

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

are continuous at $x = x_0$.

Assumption A.12: The mapping $x \mapsto E \left[\sup_{\theta \in \Theta} \|g_2^*(Y_t; \theta)\|^2 | X_t = x \right] f_X(x)$ 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 $\theta \mapsto \left(E [g_1(X_t, Y_t; \theta)]', E [g_2^*(Y_t; \theta) | X_t = x_0] \right)'$ is continuous on Θ .

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

Assumption A.16: The product kernel $K(u) = \prod_{i=1}^d \kappa(u_i)$, $u \in \mathbb{R}^d$, is such that: 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) $\int_{\mathbb{R}} v^l \kappa(v) dv = 0$ for any $l \in \mathbb{N}$ such that $l < m$, and $\int_{\mathbb{R}} \kappa(v) |v|^m dv < \infty$.

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:

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

where r is defined in Assumption A.6.

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

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:

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

Assumption A.24: *The mapping:*

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

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:*

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

are continuous on Θ .

Assumption A.26: *The instrument $Z^* = E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) \mid X \right) W(X)$, where $W(X)$ is a positive definite matrix P_0 -a.s., satisfies Assumptions A.6, A.7, A.14, A.20, A.23 and A.25.*

Assumptions A.2 and A.3 imply global and local identification with instrument Z . Assumption A.5 is a mixing condition used to prove the asymptotic normality of the finite-dimensional distributions of process Ψ_T (see Lemma A.1). Assumption A.6 is needed to prove the stochastic equicontinuity of process Ψ_T along the lines of Andrews (1991) [see the proof of Lemma A.1 in Appendix B]. Indeed, the 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 $g_{t,T}(\theta)$. Assumptions A.7-A.14 and A.20-A.25 are regularity conditions on the smoothness and the existence of moments concerning the pdf f_X and the functions g_1, g_2^* . These regularity conditions are used in the proof of Lemma A.1. Under Assumption A.16, the kernel K is a product kernel of order m . 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 the kernel regression estimator $\widehat{E}[g_2(Y; \theta) \mid x_0]$, 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 $\sqrt{Th_T^d}$, whereas other linear combinations are estimated at a parametric rate \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. 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 $\delta > 14$. Assumptions A.19 and A.20 are smoothness and moment conditions concerning the basis functions ψ_j introduced in Assumption A.6. Finally, Assumption A.26 is used to establish the asymptotic results for the kernel moment estimators with optimal instruments (see Sections A.1.6-A.1.7).

A.1.2 Asymptotic properties of the empirical process

The asymptotic distribution of the empirical process Ψ_T 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: $\Psi_T \Longrightarrow b + \Psi$, where $\Psi(\theta)$, $\theta \in \Theta$, denotes the zero-mean Gaussian stochastic process on Θ with covariance function given by:*

$$V_0(\theta, \tau) = E[\Psi(\theta)\Psi(\tau)'] = \begin{pmatrix} S_0(\theta, \tau) & 0 \\ 0 & w^2 \Sigma_0(\theta, \tau) / f_X(x_0) \end{pmatrix}, \quad \text{for } \theta, \tau \in \Theta,$$

with:

$$S_0(\theta, \tau) = \sum_{k=-\infty}^{\infty} \text{Cov}[g_1(X_t, Y_t; \theta), g_1(X_{t-k}, Y_{t-k}; \tau)], \quad \Sigma_0(\theta, \tau) = \text{Cov}[g_2^*(Y_t; \theta), g_2^*(Y_t; \tau) | X_t = x_0],$$

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

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

matrices S_0, Σ_0 are defined in Assumption A.7, and $\bar{c} := \lim Th_T^{d+2m}$.

The block diagonal elements of matrix V_0 are the standard asymptotic variance-covariance matrices of sample average and kernel regression estimators, respectively. The bias function $b(\theta)$ is zero for the unconditional moments, and is equal to the kernel regression bias for the conditional moments. Lemma A.1 implies that unconditional and conditional empirical moment restrictions are asymptotically independent, and that the convergence is uniform w.r.t. $\theta \in \Theta$.

A.1.3 Consistency of the kernel moment estimators

Let us write the criterion in Definition 4 as:

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

By using the weak convergence of process $\Psi_T(\theta)$ from Lemma A.1, and the inequality [see (B.28) in Appendix B]:

$$\inf_{\theta^* \in \Theta \times B: \|\theta^* - \theta_0^*\| \geq \varepsilon} m_T(\theta^*)' \Omega m_T(\theta^*) \geq CTh_T^d, \quad (\text{A.3})$$

for a constant $C > 0$ and any $\varepsilon > 0$ from identification Assumption A.2, the usual arguments to prove the consistency of an extremum estimator can be adapted to get:

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

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

The following Lemma A.3 is proved in Appendix B by following the approach in the proof of Lemma A1 in Stock, Wright (2000), and provides a lower bound on the rate of convergence.

Lemma A.3: Under Assumptions A.1-A.25: $\left\| \widehat{\theta}_T^* - \theta_0^* \right\| = O_p \left(1/\sqrt{Th_T^d} \right)$.

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

A.1.4 Asymptotic normality of kernel moment estimators

Let us consider the linear change of parameter from θ to $\eta = (\eta'_1, \eta'_2)'$ defined in (3.12), and let us introduce the matrix:

$$J_0 = \begin{pmatrix} E \left(\frac{\partial g_1}{\partial \theta'} \right) R_{1,Z} & 0 & 0 \\ 0 & E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_{2,Z} & 0 \\ 0 & E \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R_{2,Z} & -Id_L \end{pmatrix} = \begin{pmatrix} E \left(\frac{\partial g_1}{\partial \eta_1} \right) & 0 & 0 \\ 0 & E \left(\frac{\partial g_2}{\partial \eta_2} | x_0 \right) & 0 \\ 0 & E \left(\frac{\partial a}{\partial \eta_2} | x_0 \right) & -Id_L \end{pmatrix}. \quad (\text{A.4})$$

Under Assumption A.3, 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 invertible $(p+L, p+L)$ matrix:

$$R_T = \begin{pmatrix} T^{-1/2} R_{1,Z} & (Th_T^d)^{-1/2} R_{2,Z} & 0 \\ 0 & 0 & (Th_T^d)^{-1/2} Id_L \end{pmatrix}.$$

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 $\widehat{\theta}_T^*$ be such that $\left\| \widehat{\theta}_T^* - \theta_0^* \right\| = O_p \left(1/\sqrt{Th_T^d} \right)$. Then, under Assumptions A.1-A.25: $\text{plim} \frac{\partial \widehat{g}_T}{\partial \theta^{*'}} \left(\widehat{\theta}_T^* \right) R_T = J_0$.

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

$$\frac{\partial \widehat{g}_T'}{\partial \theta^*} \left(\widehat{\theta}_T^* \right) \Omega_{\widehat{g}_T} \left(\widehat{\theta}_T^* \right) = 0.$$

By a mean-value expansion we can write:

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

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

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

Let us define:

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

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

$$\text{plim } \hat{J}_T = \text{plim } \tilde{J}_T = J_0.$$

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

$$\left(\sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0})', \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0})', \sqrt{Th_T^d}(\hat{\beta}_T - \beta_0)' \right)' = - \left(\tilde{J}_T' \Omega \tilde{J}_T \right)^{-1} \tilde{J}_T' \Omega \hat{g}_T(\theta_0^*). \quad (\text{A.5})$$

Since $\hat{g}_T(\theta_0^*) = \Psi_T(\theta_0) \xrightarrow{d} N(\sqrt{c}b_0, V_0)$ from Lemma A.1, Proposition 1 follows.

A.1.5 Optimal weighting matrix for given instrument and bandwidth

When the bandwidth is such that $h_T = cT^{-1/(2m+d)}$, for some constant $c > 0$, from Proposition 1 the asymptotic MSE 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 $\left(J_0' \Omega J_0 \right)^{-1} J_0' \Omega M_0 \Omega J_0 \left(J_0' \Omega J_0 \right)^{-1}$, where $M_0 := V_0 + c^{2m+d} b_0 b_0'$. The optimal weighting matrix for given instrument Z and bandwidth constant c is:

$$\Omega = M_0^{-1} = \left(V_0 + c^{2m+d} b_0 b_0' \right)^{-1}. \quad (\text{A.6})$$

The corresponding minimal 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 $(\eta_2', \beta)'$, the associated asymptotic MSE of the estimator of β is:

$$M(Z, c, a) = e' \left(J_{0,Z}' \left(\frac{w^2}{c^d f_X(x_0)} \Sigma_0 + c^{2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_{0,Z} \right)^{-1} e, \quad (\text{A.7})$$

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

$$J_{0,Z} =: \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_{2,Z} & 0 \\ E \left(\frac{\partial a}{\partial \theta'} | x_0 \right) R_{2,Z} & -Id_L \end{pmatrix} = \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \eta_2} | x_0 \right) & 0 \\ E \left(\frac{\partial a}{\partial \eta_2} | x_0 \right) & -Id_L \end{pmatrix}.$$

A.1.6 Proof of Propositions 2 and 3

i) Optimal instruments

Let us first prove that instrument Z^* is admissible, that is, satisfies Assumptions A.2 and A.3. We have, for any vector $\alpha \in \mathbb{R}^p$:

$$\begin{aligned} E \left[\frac{\partial g_1}{\partial \theta'} (X, Y; \theta_0) \right] \alpha &= 0 \iff E \left[E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X) E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) \alpha \right] = 0 \\ &\iff \alpha' E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X) E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) \alpha = 0, \quad P_0\text{-a.s.}, \\ &\iff E \left(\frac{\partial g}{\partial \theta'} (Y; \theta_0) | X \right) \alpha = 0, \quad P_0\text{-a.s.} \end{aligned} \quad (\text{A.8})$$

Thus, Assumption A.3 follows from Assumption 2 in the text. Then, Assumption A.2 is also satisfied if Θ is taken small enough, which is sufficient for the validity of the asymptotic results. From Assumption A.26, the asymptotic properties in Proposition 1 applies for the kernel moment estimators with instrument Z^* .

Let us now prove that instrument Z^* is optimal. In the expression of $M(Z, c, a)$ in (A.7), instrument Z affects matrix $J_{0,Z}$ only, and the matrix $J_{0,Z}$ depends on Z through $\mathcal{J}_Z = \text{Ker } E \left[\partial g_1 (X, Y; \theta_0) / \partial \theta' \right]$, only. In particular, the larger is the null space \mathcal{J}_Z , the larger is the dimension of vector η_2 of structural parameters that are unidentifiable from the unconditional restrictions and must be estimated jointly with β . In other words, if Z and \tilde{Z} are two admissible instruments such that $\mathcal{J}_Z \subset \mathcal{J}_{\tilde{Z}}$, then $M(Z, c, a) \leq M(\tilde{Z}, c, a)$. Since $\mathcal{J}^* \subset \mathcal{J}_Z$ for any admissible instrument Z , Z^* is an optimal instrument if $\mathcal{J}^* = \mathcal{J}_{Z^*}$. The latter equality follows from (A.8).

Since $\mathcal{J}^* = \mathcal{J}_{Z^*}$, the ranges of matrix R_2 in (3.5) and matrix R_{2,Z^*} coincide. From (A.7), the asymptotic MSE for (any) optimal instrument Z^* becomes

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

where:

$$J_0^* = \begin{pmatrix} E \left(\frac{\partial g_2}{\partial \eta_2'} | x_0 \right) & 0 \\ E \left(\frac{\partial a}{\partial \eta_2'} | x_0 \right) & -Id_L \end{pmatrix}.$$

ii) Optimal bandwidth

In the rest of this proof we assume $L = \dim(a) = 1$. Then, function $M(c, a)$ is scalar, and the first-order condition for minimizing $M(c, a)$ w.r.t. c is given by:

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

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 $e = (0', 1)'$ $\in \mathbb{R}^{s^*} \times \mathbb{R}$. The solution $c^*(a)$ is $c^*(a) = \xi^{1/(2m+d)}$ where $\xi = \xi(a)$ satisfies the equation:

$$\xi = \frac{w^2 d}{2mw_m^2 f_X(x_0)} \frac{e' \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} \Sigma_0 A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e}{e' \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} b(x_0) b(x_0)' A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e}, \quad (\text{A.9})$$

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

iii) Optimal weighting matrix

From (A.6) the optimal weighting matrix is:

$$\Omega^*(a) = \left(V_0 + c^{*2m+d} b_0 b_0' \right)^{-1}, \quad (\text{A.10})$$

where V_0 is defined in (A.2) with optimal instrument Z^* . This concludes the proof of Proposition 2.

iv) Kernel nonparametric efficiency bound

The kernel nonparametric efficiency bound is $M(c^*(a), a)$ and is equal to $\mathcal{M}(x_0, a)$ in Proposition 3.

A.1.7 Proof of Corollary 4

If the bandwidth is such that $\bar{c} = \lim Th_T^{2m+d} = 0$, the optimal weighting matrix for given instrument is $\Omega = V_0^{-1}$. The proof that $Z^* = E \left(\frac{\partial g'}{\partial \theta} (Y; \theta_0) | X \right) W(X)$ is still an optimal instrument is similar to the proof of Proposition 2, replacing $M(Z, c, a)$ with $V(Z, a) = \frac{w^2}{f_X(x_0)} e' (J_{0,Z}' \Sigma_0^{-1} J_{0,Z})^{-1} e$, 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 4 follows from the block inversion formula.

APPENDIX 2

Asymptotic properties of the XMM estimator

A.2.1 Concentration with respect to the functional parameters

We first concentrate the estimation criterion in Definition 1 w.r.t. the functional parameters. Let us introduce the Lagrange multipliers $\lambda, \mu, \lambda_t, \mu_t, t = 1, \dots, T$. The Lagrangian function is given by:

$$\begin{aligned} \mathcal{L}_T &= \frac{1}{T} \sum_{t=1}^T \int \frac{[\hat{f}(y|x_t) - f_t(y)]^2}{\hat{f}(y|x_t)} dy + h_T^d \int \log [f_0(y)/\hat{f}(y|x_0)] f_0(y) dy \\ &\quad - 2 \frac{1}{T} \sum_{t=1}^T \mu_t \left(\int f_t(y) dy - 1 \right) - h_T^d \mu \left(\int f_0(y) dy - 1 \right) \\ &\quad - 2 \frac{1}{T} \sum_{t=1}^T \lambda_t' \int g(y; \theta) f_t(y) dy - h_T^d \lambda' \int g_2(y; \theta) f_0(y) dy. \end{aligned}$$

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

$$\left[f_t(y) - \hat{f}(y|x_t) \right] \frac{1}{\hat{f}(y|x_t)} - \mu_t - \lambda_t' g(y; \theta) = 0, \quad t = 1, \dots, T, \quad (\text{A.11})$$

$$1 + \log \left(f_0(y)/\hat{f}(y|x_0) \right) - \mu - \lambda' g_2(y; \theta) = 0, \quad (\text{A.12})$$

which yields:

$$f_t(y) = \hat{f}(y|x_t) + \mu_t \hat{f}(y|x_t) + \lambda_t' g(y; \theta) \hat{f}(y|x_t), \quad t = 1, \dots, T, \quad (\text{A.13})$$

$$f_0(y) = \hat{f}(y|x_0) \exp \left(\lambda' g_2(y; \theta) + \mu - 1 \right). \quad (\text{A.14})$$

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 $\int g(y; \theta) f_t(y) dy = 0$

$$\iff \lambda_t = - \left[\int g(y; \theta) g(y; \theta)' \widehat{f}(y|x_t) dy - \int g(y; \theta) \widehat{f}(y|x_t) dy \int g(y; \theta)' \widehat{f}(y|x_t) dy \right]^{-1} \int g(y; \theta) \widehat{f}(y|x_t) dy,$$

$t = 1, \dots, T$. Similarly, from (A.14) we deduce the value of the Lagrange multiplier μ :

$$\int f_0(y) dy = 1 \iff \exp(1 - \mu) = \int e^{\lambda' g_2(y; \theta)} \widehat{f}(y|x_0) dy.$$

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

$$\begin{aligned} f_t(y; \theta) &= \widehat{f}(y|x_t) - \widehat{E}(g(\theta)|x_t)' \widehat{V}(g(\theta)|x_t)^{-1} \left[g(y; \theta) - \widehat{E}(g(\theta)|x_t) \right] \widehat{f}(y|x_t), \quad t = 1, \dots, T, \\ f_0(y; \theta, \lambda) &= \frac{\exp(\lambda' g_2(y; \theta))}{\widehat{E}[\exp(\lambda' g_2(\theta)) | x_0]} \widehat{f}(y|x_0), \end{aligned} \quad (\text{A.15})$$

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

$$\mathcal{L}_T^c(\theta, \lambda) = \frac{1}{T} \sum_{t=1}^T \widehat{E}(g(\theta)|x_t)' \widehat{V}(g(\theta)|x_t)^{-1} \widehat{E}(g(\theta)|x_t) - h_T^d \log \widehat{E}(\exp(\lambda' g_2(\theta)) | x_0). \quad (\text{A.16})$$

Then, the XMM estimator is such that $\widehat{\theta}_T$ is solution of the saddle point problem [see Kitamura, Stutzer (1997) in an unconditional framework]:

$$\widehat{\theta}_T = \arg \min_{\theta} \mathcal{L}_T^c(\theta, \lambda(\theta)),$$

where

$$\lambda(\theta) = \arg \max_{\lambda} \mathcal{L}_T^c(\theta, \lambda) \iff \widehat{E} \left[g_2(\theta) \exp(\lambda(\theta)' g_2(\theta)) | x_0 \right] = 0.$$

Thus, $\widehat{\theta}_T$ is defined by the optimization problem (2.10)-(2.11). The conditional density estimator $\widehat{f}^*(\cdot|x_0)$ of $f(\cdot|x_0)$ is obtained from (A.15) by replacing θ by $\widehat{\theta}_T$, and λ by $\widehat{\lambda}_T = \lambda(\widehat{\theta}_T)$ [see (2.12)].

A.2.2 Asymptotic expansions

i) Asymptotic expansion of $\widehat{\theta}_T$ and Lagrange multiplier $\widehat{\lambda}_T$

In order to derive the asymptotic expansion of the estimator $\widehat{\theta}_T$, we have to distinguish between the linear combinations converging at a parametric rate and those converging at a nonparametric rate. For this purpose, we consider the new parameterization in (3.5). The asymptotic expansion for the estimators of the transformed parameters $\widehat{\eta}_{1,T}^*$, $\widehat{\eta}_{2,T}^*$ and the Lagrange multiplier $\widehat{\lambda}_T$ are given in Lemma A.5 below.

Lemma A.5: (i) The asymptotic expansions of $\widehat{\eta}_{1,T}^*$ and $\widehat{\eta}_{2,T}^*$ are given by:

$$\begin{aligned} \sqrt{T}(\widehat{\eta}_{1,T}^* - \eta_{1,0}^*) &\simeq - \left(R_1' E \left[E \left(\frac{\partial g'}{\partial \theta} | X \right) V(g|X)^{-1} E \left(\frac{\partial g}{\partial \theta'} | X \right) \right] R_1 \right)^{-1} \\ &\quad \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^T R_1' E \left(\frac{\partial g'}{\partial \theta} | x_t \right) V(g|x_t)^{-1} g(y_t; \theta_0), \end{aligned}$$

and:

$$\begin{aligned} \sqrt{Th_T^d}(\widehat{\eta}_{2,T}^* - \eta_{2,0}^*) &\simeq - \left[R_2' E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2|x_0)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_2 \right]^{-1} \\ &\quad \cdot R_2' E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2|x_0)^{-1} \sqrt{Th_T^d} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy, \quad (\text{A.17}) \end{aligned}$$

respectively. (ii) The asymptotic expansion of $\widehat{\lambda}_T$ is:

$$\widehat{\lambda}_T \simeq -V(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy, \quad (\text{A.18})$$

where M is the orthogonal projection matrix on the column space of $E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_2$ for the inner product defined by $V(g_2|x_0)^{-1}$:

$$M = E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_2 \left[R_2' E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2|x_0)^{-1} E \left(\frac{\partial g_2}{\partial \theta'} | x_0 \right) R_2 \right]^{-1} R_2' E \left(\frac{\partial g_2'}{\partial \theta} | x_0 \right) V(g_2|x_0)^{-1}. \quad (\text{A.19})$$

The proof of Lemma A.5 in Appendix B is based on the asymptotic expansion of the criterion (A.16). The term $\log \widehat{E} \left(\exp \left(\lambda' g_2(\theta) \right) | x_0 \right)$ in (A.16), which is induced by the KLIC component, has a weighting factor h_T^d (see Definition 1). This weighting factor ensures that the contribution of the discrepancy measure associated with the local restrictions at x_0 is asymptotically the same as for a kernel moment estimator with optimal weighting matrix and instruments.

ii) Asymptotic expansion of $\widehat{f}^*(\cdot|x_0)$

Using $\widehat{f}^*(y|x_0) = f_0(y; \widehat{\theta}_T, \widehat{\lambda}_T)$, from (A.15) we have:

$$\begin{aligned} \widehat{f}^*(y|x_0) &\simeq \frac{1 + \widehat{\lambda}_T' g_2(y; \widehat{\theta}_T)}{1 + \widehat{\lambda}_T' \widehat{E}(g_2(\widehat{\theta}_T)|x_0)} \widehat{f}(y|x_0) \simeq \left[1 + \widehat{\lambda}_T' \left(g_2(y; \widehat{\theta}_T) - \widehat{E}(g_2(\widehat{\theta}_T)|x_0) \right) \right] \widehat{f}(y|x_0) \\ &\simeq \widehat{f}(y|x_0) + \widehat{\lambda}_T' g_2(y; \theta_0) f(y|x_0). \end{aligned}$$

Then, from (A.18) we get:

$$\widehat{f}^*(y|x_0) \simeq \widehat{f}(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy. \quad (\text{A.20})$$

iii) Asymptotic expansion of $\widehat{E}^*(a|x_0)$

We have:

$$\begin{aligned}
\widehat{E}^*(a|x_0) &= \int a(y; \widehat{\theta}_T) \widehat{f}^*(y|x_0) dy \\
&\simeq \int a(y; \theta_0) f(y|x_0) dy + \int \frac{\partial a}{\partial \theta'}(y; \theta_0) f(y|x_0) dy (\widehat{\theta}_T - \theta_0) + \int a(y; \theta_0) [\widehat{f}^*(y|x_0) - f(y|x_0)] dy \\
&\simeq E(a|x_0) + E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 (\widehat{\eta}_{2,T}^* - \eta_{2,0}^*) \\
&\quad + \int a(y; \theta_0) \left\{ \widehat{f}(y|x_0) - f(y|x_0) - f(y|x_0) g_2(y; \theta_0)' V(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \right\} dy,
\end{aligned}$$

where the last asymptotic equivalence comes from (A.20) and the fact that the contribution of $\widehat{\eta}_{1,T}^* - \eta_{1,0}^*$ is asymptotically negligible. Then, from (A.17) we get:

$$\begin{aligned}
\widehat{E}^*(a|x_0) &\simeq E(a|x_0) - E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 \left[R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2 \right]^{-1} \\
&\quad \cdot R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy \\
&\quad + \int a(y; \theta_0) [\widehat{f}(y|x_0) - f(y|x_0)] dy - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} (Id - M) \int g_2(y; \theta_0) \widehat{f}(y|x_0) dy.
\end{aligned}$$

We deduce the asymptotic expansion:

$$\begin{aligned}
\widehat{E}^*(a|x_0) - E(a|x_0) &\simeq \int a(y; \theta_0) \delta \widehat{f}(y|x_0) dy - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy \\
&\quad - \left[E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 - Cov(a, g_2|x_0) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2 \right] \\
&\quad \cdot \left[R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2 \right]^{-1} R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V(g_2|x_0)^{-1} \int g_2(y; \theta_0) \delta \widehat{f}(y|x_0) dy,
\end{aligned} \tag{A.21}$$

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

A.2.3 Asymptotic distribution of the XMM estimator (Proofs of Proposition 5 and Corollary 6)

Let us derive the asymptotic distribution of the estimator $\widehat{E}^*(a|x_0)$. In the asymptotic expansion (A.21), the first two terms in the RHS correspond to 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 in the RHS. Thus, from the asymptotic normality of integrals of kernel estimators, we get:

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

where $\mathcal{B}(x_0, a)$ is given in Corollary 4. This proves Proposition 5. Corollary 6 is proved by checking that the asymptotic expansion for the XMM estimator $\widehat{\theta}_T$ in Lemma A.5 (i) corresponds to the asymptotic expansion for the kernel moment estimator in Appendix A.1.4 with $\bar{c} = \lim Th_T^{d+2m} = 0$, $\Omega = V_0^{-1}$ and $Z = E\left[\frac{\partial g'}{\partial \theta}(Y; \theta_0)|X\right] V[g(Y; \theta_0)|X]^{-1}$. The details of the derivation are given in Appendix B.

APPENDIX 3

A parametric stochastic volatility model

In this Appendix we describe and analyze a coherent parametric stochastic volatility model for both the historical and risk-neutral distributions of the state variables. This parametric stochastic volatility model is compatible with both the risk-neutral parametric specification considered in the cross-sectional calibration approach [Section 2.6 i)], and the semi-parametric model used in the XMM approach [Section 2.6 ii)]. The riskfree rate is set equal to zero for expository purpose, but the results extend to the model with deterministic riskfree rate.

A.3.1 The model

The return of the underlying asset is:

$$r_t = \gamma \sigma_t^2 + \sigma_t \varepsilon_t, \quad (\text{A.22})$$

where (ε_t) is a standard Gaussian white noise, σ_t^2 denotes the stochastic volatility, and γ measures the magnitude of the risk premium in the expected return. The stochastic volatility (σ_t^2) follows an Autoregressive Gamma (ARG) process [see Gouriéroux and Jasiak (2006), Darolles, Gouriéroux and Jasiak (2006)], and is independent of the shocks (ε_t) on returns. The transition distribution of the stochastic volatility is defined through its conditional Laplace transform:

$$E \left[\exp(-u \sigma_{t+1}^2) \mid \underline{\sigma}_t^2 \right] = \exp[-a(u) \sigma_t^2 - b(u)], \quad (\text{A.23})$$

where $a(u) = \rho \frac{u}{1+cu}$, and $b(u) = \delta \log(1+cu)$. The positive parameter ρ is the first-order autocorrelation of the volatility process (σ_t^2) , parameter δ , $\delta \geq 0$, describes its (conditional) over-/under-dispersion, and c , $c > 0$, is a scale parameter. The sdf is given by:

$$M_{t,t+1} = \exp(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 r_{t+1}),$$

where $\theta_1, \theta_2, \theta_3, \theta_4$ are parameters. The restrictions implied by the no-arbitrage assumption for the riskfree asset and the underlying asset are given by:

$$E(M_{t,t+1} | x_t) = 1, \quad E(M_{t,t+1} \exp r_{t+1} | x_t) = 1, \quad \forall x_t = (r_t, \sigma_t^2)'$$

We have the following Lemma A.6.

Lemma A.6: *The sdf is compatible with the no-arbitrage restrictions if, and only if:*

$$\theta_1 = -\delta \log[1 + c(\theta_2 + \gamma^2/2 - 1/8)], \quad \theta_3 = -\rho \frac{\theta_2 + \gamma^2/2 - 1/8}{1 + c(\theta_2 + \gamma^2/2 - 1/8)}, \quad \theta_4 = \gamma + 1/2.$$

Proof: The no-arbitrage restrictions are:

$$\begin{aligned} & \begin{cases} E_t(M_{t,t+1}) = 1, \\ E_t(M_{t,t+1} \exp r_{t+1}) = 1, \end{cases} \iff \begin{cases} E_t \exp[-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 r_{t+1}] = 1, \\ E_t \exp[-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - (\theta_4 - 1) r_{t+1}] = 1, \end{cases} \\ \iff & \begin{cases} E_t \exp\left[-\theta_1 - \left(\theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2}\right) \sigma_{t+1}^2 - \theta_3 \sigma_t^2\right] = 1, \\ E_t \exp\left[-\theta_1 - \left(\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2}\right) \sigma_{t+1}^2 - \theta_3 \sigma_t^2\right] = 1, \end{cases} \\ & \text{(by integrating } r_{t+1} \text{ conditional on } \sigma_{t+1}^2) \\ \iff & \begin{cases} \theta_1 + a\left(\theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2}\right) \sigma_t^2 + \theta_3 \sigma_t^2 + b\left(\theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2}\right) = 0, \\ \theta_1 + a\left[\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2}\right] \sigma_t^2 + \theta_3 \sigma_t^2 + b\left[\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2}\right] = 0, \end{cases} \end{aligned}$$

where E_t denotes expectation conditional on x_t . Since the above conditions have to be satisfied for any admissible value of σ_t^2 , we get:

$$\begin{cases} \theta_1 + b \left(\theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2} \right) = 0, & \theta_1 + b \left[\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2} \right] = 0, \\ \theta_3 + a \left(\theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2} \right) = 0, & \theta_3 + a \left[\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2} \right] = 0. \end{cases}$$

Since functions a and b are one-to-one, the difference between the two equations in the first line (resp. the two equations in the second line) implies $\theta_2 + (\theta_4 - 1) \gamma - \frac{(\theta_4 - 1)^2}{2} = \theta_2 + \theta_4 \gamma - \frac{\theta_4^2}{2}$, that is, $\theta_4 = \gamma + \frac{1}{2}$. The conclusion follows. ■

In particular, parameter θ_2 is unrestricted. In this incomplete market framework with a riskfree asset and a risky asset, the risk premium for the current stochastic volatility can be fixed arbitrarily, that is, the dimension of the residual market incompleteness is equal to 1. This residual incompleteness is not a consequence of the specific ARG dynamic assumed for the stochastic volatility, but is observed whenever the state variables follow an affine process. Indeed, in this setting, the specification of a parametric exponential affine sdf does not select a unique pricing kernel [Gouriéroux and Monfort (2007)].

Under the risk-neutral probability, the underlying asset return still follows an ARG stochastic volatility model with adjusted risk premium parameter $\gamma^* = -1/2$ and volatility parameters (see Appendix B on the website):

$$\rho^* = \frac{\rho}{[1 + c(\theta_2 + \gamma^2/2 - 1/8)]^2}, \quad \delta^* = \delta, \quad c^* = \frac{c}{1 + c(\theta_2 + \gamma^2/2 - 1/8)}. \quad (\text{A.24})$$

A.3.2 Identification

Let us now discuss parameter identification when the DGP P_0 is compatible with the parametric stochastic volatility model described in A.3.1. Let us consider XMM estimation with state variable $X_t = (r_t, \sigma_t^2)'$ and parametric sdf $M_{t,t+1}(\theta) = \exp(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 r_{t+1})$. This is a well-specified sdf, with true parameter value θ_0 satisfying the restrictions in Lemma A.6. We first derive the matrix R_2 characterizing full-information underidentification.

i) Computation of matrix R_2

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

$$E \left[\begin{pmatrix} 1 \\ \exp r_{t+1} \end{pmatrix} \frac{\partial M_{t,t+1}(\theta_0)}{\partial \theta'} \mid x_t \right] \alpha = 0, \quad \forall x_t. \quad (\text{A.25})$$

Since θ_0 satisfies the no-arbitrage restrictions:

$$E \left[M_{t,t+1}(\theta_0) \begin{pmatrix} 1 \\ \exp r_{t+1} \end{pmatrix} \mid x_t \right] = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \forall x_t,$$

we deduce that any $\theta = \theta_0 + \alpha \varepsilon$, where ε is small and α satisfies (A.25), is such that:

$$E \left[M_{t,t+1}(\theta) \begin{pmatrix} 1 \\ \exp r_{t+1} \end{pmatrix} \mid x_t \right] = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \forall x_t,$$

at first-order in ε . Therefore, the vectors in \mathcal{J}^* are the directions $d\theta = \theta - \theta_0$ of infinitesimal parameter changes that are compatible with no-arbitrage. From Lemma A.6, the parameters θ compatible with no-arbitrage are characterized by the nonlinear restrictions:

$$\theta_1 = -b(\theta_2 + \gamma^2/2 - 1/8), \quad \theta_3 = -a(\theta_2 + \gamma^2/2 - 1/8), \quad \theta_4 = \gamma + 1/2.$$

Thus, the tangent set at θ_0 is spanned by the vector:

$$\alpha = \begin{pmatrix} d\theta_1/d\theta_2 \\ d\theta_2/d\theta_2 \\ d\theta_3/d\theta_2 \\ d\theta_4/d\theta_2 \end{pmatrix} \Big|_{\theta=\theta_0} = \begin{pmatrix} -db(\theta_2 + \gamma^2/2 - 1/8)/d\theta_2 \\ 1 \\ -da(\theta_2 + \gamma^2/2 - 1/8)/d\theta_2 \\ 0 \end{pmatrix} \Big|_{\theta=\theta_0} = \begin{pmatrix} -\delta \frac{c}{1+c\lambda_2} \\ 1 \\ -\rho \frac{1}{(1+c\lambda_2)^2} \\ 0 \end{pmatrix},$$

where $\lambda_2 := \theta_{2,0} + \gamma^2/2 - 1/8$. We deduce that $\dim(\mathcal{J}^*) = 1$ and matrix R_2 is given by:

$$R_2 = \left(-\delta \frac{c}{1+c\lambda_2}, 1, -\rho \frac{1}{(1+c\lambda_2)^2}, 0 \right)'. \quad (\text{A.26})$$

ii) Local identification (Assumption 2)

Let us now verify that Assumption 2 is satisfied when the conditional restrictions include the observed price of a European call. To simplify, we consider a European call at time-to-maturity 1, and check that $E \left[\frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) (\exp r_{t+1} - k)^+ \mid x_t \right] R_2 \neq 0$, $\forall k > 0$, for any given x_t . We have from (A.26):

$$\begin{aligned} & E \left[\frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) (\exp r_{t+1} - k)^+ \mid x_t \right] R_2 \\ &= -E \left[M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ (1, \sigma_{t+1}^2, \sigma_t^2, r_{t+1}) R_2 \mid x_t \right] \\ &= \left(\frac{\delta c}{1+c\lambda_2} + \frac{\rho}{(1+c\lambda_2)^2} \sigma_t^2 \right) E \left[M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ \mid x_t \right] - E \left[M_{t,t+1}(\theta_0) (\exp r_{t+1} - k)^+ \sigma_{t+1}^2 \mid x_t \right]. \end{aligned}$$

From (A.24), we have:

$$\frac{\delta c}{1+c\lambda_2} + \frac{\rho}{(1+c\lambda_2)^2} \sigma_t^2 = \rho^* \sigma_t^2 + \delta^* c^* = E^Q [\sigma_{t+1}^2 \mid x_t],$$

where Q denotes the risk-neutral distribution, whereas from the Hull-White formula [Hull, White (1987)]:

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

where $BS(h, k, \sigma^2)$ is the Black-Scholes pricing formula. Thus, we get:

$$E \left[\frac{\partial M_{t,t+1}}{\partial \theta'}(\theta_0) (\exp r_{t+1} - k)^+ \mid x_t \right] R_2 = -Cov^Q [\sigma_{t+1}^2, BS(1, k, \sigma_{t+1}^2) \mid x_t],$$

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