UNIVERSITÀ DELLA SVIZZERA ITALIANA

FACULTY OF ECONOMICS - INSTITUTE OF FINANCE

# SEMI-PARAMETRIC ESTIMATION OF AMERICAN OPTION PRICES

DIEGO RONCHETTI

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN ECONOMICS
AND DEFENDED ON JUNE 24, 2011

THESIS COMMITTEE:
PROF. PATRICK GAGLIARDINI - THESIS DIRECTOR
PROF. GIOVANNI BARONE-ADESI - COMMITTEE CHAIR
PROF. JÉRÔME DETEMPLE - EXTERNAL MEMBER
PROF. CHRISTIAN GOURIÉROUX - EXTERNAL MEMBER

An American option provides the right to perform a specified financial transaction (sell, buy, exchange) on or before the contract maturity. Many different contracts traded on centralized and OTC markets are of this kind. In particular, a plain vanilla American option is a contract between two parties concerning the possibility of selling or buying a reference asset (underlying) at a specified price (strike price). Setting the contract price and choosing the best moment for its exercise are two of the most studied problems in finance during the last $40$ years. In financial markets, the behavior of the underlying is not predictable. Thus, a description of the probability law governing its stochastic evolution is necessary for the determination of the contract price and the optimal exercise decision.

The majority of the existing literature focuses on mathematical and numerical procedures for computing the option price and determining the optimal exercise policy for a given law of motion of the underlying. For these purposes, only a model for the dynamics of the underlying under the risk-neutral distribution is required. When this approach is put into practice, typically a parametric model for such distribution is adopted and the parameters are calibrated on a cross-section of available option prices. On the contrary, in this PhD thesis, that summarizes the research conducted to obtain the degree of Doctor of Philosophy in Economics at the University of Lugano, an econometric framework for the empirical pricing of American options is developed. In this framework, a statistical model for the dynamics of the underlying is specified by the researcher and estimated on available data. Data include both time series of relevant state variables and cross-sections of observed option prices. The estimated model is then used to estimate the price of contracts that are not currently actively traded on the market. The econometric approach proposed in this thesis features three major characteristics. First, it is based on a coherent specification of both historical and risk-neutral dynamics. Second, the statistical model for the dynamics of the underlying is more general than most of the models previously considered in the literature. Third, the model parameters can be consistently estimated even when the amount of option data is limited.

In the first three chapters of the thesis, the problem, the proposed solution and an empirical application of the novel method are presented. Chapter 1 introduces the price of an American option as the expected value of the contract at the most remunerative time for exercising it. Some different pricing techniques based on this representation and the way they are used to handle with real data are briefly reviewed. Chapter 2 presents the novel empirical methodology developed in the PhD research. Chapter 3 describes an application of this methodology for the analysis of IBM shares and plain vanilla American options written on them. In the last two chapters of the thesis, the regularity assumptions for the validity of the asymptotic properties of the proposed method and the proofs of propositions and technical lemmas are reported. In particular, Chapter 4 provides details on the content of Chapter 2 and Chapter 5 does it for the content of Chapter 3.

Chapter 1, titled *The Pricing of American Options: Optimal Stopping and Financial Theory*, is a brief survey on several different approaches to the pricing of American options that combine the theory of optimal stopping and asset pricing, under the hypotheses of rational agents and efficient markets. The formulations of the pricing of American options by the martingale approach and, in presence of some variables summarizing the state of economy, by the equivalent Markov approach are presented. Some general properties of put and call options, that are the most liquid American options, are considered. Some different approaches to find a solution of the optimal stopping problem are reviewed. Among these techniques, some rely on the transformation of the optimal stopping problem into a free-boundary problem, while others are directly inspired by the principle of optimality of dynamic programming. Finally the implementation of these pricing techniques to handle with real data is briefly discussed.

Chapter 2, titled *Semi-Parametric Estimation of American Option Prices*, introduces a novel semi-parametric estimation methodology for the pricing of American options, that is the result of a joint work with Prof. Patrick Gagliardini. The proposed methodology requires three data inputs: a time series of state variables for the underlying, a cross-section of prices of American options written on the underlying and a risk-free interest rate. The estimation is based on a parametric specification of the Stochastic Discount Factor (SDF) and is non-parametric w.r.t. the historical dynamics of the Markov state variables. The model parameters are the SDF parameter, that has finite dimension, and the historical transition density, that is a functional parameter. This semi-parametric setting is intermediate between fully parametric and fully non-parametric approaches. The advantage w.r.t. the former is the flexibility in modeling the historical transition density and the possibility to get a proper distribution theory for the estimators without introducing ad-hoc pricing errors. The advantage w.r.t. the latter is that the estimated pricing model is arbitrage-free by construction. In non-parametric approaches, ensuring the absence of arbitrage opportunities by imposing shape restrictions on the pricing function might be difficult, since such shape restrictions are not completely known in the general framework considered in this thesis. The proposed method exploits the no-arbitrage conditions for a short-term non-defaultable zero-coupon bond, the underlying and a cross-section of observed prices of American options. In particular, the method considers the uniform moment restrictions and some restrictions that involve nonlinear functionals of the transition density. These nonlinear functionals result from representing the American option price through a backward recursive application of a risk and time discounting operator on the option payoff (dynamic programming). The estimation in such a framework needs an extension of the Generalized Method of Moments. First, the SDF parameter is estimated by minimizing a quadratic criterion based on empirical restrictions. Then, the historical transition density of the state variables is estimated by minimizing a statistical measure based on the Kullback-Leibler divergence from a kernel-based transition density. The estimators of the model parameters can be used to estimate many functionals of SDF parameter and the transition density of the state variables, such as the prices of American options not traded in the market, historical and risk-neutral conditional Laplace transforms, skewness, kurtosis and cross-moments of the state variables. A Monte Carlo experiment shows how the proposed method outperforms a pricing methodology that exploits only the no-arbitrage conditions for the short-term risk-free bond and the underlying.

Chapter 3, titled *An Empirical Study of Stock and American Option Prices*, describes an empirical study of the information content of daily IBM share prices and American put and call option quotes about their generating process. The study focuses on daily IBM share closing prices at NYSE from January 2006 to August 2008 and closing quotes for IBM American call and put options selected among U.S. centralized markets in July and August 2008. Two results are empirically obtained considering stock return and its volatility as the risk factors and without parameterizing their historical joint dynamics. First, contemporaneous share prices and option quotes are both necessary to quantify the equity and variance premia. Second, an arbitrage-free pricing model is useful to get more precise estimates of the historical joint dynamic properties of the risk factors. As an illustration, time series of different estimates of historical conditional correlation of the risk factors, Sharpe ratio of an investment on the stock, return skewness and kurtosis are reported. The empirical results are obtained by confronting the results of different approaches to the estimation of the data generating process.

I owe my deepest gratitude to Prof. Patrick Gagliardini, who introduced me to econometrics, provided continuous encouragement for my PhD studies and supervised this thesis. The National Centre of Competence in Research FINRISK and the University of Lugano financially supported my participation to several international conferences to present the work described in this thesis.

# Contents

# 1 The Pricing of American Options: Optimal Stopping and Financial Theory

An American option is a financial contract that offers the right to perform a pre-specified financial transaction (sell, buy, exchange) on or before the contract expiration date (maturity). In the last 40 years many American-style financial products have been introduced in the financial markets.[1] Among them we find options on index, stock, interest rate, exchange rate and future contracts. Because of the large diffusion of these products, their pricing is one of the most considered problems in the history of the financial sector. Many analytical and numerical ways to determine the optimal time to exercise the contract have been proposed. In this survey I briefly review some related literature, with special attention to equity options. For further surveys on the literature of American option pricing see Broadie and Detemple [2004], Barone-Adesi [2005] and Detemple [2005].

In Section 1.1 I introduce the general formulation of the American options pricing as an optimal stopping problem and some equivalent representations of the American option price. In Section 1.2 I describe some properties of put and call options, that are the most liquid American options traded in centralized markets. In Section 1.3 I review some techniques that rely on a particular specification of the dynamics of the state variables. In Section 1.4 I describe some calibration and empirical methods to handle with real data.

## 1.1 Optimal stopping time formulation

The most general mathematical formulation of the pricing of American options combines the theory of optimal stopping and asset pricing. The theory of optimal stopping pertains the choice of the best time to take a given action. This decision is based on sequential realizations of a random process and is taken in order to maximize an expected payoff or to minimize an expected cost. The first systematic approach to the theory of optimal stopping is in the field of sequential analysis, in particular for the search of the best time to stop a sequential testing of two alternative hypotheses by not rejecting one of them at a certain confidence interval. This stopping time is such that the probability of occurrence of errors of first and second kind is lower than some given levels and its expected value is the smallest as possible, since the longer the sequential analysis is, the more it costs (see Wald and Wolfowitz [1948] and Arrow, Blackwell and Girshick [1949]). The first generalization for sequential problems without a statistical structure is in Snell [1952].

[1] The Chicago Board of Options Exchange, the largest U.S. options exchange, was opened in 1973.

Shiryaev [1978] and Peskir and Shiryaev [2006] provide extensive treatments of the theory of optimal stopping. The theory of asset pricing pertains the origin of financial asset prices. In particular the existence of a probability measure $\mathscr{Q}$, called risk-neutral probability, such that option and underlying asset price processes are local $\mathscr{Q}$-martingale, is assumed. The mathematical justification of the existence of this probability is in Cox and Ross [1976] and its complete explanation is in Harrison and Kreps [1979] and Harrison and Pliska [1981]. Bensoussan [1984], Karatzas [1988] and Karatzas [1989] highlight the connection between American option hedging and pricing and the mathematical theory of optimal stopping.

The exercise of an American option depends on the non-negative gain (or reward) at exercise. This gain depends on the state of the economy and on some option characteristics stated in the contract indenture. The American option pricing problem can be formulated as the determination of the best time to exercise the option. I consider in this section two approaches to the problem stated as an optimal stopping problem. The first is known as martingale approach and is grounded in the theory of martingales. The second is known as Markovian approach and requires the existence of a Markovian state variables vector that re-assumes the state of the economy. Any time at which the exercise is favorable is called stopping time. Both the approaches consider the marginal expectations of the gain at exercise, discounted by risk and time, at every stopping time before or at the contract maturity. The price of an American option is the maximum between all these expectations. The difference between the two approaches is that in the second approach the conditioning information set is completely spanned by the state variables. Therefore, the martingale approach involves the conditional probability density of the state variables, while the Markovian approach involves their transition density.

### 1.1.1 Martingale approach

Let us consider an American option with maturity $T$, gain $g_t$ at exercise and interest rate $r_t$ at time $t$. The maturity is constant and possibly infinite. The gain at exercise depends on the unpredictable state of the economy. The interest rate is time-dependent and possibly stochastic. Let us indicate the option price at time $t$ by $V_t$ and its time-to-maturity by $h := T - t$. Generally speaking, a stopping time $\tau$ for a sequence of random variables is an almost surely finite random variable such that its occurrence is adapted to the filtration of sigma-algebras generated by the sequence. Let us define the set of stopping times for the option exercise decision at time $t$ as

$$\mathcal{T}_t(h) := \{\tau \in [t, t+h] \;\; s.t. \;\; V_\tau = g_\tau\} \,. \tag{1.1}$$

The time- and risk-discounted gain expected at time $t$ from the exercise of the considered option at any stopping time $\tau$ is

$$\mathrm{E}^{\mathscr{Q}}\left[ e^{-\int_t^\tau r_s ds} g_\tau \middle| \mathcal{F}_t \right], \tag{1.2}$$

where $(\mathcal{F}_t)$ is the filtration such that $e^{-\int_t^\tau r_s ds} g_\tau$ is $\mathcal{F}_t$-measurable and $\mathrm{E}^{\mathscr{Q}}[. \mid \mathcal{F}_t]$ is the risk-neutral expectation conditional to the information available at time $t$. To make this expectation exist, let us assume that

$$\sup_{\tau \in \mathcal{T}_t(h)} e^{-\int_t^\tau r_s ds} g_\tau \in \mathcal{L}^p(\mathscr{Q}), \tag{1.3}$$

where $\mathcal{L}^p(\mathscr{Q})$ is the linear space of $p$-integrable functions under the measure $\mathscr{Q}$, for some $p > 1$. To prevent arbitrage opportunities, the price maker values the option at time $t$ as the maximum time- and risk-discounted gain expected from an exercise at any stopping time $\tau$. Let us introduce the stochastic process $(V_t)$, for positive $t$, such that

$$V_t(h) := \operatorname{ess} \sup_{\tau \in \mathcal{T}_t(h)} \mathrm{E}^{\mathscr{Q}}\left[ e^{-\int_t^\tau r_s ds} g_\tau \middle| \mathcal{F}_t \right], \tag{1.4}$$

for the time-to-maturity $h$. I use the essential supremum operator $\operatorname{ess}\sup$ and not simply the operator $\sup$ because the supremum of a set of random variables is random itself.[2] The process $(V_t)$ is the smallest $\mathscr{Q}$-supermartingale majorant of the time-discounted gain process $\left(e^{-\int_t^\tau r_s ds} g_\tau\right)$. The former is known as the Snell envelope of the latter (see Snell [1952]) under the measure $\mathscr{Q}$. For a discussion about optimal stopping in discrete time and the basic properties of the essential upper bound of a stochastic process see Neveu [1975].

Let us consider the stopping times included in set $\mathcal{T}_t$ defined in Equation (1.1). Between them, the optimal stopping time is the one such that, under the measure $\mathscr{Q}$, the expected gain from the exercise at that time is not lower than any expected gain at a following moment. Similarly, the exercise decision at that time is optimal. The optimality condition for the stopping time and exercise decision depend on the state of the economy and the price maker needs to determine optimal stopping time and risk-neutral expectation at any time. Let us consider the optimal stopping time at time $t$ for the period $[t, t+h]$ and interpret the American option price as the value of a fictitious European contract. This European contract, that offers the right to perform the pre-specified financial transaction only at maturity, expires at the optimal stopping time for the American option. All the other contract characteristics are the

[2] The essential supremum (or essential upper bound) for $\mathrm{E}^{\mathscr{Q}}\left[ e^{-\int_t^\tau r_s ds} g_\tau \middle| \mathcal{F}_t \right]$ is the smallest positive $\epsilon$ such that $\mathrm{E}^{\mathscr{Q}}\left[ e^{-\int_t^\tau r_s ds} g_\tau \middle| \mathcal{F}_t \right] < \epsilon$ almost surely, for any $\tau$.

same. This means that the European option price is

$$\mathrm{E}^{\mathscr{Q}}\left[e^{-\int_t^\tau r_s ds} g_\tau \middle| \mathcal{F}_t\right], \tag{1.5}$$

for the optimal stopping time $\tau$ at time $t$.

A potential is defined as a càdlàg non-negative supermartingale process with vanishing expected value as time goes to infinity (see e.g. Protter [2004]). The Riesz decomposition states that any càdlàg uniformly integrable supermartingale is equal to the sum of a potential and a càdlàg uniformly integrable martingale (see e.g. Neveu [1975]). El Karoui and Karatzas [1995] apply this result to the Snell envelope for an American put option written on an asset with continuous path (see also Myneni [1992] and Rutkowski [1994] in the case of constant and stochastic interest rates, respectively). They derive the Early Exercise Premium (EEP) representation of the option: the American option price can be written as the sum of the price of the European option with corresponding contract characteristics and the EEP, that is the present value of the gains from exercise before maturity. The EEP is an integral of the discounted payoff at exercise, that is a non-negative function. When it is not strictly positive, the American option is not optimally exercised before maturity.[3]

Let us now assume to be able to find a non-negative criterion function $Q$ that allows us to order the $\mathscr{Q}$-supermartingales majorant of the time-discounted gain process. The price expressed in Equation (1.4) is equivalent to

$$V_t(h) = \inf_{B \in \mathcal{B}} Q\left(B\right), \tag{1.6}$$

where $\mathcal{B}$ is the space of the $\mathscr{Q}$-supermartingales $B := (B_{\tilde{t}})$ majorant of process $\left(e^{-\int_t^{\tilde{t}} r_s ds} g_{\tilde{t}}\right)$. The formulation of the American option pricing problem in Equation (1.6) is defined primal. An equivalent dual representation of the price of the American option as the infimum over a class of $\mathscr{Q}$-martingales of a family of expectations can also be shown (see for instance Davis and Karatzas [1994], Rogers [2002] and Haugh and Kogan [2004]):

$$V_t(h) = \inf_{C \in \mathcal{C}} \mathrm{E}^{\mathscr{Q}}\left[\sup_{\tau \in [t, t+h]} \left(e^{-\int_t^\tau r_s ds} g_\tau - C_\tau\right) \middle| \mathcal{F}_t\right], \tag{1.7}$$

[3]For instance, this is the case of a call option written on a a stock that pays no dividend. In some empirical applications, when the EEP is smaller than the bid-ask spread or a basis point, some authors approximate the price of an American option by the one of an European option with same contract characteristics. See for instance Flesaker [1993] in the context of American options written on interest rates, Broadie, Chernov and Johannes [2007] for American options written on futures and Bikbov and Chernov [2010] for American options written on exchange rates.

where $\mathcal{C}$ is the space of the $\mathscr{Q}$-martingales $C := (C_\tau)$ null at time $t$ and such that $\sup_{\tau \in [t,t+h]} |C_\tau|$ belongs to the linear space of integrable functions under the measure $\mathscr{Q}$.[4]

Let us now make an inductive argument on the price $V_t$ in a discrete time setting. At any time the American option price is the maximum between the contemporaneous exercise gain and the expected time- and risk-discounted value of the option at the following time. The former value is called exercise value. The latter value is called continuation value and it involves an expectation conditional on the information currently available. At maturity, the price of the option equals the exercise value, since there is no any later exercise possibility. Then, we can write

$$V_t(h) = \begin{cases} \max\left[g_t, \mathrm{E}^{\mathscr{Q}}\left[V_{t+1}(h-1)\middle|\,\mathcal{F}_t\right]\right], \text{ for } h > 0, \\ \qquad\qquad\qquad\qquad g_T, \text{ for } h = 0. \end{cases} \tag{1.9}$$

This inductive argument is known as the principle of optimality of Dynamic Programming (DP) or Bellman's principle.[5] From Equation (1.9) at the contract expiration we understand that set $\mathcal{T}_t$ defined in Equation (1.1) is never empty, since it contains the option maturity.

### 1.1.2 Markovian approach

Let us assume the existence of a Markovian state variables vector $X_t$ that represents the state of the economy and takes value in set $\mathcal{X}$. In other words, let us assume that at any time the information set is completely spanned by this vector. We can write the gain at exercise as a function of this vector: $g_t \equiv g(X_t)$.[6] The value of the American option at time $t$, for time-to-maturity $h$, is $V_t(h, X_t)$. This value has the following expression:

$$V_t(h,x) := \sup_{\tau \in \mathcal{T}_t(h)} \mathrm{E}^{\mathscr{Q}}\left[e^{-\int_t^\tau r_s ds} g(X_\tau)\middle|\, x\right], \tag{1.10}$$

[4]The derivation of the dual representation is based on the Doob-Meyer decomposition of the price $V_t$ into the sum of a $\mathscr{Q}$-martingale component and a bounded variation component:

$$V_t = V_0 + M_t + A_t, \tag{1.8}$$

where $(M_t)$ is a $\mathscr{Q}$-martingale and $(A_t)$ is a decreasing predictable process (i.e. $A_t$ is measurable w.r.t the information at date $t-1$) such that $M_0$ and $A_0$ are null. This decomposition holds under Assumption (1.3) and for a càdlàg time-discounted gain process $\left(e^{-\int_t^\tau r_s ds} g_\tau\right)$.

[5]Bellman [1957] states the principle of optimality in this way: *an optimal policy has the property that whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

[6]For sake of simplicity I do not consider in the exposition gain at exercise that are also dependent on time in a deterministic way. For them, a similar reasoning applies.

where $\mathrm{E}^{\mathscr{Q}}[.\,|\,x]$ is the risk-neutral expectation conditional on the realization $x$ of the state variables process. To ensure the existence of the previous expectations, let us assume that

$$\mathrm{E}^{\mathscr{Q}}\left[\sup_{\tau\in\mathcal{T}_t(h)} e^{-\int_t^\tau r_s ds} g(X_\tau)\right| x\left.\right] < \infty, \tag{1.11}$$

for any $x \in \mathcal{X}$. At any time $t$, the set of state variable values for which the option is exercised is defined stopping region and its set-theoretical complement is defined continuation region $\mathcal{C}_t$:

$$\mathcal{C}_t(h) := \{x \in \mathcal{X} : V_t(h,x) > g(x)\},$$

for the option time-to-maturity $h$. The exercise boundary is the set of state variable values that separate the continuation from the stopping region. These values are defined critical.

Similarly to the interpretation of the American option price as an European option price in the Martingale approach of Subsection 1.1.1, we can write

$$V_t(h,x) = \mathrm{E}^{\mathscr{Q}}\left[e^{-\int_t^\tau r_s ds} g(X_\tau)\right| x\left.\right], \tag{1.12}$$

for any $x \in \mathcal{X}$ and the optimal stopping time $\tau$ at time $t$.

The principle of optimality of DP in the Markovian case gives the Wald-Bellman equations. For a time-homogeneous state variables process in discrete time these are

$$V_t(h,x) = \begin{cases} \max\left[g(x), \mathrm{E}^{\mathscr{Q}}\left[V_{t+1}(h-1, X_{t+1})|\, x\right]\right], \text{ for } h > 0, \\ g(x), \text{ for } h = 0. \end{cases} \tag{1.13}$$

The Wald-Bellman equations in discrete time are also known as Value Iteration (VI) algorithm for DP. The $Q$-VI algorithm for DP is an equivalent algorithm focused on the continuation value $Q_t(h,x) := \mathrm{E}^{\mathscr{Q}}\left[V_{t+1}(h-1, X_{t+1})|\, x\right]$ at time $t$:

$$Q_t(h,x) = \begin{cases} \mathrm{E}^{\mathscr{Q}}\left[\max\left[g(X_{t+1}), Q_{t+1}(h-1, X_{t+1})\right]|\, x\right], & \text{for } h > 0, \\ 0, & \text{for } h = 0, \end{cases} \tag{1.14}$$

see e.g. Tsitsiklis and Van Roy [2001]. The optimal stopping time depends on the state variables. Let us make this dependence explicit and consider the indicator function $\mathbf{1}_{\mathcal{C}_t(h)}$ for the continuation region at time $t$. Let us use these functions to formalize an inductive argument in discrete time for the optimal stopping time:

$$\tau_t^*(h,x) = \begin{cases} t + \left(\tau_{t+1}^*(h-1,x) - t\right) \mathbf{1}_{\mathcal{C}_t(h)}(x), & \text{for } h > 0, \\ T, & \text{for } h = 0. \end{cases} \tag{1.15}$$

for any $x \in \mathcal{X}$.

## 1.2 No arbitrage inequalities

The style of many traded options is American. For a given gain at exercise, further boundaries on the option price can be derived by no-arbitrage arguments. I consider in this section the case of the most liquid American-style options traded in centralized markets: call and put options, generally called plain vanilla options. For these options the gain at exercise depends on time only through the value of the underlying asset. An American call option provides the holder the right to purchase the underlying asset at an agreed strike price $K$ at any time on or before maturity $T$. Its gain at exercise is then $g_t = (S_t - K)^+$. Conversely, an American put option provides the holder the right to sell the underlying asset at strike price $K$ at any time on or before maturity $T$. Its gain at exercise is then $g_t = (K - S_t)^+$. Let us consider an American call option with price $C_t$ and an American put option with price $P_t$, both written on the same underlying with price $S_t$ and with the same time-to-maturity $h$. I indicate by $c_t$ and $p_t$ the value of the European call and put option with the same characteristics as the American ones. I consider the dividend yield $\delta$ paid by the underlying asset and the risk-free rate $r_f$, assumed both constant in time. By no-arbitrage arguments, the following inequalities hold (see e.g. Musiela and Rutkowski [2005]):

$$(S_t e^{-\delta h} - K e^{-r_f h})^+ \leq c_t \leq C_t \leq S_t,$$

$$(K e^{-r_f h} - S_t e^{-\delta h})^+ \leq p_t \leq P_t \leq K e^{-r_f h}.$$

A put-call inequality holds (see Merton [1973b]):

$$C_t - S_t e^{-\delta h} + K e^{-rh} \leq P_t \leq C_t - S_t e^{-\delta h} + K.$$

Moreover, for a given maturity, the call (put) option prices are convex decreasing (increasing) functions of the strike price.

## 1.3 Solution of the optimal stopping problem

In this section I review some of the most notable techniques for the pricing of an American option that rely on a specification of the Markov state variables process $(X_t)$ under the risk-neutral measure $\mathscr{Q}$. For some particular specification of this process, some properties of the price of an American plain vanilla option are known. For instance, the price $V_t$ on an American call option written on an asset with price $S_t$ following a GBM is continuous w.r.t. $S_t$ and time $t$, because of the continuity of the payoff function at exercise $g_t = (S_t - K)^+$ and any path of $S_t$. The price $V_t$ is non-decreasing and convex in $S_t$, because the payoff function at exercise has these characteristics and it is non-decreasing in the initial condition (see e.g. Detemple [2005]). Moreover, the exercise boundary for this option is convex (see Chen, Chadam, Jiang and Zheng [2008]). Some characteristics of American options written on a stock that follows a diffusive process with volatility that is dependent on both time and level are studied in Ekstrom [2004] and a jump-diffusion process in Ekstrom and Tysk [2007]. Schroder [1999] and Detemple [2001] show the equivalence between the values of an American call and put with the same maturity. The two options differ in terms of underlying and strike prices, interest and dividend rates. The diffusive processes followed by the underlying assets of the two options are different. At a given time, the strike price of an option plays as underlying price for the other, and the same happens with interest and dividend rates. Moreover, Detemple [2001] extends the results to some American-style exotic options (see also Detemple [2005]).

Once the process of the Markov state variables process is specified, some techniques rely on the transformation of the optimal stopping problem into a free-boundary problem. These techniques can lead to analytic or semi-analytic results. Other techniques are iterative and directly inspired by the DP representation. By these last techniques we can get only numerical results. I describe some techniques based on the free boundary formulation of the problem in Subsection 1.3.1 and some iterative procedures in Subsection 1.3.2. For comparative analysis of the performance of different methods see Broadie and Detemple [1996], Ait-Sahlia and Carr [1997] and Pressacco, Gaudenzi, Zanette and Ziani [2008]. I conclude this section by reporting in Subsection 1.3.3 some empirical findings against the hypotheses of rational agents and frictionless market.

### 1.3.1 Free boundary formulation

The American option price expressed in Equation (1.10) is related to a parabolic-elliptic Partial Differential Equation (PDE), when $X_t$ has continuous paths, and to a Partial Integro-Differential Equation, when $X_t$ has discontinuous paths (see e.g. Peskir and Shiryaev [2006]). No analytic expression for the price of the American option is available, even for an American-style plain vanilla written on a stock following a Geometric Brownian Motion (GBM).[7] Nonetheless, many characteristics of the price of the American options are known in some special cases.

The first attempt to price an American option is the work of McKean [1965] appeared in an appendix of Samuelson [1965], an article on the pricing of warrants. McKean [1965] considers the PDE with the free boundary conditions for a stock price following a GBM. He uses an incomplete Fourier transform to obtain an integral representation for the American call price that involves the exercise boundary. The evaluation of this expression on the exercise boundary provides an integral equation for the boundary itself. In this article there is not the risk-neutralization argument based on Ito's lemma and risk-less hedging introduced in Black and Scholes [1973]. Nonetheless McKean [1965] derives several properties of the solution, including an exact asymptotic formula. Friedman [1959] faces a similar one-dimensional free boundary problem in a study of ice melting. In this article, the existence and local uniqueness of the solution are proved by using the contraction mapping theorem for a small time interval and then extended by induction to any interval of time. For a stock following a GBM, Black and Scholes [1973] derive the PDE for the European call option price by arbitrage-free arguments on a portfolio composed by a long position in the option $V_t$ and a short position in $\frac{\partial V_t}{\partial S_t}$ shares of the underlying asset. This portfolio is immune by risk and then its rate of return is equal to the constant risk-free rate $r_f$. The Black and Scholes (BS) PDE with boundary condition at maturity then follows:

$$\begin{cases} r_f V_t - \mathcal{D} V_t = 0, \\ \qquad\qquad V_T = g_T, \end{cases}$$

for the Dynkin operator $\mathcal{D}$ associated with the GBM dynamics defined as

$$\mathcal{D} = \frac{\partial}{\partial t} + \mu \frac{\partial}{\partial S_t} + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial S_t^2}, \tag{1.16}$$

with constant drift $\mu$ and volatility $\sigma$. The BS PDE associated with the boundary condition gives rise to a linear parabolic Cauchy problem, and its solution is the famous BS formula. In the case of an American option, by a similar reasoning we get the following non-linear PDE with boundary condition at maturity:

$$\begin{cases} \min\left[r_f V_t - \mathcal{D} V_t, V_t - g_t\right] = 0, \\ \qquad\qquad\qquad\qquad V_T = g_T. \end{cases} \tag{1.17}$$

[7] Zhu [2006] uses the homotopy-analysis method to derive an asymptotic expansion of the price of an American put written on a stock following a GBM process. Anyway, to make the computation feasible the series must be truncated and then approximated.

This system represents a free boundary problem. In order to uniquely identify the frontier between exercise and continuation regions, let us add two boundary conditions. These conditions pertain the value of the option and its first derivative w.r.t. the underlying asset price on the exercise boundary. For a put option with critical stock price value $S_t^*$, the first condition, known as value matching condition, is

$$V_t(S_t^*) = K - S_t^* \tag{1.18}$$

and the second condition, known as smooth fit principle (or smooth pasting or high contact condition), is

$$\frac{\partial V_t(S_t^*)}{\partial S_t} = -1. \tag{1.19}$$

The first attempt to derive the smooth fit principle for diffusion processes in the literature of American option pricing is in McKean [1965]. Another derivation is in Bather [1970] and Van Moerbeke [1976], who consider a Taylor expansion of the exercise payoff at the exercise boundary. McKean [1965] correctly points out that the smooth fit principle in Equation (1.19) is not valid if the option underlying asset follows a process with jumps (see the appendix of Barone-Adesi [2005] for an intuitive discussion). Jamshidian [1992] shows how to move from the solution of the homogeneous PDE for plain vanilla American options inside the continuation region to the solution of an inhomogeneous PDE on an unrestricted region. He provides an intuitive explanation of the EEP representation as the weighted sum of the solutions of the two equations corresponding to the PDE in the stopping region and the continuation region respectively. The weights are given by the probabilities of being in either region at any time before maturity. Kim [1990], Jacka [1991] and Carr, Jarrow and Myneni [1992] consider a GBM for the underlying asset and derive the EEP representation. They write the price of the European option with the same contract characteristics by the BS formula and the EEP as an integrated BS formula with the stock critical value $S_t^*$ replacing the strike price. In particular the integration is w.r.t. the maturity of the option and the domain of integration is the time interval from the actual time to the maturity of the American option. The critical value $S_t^*$ is the solution of a recursive non-linear integral equation with the boundary conditions regarding its value at maturity and at the moment immediately prior to maturity. Myneni [1992] and Carr, Jarrow and Myneni [1992] derive the Delayed Exercise Premium (DEP) representation of the value of the American option for an underlying asset price that follows a GBM (see Detemple [2005] in the case of an Ito process). In this representation, the American option is written as the sum of the immediate exercise value (also called intrinsic value) and the DEP (also called time value) of the American option, that is the present value of the gains from delaying the exercise.

Many authors focus on analytic expressions for the properties of the solution in some extreme cases.

For instance, McKean [1965] derives the expression of the critical price when the American put option is perpetual (i.e. with infinite time-to-maturity). Kim [1990] derives the price for the perpetual American call option and a limit of the exercise boundary at maturity. In Wilmott, Howison and Dewynne [1993] this last limit is obtained by considering the PDE for very small times-to-maturity. Other results on the asymptotic behavior of the critical price for processes with constant volatility can be found in Barles, Burdeau, Romano and Sansoen [1995], Evans, Kuske and Keller [2002], Lamberton and Villeneuve [2003] and Chen, Chadam, Jiang and Zheng [2008]. The adaptation of the results to the case of a local volatility dependent only on the stock value is in Chevalier [2005]. Mordecki [2002] presents the results of a perpetual American option for models based on Lévy processes. Alili and Kyprianou [2005] show the necessary and sufficient condition for the smooth fit principle to hold in the case of a perpetual American option written on a stock modeled by a Lévy process and that pays no dividend.

The American option pricing problem in System (1.17) can be equivalently formulated as an obstacle problem or a variational inequality. Bensoussan and Lions [1982] develop some variational inequalities and interpret this system and other differential equations associated to different Markovian processes as weak non-linear PDE's with initial conditions on the space of distributions. If the exercise payoff function $g_t$ belongs to a suitable Sobolev space, variational methods lead to solutions in a stronger sense (see also Bensoussan [1984] for the theory of variational inequalities). Crandall, Ishii and Lions [1992] interpret these equations in the viscosity sense. Thanks to their interpretation, less regularity assumptions on the exercise payoff function $g_t$ are required. Gatarek and Świech [1999] consider a diffusion taking values on a Hilbert space and show that the value function is the unique viscosity solution of an obstacle problem for the associated parabolic PDE in the Hilbert space.

Brennan and Schwartz [1977] and Brennan and Schwartz [1978] are the first to use the finite difference method to approximate numerically the Dynkin operator in System (1.17). The time and the domain of the state variables vector are discretized on a grid and then at every grid point each partial derivative is replaced by a Taylor expansion. In an explicit finite difference scheme the variable at date $t+1$ depends explicitly on its lagged value at time $t$, while in an implicit finite difference scheme a combination of the possible value of the variable at date $t+1$ depends on the value of the variable at time $t$. The PDE approximated by one of these schemes and the boundary conditions form a set of difference equations which can be solved either directly or iteratively. Brennan and Schwartz [1978] show that the probabilities of a jump process approximation to the underlying diffusion process correspond to the coefficients of the difference equation that approximates the BS PDE. Jaillet, Lamberton and Lapeyre [1990] provide a rigorous justification of the method introduced in Brennan and Schwartz [1977] by using the theory of variational inequalities illustrated in Bensoussan and Lions [1982] in the case of diffusion models and Zhang [1997] in the case of jump-diffusion models. A recent review on the use of finite difference methods in finance is Duffy [2006]. Other authors use the finite element method for finding the approximated solution of the PDE, for a review see Topper [2005]. Johnson [1983] replaces the put value by an approximating function of the model and contract parameters, but this approximation can not be arbitrarily accurate. Geske and Johnson [1984] suggest a variation on the approach based on a extrapolation scheme. They solve the problem of the valuation of a Bermudan option (i.e with exercise possible at only a discrete number of dates) and consider the result as a discrete approximation of the solution of the American put option pricing problem. This last approach is computationally intensive because requires the computation of high dimensional multivariate normal probabilities. Bunch and Johnson [1992] improve the accuracy and computational efficiency by restricting the exercise possibility at only two dates. MacMillan [1986] and Barone-Adesi and Whaley [1987] use the EEP decomposition. They price the European option component by the BS formula and model directly the EEP. Barone-Adesi and Whaley [1987] focus on the PDE for the EEP in the case of options with very short or very long maturity. This PDE is a second order ordinary differential equation and the authors get the first order term of the correct solution. Improvement of this methodology are in Barone-Adesi and Elliott [1991] and Allegretto, Barone-Adesi and Elliott [1995]. Ju and Zhong [1999] add a second order term to the solution. Ho, Stapleton and Subrahmanyam [1997] adapt the Geske and Johnson [1984] method to the case of stochastic interest rates. Carr and Faguet [1994] discretize the time derivative in the PDE by the method of lines, approximate the exercise boundary and then compute option values and hedging parameters. Capped options are call and put options automatically exercised when the asset price closes at or above (for a call) or at or below (for a put) a predetermined level. Broadie and Detemple [1995] value these options written on a stock following a GBM. Broadie and Detemple [1996] use these options to express lower and upper bounds of the prices of plain vanilla American options. The value of these last options is then obtained by interpolation. Huang, Subrahmanyam and Yu [1996] approximate the exercise boundary using Richardson extrapolation to accelerate convergence for a stepwise approximation of the free boundary. The computation by this methodology is faster than the one by Kim [1990], but it still requires the solution of several integral equations. Ju [1998] approximates the exercise boundary with a multipiece exponential function. Pham [1998] considers the EEP representation for a jump-diffusion model for the underlying asset and shows how in this case the EEP has a component due to the eventual occurrence of jumps. Little, Pant and Hou [2000] reduce the dimensionality of the problem by an equation for the boundary with only an integral. Broadie, Detemple, Ghysels and Torrés [2000a] and Broadie, Detemple, Ghysels and Torrés [2000b] consider the EEP representation for an underlying asset with stochastic dividend yield and volatility and estimate the early exercise boundary in a nonparametric way. Detemple and

Tian [2002] considers American options written on an asset price that follows a diffusion process and with stochastic interest rate. They derive a recursive integral equation for the exercise boundary and provide a parametric representation of the American option price. Kallast and Kivinukk [2003] propose a numerical method to compute the solution of the Kim's integral equations and estimate numerically the exercise boundary. Carr [1998] computes the American option from a sequence of options with random maturity. Huang and Pang [1998] consider some American option pricing models as linear complementarity problems involving partial differential operators and study the properties of their discretizations. Hon and Mao [1999] use the radial basis function method to solve numerically System (1.17). Chiarella and Ziogas [2009] adopt the method introduced by Jamshidian [1992] for a jump-diffusion process.

### 1.3.2 Iterative procedures

Regression-based methods rely on the computation of the continuation value at each step in a discrete time framework (see Equations (1.13)). The conditional expectation is usually computed by lattice methods or quadrature methods when the state variables vector is low dimensional and by Monte Carlo (MC) or quasi-MC methods when it is high dimensional.

Lattice methods require the discretization of the state variables space to be deterministic. At each time step, the state variables vector can move toward one of a finite number of values with a certain probability. Parkinson [1977] considers the American put problem by taking series expansion of the solution given by McKean [1965] in transform space. He uses a three-jump process that approximates the process followed by the continuous log-normal underlying price. He also shows that a trinomial tree is equivalent to the explicit finite difference method and that a generalized multinomial jump process is equivalent to a complex implicit finite difference approximation. The binomial tree approach becomes widely famous with the article by Cox, Ross and Rubinstein [1979], where, even if not dealing with American-style contingent claims, the authors clarify the direct implementation of the DP principle for option pricing. This pricing technique is introduced also in Sharpe [1978] and Rendleman and Bartter [1979]. Improvement of the binomial tree are the trinomial tree of Boyle [1988], the multinomial tree of Kamrad and Ritchken [1991] and the efficient lattice algorithm of Ritchken and Trevor [1999]. Convergence results for the discrete time approximation in a general setting are in Amin [1993] and Amin and Khanna [1994]. In particular, Amin [1993] develops an extension of the binomial method to handle the inclusion of jumps, making the binomial tree applicable to jump-diffusion models. A quadrature method is a computational method to value a definite integral. This method makes use of interpolating functions that are easy to integrate.[8] It gives the value of an approximation of the considered integral. The value of this last integral corresponds to the limit case of an infinite number of considered points and can be inferred by an extrapolation methods, for instance the Richardson's extrapolation.[9] Sullivan [2000] proposes a quadrature method to approximate the value of the American put given by a recursive function of its payoff by using Chebyshev polynomials.

For high dimensional integrals MC and quasi-MC methods are preferred. In MC and quasi-MC methods the integrand is evaluated at some randomly or quasi randomly chosen points and then averaged. These methods are then based on a random discretization of the state variables space. Tilley [1993] introduces the use of MC simulations in the literature on DP approaches to the American option pricing problem. Grant, Vora and Weeks [1996] show how to price American options written on an asset that follows a pure diffusion and jump-diffusion processes by using MC simulations. Carriere [1996] considers the price of an American option expressed as the solution of the VI algorithm in a DP framework. He considers a nonparametric estimator of the continuation value based on q-splines and local polynomial smoothers as approximating functions over a sample of MC simulated paths under the risk-neutral probability measure. Broadie and Glasserman [1997b] develop algorithms that use simulated trees and backwards recursion to obtain a biased high and a biased low estimator of the American option price that are convergent and asymptotically unbiased in the computational burden. Longstaff and Schwartz [2001] introduce the MC Least Squares method for American option pricing. At each point in time they regress the continuation value on a set of basis functions dependent of the state variables vector. They simulate many MC paths of the state variables vector, consider only the value of the state variables vector such that the option is in the money and estimate the expectations by MC averages. The authors outline a convergence proof and Clément, Lamberton and Protter [2002] provide a complete convergence proof and a central limit theorem for the algorithm. Tsitsiklis and Van Roy [2001] consider payoffs dependent on many state variables, for example options written on many underlyings, as a basket option. In order to handle with the curse of dimensionality, they use an Approximate VI algorithm. It is basically the same algorithm as the VI algorithm of Equations (1.13), but it relies on the computation of the option value only at some points of the state space. The entire value function is reconstructed by using a linear combination of basis functions to fit the data via least squares regression. At any step a projection operator minimizes a least squares criterion. The authors study how the approximation errors propagate in the algorithm. Rogers [2002] considers the dual representation of

[8] For instance, the Newton-Cotes quadrature methods express the integral as the sum of integrals on smaller integration domains and approximate each of these integrals by the area of geometric figures. The Gaussian and Clenshaw-Curtis quadrature methods express the integral as a weighted sum of the value of the integrand function at some special points.

[9] The quadrature rule can be applied to high dimensional integrals by computing an iterated one-dimensional integration thanks to the Fubini's theorem, but the function evaluations grow exponentially in the number of dimensions (curse of dimensionality).

the price of the American option in Equation (1.7), chooses a $\mathcal{Q}$-martingale and computes the conditional expectation by simulation. Andersen and Broadie [2004] and Haugh and Kogan [2004] adopt a similar approach and construct upper and lower bounds on the true price of the option. Bally and Pagès [2003] and Bally, Pagès and Printems [2005] introduce a quantization method for the computation of a large number of conditional expectations that consists in the projection of the underlying process on a special grid designed to minimize a projection error. Glasserman and Yu [2004] and Stentoft [2004] study the convergence of the different DP algorithms as the number of basis functions and MC samples increase. In the stochastic mesh approach proposed in Broadie and Glasserman [2004], many paths for of the future values of the state variables vector are simulated. In this way a set of nodes is created at each time. The estimation of the continuation value is then performed backwards in time: the continuation value at time $t$ is approximated by a weighted sum of the possible values at time $t+1$. The weight of these values is connected to their probability of occurrence. An additional study on this method is in Avramidis and Matzinger [2004] and a modification is in Liu and Hong [2009]. Barty, Girardeau, Roy and Strugarek [2008] consider recursive kernel regression estimates of the continuation value on simulated data. In particular they use Gaussian kernels as mollifiers and approximate the value functions as a sum of Gaussian kernels, that can be rapidly computed by a Gauss Transform. Egloff, Kohler and Todorovic [2007] consider least squares splines MC regressions. Kohler, Krzyżak and Todorovic [2010] consider least squares neural network MC regressions. Laprise, Fu, Marcus, Lim and Zhang [2006] price American-style derivatives in a Markovian setting. They approximate the value function with an interpolation function and convert the pricing of American-style derivative with an arbitrary payoff function to the pricing of a portfolio of European Call options. For extended surveys on Monte Carlo methods for option pricing see Dupire [1998], Fu, Laprise, Madan, Su and Wu [2001],Glasserman [2004] and Kohler [2010].

### 1.3.3 Empirical findings against some assumed hypotheses

The optimal stopping formulation of the American option pricing problem described in Section 1.1 and consequently all the methods considered in this survey are based on the assumptions of rational investors and frictionless market. In real markets investors play in a more or less rational way on the base of preferences and transaction costs. Many studies stress the limits of these assumptions. Tax depend on the investor's tax status and on the type of underlying asset. They can therefore inhibit a rational strategy. Diz and Finucane [1993] show that transaction costs and other market frictions can induce early exercise of index options. Overdahl and Martin [1994], Finucane [1997] and Engstrom, Nordén and Stromberg [2000] find similar results for markets of equity options. Carpenter [1998] explain the irrational exercise of not tradeable Employee Stock Options (also Detemple and Sundaresan [1999] takes this fact into account). Poteshman and Serbin [2003] find that sometimes public customers of full-service and discount brokers exercise American options in an irrational way, while firm proprietary traders never do so. In particular this irrational exercise is triggered by exceptional levels of the underlying stock price (see also Duffie, Liu and Poteshman [2005]). Alpert [2010] show how taxes can trigger an early exercise. Some models recently presented account for market frictions. For instance, Costantidines and Perrakis [2007] and Roux and Zastawniak [2009] consider transaction costs.

## 1.4 Model calibration and empirical methods

In this section I describe some calibration and empirical methods. The aim of these methods is bringing the theoretical models to data. They provide us with the value of the model risk-neutral parameters that best fit some cross-sections of observed option prices on the base of a criterion. The risk-neutral density, or at least some of its moments, is backed out of observed option data and then used to price other contracts. This density is called implied density. Due to the complexity introduced by the early exercise possibility, models are preferably calibrated to some observed European option prices. In this section I focus on some methods that allow for a calibration to American option prices.

Achdou [2005] considers the calibration of a GBM with American plain vanilla options. Achdou [2008] does the same for an underlying asset following a Lévy process. Rubinstein [1994] introduces the method of implied binomial trees. He extracts the risk-neutral transition density from a cross section of European options (see also Jackwerth and Rubinstein [1996], Jackwerth [1997], Jackwerth [1999], Jackwerth [2000]) by considering the closest transition density to a reference density, e.g. a log-normal one. Tian [2011] adapts the method of implied binomial trees to American options. Stutzer [1996] introduces the canonical valuation methodology in the context of European option pricing: the risk-neutral transition density is estimated by the minimizer of the Kullback Leibler distance from a nonparametric estimator of the historical transition density, subject to the no-arbitrage restriction on the underlying asset. Alcock and Carmichael [2008] adopts the canonical valuation methodology with no-arbitrage restrictions on an American option at every possible exercise date. The estimated risk-neutral transition density is then used to price the option itself by the Longstaff and Schwartz [2001] algorithm (see also Liu [2010]). Alcock and Auerswald [2010] consider the same approach and add the no-arbitrage restrictions on a cross section of observed European options. Duan [2002] considers the series of returns on the underlying and divides the difference between each return and the conditional mean by the conditional volatility. He obtains in this way an independent identically distributed (i.i.d.) sequence and by a transformation of their marginal distribution he gets an i.i.d. standard normal sequence. He then considers the minimizer of the relative entropy distance from the standard normal distribution of these normalized returns, subject to the no-arbitrage restriction on the underlying asset, as estimator for the risk-neutral distribution.

The stochastic discount factor embodies the aversion of the investors for risk. If they are risk-averse, the risk-neutral transition density has more probability mass in extreme events than the historical one (see for instance the review on pricing kernels in Hansen and Renault [2010]). Only in absence of risk aversion, i.e. when in the price formation there is no risk-discount, the two transition densities coincide. The expected return under the risk-neutral measure is the risk-free rate and it is usually a different value under the historical measure, since investors ask for a premium for bearing the risks associated to holding the stock. They require a premium for the risks of a lower excess return and an higher return variance than expected (see e.g. Mehra and Prescott [1985] for the equity premium in the CAPM, Lamoureux and Lastrapes [1993] for the volatility premium and Carr and Wu [2009] for the variance premium). Many empirical studies show structural differences between the probability densities of the same variables under the historical and risk-neutral measures. For instance, Canina and Figlewski [1993], Bakshi, Cao and Chen [2000], Jackwerth [2000], Jiang and Tian [2005], Bakshi and Madan [2006], Christoffersen, Heston and Jacobs [2006], Carr and Wu [2009] and Bollerslev, Gibson and Zhou [2011] study the differences in the case of index returns and Dennis and Mayhew [2002], Bakshi, Kapadia and Madan [2003] and Duan and Wei [2009] study the differences in the case of equity returns. We assume risk-averse agents and consequently we consider some risk-neutralization arguments in asset pricing. The reason why the risk-neutral transition density should be the closest transition density inside a certain class of distributions is not evident.

Daglish [2003] compare the in-sample and the out-of-sample pricing and hedging performances of some parametric and nonparametric American option pricing techniques. He considers the BS model and the stochastic volatility model considered in Heston [1993], the practitioners' BS model introduced in the context of European option pricing in Aït-Sahalia and Lo [1998], a kernel regression nonparametric model and a technique based on spline fitting. Daglish [2003] finds a superior performance of the nonparametric techniques for in-sample pricing and of the parametric techniques for forecasting and hedging. Nonparametric techniques are considered particularly useful to determine the actual state price density from a cross-section of observed options.[10] The nonparametric techniques do not take into account the no-arbitrage restrictions on the considered assets, but they do not suffer of the

[10] Daglish [2003] writes that *nonparametric methods can be a very effective tool for in-sample pricing of options. Given this, they may represent a very valuable tool for extracting the state price density, which can then be used to perform arbitrage-based pricing of other securities. In light of the work by Broadie, Detemple, Ghysels and Torrés [2000b], I might expect to see nonparametric methods outperforming parametric models when pricing American options for two reasons: improvement in modeling of the underlying asset process, and also being able to fit the optimal exercise boundary more consistently.*

risk of a misspecification of the state variables dynamics, that could lead to mispricing. See Longstaff, Santa-Clara and Schwartz [2001] for a study on the costs of applying single-factor exercise strategies to American-style swap options when the true term structure is driven by several factors.

Some authors face the American option pricing problem as a learning problem. For instance, Chen and Magdon-Ismail [2006] combine an artificial neural network with a multinomial tree to derive the risk-neutral measure from a cross-section of American options. On the one hand, a neural network algorithm offers the possibility of choosing any economic variable as state variable, on the other hand it requires a large cross-section of options to form the algorithm training set.

Using the EEP representation, we could in principle compute the EEP as the difference between the prices of American and European options written on the same asset and with the same contract characteristics. Unfortunately simultaneous liquid market for such pairs of options are extremely rare.[11] Since the put-call parity (see e.g. Stoll [1969] and Klemkosky and Resnick [1979]) does not hold for American options and only the weaker put-call relationship reported in Section 1.2 holds, we could take the deviations from this parity as a measure of the EEP (see Evnine and Rudd [1985] and Zivney [1991] for empirical studies based on this idea).

[11] The comparison in Dueker and Miller [2003] is based on the exceptional quote of both European and American plain vanilla options by CBOE during just few months in 1986.

## 2 Semi-Parametric Estimation of American Option Prices

This chapter deals with the estimation of American option prices in a discrete time, incomplete market, Markovian framework. The Markov state variables vector includes the return on the fundamental asset, as well as other relevant pricing factors, such as the asset stochastic volatility and the discount rate. An American option differs from the corresponding European security since the holder has the right to exercise the option on or before the maturity date (see Broadie and Detemple [2004] and Detemple [2005] for reviews on valuation of American-style derivatives). Thus, the American option valuation problem can be faced as an optimal stopping time problem (see Bensoussan [1984], Karatzas [1988] and Karatzas [1989]).[12] Equivalently, at each date the option value is the maximum between the exercise payoff and the continuation value, that is, the risk adjusted and time discounted conditional expectation of the option value one day ahead. This dynamic programming argument suggests that, in a discrete time framework, the pricing of an American option can be represented by a backward recursive application of a valuation operator that embodies both the exercise decision and the computation of the continuation value.

The literature on dynamic programming approaches to American option pricing has mostly focused on parametric models for the risk-neutral dynamics of the state variables vector, such as the Black-Scholes, stochastic volatility and jump-diffusion models. The time is discretized and, for given values of the model parameters, the backward recursive option valuation is performed assuming a finite set of possible values for the state variables at each date. In lattice methods the state variables domain is discretized in a deterministic way depending on the model (see e.g. the binomial tree of Cox, Ross and Rubinstein [1979], the trinomial tree of Boyle [1988], the multinomial tree of Kamrad and Ritchken [1991] and the efficient lattice algorithm in Ritchken and Trevor [1999]). In Monte Carlo methods the state variables domain is discretized in a stochastic way based on a special choice of the space sampling (see e.g. the random tree of Broadie and Glasserman [1997b], the regression-based Monte Carlo methods of Carriere [1996], Longstaff and Schwartz [2001] and Tsitsiklis and Van Roy [2001] and the stochastic mesh of Broadie and Glasserman [2004]). For instance, in regression-based Monte Carlo methods a sample of state variables paths is artificially generated from the model. The conditional expectation that gives the continuation value at a given date and state is approximated by using nonparametric regression methods applied to the simulated cash-flows or option values at the future dates. Glasserman [2004] explains how regression-based Monte Carlo methods can be interpreted as stochastic mesh approaches. Finally, Sullivan [2000] uses a Gaussian quadrature to

[12] Alternative characterizations of the American option pricing problem for special parametrizations of the state variables process include for instance the free boundary formulation (see e.g. McKean [1965], Brennan and Schwartz [1977], Barone-Adesi and Whaley [1987] and Huang, Subrahmanyam and Yu [1996]).

compute the continuation value.

Despite this huge body of literature on valuation, the analysis of statistical estimation methods with American option price data is very limited, likely because of the complexity induced by the pricing problem. Nonparametric estimation methods are particularly convenient in this respect, since they allow to bypass this complexity by postulating a flexible link function relating the American option price with observable contract characteristics and state variables. For instance, Broadie, Detemple, Ghysels and Torrés [2000a] and Broadie, Detemple, Ghysels and Torrés [2000b] consider kernel-based regression methods including the moneyness strike, the time-to-maturity, the asset stochastic volatility and dividend yield among the regressors. In an empirical study, these authors find that both dividend yield and stochastic volatility are important determinants of the American option price. Other nonparametric approaches, such as splines and neural networks, are also possible (see Daglish [2003] for a comparative study as well as Hutchinson, Lo and Poggio [1994] and Garcia and Gencay [2000] for the use of neural networks to price European options).

We depart from this literature by combining the dynamic programming formulation with a semi-parametric specification of the risk-neutral distribution in discrete time. Specifically, the historical transition density $f$ of the Markov state is left unconstrained and treated as a functional parameter, while the stochastic discount factor (SDF) is assumed in a parametric family indexed by the finite-dimensional parameter $\theta$. The goal is to estimate the true values of both parameters $f_0$ and $\theta_0$ by the information in a time-series of state variables observations and a cross-section of observed American option prices at the current date. The estimates of $\theta_0$ and $f_0$ are then used to estimate the prices of American options at the current date that are not actively traded on the market. We also propose new semi-parametric estimators for a class of linear or nonlinear functionals of $\theta$ and $f$ that include historical and risk-neutral conditional cross-moments of the state variables, such as leverage effects (see Black [1977]) and term structures of skewness and kurtosis measures (see e.g. Bakshi, Kapadia and Madan [2003]).

The semi-parametric setting introduced in this chapter is intermediate between fully parametric and fully nonparametric approaches. The advantage w.r.t. the former approach is the flexibility in modeling the historical transition density, which allows to get estimators of the option prices and exercise boundary in a rather general model setting. Moreover, we get a proper distribution theory for the estimators without introducing ad-hoc pricing errors. The advantage w.r.t. the latter approach is that the estimated pricing model is arbitrage-free. In nonparametric approaches, ensuring the absence of arbitrage opportunities by imposing shape restrictions on the pricing function might be difficult, since such shape restrictions are not completely known for American options in a general framework (see e.g. Aït-Sahalia and Duarte [2003], Yatchev and Härdle [2006] and Birke and Pilz [2009] for constrained nonparametric estimation of the state price density from European option data).

The information contained in the historical state variables and cross-sectional option data is exploited through the associated no-arbitrage restrictions. In our framework these restrictions are multi-day and involve the recursive valuation operator for American options. The resulting constraints on $\theta_0$ and $f_0$ are nonlinear w.r.t. both parameters and do not correspond to moment restrictions. This feature yields a setting that is different from the ones of the Generalized Method of Moments (GMM, see Hansen [1982] and Hansen and Singleton [1982]), the Extended Method of Moments (XMM, see Gagliardini, Gouriéroux and Renault [2011]) and other semi-parametric settings considered in the literature (see e.g. Ai and Chen [2003], Powell [1994] and Ichimura and Todd [2007]). This difference explains the methodological novelty introduced in this chapter. To get numerically tractable estimators, we consider a two-step approach. First, the SDF parameter $\theta_0$ is estimated by minimizing a distance criterion that corresponds to a quadratic form of the empirical constraint vector. Second, the historical transition density $f_0$ is estimated by minimizing an information-theoretic criterion subject to the set of no-arbitrage restrictions with estimated SDF parameter. The information criterion is based on the Kullback-Leibler distance of $f_0$ from a kernel density estimator (see Kitamura and Stutzer [1997] and Kitamura, Tripathi and Ahn [2004]).

Despite the differences in terms of model specification and data usage, comparing our estimation methodology with the existing literature on dynamic programming valuation gives interesting insights. Indeed, for any given value of the SDF parameter $\theta$, we compute the conditional expectation that gives the continuation value as a weighted average over the sample observations of the state variables. Thus, our approach is closer in spirit to stochastic mesh than to lattice methods, with the historical realization of the state variables vector process taken as a mesh. The weights turn out to be kernel weights adjusted by a tilting factor accounting for the no-arbitrage restrictions, and multiplied by the SDF to pass from the historical to the risk-neutral distribution.

In Section 2.1 we describe the discrete time Markovian framework and define the American option pricing operator for recursive valuation. In Section 2.2 we introduce the semi-parametric specification with historical transition density $f$ of the Markov state and SDF parameter $\theta$. We discuss the no-arbitrage restrictions from the available historical and option data. We investigate the local sensitivity of the no-arbitrage constraint vector to the model parameters by computing the gradient of the constraints w.r.t. $\theta$ and their Fréchet derivative w.r.t. $f$. In Section 2.3 we introduce the semi-parametric estimators of the true SDF parameter $\theta_0$, the true historical transition density $f_0$ and a class of their functionals, including the American option prices. We study the large sample properties of these estimators in Section 2.4. The asymptotics is for a long time-series of state variables observations and a fixed number of cross-sectionally observed option prices. We make a link between the asymptotic properties of the proposed estimators and the ones of information-theoretic GMM estimators, by interpreting the Fréchet derivative of the constraint vector as a moment function locally around the true transition density $f_0$. In Section 2.5 we present the results of a Monte Carlo experiment to study the finite-sample properties of the estimators. In Appendix 4.1 we list the set of regularity assumptions for the validity of the asymptotic properties. Proofs of the propositions are gathered in Appendices 4.2-4.6.

## 2.1 Valuation of American options

In this section we define the dynamics of the state variables and asset prices. We first consider the state variables and the SDF in Section 2.1.1. We then state an homogeneity property w.r.t. the underlying asset price for a class of American options in Section 2.1.2. Finally in Section 2.1.3 we introduce an operator formulation for the American option price useful for the derivation of the theoretical results.

### 2.1.1 The framework

We consider an incomplete market framework in discrete time. The time index $t$, with $t \in \mathbb{N}$, identifies a trading day. A fundamental asset (a stock, say) with price $S_t$, a short-term non-defaultable zero-coupon bond and a set of American options with different contract characteristics written on the fundamental asset are traded on the market. The state variables are the daily geometric return on the fundamental asset $r_t := \log(S_t/S_{t-1})$ and a $(d-1)$-dimensional stochastic vector $\sigma_t$ of relevant pricing factors, with $d \geq 2$. The vector $\sigma_t$ can include the daily volatility of the stock return, the stock dividend yield and the discount rate. To simplify, we refer generically to $\sigma_t$ as the volatility factor. We collect the state variables in the vector $X_t := [r_t\ \sigma_t']'$. The filtration generated by the process $(X_t)$ represents the flow of information available to the investor and coincides with the filtration generated by the sequence of $[S_t\ \sigma_t']'$, given the initial asset value $S_0$.

**Assumption 1.** *Under the physical probability measure $\mathscr{P}$, the process $(X_t)$ is stationary, time-homogeneous and Markov of order $1$ in $\mathcal{X} = \mathcal{R} \times \mathcal{S} \subset \mathbb{R} \times \mathbb{R}^{d-1}$ with transition density $f(x_t|x_{t-1})$.*

When the underlying asset volatility is included in vector $\sigma_t$, Assumption 1 is compatible with the usual discrete time stochastic volatility models and multivariate volatility factor models.[13] Assumption

[13] In a standard discrete time one-factor stochastic volatility model $\sigma_t$ is a scalar ($d = 2$) and represents the volatility of the stock return. We have $r_t = \mu(\sigma_t) + \sigma_t \varepsilon_t$ , $\sigma_t = a(\sigma_{t-1}, u_t)$, where $[\varepsilon_t\ u_t]' \sim IIN\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. This model allows for a leverage effect through the contemporaneous correlation $\rho$ between the shocks on the geometric return and volatility of the stock, and is compatible with Assumption 1. Markov processes of order $m > 1$ for the volatility $\sigma_t$ are compatible with Assumption 1 if we extend the state variables vector as $X_t := [r_t\ \sigma_t\ \ldots\ \sigma_{t-m+1}]'$ and $d = m + 1$.

1 allows for both a contemporaneous leverage effect, through the dependence between $r_t$ and the underlying asset volatility conditional on $X_{t-1}$, and a lagged leverage effect, through the dependence of the underlying asset volatility on $r_{t-1}$. Since the state variables are assumed observable by the econometrician, the underlying asset volatility has to be replaced by an observable proxy such as a realized volatility measure (see Broadie, Detemple, Ghysels and Torrés [2000a]). Note that the underlying asset return $r_t$, and not its price $S_t$, is included in the state variables vector $X_t$ since we invoke stationarity and ergodicity conditions for $X_t$ to prove consistency and asymptotic normality of the estimators in Sections 2.3 and 2.4.

We assume that the prices of all traded assets are compatible with a (not necessarily unique) risk-neutral probability measure $\mathscr{Q}$ associated with a SDF (Hansen and Richard [1987] and Gouriéroux and Monfort [2007]) satisfying the next Assumption 2.

**Assumption 2.** *The one-day SDF $M_{t,t+1}$ between date $t$ and date $t+1$ is a function of the value of the state variables at date $t+1$, i.e. $M_{t,t+1} = m(X_{t+1})$.*

Under Assumptions 1 and 2 the sequence of $X_t$ is a time-homogeneous Markov process of order $1$ also under the risk-neutral probability measure $\mathscr{Q}$.

For expository purpose, in Sections 2.1.2-2.4 we consider null risk-free rate and dividend yield on the stock. In this case, the American option price is equivalent to the price of an European option written on the same underlying and with the same contract characteristics. We do not use this equivalence to derive our theoretical results. They can be extended to stochastic risk-free rate and dividend yield by including them in vector $\sigma_t$ and considering cum-dividend stock returns. We use a constant non-zero risk-free rate in Section 2.5 for our Monte Carlo experiment.

### 2.1.2 The American put options

Let us consider an American put option with payoff at exercise $(K-S)^+ := \max[K-S, 0]$ for strike price $K > 0$.[14] By the principle of dynamic programming and Assumption 1, the price $V_t(h, K)$ at date $t$ of the American put option with time-to-maturity $h$ and strike price $K$ is such that

$$V_t(h,K) = \begin{cases} \max\left[(K-S_t)^+, \mathrm{E}_t^{\mathscr{Q}}\left[V_{t+1}(h-1,K)\right]\right], \text{ for } h>0, \\ (K-S_t)^+, \text{ for } h=0, \end{cases} \tag{2.1}$$

[14] The results in this chapter extend to options with payoff at exercise $\phi(S_t, K)$ that is linearly homogeneous w.r.t. the stock price, i.e., $\phi(S_t, K) = S_t\phi(1, K/S_t)$. For instance, an American chooser option has payoff at exercise $\phi(S_t, K) = \max\left[(K-S_t)^+, (S_t-K)^+\right]$. When the homogeneity property is not satisfied, the approach in this chapter adapts by defining $Y_t := [S_t\ X_t']'$ in Equation (2.2). Moreover, when the option is written on a different underlying than stocks, such as volatility options, this underlying plays the role of the fundamental asset in this chapter.

where $\mathrm{E}_t^{\mathscr{Q}}[.]$ denotes the conditional expectation operator under the risk-neutral probability measure $\mathscr{Q}$ given the investors' information at date $t$. The quantities $(K - S_t)^+$ and $\mathrm{E}_t^{\mathscr{Q}}[V_{t+1}(h-1, K)]$ are the early exercise payoff (or intrinsic value) and the continuation (or holding) value of the option at date $t$, respectively. The former is the value of the option if it is exercised at date $t$, the latter if it is not. The American option price is the maximum between the intrinsic and continuation values. Equation (2.1) corresponds to the value iteration algorithm (see Carriere [1996] and Tsitsiklis and Van Roy [2001]).

Since the state variables vector $X_t$ does not include the stock price $S_t$ while the option exercise payoff is written on $S_t$, we have to augment the state space for the option valuation. More specifically, for a given strike $K > 0$ let us introduce the process of the moneyness strike $k_t := K/S_t$ associated with $S_t$. From Assumptions 1 and 2, the process of the variable

$$Y_t := [k_t \; X_t']' \tag{2.2}$$

in $\mathcal{Y} := \mathbb{R}_+ \times \mathcal{X}$ is time-homogeneous and Markov of order 1 under both $\mathscr{P}$ and $\mathscr{Q}$. Its transition law is independent of the strike $K$ under both $\mathscr{P}$ and $\mathscr{Q}$. By the Markovianity of process $(Y_t)$ under $\mathscr{Q}$, we deduce the next Proposition 1, which states an homogeneity property of the American option price similar to Merton [1973a] and Merton [1990].[15]

**Proposition 1.** *Under Assumptions 1 and 2, the American put option price $V_t(h, K)$ is a linearly homogeneous function of the underlying asset price:*

$$V_t(h, K) = S_t v(h, Y_t),$$

*where the American put option-to-stock price ratio $v$ is such that*

$$v(h, y_t) = \begin{cases} \max\left[(k_t - 1)^+, \mathrm{E}^{\mathscr{Q}}\left[\left.\dfrac{k_t}{k_{t+1}} v(h-1, Y_{t+1})\right| Y_t = y_t\right]\right], & \text{for } h > 0, \\ (k_t - 1)^+, & \text{for } h = 0, \end{cases} \tag{2.3}$$

*for any $y_t = [k_t \; x_t']' \in \mathcal{Y}$, and $\mathrm{E}^{\mathscr{Q}}[.| Y_t = y_t]$ denotes the conditional expectation under the risk-neutral probability measure $\mathscr{Q}$ given $Y_t = y_t$.*

*Proof.* See Appendix 4.2. □

From Proposition 1, the American put option-to-stock price ratio $V_t(h, K)/S_t$ is a function of the time-

[15]Theorem 9 in Merton [1973a] and Theorem 8.6 in Merton [1990] show that the American call price function is homogeneous of degree 1 in the stock price and in the strike price, when the stock returns are independent and identically distributed or follow an autonomous diffusion process, respectively.

to-maturity $h$, the moneyness strike $k_t$ and the state variables vector $X_t$ only. Since the risk-neutral transition law of the Markov process $(Y_t)$ is independent of strike $K$, the option-to-stock price ratio at date $t$ is independent of $K$ when the moneyness strike $k_t$ is given. Thus, the homogeneity property in Proposition 1 reduces the dimensionality of the valuation problem, since function $v(h, .)$ gives the option-to-stock price ratio at time-to-maturity $h$ for any strike $K$, stock price $S$ and state variables vector $X$.[16] Finally, the daily stock gross return $k_t/k_{t+1} = S_{t+1}/S_t$ in the conditional expectation in Equation (2.3) accounts for the fact that we consider option-to-stock price ratios.

The function $v$ determines the optimal exercise policy. More precisely, the continuation region at time-to-maturity $h \geq 1$ is defined as

$$\mathcal{C}(h) := \left\{ y = [k\ x']' \in \mathcal{Y} : v(h, y) > (k-1)^+ \right\}. \tag{2.4}$$

Equivalently, the continuation region $\mathcal{C}(h)$ is the set of pairs of moneyness strike $k_t$ and state variable value $X_t$ for which the holding-to-stock price ratio $u(h, y_t) := \mathrm{E}^{\mathcal{Q}}\left[ \frac{k_t}{k_{t+1}} v(h-1, Y_{t+1}) \middle| Y_t = y_t \right]$ is strictly larger than the exercise-to-stock price ratio $(k_t - 1)^+$. The set-theoretical complement of $\mathcal{C}(h)$ in $\mathcal{Y}$ is the exercise (or stopping) region. The frontier between the two regions is known as exercise boundary, and the values assumed by $y$ on this frontier are called critical values.

### 2.1.3 The American put pricing operator

Following Proposition 1 we compute the American put option-to-stock price ratio $v(h, y)$ recursively backward w.r.t. the time-to-maturity $h$. This recursion can be expressed in terms of a pricing operator acting on $L_2(\mathcal{Y})$, that is the linear space of functions $\varphi$ on $\mathcal{Y}$ such that $\int_{\mathcal{Y}} \varphi(y)^2 \frac{f_X(x)}{k^2} dy < \infty$, where $f_X$ denotes the stationary density of $X_t$.[17]

**Definition 1.** *The American put pricing operator $\mathcal{A} : L_2(\mathcal{Y}) \to L_2(\mathcal{Y})$ applied to a payoff-to-stock price ratio $\varphi \in L_2(\mathcal{Y})$ and evaluated at $y_t = [k_t\ x_t']' \in \mathcal{Y}$ is*

$$\mathcal{A}[\varphi](y_t) := \max\left[ (k_t - 1)^+, \mathrm{E}^{\mathcal{Q}}\left[ \frac{k_t}{k_{t+1}} \varphi(Y_{t+1}) \middle| Y_t = y_t \right] \right].$$

The linear operator that maps $\varphi(y_t)$ into $\mathrm{E}^{\mathcal{Q}}\left[ \frac{k_t}{k_{t+1}} \varphi(Y_{t+1}) \middle| Y_t = y_t \right]$ is the one-day adjusted condi-

[16] While following Merton [1973a] and Merton [1990] we consider option-to-stock price ratios, lowering the problem dimensionality by considering price-to-strike ratios is quite common in the American option pricing literature (see e.g. Wilmott, Howison and Dewynne [1993] for the Black-Scholes setting, and Broadie, Detemple, Ghysels and Torrés [2000a] and Broadie, Detemple, Ghysels and Torrés [2000b] for a diffusion setting).

[17] We prove that the American put pricing operator maps $L_2(\mathcal{Y})$ into itself in Appendix 4.3. See Peskir and Shiryaev [2006], p.15, for a similar operator representation of the Wald-Bellman equations.

tional expectation operator for Markov process $(Y_t)$ under the risk-neutral probability measure $\mathscr{Q}$ given $Y_t = y_t$. This operator acts on a payoff-to-stock price ratio and returns another payoff-to-stock price ratio. By a change of variable and Assumption 2, we can rewrite this operator through the historical transition density of $X_t$ and the SDF:

$$\mathrm{E}^{\mathscr{Q}}\left[\frac{k_t}{k_{t+1}}\varphi(Y_{t+1})\middle| Y_t = y_t\right] = \int_{\mathcal{X}} m(x)e^r\varphi(k_t e^{-r}, x)f(x|x_t)dx, \quad y_t \in \mathcal{Y}. \tag{2.5}$$

From Proposition 1 the option-to-stock price ratio function satisfies the backward recursion:

$$v(h, y_t) = \mathcal{A}[v(h-1, .)](y_t), \tag{2.6}$$

with value at maturity $v(0, y_t) = (k_t - 1)^+$. Thus, we get

$$v(h, y_t) = \mathcal{A}^h[v(0, .)](y_t), \text{ for all } h \in \mathbb{N}, \tag{2.7}$$

where $\mathcal{A}^h$ denotes the $h$-fold application of operator $\mathcal{A}$.

## 2.2 A semi-parametric option pricing model

Building on the framework of Section 2.1, we now introduce a semi-parametric option pricing model. We consider the parametrization of the SDF in Section 2.2.1 and describe the restrictions on the parameters induced by the no-arbitrage assumption in Section 2.2.2. Finally in Section 2.2.3 we derive the sensitivity of the American option-to-stock price ratios to a change in the model parameters.

### 2.2.1 The historical and risk neutral parameters

The SDF is parametrized by a finite-dimensional parameter, while the historical transition density $f$ of process $X_t$ in Assumption 1 is left unconstrained.

**Assumption 3.** *The single-day SDF $M_{t,t+1}$ between date $t$ and date $t+1$ is a function of the unknown parameter vector $\theta_0 \in \Theta$, i.e. $M_{t,t+1} = m(X_{t+1}; \theta_0)$, where $m$ is a known function and $\Theta \subset \mathbb{R}^p$ is the SDF parameter set.*

The parameter vector $\theta$ includes the risk premia associated with the priced risk factors. In an incomplete market framework, a multiplicity of admissible SDF's may exist. Here we implicitly assume that only one valid SDF admits the parametric specification in Assumption 3. This is made explicit by the identification conditions for parameter $\theta$ in Section 2.4 (see Assumptions 5 and 7).

From Equation (2.5) and Assumption 3 the pricing operator $\mathcal{A}$ in Definition 1 involves both the finite-dimensional parameter $\theta$ and the infinite-dimensional parameter $f$. We denote by $\mathcal{A}_{\theta,f}$ the pricing operator $\mathcal{A}$ defined for generic parameters $\theta$ and $f$. This operator yields a semi-parametric pricing model for American options through Equation (2.7). The goal is to estimate the true SDF parameter $\theta_0$ and the true historical transition density $f_0$. Then, by the plug-in principle, we can estimate the American put option-to-stock price ratio $\mathcal{A}^{h^*}_{\theta_0,f_0}[v(0,.)](k^*, x_{t_0})$ at the current date $t_0$ for any given moneyness strike $k^*$ and time-to-maturity $h^*$, as well as other functionals of interest that depend on the true parameters $(\theta_0, f_0)$.

### 2.2.2 The no-arbitrage restrictions

The true values $\theta_0$ and $f_0$ of the model parameters are estimated from the information contained in the no-arbitrage restrictions implied by the market price data. The data consist of two sets of observations. First, we have at the current date $t_0$ a sample of $N$ cross-sectionally observed trading prices of American put options with times-to-maturity $h_j$ and moneyness strikes $k_j := k_{t_0,j}$, where $j = 1, \dots, N$. The corresponding option-to-stock price ratios are denoted by $v_j$, for $j = 1, \dots, N$. Second, we have a sample of $T$ historical observations $x_t$, where $t = t_0 - T + 1, \dots, t_0$, for the state variables vector previous to date $t_0$.

The observational design for the options reflects the common practice of cross-sectional calibration. This practice conveniently accounts for the fact that the set of actively traded options typically changes from one trading day to the next one. The results of this chapter can be extended to include a few cross-sections of observed option prices with minor modifications. The extension of the asymptotic analysis to include a full panel of observed option prices at every trading day in the sample is more difficult because of the time-varying random characteristics of the actively traded options (time-to-maturity and moneyness strike) and is beyond the scope of this chapter.

The one-day no-arbitrage restrictions on the underlying stock and on the short-term non-defaultable bond are

$$\begin{cases} \mathrm{E}_0\left[m(X_{t+1};\theta_0)e^{r_{t+1}} \middle| X_t = x\right] = 1, \\ \\ \mathrm{E}_0\left[m(X_{t+1};\theta_0) \middle| X_t = x\right] = 1, \end{cases} \qquad \text{for almost every (a.e.) } x \in \mathcal{X}, \tag{2.8}$$

respectively, where $\mathrm{E}_0\left[\cdot \middle| X_t = x\right]$ denotes the conditional expectation under the true historical probability measure given $X_t = x$. The conditional moment restrictions (2.8) are valid uniformly in the conditioning value of the state variables vector. We refer to them as uniform capital market restrictions.

The no-arbitrage restrictions on the cross-sectionally observed American option prices at date $t_0$, that we call derivative market restrictions, are given by

$$g(\theta_0, f_0) = 0, \tag{2.9}$$

where the vector function $g = [g_1 \ \dots \ g_N]'$ with argument $(\theta, f)$ is defined by

$$g_j(\theta, f) := \mathcal{A}^{h_j}_{\theta,f}[v(0,.)](y_j) - v_j, \qquad j = 1, \cdots, N, \tag{2.10}$$

for $y_j := [k_j \ x_0']'$ and $x_0 := x_{t_0}$. The derivative market restrictions (2.9) are not moment restrictions, since we cannot write them as an expectation under $f_0$ of a known function of the unknown parameter $\theta_0$ and the data. Indeed, the restriction vector $g$ depends nonlinearly on $f$ because of the multi-day nature of the constraints and the exercise decision embodied in the pricing operator. Moreover the derivative market restrictions (2.9) are local in nature, holding for the value $x_0$ of the state variables vector at date $t_0$ only. These features explain why our framework differs from the standard GMM setting (Hansen [1982] and Hansen and Singleton [1982]) as well as from the XMM setting (Gagliardini, Gouriéroux and Renault [2011]).

The total set of no-arbitrage restrictions is given by System (2.8) and Equation (2.9). For the definition and interpretation of the estimators in Section 4, it is useful to rewrite these restrictions in an equivalent form. Since the $N$ traded options at $t_0$ are in the continuation region, their prices equal the holding values. Thus, by using Definition 1 and Equation (2.5), the restrictions (2.8) and (2.9) can be rewritten as

$$\begin{cases} \mathrm{E}_0\left[\Gamma_U(X_{t+1};\theta_0)|X_t = x\right] = 0, \quad \text{for a.e. } x \in \mathcal{X}, \\ \\ \mathrm{E}_0\left[\gamma_S(X_{t+1};\theta_0, f_0)|X_t = x_0\right] = 0, \end{cases} \tag{2.11}$$

where

$$\Gamma_U(x;\theta) := m(x;\theta)[e^r \ \ 1]' - [1 \ \ 1]' \tag{2.12}$$

and the vector function $\gamma_S = [\gamma_{S,1} \ \dots \ \gamma_{S,N}]'$ is defined as

$$\gamma_{S,j}(x;\theta,f) := m(x;\theta)\gamma_{1,j}(x;\theta,f) - v_j, \qquad \gamma_{1,j}(x;\theta,f) := e^r \mathcal{A}^{h_j-1}_{\theta,f}[v(0,.)](k_j e^{-r}, x), \tag{2.13}$$

for $j = 1, \cdots, N$ and any $x \in \mathcal{X}$. Vector $\Gamma_U$ is the moment function for the capital market restrictions. Vector $\gamma_S$ defines a short-term quasi moment function for the derivative market restrictions. Vector $\gamma_S$ is not a feasible moment function since, when $h_j > 1$ for some option $j$, it involves the unknown transition density $f_0$ through $\gamma_{1,j}$, that is, the one-day ahead price of option $j$ in units of the current underlying asset price. We could consider $\gamma_S$ as a moment function involving both a finite-dimensional parameter $\theta$ and an infinite-dimensional parameter $f$ as in Ai and Chen [2003]. However, their estimation approach cannot be applied here since the restriction is local and not uniform w.r.t. the conditioning value.

### 2.2.3 Sensitivity of the derivative market constraints to the model parameters

The informational content of the derivative market restrictions (2.9) depends on the sensitivity of vector function $g$ to an infinitesimal change in the parameters $\theta$ and $f$. In Proposition 2 below we compute the gradient $\nabla_{\theta'} g_j$ of function $g_j$ w.r.t. the finite-dimensional parameter $\theta$, and the Fréchet derivative of $g_j$ w.r.t. the infinite-dimensional parameter $f$, for $j = 1, \ldots, N$. The Fréchet derivative of function $g_j(\theta, \cdot)$ at $f$ in the direction $\Delta f$, denoted by $\langle Dg_j(\theta, f), \Delta f\rangle$, measures the first-order variation of $g_j(\theta, \cdot)$ when we perturb the transition density from $f$ to $f + \Delta f$, holding parameter $\theta$ fixed. Hence

$$g_j(\theta, f + \Delta f) = g_j(\theta, f) + \langle Dg_j(\theta, f), \Delta f\rangle + O\left(\|\Delta f\|_\infty^2\right), \tag{2.14}$$

where $\|\Delta f\|_\infty$ denotes the supremum norm of $\Delta f$ (see e.g. Ichimura and Todd [2007] for the use of the Fréchet derivative in nonparametric and semi-parametric methods).

**Proposition 2.** *Let parameters $(\theta, f)$ satisfy the no-arbitrage restrictions $g(\theta, f) = 0$ and $\mathrm{E}_f[\Gamma_U(X_{t+1}; \theta)|X_t = x] = 0$, for a.e. $x \in \mathcal{X}$, where $\mathrm{E}_f[\cdot|X_t = x]$ denotes the expectation w.r.t. the pdf $f(\cdot|x)$. Moreover, assume that $y_j$ is in the interior of the continuation region $\mathcal{C}_{\theta,f}(h_j)$ for time-to-maturity $h_j$ and parameters $(\theta, f)$, for all $j = 1, \ldots, N$. Then, under Assumptions 1-3, and A 2 and A 8 in Appendix 4.1, the Fréchet derivative of $g_j(\theta, \cdot)$ at $f$ in the direction $\Delta f$ is*

$$\begin{aligned}\langle Dg_j(\theta, f), \Delta f\rangle &= \int_{\mathcal{X}} m(x; \theta)\gamma_{1,j}(x; \theta, f)\Delta f(x|x_0)dx \\ &\quad + \int_{\mathcal{X}}\int_{\mathcal{X}} m(x; \theta)\gamma_{2,j}(x, \tilde{x}; \theta, f)\Delta f(x|\tilde{x})dxd\tilde{x},\end{aligned} \tag{2.15}$$

*and the gradient of $g_j$ w.r.t. $\theta$ is*

$$\begin{aligned}\nabla_{\theta'} g_j(\theta, f) &= \mathrm{E}_f\left[\left(\nabla_{\theta'} m(X_{t+1}; \theta)\right)\gamma_{1,j}(X_{t+1}; \theta, f)|X_t = x_0\right] \\ &\quad + \int_{\mathcal{X}} \mathrm{E}_f\left[\left(\nabla_{\theta'} m(X_{t+1}; \theta)\right)\gamma_{2,j}(X_{t+1}, \tilde{x}; \theta, f)|X_t = \tilde{x}\right] d\tilde{x},\end{aligned} \tag{2.16}$$

*for $j = 1, \ldots, N$, where functions $\gamma_{1,j}(x; \theta, f)$ are given in Equation (2.13) and*

$$\gamma_{2,j}(x, \tilde{x}; \theta, f) := e^r \sum_{l=2}^{h_j} f^{\mathscr{Q}}_{\theta,l-1}(\tilde{x}|x_0) \mathrm{E}^{\mathscr{Q}}_{\theta,f} \Bigg[ \mathbf{1}_{\mathcal{C}_{\theta,f(h_j-1)}}(Y_{t+1}) \ldots \mathbf{1}_{\mathcal{C}_{\theta,f(h_j-l+1)}}(Y_{t+l-1}) \cdot \frac{k_t}{k_{t+l-1}} \mathcal{A}^{h_j-l}_{\theta,f}[v(0,.)](k_{t+l-1}e^{-r}, x) \Bigg| X_{t+l-1} = \tilde{x}, Y_t = y_j \Bigg], \tag{2.17}$$

*and where $\mathbf{1}_{\mathcal{C}_{\theta,f(h)}}$ is the indicator of the continuation region for time-to-maturity $h$ and parameters $(\theta, f)$, the conditional expectation $\mathrm{E}^{\mathscr{Q}}_{\theta,f}[.|.]$ is taken under the risk-neutral probability measure of $(Y_t)$ for parameters $(\theta, f)$, and $f^{\mathscr{Q}}_{\theta,l-1}$ is the $(l-1)$-day risk-neutral transition density of $(X_t)$ for parameters $(\theta, f)$.*

*Proof.* See Appendix 4.4. □

The Fréchet derivative in Equation (2.15) involves two components. The first one yields the sensitivity to infinitesimal perturbations $\Delta f(\cdot|x_0)$ of the transition density for the conditioning value $x_0$ of the state variables vector at $t_0$. The second one yields the integrated sensitivity to infinitesimal perturbations $\Delta f(\cdot|\tilde{x})$ of the transition densities for the conditioning values $\tilde{x} \in \mathcal{X}$ of the state variables vector. This decomposition of the Fréchet derivative results from the multi-day nature of the constraint vector $g$ and an application of a functional version of the product rule for differentiation. Indeed, since in Proposition 2 the options are assumed to be in the continuation region at date $t_0$ for parameters $(\theta, f)$, we have

$$g_j(\theta, f) = \int_{\mathcal{X}} m(x; \theta) \gamma_{1,j}(x; \theta, f) f(x|x_0) dx - v_j, \tag{2.18}$$

in a neighborhood of parameters values, for $j = 1, \ldots, N$. Thus, if we hold the transition density $f$ in the normalized future option-to-stock price ratio $\gamma_{1,j}(x; \theta, f)$ fixed, the quantity $g_j(\theta, f)$ is sensitive to an infinitesimal perturbation in parameter $f$ only through the perturbation in the pdf $f(\cdot|x_0)$. The associated short-term sensitivity is measured by function $m \cdot \gamma_{1,j}$, which yields the first term in the RHS of Equation (2.15). The dependence of the normalized future option-to-stock price ratio $\gamma_{1,j}(x; \theta, f)$ on the transition density $f$ explains the second term in the RHS of Equation (2.15). Since $\gamma_{1,j}(x; \theta, f)$ involves a $(h_j - 1)$-fold application of the pricing operator $\mathcal{A}_{\theta,f}$, function $\gamma_{2,j}(x, \tilde{x}; \theta, f)$ in the long-term sensitivity consists of a sum over $h_j - 1$ terms. The term for index $l$, with $2 \leq l \leq h_j$, involves a conditional expectation under the risk-neutral probability measure of the option price at date $t_0 + l$ in units of the stock price at date $t_0$, keeping fixed the state variables vector $x = [r\ \sigma']'$ at date $t_0 + l$. The expectation is w.r.t. the pathes of $Y_t$ that lie in the continuation region between $t_0$ and $t_0 + l - 1$, and is conditional on $X_{t_0+l-1} = \tilde{x}$ and $Y_{t_0} = y_j$. The weight $f^{\mathscr{Q}}_{\theta,l-1}(\tilde{x}|x_0)$ accounts for the risk-neutral likelihood of the state variables vector transition between $x_{t_0} = x_0$ and $x_{t_0+l-1} = \tilde{x}$. Function $\gamma_{2,j}$ is equal to zero if the $j$-th option has time-to-maturity $h_j = 1$.[18]

Finally, the gradient of the local constraint vector $g$ w.r.t. $\theta$ in Equation (2.16) also involves two components, that are a conditional expectation given $X_t = x_0$ and a conditional expectation integrated over the conditioning value $\tilde{x} \in \mathcal{X}$, respectively. These two components come from the application of the product rule for differentiation w.r.t. $\theta$ in the RHS of Equation (2.18).

## 2.3 Semi-parametric estimation

In this section we introduce semi-parametric estimators of the true SDF parameter $\theta_0$, of the true historical transition density $f_0$ and of some of their functionals. To get numerically tractable estimators, we focus on a two-step estimation procedure. It consists in first getting an estimator of the SDF parameter, and then using it to derive an estimator of the historical transition density. We consider a minimum-distance estimator of the SDF parameter that exploits the information in the local no-arbitrage restrictions at the current date only (Section 2.3.1), and another one that exploits the full set of no-arbitrage restrictions (Section 2.3.2). We then introduce an estimator of the transition density that minimizes an information-theoretic criterion subject to the full set of no-arbitrage restrictions (Section 2.3.3). Finally, we introduce an estimator for a class of functionals of $\theta_0$ and $f_0$ that includes the prices of American options (Section 2.3.4).

### 2.3.1 The cross-sectional estimator of the SDF parameter

The estimators we consider require as input nonparametric estimators of the historical transition density of process $(X_t)$ and of its stationary density. For this purpose, we use kernel density estimators. We need some standard assumptions on the serial dependence of process $(X_t)$ (see Bosq [1998]).

**Assumption 4.** *Under the physical probability measure $\mathscr{P}$, the process $(X_t)$ is geometrically strong mixing, that is, the $\alpha$-mixing coefficients $\alpha_j$, for $j \in \mathbb{N}$, are such that $\alpha_j = O(\varrho^j)$, as $j \to \infty$, for a scalar $\varrho \in (0, 1)$.*

Under Assumption 4 the serial dependence between $X_t$ and $X_{t-j}$, for $j \in \mathbb{N}$, decays geometrically fast as the lag $j$ increases. Assumption 4 is satisfied by a wide class of commonly used discrete-time processes (see e.g. Carrasco and Chen [2002]) and discretely-sampled continuous time diffusion

[18] The $\max$ operator in $\mathcal{A}$ does not prevent differentiability of $g(\theta, .)$. Indeed, the kinks induced by the exercise decisions at $t_0 + l - 1$, for $2 \leq l \leq h_j$, are smoothed by a subsequent application of the conditional expectation operator (see the proof of Proposition 2 in Appendix 4.4), while the kink for the exercise decision at $t_0$ is irrelevant as long as the option is in the continuation region at $t_0$.

processes (see e.g. Chen, Hansen and Carrasco [2009] and references therein). The kernel estimator of the historical transition density of process $(X_t)$ is

$$\hat{f}(x|\tilde{x}) := \frac{1}{h_T^d} \sum_{t=2}^{T} K\left(\frac{x_t - x}{h_T}\right) K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right) \Big/ \sum_{t=2}^{T} K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right) \tag{2.19}$$

and the kernel estimator of the historical stationary density $f_X$ is

$$\hat{f}_X(x) := \frac{1}{T h_T^d} \sum_{t=1}^{T} K\left(\frac{x_t - x}{h_T}\right), \tag{2.20}$$

where $K$ is a $d$-dimensional kernel, $h_T$ is the bandwidth (Bosq [1998]) and we have switched to the simpler notation $x_1 := x_{t_0-T+1}, \ldots, x_T := x_{t_0}$.[19]

The full set of no-arbitrage restrictions at date $t_0$ includes the derivative market restrictions (2.9) and the capital market restrictions (2.8) for the state value $x_0$. This set of local restrictions can be written as

$$G(\theta_0, f_0) = 0, \tag{2.21}$$

where the $(N+2) \times 1$ vector function $G(\theta, f)$ is defined by

$$G(\theta, f) = [g(\theta, f)' \quad \mathrm{E}_f[\Gamma_U(X_{t+1}; \theta)|X_t = x_0]']'. \tag{2.22}$$

We follow the minimum distance principle and estimate parameter $\theta$ by minimizing a quadratic criterion based on the sample counterpart $G(\theta, \hat{f})$ of the local restrictions at date $t_0$. This sample counterpart is defined by replacing the transition density $f$ with the kernel estimator $\hat{f}$ into Equation (2.22). Then, the vector $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_0] := \int_{\mathcal{X}} \Gamma_U(x; \theta) \hat{f}(x|x_0) dx$ is the conditional expectation of the moment function $\Gamma_U(.; \theta)$ w.r.t. the kernel density $\hat{f}(.|x_0)$, while vector $g(\theta, \hat{f})$ involves the empirical American put pricing operator

$$\mathcal{A}_{\theta, \hat{f}}[\varphi](y) = \max\left[(k-1)^+, \mathrm{E}_{\hat{f}}[m(X_{t+1}; \theta) e^{r_{t+1}} \varphi(k e^{-r_{t+1}}, X_{t+1})|X_t = x]\right], \tag{2.23}$$

for $\varphi \in L_2(\mathcal{Y})$ and $y \in \mathcal{Y}$, in which the continuation value is computed as a risk-adjusted conditional expectation under the kernel probability measure.

[19] In the Monte-Carlo experiment in Section 6, the different components of vector $X_t$ are rescaled before applying the common bandwidth $h_T$.

**Definition 2.** *The cross-sectional semi-parametric estimator of the SDF parameter* $\theta_0$ *is*

$$\hat{\theta} := \arg\min_{\theta\in\Theta} \mathcal{Q}_T(\theta), \qquad \mathcal{Q}_T(\theta) := G(\theta, \hat{f})'\Omega_T G(\theta, \hat{f}),$$

*where* $\Omega_T$ *is a positive-definite* $(N+2)\times(N+2)$ *weighting matrix for all* $T$*,* $P$*-a.s.*

The estimator $\hat{\theta}$ yields the SDF parameter that minimizes a weighted sum of squared errors on price ratios at date $t_0$ for the options, the stock and the short-term non-defaultable bond.

### 2.3.2 The XMM estimator of the SDF parameter

The estimator of the SDF parameter introduced in the previous section can be improved by extending the set of calibrated constraints to accommodate both the local restrictions at date $t_0$ and the uniform moment restrictions on the bond and stock at all dates. In this section we build on the Extended Method of Moments (XMM) estimation for efficient pricing of European derivatives developed in Gagliardini, Gouriéroux and Renault [2011] and we introduce a second estimator of the SDF parameter.

**Definition 3.** *The XMM semi-parametric estimator of the SDF parameter* $\theta_0$ *is*

$$\hat{\theta}^* := \arg\min_{\theta\in\Theta} \mathcal{Q}_T^*(\theta),$$

*for the criterion*

$$\mathcal{Q}_T^*(\theta) := h_T^d G(\theta, \hat{f})'\Omega_T G(\theta, \hat{f}) + \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_t]'\tilde{\Omega}_T(x_t)\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_t],$$

*where* $\tilde{\Omega}_T(x)$ *is a positive-definite* $2\times 2$ *weighting matrix for all* $T$ *and* $x \in \mathcal{X}$*,* $P$*-a.s., and matrix* $\Omega_T$ *is as in Definition 2.*

The objective function $\mathcal{Q}_T^*$ in Definition 3 involves two components. The first one is a quadratic form in the estimated local no-arbitrage restrictions at date $t_0$. It corresponds to the objective function $\mathcal{Q}_T$ of the cross-sectional estimator in Definition 2 multiplied by $h_T^d$. The second component in $\mathcal{Q}_T^*$ is a sample average of quadratic forms in the vectors $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_t]$ with weighting matrices $\tilde{\Omega}_T(x_t)$, for $t = 1,\ldots,T$. The sample average is over the state variables observations. The vector $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_t]$ is an empirical counterpart of the no-arbitrage restriction vector for the stock and the bond at state variables vector $x_t$, which is asymptotically equivalent to a Nadaraya-Watson kernel regression estimator. Thus, the second component of $\mathcal{Q}_T^*(\theta)$ is similar to the minimum distance criterion introduced in Ai and Chen [2003] to estimate conditional moment restrictions models (see also Nagel and Singleton [2011] for an application to conditional asset pricing models). In the local component of criterion $\mathcal{Q}_T^*(\theta)$ we single out the factor $h_T^d$ to get the asymptotic distribution of estimator $\hat{\theta}^*$ in Section 2.4 when $\Omega_T$ and $\tilde{\Omega}_T(\tilde{x})$ converge to limit positive-definite weighting matrices.

### 2.3.3 The semi-parametric estimator of the historical transition density

Let us now consider the estimation of the historical transition density $f_0$ of the state variables. The nonparametric kernel estimator $\hat{f}$ in Equation (2.19) does not take into account the information contained in the no-arbitrage restrictions. We propose to estimate $f_0$ by the transition density that satisfies the no-arbitrage restrictions and is the closest to $\hat{f}$ in the sense of a particular statistical measure. This measure is based on the Kullback-Leibler divergence of the transition density $f$ from the kernel density estimator $\hat{f}$ for given $\tilde{x} \in \mathcal{X}$, that is defined as

$$d_{KL}(f, \hat{f}|\tilde{x}) := \int_{\mathcal{X}} \log \left( \frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})} \right) f(x|\tilde{x}) dx. \tag{2.24}$$

**Definition 4.** *The semi-parametric estimator of the historical transition density $f_0$ is*

$$\begin{aligned} \hat{f}^* \quad &:= \quad \arg\min_{f \in \mathcal{F}} D_T(f, \hat{f}) \\ & \qquad s.t. \begin{cases} G(\hat{\theta}^*, f) = 0, \\ \mathrm{E}_f[\Gamma_U(X_{t+1}; \hat{\theta}^*)|X_t = x] = 0, \text{ for a.e. } x \in \mathcal{X}, \end{cases} \end{aligned}$$

*where*

$$D_T(f, \hat{f}) := \int_{\mathcal{X}} d_{KL}(f, \hat{f}|x) \hat{f}_X(x) dx + \omega_T d_{KL}(f, \hat{f}|x_0), \tag{2.25}$$

*estimators $\hat{f}$, $\hat{f}_X$ and $\hat{\theta}^*$ are defined in Equations (2.19)-(2.20) and Definition 3, set $\mathcal{F}$ is the set of conditional densities of $X_{t+1}$ given $X_t$ and the weight $\omega_T$ is such that $\omega_T > 0$, $P$-a.s.*

The first component in criterion $D_T$ is the average Kullback-Leibler divergence over $\mathcal{X}$ weighted by the kernel density estimator $\hat{f}_X$. The second component is the local Kullback-Leibler divergence at $x_0$ weighted by $\omega_T$. This local component ensures that the minimization admits a unique solution for $\hat{f}^*(\cdot|x_0)$. The constraints involve both the local and the uniform restrictions, written for the SDF parameter estimate $\hat{\theta}^*$.

Let us now characterize estimator $\hat{f}^*$ in terms of the first-order condition. We start by defining the functional Lagrangian corresponding to the criterion and the restrictions:

$$
\begin{aligned}
\mathcal{L} \;:=\;& D_T(f,\hat{f}) - \omega_T \lambda' g(\hat{\theta}^*, f) - \omega_T \nu_0' \mathrm{E}_f[\Gamma_U(X_{t+1};\hat{\theta}^*)|X_t = x_0] - \omega_T \mu_0 \int_{\mathcal{X}} f(x|x_0)dx \\
& - \int_{\mathcal{X}} \hat{f}_X(x)\nu(x)' \mathrm{E}_f[\Gamma_U(X_{t+1};\hat{\theta}^*)|X_t = x]dx - \int_{\mathcal{X}} \hat{f}_X(\tilde{x})\mu(\tilde{x}) \int_{\mathcal{X}} f(x|\tilde{x})dx d\tilde{x}. \quad (2.26)
\end{aligned}
$$

Vectors $\lambda := [\lambda_1 \;\ldots\; \lambda_N]' \in \mathbb{R}^N$ and $\nu_0 := [\nu_{0,1}\; \nu_{0,2}]' \in \mathbb{R}^2$ are the Lagrange multiplier vectors for the local derivative and capital market restrictions at $t_0$, respectively, while $\nu(.) := [\nu_1(.)\; \nu_2(.)]'$ is a bivariate functional Lagrange multiplier vector for the uniform no-arbitrage restrictions. The scalar $\mu_0$ is a Lagrange multiplier for the local unit mass constraint $\int_{\mathcal{X}} f(x|x_0)dx = 1$ and the Lagrange multiplier scalar function $\mu$ accounts for the unit mass constraint $\int_{\mathcal{X}} f(x|\tilde{x})dx = 1$, that holds for all $\tilde{x} \in \mathcal{X}$. The multipliers $\lambda$, $\nu_0$ and $\mu_0$ in Equation (2.26) are multiplied by the weight $\omega_T$, and functions $\nu$ and $\mu$ by $\hat{f}_X$, to simplify the expressions of the estimators. The differential of the functional Lagrangian $\mathcal{L}$ w.r.t. the historical transition density $f$ is equal to zero at $\hat{f}^*$:

$$
\delta\mathcal{L}\big|_{f=\hat{f}^*} = 0. \quad (2.27)
$$

The differential of the functional Lagrangian is derived in Appendix 4.5 by using Proposition 2. By solving the first-order condition in Equation (2.27), we deduce the next Proposition 3.

**Proposition 3.** *Under Assumptions 1-4, the estimator $\hat{f}^*$ of the historical transition density and the estimators $\hat{\lambda}$, $\hat{\nu}_0$ and $\hat{\nu}(.)$ of the Lagrange multiplier vectors are such that*

$$
\hat{f}^*(x|\tilde{x}) = \begin{cases}
\dfrac{\hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)}{\int_{\mathcal{X}} \hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)dx}, & \text{if } \tilde{x} = x_0, \\
\dfrac{\hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\big/\hat{f}_X(\tilde{x})\right)}{\int_{\mathcal{X}} \hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\big/\hat{f}_X(\tilde{x})\right)dx}, & \text{if } \tilde{x} \neq x_0,
\end{cases} \quad (2.28)
$$

*and*

$$
\begin{cases}
\mathrm{E}_{\hat{f}^*}\left[\gamma_S(X_{t+1};\hat{\theta}^*,\hat{f}^*)\Big| X_t = x_0\right] = 0, \\
\mathrm{E}_{\hat{f}^*}\left[\Gamma_U(X_{t+1};\hat{\theta}^*)\Big| X_t = x\right] = 0, \text{ for a.e. } x \in \mathcal{X},
\end{cases} \quad (2.29)
$$

*where the vector function $\gamma_L$ is defined by*

$$
\gamma_L(x,\tilde{x};\theta,f) := m(x;\theta)\cdot[\gamma_{2,1}(x,\tilde{x};\theta,f)\;\ldots\;\gamma_{2,N}(x,\tilde{x};\theta,f)]', \quad (2.30)
$$

*for functions $\gamma_{2,j}$ in Equation (2.17).*

*Proof.* See Appendix 4.5. $\square$

The estimator $\hat{f}^*$ of the historical transition density in Proposition 3 is an exponential tilting transformation of the kernel estimator $\hat{f}$. When the conditioning value for the historical transition density is $x_0$, the tilting in Equation (2.28) involves the moment function $\Gamma_U$ of the uniform capital market restrictions as well as the vector $\gamma_S$ with the short-term components of the Frechét derivatives of the constraints for the options. Otherwise, the tilting involves moment vector $\Gamma_U$ and vector $\gamma_L$, which is the analogue of vector $\gamma_S$ for the long-term components of the Frechét derivatives of the constraints for the options. The constraints in System (2.29) are empirical counterparts of the constraints in System (2.11). Moreover, the estimator $\hat{f}^*$ is defined implicitly by Equation (2.28) and System (2.29). Indeed, the vector functions $\gamma_S$ and $\gamma_L$ involve the estimator $\hat{f}^*$ itself. Proposition 3 extends the results in Kitamura and Stutzer [1997] and Kitamura, Tripathi and Ahn [2004], where information based GMM estimators for models with unconditional, respectively conditional, moment restrictions are considered. In these articles, the tilting function involves the orthogonality function defining the (conditional) moment restrictions, which is independent of the transition $f$.

Proposition 3 suggests an iterative algorithm to compute numerically estimator $\hat{f}^*$ and the estimators $\hat{\lambda}$, $\hat{\nu}_0$ and $\hat{\nu}(\cdot)$ of the Lagrange multipliers. The algorithm is as follows:

**I)** In a preliminary step, we select an initial consistent estimator for $f$, e.g. $\hat{f}^{*(0)} = \hat{f}$ based on $\hat{\lambda}^{(0)} = 0$, $\hat{\nu}_0^{(0)} = 0$ and $\hat{\nu}^{(0)} = 0$.

**II)** We compute functions $\gamma_S(x; \hat{\theta}^*, \hat{f}^{*(0)})$ and $\gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^{*(0)})$.

**III)** We compute $\hat{\lambda}^{(1)}$ and $\hat{\nu}_0^{(1)}$ as

$$\left[\hat{\lambda}^{(1)\prime}\ \hat{\nu}_0^{(1)\prime}\right]' = \arg\min_{\lambda,\nu_0} \log \mathrm{E}_{\hat{f}} \left[ \exp\left( \nu_0' \Gamma_U(X_{t+1}; \hat{\theta}^*) + \lambda' \gamma_S(X_{t+1}; \hat{\theta}^*, \hat{f}^{*(0)}) \right) \middle| X_t = x_0 \right].$$

**IV)** We compute $\hat{\nu}^{(1)}(\tilde{x})$ for any $\tilde{x} \neq x_0$ as

$$\hat{\nu}^{(1)}(\tilde{x}) = \arg\min_{\nu} \log \mathrm{E}_{\hat{f}} \left[ \exp\left( \nu' \Gamma_U(X_{t+1}; \hat{\theta}^*) + \frac{\omega_T}{\hat{f}_X(\tilde{x})} \hat{\lambda}^{(1)\prime} \gamma_L(X_{t+1}, X_t; \hat{\theta}^*, \hat{f}^{*(0)}) \right) \middle| X_t = \tilde{x} \right].$$

**V)** We derive an updated estimator $\hat{f}^{*(1)}$ for $f$ from Equation (2.28) using $\hat{\lambda}^{(1)}$, $\hat{\nu}_0^{(1)}$ and $\hat{\nu}^{(1)}$.

**VI)** We repeat steps **II)-V)**, by replacing $\hat{f}^{*(0)}$, $\hat{\lambda}^{(0)}$, $\hat{\nu}_0^{(0)}$, $\hat{\nu}^{(0)}$ with $\hat{f}^{*(1)}$, $\hat{\lambda}^{(1)}$, $\hat{\nu}_0^{(1)}$, $\hat{\nu}^{(1)}$, and then iterate the algorithm until convergence.

The steps III) and IV) are similar to the computation of the Lagrange multipliers in information theoretic estimation of moment restrictions models (see e.g. Kitamura and Stutzer [1997] and Kitamura, Tripathi and Ahn [2004]). The Lagrange multipliers $(\hat{\lambda}, \hat{\nu}_0)$ and $\hat{\nu}$ are updated sequentially to ease the computation. The proof of the numerical convergence of this algorithm is beyond the scope of this chapter. In the Monte Carlo experiment in Section 2.5 we observe convergence after a few iterations in most of the replications.

The estimator defined in Proposition 3 can be extended to the case where $\omega_T = 0$, that is, when the local component in criterion (2.25) gets a zero weight. In such a case, the estimator in Systems (2.28)-(2.29) admits a simple interpretation. Estimate $\hat{f}^*(\cdot|\tilde{x})$ is the conditional density which is the closest to the kernel estimator $\hat{f}(\cdot|\tilde{x})$ in terms of distance $d_{KL}(\cdot,\cdot|\tilde{x})$ and satisfies the capital and derivative market restrictions at $\tilde{x}$, if $\tilde{x} = x_0$, and the capital market restrictions at $\tilde{x}$, otherwise.[20] The computation of the estimated conditional densities at different conditioning points $\tilde{x}$ can be done separately.

Finally, while our two-step approach may yield asymptotically inefficient estimates, the joint optimization w.r.t. $\theta$ and $f$ combined with the grid methods used to evaluate the constraint vector (see Section 2.5.2) is numerically challenging.

### 2.3.4 The estimators of functionals of the historical transition density

By the plug-in principle, the estimators $\hat{\theta}^*$ and $\hat{f}^*$ in Definitions 3 and 4 can be used to introduce semi-parametric estimators for a class of $\mathbb{R}^r$-valued Fréchet differentiable functionals $a$ of the SDF parameter $\theta$ and the historical transition density $f$. These functionals are characterized by the first-order expansion around the true parameters value $(\theta_0, f_0)$:

$$a(\theta, f) = a(\theta_0, f_0) + \nabla_{\theta'} a(\theta_0, f_0)\,(\theta - \theta_0) + \langle Da(\theta_0, f_0), \Delta f\rangle + O\left(\|\Delta f\|_\infty^2 + \|\theta - \theta_0\|^2\right), \tag{2.31}$$

for $\Delta f = f - f_0$, such that the Fréchet derivative of $a(\theta_0, .)$ w.r.t. $f$ in direction $\Delta f$ at $f_0$ can be written in the form

$$\langle Da(\theta_0, f_0), \Delta f\rangle = \int_{\mathcal{X}} \alpha_S(x)\Delta f(x|x^\star)dx + \int_{\mathcal{X}} f_X(\tilde{x}) \int_{\mathcal{X}} \alpha_L(x, \tilde{x})\Delta f(x|\tilde{x})dx d\tilde{x}, \tag{2.32}$$

for some given state variables vector $x^\star \in \mathcal{X}$ and $\mathbb{R}^r$-valued functions $\alpha_S$ and $\alpha_L$.

[20]This estimator corresponds to a particular solution of the minimization problem in Definition 4.

**Definition 5.** *The semi-parametric estimator of the true value $a_0 := a(\theta_0, f_0)$ of the $\mathbb{R}^r$-valued functional $a$ is defined as $\hat{a}^* := a(\hat{\theta}^*, \hat{f}^*)$, where $\hat{\theta}^*$ is given in Definition 3 and $\hat{f}^*$ in Definition 4.*

We exploit Equations (2.31)-(2.32) to derive the large sample properties of estimator $\hat{a}^*$ in Section 2.4.

The class of functionals defined by Equations (2.31)-(2.32) contains several functionals of interest for financial applications. We provide three examples for which we characterize functions $\alpha_S$ and $\alpha_L$.

**i) The American put option-to-stock price ratio**

From Equation (2.7) we write the American put option-to-stock price ratio for given time-to-maturity $h^\star$, moneyness strike $k^\star$ and state variables vector $x^\star$ as

$$a(\theta, f) = \mathcal{A}_{\theta,f}^{h^\star}[v(0,.)](y^\star), \qquad y^\star = [k^\star \; x^{\star\prime}]'.$$

Proposition 2 shows that this functional satisfies Equations (2.31) and (2.32) with

$$\alpha_S(x) = m(x;\theta_0)\gamma_1^\star(x;\theta_0,f_0), \quad \alpha_L(x,\tilde{x}) = m(x;\theta_0)\gamma_2^\star(x,\tilde{x};\theta_0,f_0)/\, f_X(\tilde{x}), \tag{2.33}$$

where functions $\gamma_1^\star$ and $\gamma_2^\star$ are defined as $\gamma_{1,j}$ and $\gamma_{2,j}$ in Equations (2.13) and (2.17) by setting $j = 1$, $h_1 = h^\star$ and $y_1 = y^\star$. Then, Definition 5 gives the estimator of the American put option-to-stock price ratio. The continuation value is computed through a nonparametric regression w.r.t. the transition density $\hat{f}^*$ adjusted for risk by means of the SDF $m(\cdot;\hat{\theta}^*)$.

**ii) The exercise boundary**

For given time-to-maturity $h^\star$ and state variables vector $x^\star$, the critical moneyness $k_{\theta,f}^\star$ is the solution of the equation $\mathcal{A}_{\theta,f}^{h^\star}[v(0,.)](k_{\theta,f}^\star, x^\star) = \left(k_{\theta,f}^\star - 1\right)^+$ and depends on $(\theta, f)$. This defines a functional $a(\theta,f) = k_{\theta,f}^\star$, which satisfies Equations (2.31) and (2.32) with

$$\alpha_S(x) = \frac{m(x;\theta_0)\gamma_1^\star(x;\theta_0,f_0)}{1 - \nabla_k v(h^\star, y^\star)}, \quad \alpha_L(x,\tilde{x}) = \frac{m(x;\theta_0)\gamma_2^\star(x,\tilde{x};\theta_0,f_0)}{(1 - \nabla_k v(h^\star,y^\star)) f_X(\tilde{x})},$$

where functions $\gamma_1^\star$ and $\gamma_2^\star$ are as in Equations (2.33) and $y^\star = [k_{\theta_0,f_0}^\star \; x^{\star\prime}]'$. By considering the estimator $a(\hat{\theta}^*, \hat{f}^*)$ for different values of $x^\star$, we get an estimator of the critical region.

**iii) Term structure of conditional historical and risk-neutral moments**

Let $\psi(X_{t+h^\star};\theta)$ be a function of the state variables at horizon $h^\star$ and of the SDF parameter. Let us consider the functional defined by

$$a(\theta, f) = \mathrm{E}_f\left[\psi(X_{t+h^\star};\theta) | X_t = x^\star\right].$$

The $h^{\star}$-day conditional expectation in the RHS involves the one-day transition density $f$ only because of the Law of Iterated Expectations and the Markov property. Functional $a$ satisfies Equations (2.31) and (2.32) with

$$\alpha_S(x) = \mathrm{E}_0\left[\psi(X_{t+h^\star};\theta_0)|X_{t+1}=x\right], \; \alpha_L(x,\tilde{x}) = \sum_{l=2}^{h^\star} \mathrm{E}_0\left[\psi(X_{t+h^\star};\theta_0)|X_{t+l}=x\right] \frac{f_{X_{t+l-1}|X_t}(\tilde{x}|x^\star)}{f_X(\tilde{x})}.$$

The historical conditional moment generating function corresponds to $\psi(X_{t+h^\star};\theta) = \exp\left(ur_{t+h^\star} + v'\sigma_{t+h^\star}\right)$, with $u \in \mathbb{R}$ and $v \in \mathbb{R}^{d-1}$. The historical first conditional moments and cross-moments of the one-day stock return and volatility factor correspond to $\psi(X_{t+h^\star};\theta) = r_{t+h^\star}^m \sigma_{t+h^\star}^n$, with $m \in \mathbb{N}$ and multi-index $n \in \mathbb{N}^{d-1}$. The risk-neutral counterparts of these functionals are obtained when the functions $\psi(X_{t+h^\star};\theta)$ above are multiplied by the multi-day SDF $M_{t,t+h^\star} = M_{t,t+1}\cdots M_{t+h^\star-1,t+h^\star}$. In particular, when the underlying asset volatility is included in vector $\sigma_t$, the conditional historical (resp. risk-neutral) cross-moments are the basis for the estimation of the conditional historical (resp. risk-neutral) leverage effects.

## 2.4 Large sample properties of the estimators

In this section we study the large sample properties of the semi-parametric estimators introduced in Section 2.3. The asymptotics is for a long time-series of observations of the state variables, i.e. $T \to \infty$, and a fixed number $N$ of cross-sectionally observed option prices. We use the following notation:

$$\begin{aligned}
&\bar{\Gamma}_U(x) := m(x;\theta_0)[e^r \;\; 1]', \qquad \bar{\gamma}_S(x) := m(x;\theta_0)[\gamma_{1,1}(x;\theta_0,f_0) \; \ldots \; \gamma_{1,N}(x;\theta_0,f_0)]', \\
&\bar{\Gamma}_S(x) := [\bar{\gamma}_S(x)' \;\; \bar{\Gamma}_U(x)']', \qquad \bar{\Gamma}_L(x,\tilde{x}) := [\gamma_L(x,\tilde{x};\theta_0,f_0)' \;\; 0 \;\; 0]'/f_X(\tilde{x}), \\
&\Gamma_S(x) := [\gamma_S(x;\theta_0,f_0)' \;\; \Gamma_U(x;\theta_0)']', \qquad \Gamma_L(x,\tilde{x}) := \bar{\Gamma}_L(x,\tilde{x}) - \mathrm{E}_0\left[\bar{\Gamma}_L(X_{t+1},X_t)\middle| X_t = \tilde{x}\right],
\end{aligned} \tag{2.34}$$

where functions $\Gamma_U$, $\gamma_S$, $\gamma_{1,j}$, for $j = 1,...,N$, and $\gamma_L$ are defined in Equations (2.12), (2.13) and (2.30).

### 2.4.1 The cross-sectional estimator of the SDF parameter

Let us consider the cross-sectional estimator $\hat{\theta}$ in Definition 2. Under the regularity conditions in Appendix 4.1, the criterion $\mathcal{Q}_T(\theta)$ converges uniformly to the limit criterion $\mathcal{Q}_0(\theta) = G(\theta,f_0)'\Omega_0 G(\theta,f_0)$, where $\Omega_0 := \operatorname*{plim}_{T\to\infty} \Omega_T$ is a symmetric $(N+2)\times(N+2)$ matrix assumed to be positive-definite. Let us assume the global identification of parameter $\theta_0$ w.r.t. the population constraint vector $G(\theta,f_0)$.

**Assumption 5.** *The unique element $\theta \in \Theta$ such that $G(\theta,f_0) = 0$ is $\theta = \theta_0$.*

Under Assumption 5 the limit criterion $\mathcal{Q}_0$ is uniquely minimized by $\theta_0$. By the consistency theorem for minimum distance estimators (see Theorem 2.1 in Newey and McFadden [1999]) we get the following result.

**Proposition 4.** *Under Assumptions 1-5 and A 1-10 in Appendix 4.1, estimator $\hat{\theta}$ is consistent, i.e. $\hat{\theta} \xrightarrow{P} \theta_0$.*

*Proof.* See Appendix 4.6.1. □

Let us now prove the asymptotic normality of estimator $\hat{\theta}$. The criterion function $\mathcal{Q}_T(\theta)$ is not everywhere differentiable on $\Theta$ because of the maximum operator in $\mathcal{A}_{\theta,\hat{f}}$ (see Equation (2.23)). However, by using Proposition 2, the consistency of kernel estimator $\hat{f}$ and the fact that the $N$ options are in the continuation region at $t_0$, we show in Appendix 4.6.2 that the criterion $\mathcal{Q}_T(\theta)$ is differentiable w.r.t. any $\theta$ in an open neighborhood of $\theta_0$, with probability approaching 1 (w.p.a. 1). Since estimator $\hat{\theta}$ is consistent (Proposition 4), this is enough to apply the standard approach to prove the asymptotic normality of extremum estimators as in Newey and McFadden [1999]. For this purpose, we assume local identification of parameter $\theta_0$ w.r.t. the population constraint vector $G(\theta, f_0)$.

**Assumption 6.** *The $(N+2) \times p$ matrix $J_0 := \nabla_{\theta'} G(\theta_0, f_0)$ is full column-rank.*

From Equation (2.22) and Proposition 2 the Jacobian matrix is $J_0 = J_S + J_L$, where

$$
\begin{aligned}
J_S &:= \mathrm{E}_0 \left[ \bar{\Gamma}_S(X_{t+1}) \nabla_{\theta'} \log\left(m(X_{t+1};\theta_0)\right) | X_t = x_0 \right], \\
J_L &:= \mathrm{E}_0 \left[ \bar{\Gamma}_L\left(X_{t+1}, X_t\right) \nabla_{\theta'} \log\left(m(X_{t+1};\theta_0)\right) \right].
\end{aligned}
\tag{2.35}
$$

Moreover, in Appendix 4.6.2 we derive the following asymptotic expansion of the estimator $\hat{\theta}$:

$$
\sqrt{Th_T^d}\left(\hat{\theta} - \theta_0\right) = \left(J_0'\Omega_0 J_0\right)^{-1} J_0'\Omega_0 \sqrt{Th_T^d} G(\theta_0, \hat{f}) + o_p(1). \tag{2.36}
$$

The last two components of vector $G(\theta_0, \hat{f})$ are equal to $\int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|x_0) dx$. We derive an asymptotic expansion for the other components by using Equation (2.14) and Proposition 2. We get (see Appendix 4.6.2)

$$
\begin{aligned}
\sqrt{Th_T^d} G(\theta_0, \hat{f}) &= \sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}(x|x_0) dx \\
&\quad + \sqrt{Th_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} + o_p(1).
\end{aligned}
\tag{2.37}
$$

We plug Equation (2.37) into Equation (2.36) and use the asymptotic normality of the integrals of kernel estimators (see Aït-Sahalia [1992]) to deduce the next Proposition 5.

**Proposition 5.** *Under Assumptions 1-6 and A 1-10 in Appendix 4.1, estimator $\hat{\theta}$ is asymptotically normal with $\sqrt{Th_T^d}$-rate of convergence:*

$$\sqrt{Th_T^d}\left(\hat{\theta}-\theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)}\Sigma_\theta\right),$$

*for the constant $\mathcal{K} := \int_{\mathcal{X}} K^2(x)dx$ and where the $p \times p$ matrix $\Sigma_\theta$ is defined as*

$$\Sigma_\theta := \left(J_0'\Omega_0 J_0\right)^{-1} J_0'\Omega_0\Sigma_S(x_0)\Omega_0 J_0 \left(J_0'\Omega_0 J_0\right)^{-1} \tag{2.38}$$

*and the $(N+2)\times(N+2)$ matrix $\Sigma_S(x_0)$ as $\Sigma_S(x_0) := \mathrm{V}_0\left[\bar{\Gamma}_S(X_{t+1})|X_t = x_0\right]$, with $\mathrm{V}_0\left[.|X_t = x_0\right]$ denoting the conditional variance under the true historical probability measure given $X_t = x_0$.*

*Proof.* See Appendix 4.6.2. □

The convergence rate of estimator $\hat{\theta}$ is $d$-dimensional nonparametric due to the conditioning on $X_t = x_0$ in the constraints. Moreover, the bias in the asymptotic distribution is negligible under the bandwidth conditions in Assumption A 6 in Appendix 4.1. The matrix $J_0$, that is the sum of the matrices defined in Equations (2.35), and the matrix $\Sigma_\theta$ in Equation (2.38) are reminiscent of the Jacobian and the asymptotic variance-covariance matrices of the moment function in the classical GMM setting. The matrix $\Sigma_S(x_0)$ is the conditional variance-covariance matrix of vector function $\bar{\Gamma}_S$ or, equivalently, of $\Gamma_S$. This matrix does not involve vector function $\bar{\Gamma}_L$ since the second term in the RHS of Equation (2.37) is asymptotically negligible. From the analogy with the classical GMM setting, Corollary 6 follows.

**Corollary 6.** *The weighting matrix that minimizes the asymptotic variance-covariance matrix of $\hat{\theta}$ is $\Omega_0 = \Sigma_S(x_0)^{-1}$. The minimal asymptotic variance-covariance matrix is $\dfrac{\mathcal{K}}{f_X(x_0)}(J_0'\Sigma_S(x_0)^{-1}J_0)^{-1}$.*

### 2.4.2 The XMM estimator of the SDF parameter

The XMM criterion in Definition 3 exploits both the uniform restrictions (2.8) and the restrictions (2.21) at $x_0$. The global and local identification conditions for parameter $\theta_0$ based on this extended set of restrictions are given below in Assumptions 7 and 8, respectively.

**Assumption 7.** *The unique $\theta \in \Theta$, such that $G(\theta, f_0) = 0$ and $\mathrm{E}_0[\Gamma_U(X_{t+1};\theta)|X_t = x] = 0$ for a.e. $x \in \mathcal{X}$, is $\theta = \theta_0$.*

**Assumption 8.** *The unique* $\beta \in \mathbb{R}^p$*, such that* $\nabla_{\theta'} G(\theta_0, f_0)\beta = 0$ *and* $\mathrm{E}_0 [\nabla_{\theta'} \Gamma_U (X_{t+1}; \theta_0) | X_t = x] \beta = 0$ *for a.e.* $x \in \mathcal{X}$*, is* $\beta = 0$*.*

Building on Gagliardini, Gouriéroux and Renault [2011], we distinguish between the linear transformations of $\theta_0$ that are identifiable from the uniform restrictions (2.8) alone, and the linear transformations of $\theta_0$ that are identifiable only when the local restrictions (2.21) at $x_0$ are also taken into account. The former are called full-information identifiable, the latter full-information unidentifiable. More precisely, let us define the linear space:

$$\mathcal{J} := \{\beta \in \mathbb{R}^p : \mathrm{E}_0 [\nabla_{\theta'} \Gamma_U (X_{t+1}; \theta_0) | X_t = x] \beta = 0, \text{ for a.e. } x \in \mathcal{X}\}, \tag{2.39}$$

and let $s \leq p$ be the dimension of $\mathcal{J}$. Let $R = [R_1 \;\; R_2]$ be an orthogonal $p \times p$ matrix, such that the columns of the $p \times s$ matrix $R_2$ span $\mathcal{J}$. Then, the invertible parameter transformation from $\theta$ to $\eta := [\eta_1' \; \eta_2']'$, defined by

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} R_1'\theta \\ R_2'\theta \end{pmatrix}, \tag{2.40}$$

is such that the $(p-s)$-dimensional vector $\eta_1$ involves full-information identifiable parameters only, while the $s$-dimensional vector $\eta_2$ involves full-information unidentifiable parameters only.

The asymptotic distribution of estimator $\hat{\theta}^*$ in Definition 3 is given in Proposition 7 below in terms of the estimators $\hat{\eta}_1^* = R_1'\hat{\theta}^*$ and $\hat{\eta}_2^* = R_2'\hat{\theta}^*$ of the transformed parameters. Let $\Omega_0 := \underset{T\to\infty}{\text{plim}}\, \Omega_T$ and $\tilde{\Omega}_0(x) := \underset{T\to\infty}{\text{plim}}\, \tilde{\Omega}_T(x)$, for any $x \in \mathcal{X}$, be the limit weighting matrices. We prove in Appendix 4.6.3 that the asymptotically optimal weighting matrices are $\Omega_0 = \Sigma_S(x_0)^{-1}$ and $\tilde{\Omega}_0(x) = \Sigma_U(x)^{-1}$, where $\Sigma_U(x) := \mathrm{V}_0[\Gamma_U(X_{t+1}; \theta_0) | X_t = x]$, for any $x \in \mathcal{X}$. We state the result directly for this choice.

**Proposition 7.** *Under Assumptions 1-4, 7, 8 and A 1-11 in Appendix 4.1, estimators* $\hat{\eta}_1^*$ *and* $\hat{\eta}_2^*$ *with* $\Omega_0 = \Sigma_S(x_0)^{-1}$ *and* $\tilde{\Omega}_0(x) = \Sigma_U(x)^{-1}$*, for any* $x \in \mathcal{X}$*, are consistent, asymptotically normal and independent, such that*

$$\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right) \xrightarrow{D} \mathcal{N}\left(0, \left(R_1' \mathrm{E}_0 \left[\tilde{J}_0(X_t)' \Sigma_U(X_t)^{-1} \tilde{J}_0(X_t)\right] R_1\right)^{-1}\right),$$

*and*

$$\sqrt{T h_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)} \left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2\right)^{-1}\right),$$

*where* $\tilde{J}_0(x) := \mathrm{E}_0 [\nabla_{\theta'} \Gamma_U (X_{t+1}; \theta_0) | X_t = x]$*, matrices* $R_1$ *and* $R_2$ *are defined in (2.40) and* $\eta_{1,0}$*,* $\eta_{2,0}$ *denote the true values of parameters* $\eta_1$*,* $\eta_2$*.*

*Proof.* See Appendix 4.6.3. ∎

The components of estimator $\hat{\theta}^*$ feature different rates of convergence, that are the parametric rate $\sqrt{T}$ for the full-information identifiable components, and the nonparametric rate $\sqrt{Th_T^d}$ for the full-information unidentifiable components. Mixed-rates asymptotics are obtained also in a conditional moment restrictions setting with weak identification (see Stock and Wright [2000] and Antoine and Renault [2010]). The asymptotic variance-covariance matrix of estimator $\hat{\eta}_1^*$ is the asymptotic efficiency bound for estimating parameter $\eta_{1,0}$ from the uniform moment restrictions assuming $\eta_{2,0}$ known (see Chamberlain [1987]). The asymptotic variance-covariance matrix of estimator $\hat{\eta}_2^*$ equals the minimal asymptotic variance-covariance matrix of the unfeasible cross-sectional estimator of parameter $\eta_{2,0}$ assuming $\eta_{1,0}$ known (see Corollary 6). Moreover, the estimators of the parameters $\eta_{1,0}$ and $\eta_{2,0}$ are asymptotically independent. Comparing Corollary 6 and Proposition 7 we understand that accounting for the uniform moment restrictions (2.8) allows us to increase the rate of convergence of the full-information identifiable parameters and to decrease in general the asymptotic variance of the full-information unidentifiable parameters.

#### 2.4.3 The estimator of the historical transition density and of its functionals

Let us now consider the estimator $\hat{f}^*$ in Definition 4. We derive its asymptotic distribution by considering a linearization of the tilting function in Equation (2.28) in a neighborhood of $(\theta_0, f_0)$. Under Assumption A 12 in Appendix 4.1 the weight $\omega_T$ converges to the non-negative scalar $\omega$. We get (see Appendix 4.6.4)

$$\hat{f}^*(x|\tilde{x}) \simeq \begin{cases} \hat{f}(x|x_0) + f_0(x|x_0)\hat{\Lambda}'\Gamma_S(x), \text{ if } \tilde{x} = x_0, \\ \hat{f}(x|\tilde{x}) + f_0(x|\tilde{x})\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\theta_0) + \omega\hat{\Lambda}'\Gamma_L(x,\tilde{x})\right), \text{ if } \tilde{x} \neq x_0, \end{cases} \tag{2.41}$$

where $\hat{\Lambda} = [\hat{\lambda}' \;\; \hat{\nu}_0']'$. We prove in Appendix 4.6.4 that the estimators of the Lagrange multipliers $\hat{\Lambda}$ and $\hat{\nu}(\tilde{x})$ for $\tilde{x} \neq x_0$ converge in probability to zero at rate $1/\sqrt{Th_T^d}$. Thus, we get

$$\hat{f}^*(x|\tilde{x}) = \hat{f}(x|\tilde{x}) + O_p\left(1/\sqrt{Th_T^d}\right), \tag{2.42}$$

for any $x, \tilde{x} \in \mathcal{X}$. The remainder term is dominated by the convergence rate $1/\sqrt{Th_T^{2d}}$ of the kernel estimator. Hence, estimators $\hat{f}^*$ and $\hat{f}$ are pointwise asymptotically equivalent, and we get the following Proposition 8.

**Proposition 8.** *Under Assumptions 1-4, 7, 8 and A 1-12 in Appendix 4.1, the estimator $\hat{f}^*$ is pointwise* *asymptotically normal with $\sqrt{Th_T^{2d}}$-rate of convergence:*

$$\sqrt{Th_T^{2d}}\left(\hat{f}^*(x|\tilde{x}) - f_0(x|\tilde{x})\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}^2 f_0(x|\tilde{x})}{f_X(\tilde{x})}\right), \tag{2.43}$$

*for any $x, \tilde{x} \in \mathcal{X}$, where the kernel constant $\mathcal{K}$ is defined in Proposition 5.*

*Proof.* See Appendix 4.6.4. □

The asymptotic distribution of the estimators of smooth functionals of $f_0$ and $\theta_0$ based on $\hat{f}^*$ and $\hat{f}$ differ. We give below the asymptotic distribution of estimator $\hat{a}^*$ introduced in Definition 5 for the case where $x^\star = x_0$ in Equation (2.32). This corresponds for instance to American put option-to-stock price ratios, exercise boundary and conditional cross-moments of the state variables for the state variables vector $x_0$ at the current date (see the examples i)-iii) in Section 2.3.4). The derivation of this asymptotic distribution is based on the asymptotic expansion obtained from Equation (2.31):

$$\hat{a}^* - a_0 = \nabla_{\theta'} a(\theta_0, f_0)\left(\hat{\theta}^* - \theta_0\right) + \left\langle Da(\theta_0, f_0), \Delta\hat{f}^*\right\rangle + O_p\left(\|\Delta\hat{f}^*\|_\infty^2\right) + O_p\left(\|\hat{\theta}^* - \theta_0\|^2\right), \tag{2.44}$$

where the Fréchet derivative $\left\langle Da(\theta_0, f_0), \Delta\hat{f}^*\right\rangle$ is given in Equation (2.32) with $\Delta f = \Delta\hat{f}^* := \hat{f}^* - f_0$. Since we expect a nonparametric convergence rate for $\hat{a}^*$, estimation of the SDF parameter affects the asymptotic distribution of $\hat{a}^*$ only through estimation of the full-information unidentifiable component $\eta_2$ (see Proposition 7). The relevant asymptotic expansion is

$$\sqrt{Th_T^d}\left(\hat{\theta}^* - \theta_0\right) = -R_2\left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2\right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1} \sqrt{Th_T^d}\, G(\theta_0, \hat{f}) + o_p(1), \tag{2.45}$$

where we use the asymptotically optimal weighting matrix $\Sigma_S(x_0)$ (see Appendix 4.6.4). We plug Expansions (2.37), (2.41) and (2.45) into Equation (2.44), and use the asymptotic normality of integral transformations of kernel estimators (see Aït-Sahalia [1992]). To state the result, we introduce some notation. We define the following conditional variance-covariance matrices under the true historical probability measure:

$$\begin{aligned} \Sigma_{\alpha_j,i}(x) &:= \mathrm{Cov}_0\left[\alpha_j(X_{t+1}), \Gamma_i(X_{t+1}, X_t) | X_t = x\right], \\ \Sigma_{i,l}(x) &:= \mathrm{Cov}_0\left[\Gamma_i(X_{t+1}, X_t), \Gamma_l(X_{t+1}, X_t) | X_t = x\right], \end{aligned} \tag{2.46}$$

for the subscript $j = S, L$, the subscripts $i, l = S, L, U$ and the state variables vector $x \in \mathcal{X}$.[21] We

[21] Even if functions $\Gamma_S$ and $\Gamma_U$ are independent of the lagged value of the state variables, we use Equations (2.46) for a compact notation. We also omit the dependence of $\Gamma_U$ on $\theta_0$.

further define the matrix

$$\Sigma_{i,j\perp l}(x) := \Sigma_{i,j}(x) - \Sigma_{i,l}(x)\Sigma_l(x)^{-1}\Sigma_{l,j}(x), \tag{2.47}$$

for the subscripts $i, j, l = \alpha_S, \alpha_L, S, L, U$ and $x \in \mathcal{X}$, that is the conditional covariance between the vector subscripted by $i$ and the residual of the projection of the vector subscripted by $j$ onto the vector subscripted by $l$. We set $\Sigma_i \equiv \Sigma_{i,i}$ and $\Sigma_{i\perp j} \equiv \Sigma_{i,i\perp j}$ for the conditional variances and the conditional variances of the projection residuals, respectively. Moreover we consider the Jacobian matrix

$$J_{\alpha_L\|U} := \mathrm{E}_0\Big[\Sigma_{\alpha_L,U}(X_t)\Sigma_U(X_t)^{-1}\bar{\Gamma}_U(X_{t+1})\nabla_{\theta'}\log\left(m(X_{t+1};\theta_0)\right)\Big],$$

that corresponds to the unconditional cross-second moment between $\nabla_{\theta'}\log m$ and the conditional orthogonal projection of $\alpha_L$ onto $\bar{\Gamma}_U$, and the Jacobian matrix

$$J_{L\perp U} := \mathrm{E}_0\big[\big(\bar{\Gamma}_L\left(X_{t+1}, X_t\right) - \Sigma_{L,U}(X_t)\Sigma_U(X_t)^{-1}\bar{\Gamma}_U\left(X_{t+1}\right)\big)\nabla_{\theta'}\log\left(m(X_{t+1};\theta_0)\right)\big],$$

that corresponds to the unconditional cross-second moment between $\nabla_{\theta'}\log m$ and the residual of the conditional orthogonal projection of $\bar{\Gamma}_L$ onto $\bar{\Gamma}_U$.

**Proposition 9.** *Under Assumptions 1-4, 7, 8 and A 1-12 in Appendix 4.1, the estimator $\hat{a}^*$ for $x^\star = x_0$ is asymptotically normal with $\sqrt{Th_T^d}$-rate of convergence:*

$$\sqrt{Th_T^d}\left(\hat{a}^* - a_0\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)}\Sigma_a\right),$$

*where the $r \times r$ matrix $\Sigma_a$ is defined as*

$$\Sigma_a := \Sigma_{\alpha_S\perp S}(x_0) + M_0(\omega)\Sigma_S(x_0)M_0(\omega)', \tag{2.48}$$

*constant $\mathcal{K}$ is defined in Proposition 5, and matrix $M_0(\omega)$ is defined as*

$$
\begin{aligned}
M_0(\omega) \quad := \quad & \omega\Bigg(\Sigma_{\alpha_S,S}\left(x_0\right)\Big(\Sigma_S\left(x_0\right)+\omega \mathrm{E}_0\left[\Sigma_{L\perp U}\left(X_t\right)\right]\Big)^{-1}\mathrm{E}_0\left[\Sigma_{L\perp U}(X_t)\right]\Sigma_S\left(x_0\right)^{-1} \\
& -\mathrm{E}_0\left[\Sigma_{\alpha_L,L\perp U}(X_t)\right]\Big(\Sigma_S\left(x_0\right)+\omega \mathrm{E}_0\left[\Sigma_{L\perp U}(X_t)\right]\Big)^{-1}\Bigg) \\
& +\Bigg(\Big(\Sigma_{\alpha_S,S}\left(x_0\right)+\omega \mathrm{E}_0\left[\Sigma_{\alpha_L,L\perp U}\left(X_t\right)\right]\Big)\Big(\Sigma_S\left(x_0\right)+\omega \mathrm{E}_0\left[\Sigma_{L\perp U}\left(X_t\right)\right]\Big)^{-1} \\
& \cdot(J_S+J_{L\perp U})+J_{\alpha_L\|U}-\nabla_{\theta'}a(\theta_0,f_0)\Bigg)R_2\left(R_2'J_0'\Sigma_S(x_0)^{-1}J_0R_2\right)^{-1}R_2'J_0'\Sigma_S(x_0)^{-1},
\end{aligned}
\tag{2.49}
$$

*where $\omega$ is the probability limit of weight $\omega_T$ in Definition 4.*

*Proof.* See Appendix 4.6.4. □

If the SDF parameter $\theta_0$ is full-information identifiable, that is, the linear space $\mathcal{J}$ is null and $R_2 = 0$, the term in the third and fourth line in the RHS of Equation (2.49) is zero. Then, the asymptotic variance of estimator $\hat{a}^*$ is minimized for $\omega = 0$, that is, when the criterion $D_T$ in Equation (2.25) does not account asymptotically for the local Kullback-Leibler divergence at $x_0$. We get $\Sigma_a = \Sigma_{\alpha_S\perp S}(x_0)$, which is the conditional variance of the residual of the orthogonal projection of $\alpha_S$ onto $\Gamma_S$ given $x_0$. To get the intuition, suppose that functional $a$ is the conditional expectation of function $\alpha_S$ with true value $a_0 = \mathrm{E}_0[\alpha_S(X_{t+1})|X_t = x_0]$. Then, when $\omega = 0$ the estimator $\hat{a}^*$ is asymptotically equivalent to the unfeasible estimator $\int_{\mathcal{X}} \alpha_S(x)\tilde{f}^*(x|x_0)dx$, where

$$
\tilde{f}^*(\cdot|x_0) \quad = \quad \arg\min_{f\in\mathcal{F}_0} d_{KL}(f,\hat{f}|x_0) \quad s.t. \ \int_{\mathcal{X}} \Gamma_S(x)f(x|x_0)dx = 0,
$$

and $\mathcal{F}_0$ denotes the set of transition densities given $x_0$. A similar interpretation is given for the estimation of a moment under a moment restriction by Brown and Newey [1998] in an unconditional setting, and by Antoine, Bonnal and Renault [2007] in a conditional setting. The matrix $\frac{\mathcal{K}}{f_X(x_0)}[\Sigma_{\alpha_S}(x_0) - \Sigma_{\alpha_S\perp S}(x_0)]$ is the efficiency gain from the information in the local no-arbitrage restrictions. Moreover, estimation of parameter $\theta_0$ has no effect on the accuracy of estimator $\hat{a}^*$.

If some components of the SDF parameter $\theta_0$ are full-information unidentifiable and $\omega > 0$, matrix $M_0(\omega)\Sigma_S(x_0)M_0(\omega)'$ in the RHS of Equation (2.48) is the contribution to the asymptotic variance of estimator $\hat{a}^*$ from including the local Kullback-Leibler divergence at $x_0$ in the criterion $D_T$ and estimating the SDF parameter $\theta_0$. The matrix $M_0(\omega)\Sigma_S(x_0)M_0(\omega)'$ involves conditional variances and covariances of the residual of the orthogonal projection of $\Gamma_L$ onto $\Gamma_U$ because of the interaction between local and uniform restrictions in the constrained optimization of criterion $D_T$. For a scalar functional $a$, the asymptotic weight $\omega$ can be selected in order to minimize the asymptotic variance matrix $\frac{\mathcal{K}}{f_X(x_0)}\Sigma_a$.[22] This optimal choice for $\omega$ is the solution of the problem $\min_{\omega \geq 0} M_0(\omega)\Sigma_S(x_0)M_0(\omega)'$, and depends in general on the functional of interest $a$.

Finally, let us apply Proposition 9 when the functional of interest $a$ corresponds to the option-to-stock price ratio at date $t_0$ of an American put option with time-to-maturity $h^\star$ and moneyness strike $k^\star$. From example i) in Section 2.3.4, the asymptotic variance of the estimator $\hat{a}^*$ is obtained by using $\alpha_S$ and $\alpha_L$ defined in Equations (2.33), and setting $\nabla_{\theta'} a(\theta_0, f_0) = J_{S^\star} + J_{L^\star}$, where matrices $J_{S^\star}$, $J_{L^\star}$ are defined as in Equations (2.35) by replacing $\bar{\Gamma}_S$ and $\bar{\Gamma}_L$ by $\alpha_S$ and $\alpha_L$, respectively.

## 2.5 Monte Carlo experiment

In this section we investigate the finite sample properties of the estimators in a Monte Carlo experiment. We consider a scalar volatility factor $\sigma_t$ (i.e. $d = 2$) representing the volatility of the stock return. We describe the DGP in Section 2.5.1, the numerical implementation in Section 2.5.2 and the results in Section 2.5.3.

### 2.5.1 The design

Under the historical probability measure $\mathscr{P}$, the process $(r_t)$ is such that

$$r_t = r_f + \gamma\sigma_t^2 + \sigma_t\varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}\big(0, 1\big), \tag{2.50}$$

where $\gamma \geq 0$ is the variance-in-mean parameter. The daily risk-free rate $r_f$ is constant and equal to $2 \cdot 10^{-4}$. The stochastic variance $\sigma_t^2$ follows an Autoregressive Gamma (ARG) Markov process of order 1 (Gouriéroux and Jasiak [2006]), which is the discrete-time counterpart of the Cox-Ingersoll-Ross process (Cox, Ingersoll and Ross [1985]). The historical transition density of $\sigma_t^2$ is defined by the conditional Laplace transform

$$\mathrm{E}_0[\exp\left(-u\sigma_t^2\right) \big| \, \sigma_{t-1}^2] = \exp\big(-\varphi_1(u)\sigma_{t-1}^2 - \varphi_2(u)\big), \qquad u \geq 0, \tag{2.51}$$

[22] Weighting matrix $\Omega_0$ is considered as given and equal to $\Omega_0 = \Sigma_S(x_0)^{-1}$, which is asymptotically optimal for the estimation of $\theta_0$. The asymptotic variance $\frac{\mathcal{K}}{f_X(x_0)}\Sigma_a$ could be minimized by optimizing jointly w.r.t. $\omega$ and $\Omega_0$, but the optimization problem becomes more difficult. We do not consider this alternative approach.

| $c$ | $\rho$ | $\delta$ | $\gamma$ |
|---|---|---|---|
| $3.65 \cdot 10^{-6}$ | $9.60 \cdot 10^{-1}$ | $1.05$ | $3.60 \cdot 10^{-1}$ |
| $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
| $4.55 \cdot 10^{-7}$ | $-5.90 \cdot 10^{-2}$ | $1.14 \cdot 10^{-1}$ | $8.60 \cdot 10^{-1}$ |

Table 1: The values of the historical and SDF parameters of the DGP.

where the functions $\varphi_1$ and $\varphi_2$ are defined as $\varphi_1(u) = \rho u / (1 + cu)$ and $\varphi_2(u) = \delta \log (1 + cu)$ for parameters $c, \delta > 0$ and $\rho \in [0, 1)$. We consider a 4-dimensional SDF parameter $\theta = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]'$ and an exponential affine one-day SDF:

$$M_{t,t+1}(\theta) = \exp(-r_f) \exp\left(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 (r_{t+1} - r_f)\right). \tag{2.52}$$

Parameters $\theta_2$ and $\theta_4$ are related to the risk premia associated with the stochastic volatility and the excess return of the stock, respectively. Exponential affine SDF specifications are common in reduced-form modeling (see e.g. Duffie, Pan and Singleton [2000], Duffie, Filipovic and Schachermayer [2003] and Gouriéroux and Monfort [2007]). Under the above DGP, the historical transition density of $X_t$ given $X_{t-1}$ is independent of $r_{t-1}$. In this case, the conditioning set for option valuation gets smaller.

**Corollary 10.** *When the density of $X_t$ given $X_{t-1}$ is independent of $r_{t-1}$ under $\mathscr{P}$, Proposition 1 holds with $Y_t = [k_t \ \ \sigma_t']'$.*

Thus, under the above DGP, the option-to-stock price ratio at time $t$ depends on time-to-maturity $h$, moneyness strike $k_t$ and volatility $\sigma_t$ only. Moreover, in the Definitions 2-3 and in Equations (2.28)-(2.29), the conditioning variable $X_t$ is replaced by $\sigma_t$. The parametric specification for the DGP is similar to the example considered in Gagliardini, Gouriéroux and Renault [2011]. They show that the SDF in Equation (2.52) is admissible for the DGP defined in (2.50)-(2.51). More specifically, the no-arbitrage conditions for the stock and the non-defaultable bond are satisfied, i.e.

$$\mathrm{E}_0 \left[ M_{t,t+1}(\theta_0) e^{r_{t+1}} | \sigma_t = \sigma \right] = 1, \quad \mathrm{E}_0 \left[ M_{t,t+1}(\theta_0) e^{r_f} | \sigma_t = \sigma \right] = 1, \quad \text{for all } \sigma \in \mathbb{R}_+,$$

if, and only if, the true parameter value $\theta_0 = [\theta_1^0 \ \ \theta_2^0 \ \ \theta_3^0 \ \ \theta_4^0]'$ is such that $\theta_1^0 = -\varphi_2(\xi)$, $\theta_3^0 = -\varphi_1(\xi)$ and $\theta_4^0 = 1/2 + \gamma$, where $\xi = \theta_2^0 + \gamma^2/2 - 1/8$. We report in Table 1 the chosen values of the historical and SDF parameters. They satisfy the constraints described above and are chosen on the base of a calibration to real data on liquid assets.

Having presented the parametric family to which the DGP belongs, let us now describe the data we create. We generate 1000 time series of returns and volatility with length $T = 1000$ from date

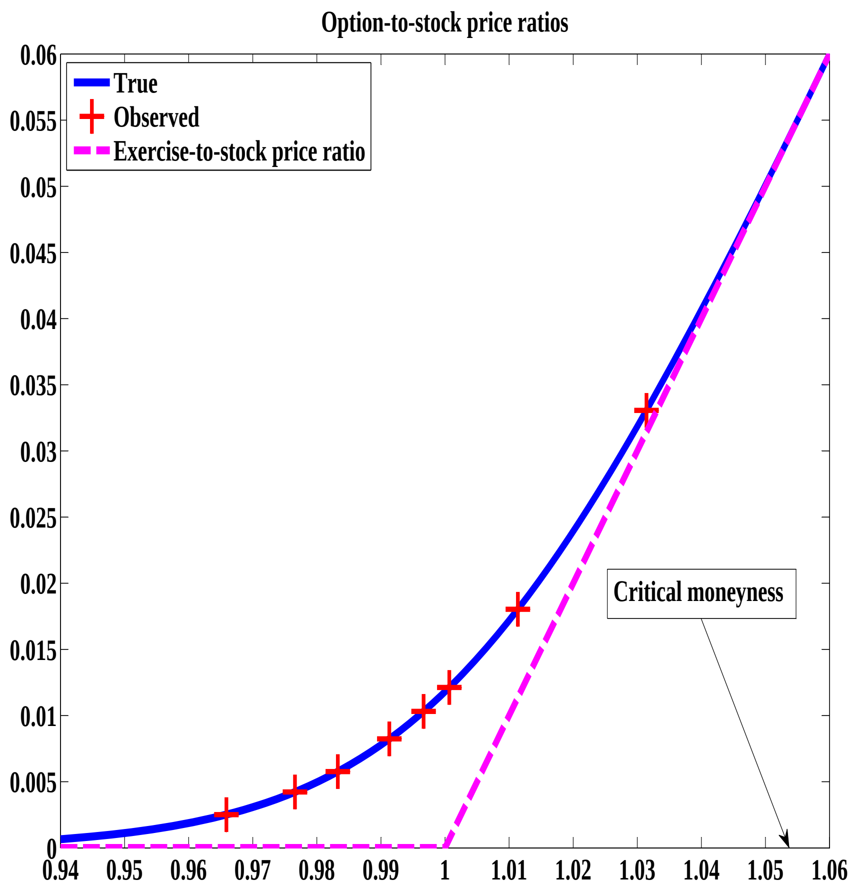


Figure 2.1: The cross-section of American option-to-stock price ratios at $t_0$ as a function of the moneyness strike, for time-to-maturity $h = 20$ days. The value of the volatility of the stock at the current date is $6.50 \cdot 10^{-3}$. The values of the historical and SDF parameters are given in Table 1. The solid line is the true American option-to-stock price ratio function for the selected DGP. The dashed line is the early exercise-to-stock price ratio. The crosses are the observed American option-to-stock price ratios.

| $k_1$ | $k_2$ | $k_3$ | $k_4$ | $k_5$ | $k_6$ | $k_7$ | $k_8$ |
|---|---|---|---|---|---|---|---|
| 0.966 | 0.976 | 0.983 | 0.991 | 0.997 | 1.007 | 1.011 | 1.031 |

Table 2: The values of the moneyness strikes for the available options at $t_0$.

$t_0 - T + 1$ to current date $t_0$. The volatility $\sigma_0$ at date $t_0$ is the same across simulations and is equal to $6.50 \cdot 10^{-3}$. For this value of volatility, we consider the cross-section of American put option-to-stock price ratios with time-to-maturity $h = 20$. We display this cross-section as a function of the moneyness strike by a solid line in Figure 2.1. We compute it by recursive valuation, using the estimate of the transition density of the state variables obtained by kernel estimation on a very long simulated time series of the state variables. From the full cross-section of American put option-to-stock price ratios, we select $N = 8$ values, with different moneyness strike $k_j$, for $j = 1, \ldots, 8$, as reported in Table 2. We display these American put option-to-stock price ratios by crosses in Figure 2.1. For each Monte Carlo replication, the data available to the econometrician are a different time series of state variables and the same $8$ selected American put option-to-stock price ratios. This simulation design reflects the analysis in previous sections, where $x_0$ (that in the Monte Carlo experiment reduces to $\sigma_0$) is assumed constant and given.

We assume that the econometrician does not know the true DGP under $\mathscr{P}$ described in (2.50)-(2.51) but is aware of the parametric specification of the SDF in Equation (2.52) and the Granger non-causality of $r_{t-1}$ on $X_t$, so that she can use Corollary 10. We then estimate the model parameters and some American put option-to-stock price ratios for each Monte Carlo replications. We start with the estimation of parameter $\theta$. In this semi-parametric setting, the full SDF parameter vector $\theta$ is not full-information identifiable (see Gagliardini, Gouriéroux and Renault [2011]). The linear space $\mathscr{J}$ defined in (2.39) is one-dimensional and spanned by vector $r_2 = [-\delta c/(1+c\xi), 1, -\rho/(1+c\xi)^2, 0]'$. Thus, the SDF parameter $\theta_4$ is full-information identifiable, while parameters $\theta_1$, $\theta_2$ and $\theta_3$ are not. We consider the cross-sectional and XMM estimators of the SDF parameter in Definitions 2-3 with identity weighting matrices $\Omega_T = I_{N+2}$ and $\tilde{\Omega}_T = I_2$. The XMM estimator becomes

$$\hat{\theta}^* = \arg\min_{\theta \in \Theta} \left( h_T \left\| G(\theta, \hat{f}) \right\|^2 + \frac{1}{T} \sum_{t=1}^{T} \left\| \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta) | \sigma_t] \right\|^2 \right). \tag{2.53}$$

The cross-sectional estimator minimizes the first component of the criterion in the RHS of Equation (2.53). We then pass to the estimation of the transition density of the state variables and compute the estimator $\hat{f}^*$ defined in Equations (2.28)-(2.29). Finally, we use the estimators $\hat{\theta}^*$ and $\hat{f}^*$ to compute the American put option-to-stock price ratios $\mathcal{A}^{h^\star}_{\hat{\theta}^*, \hat{f}^*}[v(0,.)](y^\star)$ for time-to-maturity $h^\star = 20$, volatility

$\sigma^\star = \sigma_0$ and moneyness strikes $k^\star = 0.972, 0.986, 1.000, 1.030$.

### 2.5.2 The numerical implementation

For given parameter $\theta$, the computation of vector $g(\theta, \hat{f})$ in the criterion functions minimized by $\hat{\theta}$ and $\hat{\theta}^*$ involves recursive applications of the pricing operator $\mathcal{A}_{\theta,\hat{f}}$ to functions $\varphi$ on the bi-dimensional moneyness-volatility space $\mathcal{Y} \subset \mathbb{R}^2_+$. To make the estimation procedure feasible, functions $\varphi$ are evaluated on a finite grid with $N_k \times N_\sigma$ grid points on the subset $[k_{low},\ k_{high}] \times [\sigma_{low},\ \sigma_{high}]$ of $\mathcal{Y}$. The conditional expectation w.r.t. $\hat{f}$ in the definition of operator $\mathcal{A}_{\theta,\hat{f}}$ (see Equation (2.23) ) is computed by a Nadaraya-Watson estimator. We take the Gaussian kernel with bandwidth $h_T = 0.9 \min\{s, R_q/1.34\} T^{-\frac{1}{5}}$ as suggested in Silverman [1986], where $s$ and $R_q$ denote the sample volatility and interquartile range of the observations $\sigma_t$, respectively. When the computation of $\mathcal{A}_{\theta,\hat{f}}[\varphi]$ requires to evaluate function $\varphi$ on a point $(k, \sigma)$ within $[k_{low},\ k_{high}] \times [\sigma_{low},\ \sigma_{high}]$ but outside the grid, the nearest grid point is selected. When $k < k_{low}$ we set $\varphi(k, \sigma) = 0$ and when $\sigma < \sigma_{low}$ the nearest grid point is selected. When $k > k_{high}$ and/or $\sigma > \sigma_{high}$ we use a linear extrapolation procedure. The use of a finite subset of $\mathcal{Y}$ and a finite grid introduce a numerical error, that becomes negligible as $\sigma_{low}, k_{low} \to 0$, $\sigma_{high}, k_{high} \to \infty$ and $N_k, N_\sigma \to \infty$. In the Monte Carlo experiment, we use $N_k = 300$ and $N_\sigma = 30$ grid points and we set $k_{low} = 0.8$ and $k_{high} = 1.2$ for the moneyness strike domain, while $\sigma_{low}$ and $\sigma_{high}$ are set equal to the $1\%$ and $99\%$ quantiles of the volatility realizations in the Monte Carlo repetition. The grid is homogeneous and such that the volatility at date $t_0$ coincides with one of the $N_\sigma$ points that discretize $[\sigma_{low},\ \sigma_{high}]$. Increasing the domain or the fineness of the grid w.r.t. our choice does not yield substantial accuracy improvements. We have implemented our routines in FORTRAN. A commercial 2 GHz processor takes less than a second to evaluate the American put-option-to stock ratios for $h = 20$ at all grid points. A numerical minimization of the criterion in Equation (2.53) is then feasible in less than five minutes. We compute the estimator $\hat{f}^*$ by the iterative algorithm described in Section 2.3.3 with $\omega_T = 0$. At each iteration we use $\hat{f}^*$ for the estimation of the option-to-stock price ratios of interest. We take as convergence criterion the stability of these ratios up to $10^{-5}$. Less than 10 iterations are enough in most of the Monte Carlo repetitions, making the procedure feasible in less than five minutes.

### 2.5.3 The results

We show in Figure 2.2 the kernel smoothed density functions of the XMM estimators of the SDF parameters. The estimators of parameters $\theta_1$, $\theta_2$ and $\theta_3$ feature small bias and their distributions are slightly skewed. The skew is more pronounced for parameter $\theta_3$. The estimator of parameter $\theta_4$ is

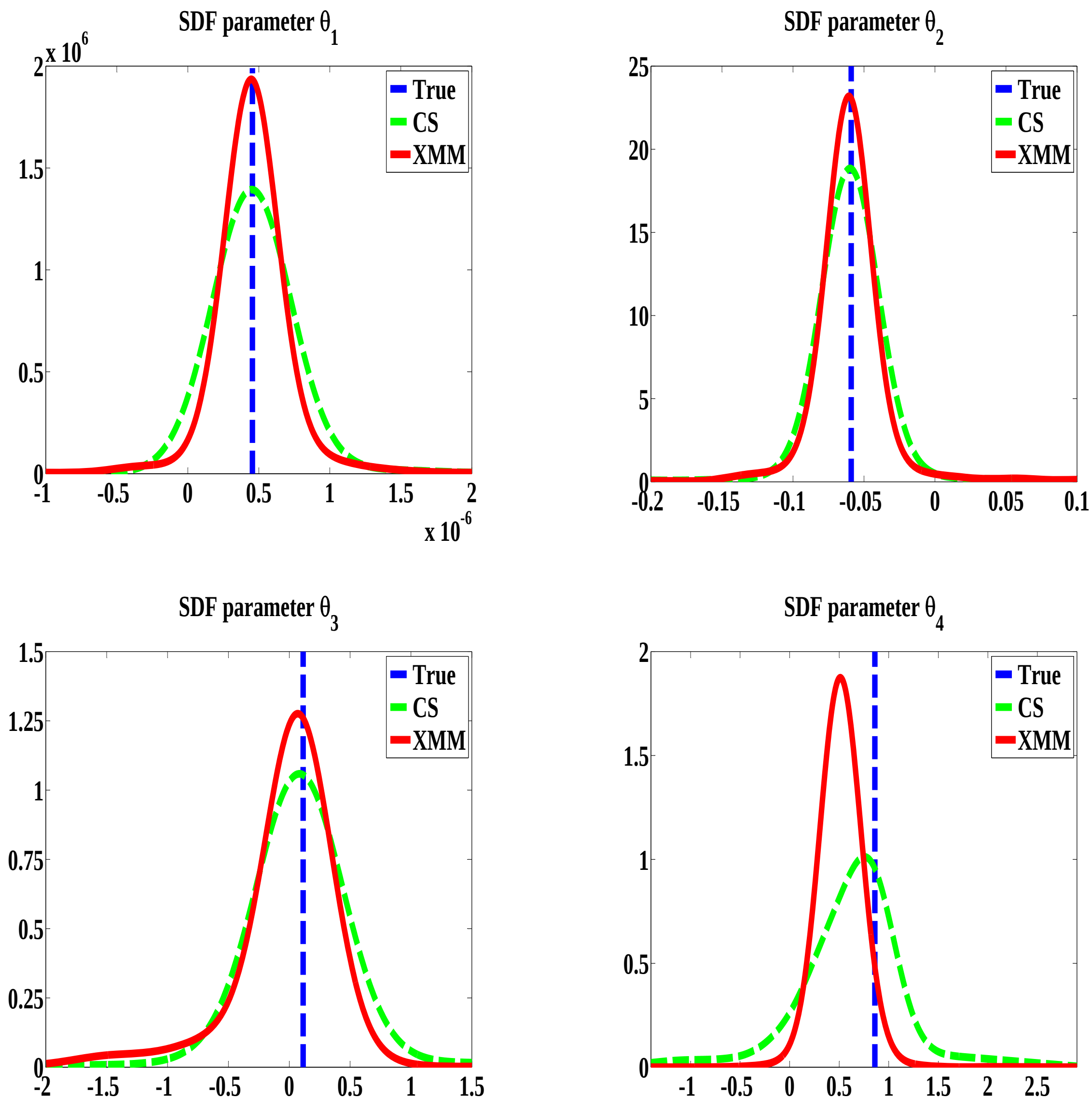


Figure 2.2: The distributions of the estimated SDF parameters. In each panel, the solid line corresponds to the XMM estimator $\hat{\theta}^*$ with weighting matrices $\Omega_T = I_{N+2}$ and $\tilde{\Omega}_T(\tilde{x}) = I_2$, the dashed line to the cross-sectional (CS) estimator $\hat{\theta}$ with weighting matrix $\Omega_T = I_{N+2}$. The true parameter values are displayed by the dashed vertical lines.

downward biased. The estimated values of the parameters have the same sign as the true parameter values (see Table 1) in most of the Monte Carlo repetitions. For comparison purpose we display in Figure 2.2 also the smoothed density functions for the components of the cross-sectional estimator in Definition 2 with identity weighting matrix $\Omega_T = I_{N+2}$. The cross-sectional estimates feature larger standard deviations than the XMM estimates. Hence, accounting for the uniform restrictions (2.8) improves the accuracy of the SDF parameter estimator also in finite sample. The difference between the XMM and cross-sectional estimators is larger for the full-information identifiable parameter $\theta_4$. The two estimators have similar biases, but the distribution of the cross-sectional estimator features larger variance and is more skewed and leptokurtic.[23] These findings are compatible with the different rates of convergence of the estimators of $\theta_4$, that are parametric for the XMM estimator and nonparametric for the cross-sectional estimator (see Sections 2.4.1-2.4.2).

We show in Figure 2.3 the kernel smoothed density functions of the estimates of the American option-to-stock prices for four moneyness strikes $k^\star$ of interest. For $k^\star = 0.972, 0.986, 1$ the bias is very small and the distribution is close to a Gaussian distribution. For moneyness strike $k^\star = 1.030$, the distribution of the estimated option-to-stock price ratio still admits a peak close to the true value but is truncated at the exercise value $k^\star - 1 = 0.03$. This truncation effect arises because some estimated continuation values are below the exercise value. Truncation is negligible for the other moneyness strikes that are far from the critical moneyness strike. For comparison purpose, we display in Figure 2.3 also the smoothed density functions of the estimates of the American option-to-stock price ratios obtained using the kernel density $\hat{f}$ as an estimator for the historical transition of the state variables. This estimator accounts neither for the available option prices nor for the no-arbitrage restrictions on stock and bond returns. The biases of the two estimators based on $\hat{f}^*$ and $\hat{f}$ are similar. However, for each considered moneyness strike, the option price estimator based on $\hat{f}^*$ features a smaller variance than the estimator based on $\hat{f}$. This finding shows that incorporating the informational content of cross-sectionally observed option prices and imposing the no-arbitrage restrictions for all assets improve substantially the accuracy of the estimators of the option prices which are not currently observed on the market.

[23] The bias of the XMM estimator of $\theta_4$ is $-3.45 \cdot 10^{-1}$, that of the cross-sectional estimator is $-5.69 \cdot 10^{-1}$.

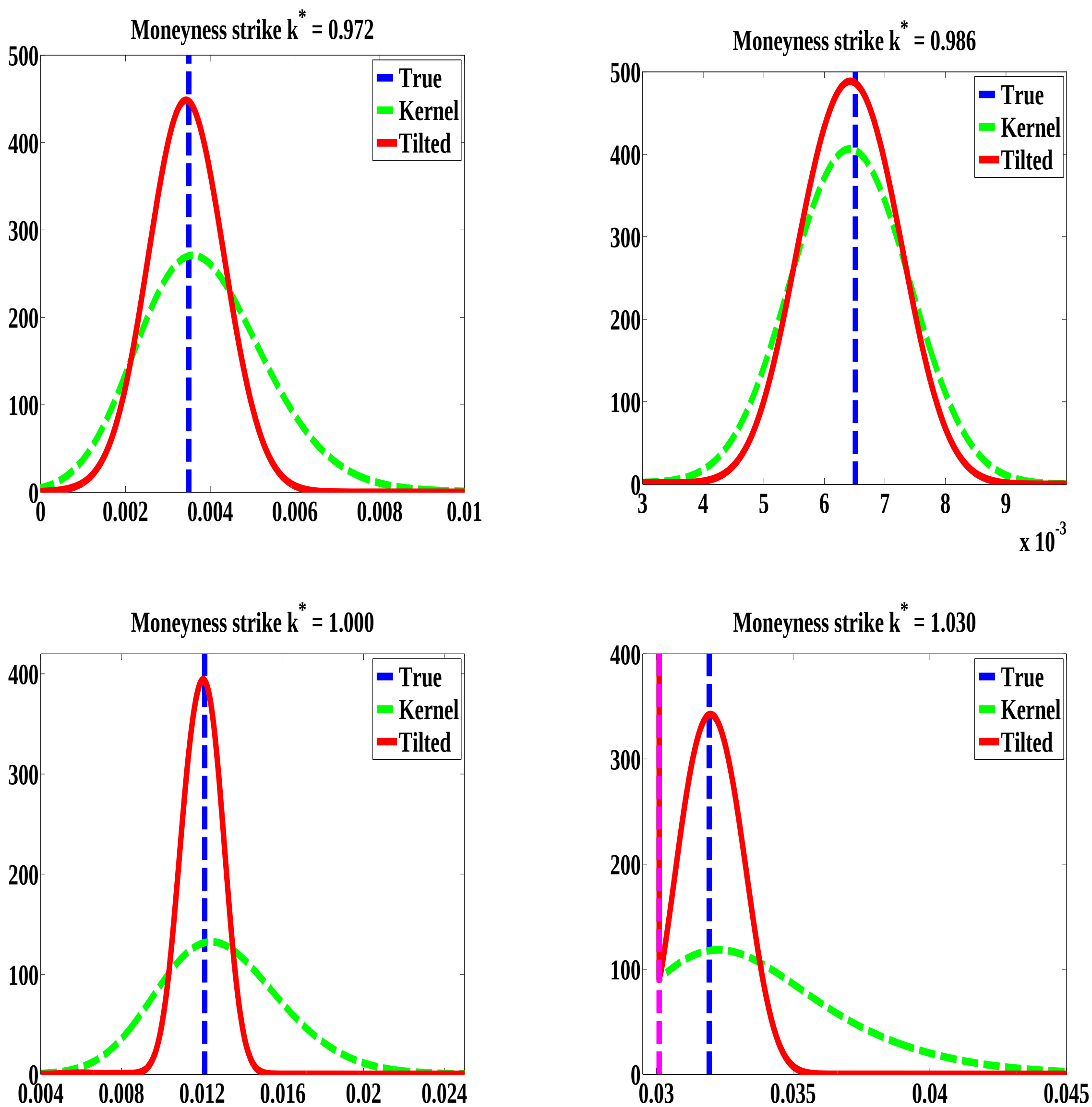


Figure 2.3: The distribution of the estimated American option-to-stock price ratios at $t_0$ for time-to-maturity $h^\star = 20$ days and four different moneyness strikes $k^\star$. In each panel, the solid line is the distribution of the estimates when we use $\hat{f}^*$ defined in Equations (2.28)-(2.29) for the estimation of the American put pricing operator with $\omega_T = 0$. The dashed line is the distribution of the estimates when we use $\hat{f}$ defined in Equation (2.19). For $k^\star = 0.972, 0.986, 1.000$ the dashed vertical line indicates the true value. For $k^\star = 1.030$ the dashed vertical line on the left indicates the exercise value and the dashed vertical line on the right the true value. The peaks at the left vertical line correspond to estimated option-to-stock price ratios equal to the exercise value.

# 3 An Empirical Study of Stock and American Option Prices

Stock return and its volatility are stochastic. Investors try to understand the properties of their joint process and require a compensation for the risks of a lower stock excess return and an higher stock return volatility than expected. The research reported in this paper consists in an empirical analysis of share prices and quotes for American options written on the shares. The study reaches two results related to the investors' behavior, without relying on any parametric specification of the dynamics of stock return and its volatility. First, share prices and option quotes are both necessary to identify at the same time the way investors discount future stock excess return and return variance to create prices. Second, the estimates of some dynamic properties of stock return and its volatility are more stable over time when an arbitrage-free pricing model is considered in the estimation procedure. The first finding is consistent with the idea that equity traders are mostly interested in the stock excess return and option traders in the underlying return volatility. The second finding is the result of adding structure to the economic model considered in the estimation procedure. The novelty of the method employed in the study is the fact that it does not depend on any parametric specification of the dynamics of stock return and its volatility. Therefore, the study does not bear the risk of a wrong specification of this dynamics.

Share and option markets are analyzed as they were arbitrage-free and free of frictions, except for the option bid-ask spread, with stock return and its volatility as state variables. Lower trading frequency and volume in the derivative market than in the share one prevent from considering option trade prices. The share and option markets are possibly incomplete and with several admissible stochastic discount factors. To characterize the way investors discount the future realizations of stock excess return and return variance and to uniquely identify equity and variance premia, only one Stochastic Discount Factor (SDF) is assumed to be an exponential-affine function of the state variables. The study is based on the estimation of the parameters appearing in this function and the historical dynamics of the state variables in a discrete-time framework. Different estimation procedures, that do not rely on any parametric specification of the state variables dynamics, either under the risk-neutral or historical probability measure, are considered. The research focuses on daily IBM share closing prices at NYSE from January 2006 to August 2008 and closing quotes for IBM American call and put options selected among U.S. centralized markets in July and August 2008. The daily 1-month T-bill rate is considered as the reference risk-free rate used by investors to compute excess stock returns. The data generating process is assumed to be the same for the entire period, characterized by two distinct phases: a (relatively) stable one, before July 2007, and a subsequent (relatively) volatile one. The two phases reflect two distinct situations of (relative) stability and turmoil in financial markets. The period considered in the study ends in September 2008 because the plunge of the IBM stock price, occurred during the stock markets crash in the fall of 2008, makes questionable the idea of the same data generating process as before.

The considered estimators of the SDF parameters and the historical dynamics of the state variables are functions of the realizations of the reference risk-free rate, the stock return and a proxy of its volatility. In particular, a measure of daily realized volatility, obtained by high-frequency returns, is taken as a proxy of the spot return volatility. A kernel nonparametric estimator of the joint transition density of daily equity return and realized volatility is taken as the reference estimate for the transition density of the state variables. This estimate enters the definition of all the estimators considered in the study. Nonparametric estimation methods of a stochastic process allow to identify its main empirical features without assuming a parametric model. These methods have been largely used in financial applications (see Cai and Hong [2009] for a review on nonparametric methods in finance). The study is composed by two parts, both made up of a comparison between estimates of the same quantity based on distinct sets of asset prices. The comparison is between estimates of the SDF parameters in the first part and between estimates of some properties of the historical dynamics of the state variables in the second part. The first comparison shows the need of contemporaneous share prices and option quotes to quantify the equity and variance premia. The second comparison shows that the estimation of the historical joint dynamic properties of the risk factors is more precise when an arbitrage-free pricing model is considered.

At every day in July and August 2008, the estimation of the SDF parameters is done in three different ways: using only a time series of share and risk-free asset prices, using only a cross-section composed by option mid-quotes and share and risk-free asset prices, or using both time series and cross-section. A different estimation procedure is considered for each set of data. The three techniques are a Generalized Method of Moments (GMM) estimation (see Hansen [1982] and Hansen and Singleton [1982]), a cross-sectional calibration and an Extended Method of Moments (XMM) estimation (see Gagliardini, Gouriéroux and Renault [2011] and Gagliardini and Ronchetti [2010]). Each method finds the values for the SDF parameters that best satisfy the empirical counterparts of a set of no-arbitrage restrictions on the base of a particular criterion. The GMM method considers only the no-arbitrage restrictions for IBM share and T-bill over time. The calibration technique considers only the no-arbitrage restrictions for IBM share, T-bill and IBM American option mid-quotes at a given date. The XMM estimation considers all the restrictions over time and across different assets. For each SDF parameter, the XMM method is the only one that provides estimates that are similar, in terms of mean and standard deviation over time, to the ones obtained by at least one of the other two methods. The results are in accordance with the idea that the informations on the SDF parameters in a time series of share and risk-free asset prices and a cross-section of option quotes are not redundant. Specifically, the time series of share prices and the cross-section of option quotes are informative mostly on the discount for uncertain stock excess return and return variance, respectively.

In this paper two distinct estimation approaches to the historical dynamics of the state variables are considered. Both the approaches are nonparametric w.r.t. this dynamics. The first approach is a kernel estimation. The second approach is an estimation constrained by the no-arbitrage restrictions. Although the considered estimators share the asymptotic properties, the structure imposed in the second estimation procedure leads to a description of the state variables process that is more precise. As an illustration, estimates of some characteristics of the historical joint dynamics of the state variables obtained by the two approaches at any day in July and August $2008$ are compared. These characteristics are the historical conditional correlation between the state variables, Sharpe ratio of an investment on the stock, skewness and kurtosis of the returns. The estimates of these characteristics are more precise when an arbitrage-free pricing model is imposed. In particular, the variation over time of the point estimates is smaller.

Some empirical studies show that the assumptions of rational investors and absence of frictions in American option markets are not always met (see e.g. Diz and Finucane [1993] for index options and Carpenter [1998] and Poteshman and Serbin [2003] for stock options). These assumptions are necessary to conduct the analysis based on observable arbitrage-free prices reported in the present paper. The assumptions are justified by the relatively high liquidity of the considered assets and by ending the analysis to August $2008$.

Section 3.1 contains the description of the asset pricing model. Section 3.2 deals with the implications of absence of arbitrage opportunities and introduces the estimators of the model parameters. The criteria to select the data are discussed in Section 3.3. Finally, Section 3.4 contains the description and interpretation of the estimates of SDF parameters and some historical properties of the joint one-day historical dynamics of the state variables over a one-day horizon. Specifically, the transition density and the conditional correlation of the state variables, the conditional Sharpe ratio of an investment in the IBM stock and the conditional skewness and kurtosis of the returns are considered.

## 3.1 Model

This section introduces the description of the state variables dynamics and the asset price formation. Section 3.1.1 deals with the representation of the state variables process and the SDF. Section 3.1.2 discusses the way American option prices are generated.

### 3.1.1 State variables and stochastic discount factor

The state variables are the daily stock cum-dividend geometric return and return volatility. The former is denoted by $r_t$, the latter by $\sigma_t$. These variables are gathered in vector $X_t = [r_t \ \sigma_t]'$ and their joint process is stationary, time-homogeneous and Markovian of order 1.[24] The historical dynamics of this process is described by means of the transition density $f(x_{t+1}|x_t)$, that is the probability density for the state variables vector to assume value $x_{t+1}$ at day $t+1$ after assuming value $x_t$ at day $t$. The causal relationships between the four arguments of this function have been widely studied in the financial literature. A causal effect of the return on the volatility is known as leverage effect (see e.g. Black [1977], Christie [1982] and Nelson [1991]). A causal effect of the volatility on the return is known as volatility feedback (see e.g. French, Schwert and Stambaugh [1987], Campbell and Hentschel [1992] and Bekaert and Wu [2000]). Both effects can be between variables either at the same day (contemporaneous effect) or at distinct days (delayed effect). Standard stochastic volatility models used in asset pricing consider stock return and its volatility as state variables (see e.g. Hull and White [1987], Stein and Stein [1991], Heston [1993], Bates [1996] and Bakshi, Cao and Chen [1997] for a continuous time setting and Heston and Nandi [2000] and Christoffersen, Heston and Jacobs [2006] for a discrete time one). Differently than in most of the literature on American option pricing, in this paper no fully parametric form for the transition density of the state variables is adopted, either under the historical or risk-neutral probability measure. Since the true volatility is not observable and the method relies on empirical realizations of the state variables, a measure of realized volatility is taken as a proxy for the spot daily volatility (see e.g. Andersen, Bollerslev, Diebold and Ebens [2001] and Andersen, Bollerslev, Diebold and Labys [2003]).

The stockholders need to be compensated for the risks assumed by investing in the stock. They require some premia for the risks of a lower excess return and an higher volatility than expected. Several methods to quantify these premia have been proposed in the literature (see e.g. Mehra and Prescott [1985] for the equity premium in the CAPM, Lamoureux and Lastrapes [1993] for the volatility premium and Carr and Wu [2009] for the variance premium). To include these premia in the asset prices, investors distort the historical transition density and use a risk-neutral one for pricing purposes. The link between the historical and any risk-neutral transition density is provided by an SDF (see e.g. Duffie [2001]). While the market is not assumed to be complete and the SDF to be unique, only one admissible SDF from day $t$ to day $t+1$ is supposed to admit the following parametrization:

$$M_{t,t+1}(\theta) = \exp\left(-r_{f,t+1}\right) \exp\left(-\theta_1 - \theta_2 \left(r_{t+1} - r_{f,t+1}\right) - \theta_3 \left(r_t - r_{f,t}\right) - \theta_4 \sigma_{t+1}^2 - \theta_5 \sigma_t^2\right), \quad (3.1)$$

[24] When the state variables are assumed Markovian of higher order, a study similar to the one reported in this paper can be conducted by extending the state variables vector.

for the unknown SDF parameters vector $\theta = [\theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5]'$ and the daily risk-free rate $r_{f,t+1}$ from day $t$ to day $t+1$, that is assumed to be known at day $t$. The functional form of the SDF $M_{t,t+1}(\theta)$ from day $t$ to day $t+1$ is exponential-affine in the state variables at day $t$ and day $t+1$. The parameters $\theta_2$ and $\theta_4$ are the coefficients of the stock excess return and the return variance at day $t+1$. These parameters measure how investors discount the future realizations of the stock excess return and return variance, respectively, and therefore are the sources of the equity and variance premia. The parameter $\theta_1$ represents the constant part of the SDF. It measures any additional fixed discounting done by the investors, as the discount for time or to account for sample biases in the estimation of the true return volatility by a measure of realized volatility. The parameters $\theta_3$ and $\theta_5$ are the coefficients of the stock excess return and return variance at day $t$. The presence of these two variables at day $t$ increases the flexibility of the SDF. For some common parametric specifications of the historical dynamics, including the state variables at day $t$ in the SDF $M_{t,t+1}$ is necessary for considering the risk premia as free parameters. For instance, when the underlying asset return follows a Cox-Ingersoll-Ross process under the historical probability measure, if we set $\theta_3 = \theta_5 = 0$ in Equation (3.1) the no-arbitrage restrictions pin down the value of parameters $\theta_1, \theta_2, \theta_4$ uniquely as functions of the historical parameters (see Gagliardini, Gouriéroux and Renault [2011]). This degeneracy is avoided by including $r_t$ and $\sigma_t^2$ in the specification of the SDF and estimating the value of the parameters $\theta_3$ and $\theta_5$. Since investors can trade a risk-free asset, they discount future stock returns w.r.t. to the risk-free rate. Therefore, the chosen parametrization of the SDF given in Equation (3.1) involves the excess return. If investors could trade assets with a reference level of return variance, they would discount future return variance w.r.t. this level. Daily volatility swap rates are usually taken as risk-free levels of volatility.[25] These rates are derived from volatility swap contracts, that are OTC contracts, or approximated by using some option portfolios (see Carr and Wu [2009]). Volatility swap contracts are less liquid than the shares and options considered in this paper and approximations due to the estimation of volatility swap rates are avoided in the study. Therefore, the SDF is parametrized as a function of variances and not excess variances w.r.t. risk-free variance levels.

The exponential-affine specification of Equation (3.1) ensures the positivity of the SDF and then of the risk-neutral transition density. This parametrization of the SDF is common in the asset pricing literature. In continuous time, when coupled with affine specifications of the differential equation for the Markov process of the state variables, an exponential-affine specification of the SDF offers analytical tractability. A first example is in option pricing. An exponential-affine specification of

[25]The payoff of a volatility swap contract is the volatility premium converted to monetary units. This contract has zero market value at initiation. The absence of arbitrage in the market for these contracts makes the volatility swap rate equal to the true risk-neutral expectation of the value of the volatility multiplied by the gross return on the risk-free asset one day ahead (see e.g. Section 1.2 in Carr and Wu [2009]).

the SDF combined with a jump-diffusion state variables dynamics makes the computations of some transforms of the state variables feasible in closed form (see e.g. Hull and White [1987], Heston [1993], Duffie, Pan and Singleton [2000] and Duffie, Filipovic and Schachermayer [2003]). A second example is in equilibrium models. When the representative agent in the CCAPM model (see Lucas [1978]) has a power or CARA utility function, the implicit SDF is exponential-affine. The same happens in consumption-based asset pricing models with recursive utility (see e.g. Epstein and Zin [1989], Campbell and Cochrane [1999] and Bansal and Yaron [2004]). Similar manageability benefits are offered also in discrete time (see e.g. Gouriéroux, Monfort and Polimenis [2006] and Gouriéroux and Monfort [2007]).

### 3.1.2 American options

Let us express an American equity option price by means of the principle of optimality of dynamic programming (also known as Bellman's principle). Consider an American put option with strike price $K$ and written on a share with price $S$. At its expiration, i.e. if its time-to-maturity $h$ is null, the price of this contract is $(K - S)^+$. Otherwise, when $h \geq 1$, its price is the maximum between the early exercise payoff $(K - S)^+$ and the discounted risk-neutral expectation of the option price at the following day, conditional on the current information. The former is the value of the option price if it is exercised, the second if it is kept alive. Similar equations and definitions hold also for American call options. This way of representing the value of the option price is the same as in lattice methods (see e.g. Cox, Ross and Rubinstein [1979], Boyle [1988] and Ritchken and Trevor [1999]), regression-based Monte Carlo methods (see e.g. Longstaff and Schwartz [2001]) and other iterative integration methods (see e.g. Sullivan [2000]). The model-implied American option prices depend on both the SDF parameters vector and the transition density of the state variables. Let us use a notation for the option prices that highlights this dependence. The model-implied option price at day $t$ of an American put option with time-to-maturity $h$ and strike price $K$ computed by taking the SDF parameters vector $\theta$ and the transition density $f$ of the state variables is denoted by $P_t(h, K; \theta, f)$. If the contract is an American call option with the same option characteristics, the notation is $C_t(h, K; \theta, f)$.

The American put option price for any time-to-maturity and strike price can be expressed as the product of the underlying share price and the American put option-to-share price ratio $p$. To have this expression, let us use a result in Gagliardini and Ronchetti [2010] (GR). They show that, in the framework of the present paper, American-style options with payoff at exercise that is linearly homogeneous w.r.t. the underlying asset price are linearly homogeneous.[26] Hence, the American put option-to-share

[26] This property is shown in Merton [1973a] and Merton [1990] for more specific settings.

price ratio depends on the share price $S_t$ and the strike price $K$ by means of the moneyness strike $k_t = K/S_t$ only:

$$P_t(h, K; \theta, f) = S_t p(h, k_t, X_t; \theta, f), \tag{3.2}$$

for the value $X_t$ of the state variables vector at day $t$. While the American option price $P_t$ depends on time, the functional form of the American put option-to-share price ratio is time-invariant and allows for the description of the price of any American put option written on the same underlying asset. By using this time-homogeneous ratio we have the advantage of a common representation of the price of different financial assets. Specifically, if the only options written on the given share are put and call options, we can express all the option class prices by the share price and the American put and call option-to-share price ratios. Let us now make the value of the American put option-to-share price ratio $p$ explicit. Let us consider it for the moneyness strike $k$ and state variables $x$. At maturity, when the time-to-maturity $h$ is null, this value is just the exercise-to-share price ratio, i.e.:

$$p(0, k, x; \theta, f) = (k-1)^+ . \tag{3.3}$$

As shown in GR, at any day before maturity, when $h \geq 1$, this value is the maximum between the exercise-to-share price ratio and a discounted expected value of the American put option-to-share price ratio one day ahead:

$$p(h, k, x; \theta, f) = \max \left[ (k-1)^+, \mathrm{E}_f \left[ M_{t,t+1}(\theta) e^{r_{t+1}} p(h-1, k e^{-r_{t+1}}, X_{t+1}; \theta, f) \middle| X_t = x \right] \right], \tag{3.4}$$

where $\mathrm{E}_f \left[ \cdot | \, X_t = x \right]$ is the conditional expectation w.r.t. the transition density $f$ given the value $x$ of the state variables. These quantities are the counterparts in ratio terms of the early exercise payoff and the continuation value of an American put option at day $t$. The daily share gross return $e^{r_{t+1}} = S_{t+1}/S_t$ in the continuation value-to-share price ratio accounts for the fact that we deal with option-to-share price ratios and not just with prices. Similar equations and definitions hold for the American call option price $C_t$ and the American call option-to-share price ratio $c$.

## 3.2 Estimation approaches

In this section the employed estimation approaches are discussed. Section 3.2.1 introduces the no-arbitrage restrictions on share, risk-free asset and American options. Section 3.2.2 describes how they are taken into account in the estimation of the model parameters.

### 3.2.1 No-arbitrage restrictions

Let us say that at the current day we observe the price of $M$ American put options and $N$ American call options written on a single share of the considered stock.[27] Let us assume that these prices are consistent with the absence of arbitrage opportunities. This fact and the correct model specification ensure that the observed option prices coincide with the model-implied ones when the true value of the model parameters are used for the pricing of the options. Any calibration method is based on this match. From this relation and the homogeneity property of the American option price (expressed in Equation (3.2) for a put option) we get some restrictions on the true model parameters $\theta_0$ and $f_0$. To illustrate this concept, let us focus on the $j$-th observed put option, with moneyness strike $k_j^p$, time-to-maturity $h_j^p$ and option-to-stock price ratio $p_j$. We can compute the put option-to-stock price ratio evaluated at the value $k_j^p$ of the moneyness strike, at the value $h_j^p$ of the time-to-maturity and at the current value $x_0$ of the state variables. If we would use the true model parameters, the computed ratio would coincide with the observed ones:

$$p(h_j^p, k_j^p, x_0; \theta_0, f_0) = p_j, \tag{3.5}$$

for $j = 1, \dots M$. A similar match would be satisfied for the $i$-th observed call option with moneyness strike $k_i^c$, time-to-maturity $h_i^c$ and option-to-stock price ratio $c_i$:

$$c(h_i^c, k_i^c, x_0; \theta_0, f_0) = c_i, \tag{3.6}$$

for $i = 1, \dots N$. Equations (3.5)-(3.6) provide model restrictions, and since the put-call parity does not hold for American options (for whom only a weaker put-call relationship holds) there is no redundancy between them.

In addition to the restrictions on option prices, we must impose the infeasibility of any arbitrage strategy based on trades of the underlying share and a short-term non-defaultable bond. In other words, we must impose the martingale property for these two assets. The restrictions for the share and a non-defaultable zero-coupon bond that matures after a day are

$$\begin{cases} \mathrm{E}_{f_0}\left[M_{t,t+1}(\theta_0) e^{r_{t+1}} \mid X_t = x\right] = 1, \\ \\ \mathrm{E}_{f_0}\left[M_{t,t+1}(\theta_0) e^{r_{f,t+1}} \mid X_t = x\right] = 1, \end{cases} \tag{3.7}$$

respectively, for any conditioning value $x$ of the state variables.

[27] The discussion for options written on a lot of shares is equivalent.

Similar restrictions as Equations (3.5)-(3.6) and System (3.7) are adopted in many option pricing methodologies. For instance, let us consider a standard binomial tree for the risk-neutral dynamics of a share price with null risk-free rate and dividend yield and with the share price as unique state variable. In consecutive days, the share price can move from $S$ to $Su$, with probability $\tilde{p}$, or to $Sd$, with probability $1-\tilde{p}$. We exclude arbitrage opportunities on the share and impose $u\tilde{p}+d(1-\tilde{p}) = 1$. This last martingale restriction plays the role of the first equation in System (3.7) with parameters $\tilde{p}, u, d$. We then calibrate these parameters to the market price of a cross-section of financial derivatives written on the share. This last idea is the same as the one expressed by Equations (3.5)-(3.6).

Although all the model restrictions hold for any value of the state variables, not for any value of the state variables we can find empirical counterparts of all the restrictions. This is the reason why the restrictions for bond and share price in System (3.7) are introduced for any value $x$ of the state variables and the restrictions for the options in Equations (3.5)-(3.6) only for the value $x_0$. This difference is due to the fact that shares, bond and options have different trading frequency and volume, as explained in Section 3.3. At the current day we have at disposal a time series of arbitrage-free prices only for the share and the non-defaultable zero-coupon bond. Over the period covered by the time series a relatively large part of the domain of the state variables is realized. Differently, we have at disposal prices of the options only for the current value of the state variables. Therefore, we can build empirical counterparts of the model restrictions holding for several realized values in the case of share and bond and just for the current value of the state variables in the case of options.

### 3.2.2 Estimators

In this section the estimation approaches for the true value of the SDF parameters vector and transition density of the state variables are introduced. These methodologies make use of the empirical counterparts of the model restrictions given by Equations (3.5)-(3.6) and System (3.7) in different ways.

All the methodologies need a nonparametric kernel estimator of the historical transition density of the state variables. This estimator, for a time series sample of length $T$, is defined as

$$\hat{f}(x|\tilde{x}) = \frac{1}{h_T^2} \sum_{t=2}^{T} K\left(\frac{x_t - x}{h_T}\right) K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right) \Bigg/ \sum_{t=2}^{T} K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right), \tag{3.8}$$

where $K$ is a kernel function, $h_T$ is the bandwidth (see e.g. Bosq [1998]) and $x, \tilde{x}$ are generic values of the state variables vector.[28] Kernel estimators are largely used in financial applications (see e.g.

[28]Since the four arguments of function $\hat{f}$ are correlated, in the empirical application 2- and 4-dimensional kernel functions with 2- and 4-dimensional bandwidth matrices are used (see Appendix 5.1.1). These matrices are related to the variance-covariance matrix of the bivariate state variables process $(X_t)$.

Aït-Sahalia [1996a], Aït-Sahalia [1996b], Aït-Sahalia and Lo [1998], Pritsker [1998], Chapman and Pearson [2000], Hong and Li [2005], Hong, Tu and Zhou [2007] and Li and Zhao [2009]). Moreover, all the estimation methodologies use in different ways the no-arbitrage restrictions. In order to present the estimators in a compact form, let us introduce two vectors that collect these restrictions. The vector $U$ is defined as

$$U(x;\theta,f) = \begin{bmatrix} \mathrm{E}_f\left[M_{t,t+1}(\theta)e^{r_{t+1}} \middle| X_t = x\right] - 1 \\ \mathrm{E}_f\left[M_{t,t+1}(\theta)e^{r_{f,t+1}} \middle| X_t = x\right] - 1 \end{bmatrix}, \tag{3.9}$$

for any value $x$ of the state variables. The components of vector $U$ are the differences between LHS and RHS of the equations in System (3.7) computed at the generic value $(\theta, f)$ of the model parameters instead of $(\theta_0, f_0)$. The vector $L$ is defined as

$$L(\theta,f) = \begin{bmatrix} p(h_j^p, k_j^p, x_0; \theta, f) - p_j, \qquad \text{for } j = 1,\ldots,M \\ c(h_i^c, k_i^c, x_0; \theta, f) - c_i, \qquad \text{for } i = 1,\ldots,N \\ \mathrm{E}_f\left[M_{t,t+1}(\theta)e^{r_{t+1}} \middle| X_t = x_0\right] - 1 \\ \mathrm{E}_f\left[M_{t,t+1}(\theta)e^{r_{f,t+1}} \middle| X_t = x_0\right] - 1 \end{bmatrix}, \tag{3.10}$$

where the option-to-share price ratios are denoted in the same way as the option-to-stock price ratio in Section 3.2.1. The first $M + N$ components of vector $L$ are the differences between each model-implied and observed American option-to-share price ratios, for precise values of moneyness strike and time-to-maturity. The ratios are observed when the state variables have the current value $x_0$. The last 2 components of vector $L$ are the components of vector $U$ for the conditioning value $x_0$ of the state variables. In this way the restrictions that hold for any value of the conditioning state variables are gathered in vector $U$ and the restrictions that hold just for the current value $x_0$ of the state variables in vector $L$. Valuing vectors $U$ and $L$ at the generic value $(\theta, f)$ of the model parameters means that the model-implied American option-to-share price ratios and the conditional risk-neutral expectations are computed by using this value of the model parameters. From Equations (3.5)-(3.6) and System (3.7), vectors $U$ and $L$ are null for the true values $(\theta_0, f_0)$ of the model parameters. When they are valued at $(\theta, \hat{f})$, i.e. they are computed for a generic value $\theta$ and the kernel estimator $\hat{f}$, they collect the empirical restrictions for the value $\theta$ of the SDF parameters vector. For any given $\theta$, the model-implied American option-to-share price ratios in the first $M + N$ components of vector $L(\theta, \hat{f})$ are computed by a dynamic programming approach with kernel regressions.[29]

Let us now consider the three estimation methods for the SDF parameter $\theta$. The GMM method

[29]Each regression function in the computation of the continuation value is estimated by a Nadaraya-Watson estimator. This estimator is asymptotically equivalent to the conditional expectation operator computed by using the transition density defined in Equation (3.8).

minimizes a quadratic form in the empirical counterparts of the restrictions that hold for any value of the conditioning state variables, i.e. in $U(x;\theta,\hat{f})$ for any $x$. The calibration method does similarly considering the restrictions that hold just for the current value of the state variables, i.e. $L(\theta,\hat{f})$. The XMM method does similarly for all the restrictions. The XMM estimator $\hat{\theta}^{\star}$ of the SDF parameters vector is defined as

$$\hat{\theta}^{\star}=\arg\min_{\theta}\left[h_T^2 L(\theta,\hat{f})' L(\theta,\hat{f})+\frac{1}{T}\sum_{t=1}^{T}U(x_t;\theta,\hat{f})' U(x_t;\theta,\hat{f})\right], \tag{3.11}$$

where the time series of the state variables is up to the actual day, so that $x_T = x_0$. The criterion minimized in Equation (3.11) is a weighted sum of a cross-sectional calibration criterion (the first scalar product) and a GMM criterion (the time-averaged scalar product). The former takes into account the information contained in the data considered at the current date, the latter exploits the information contained in the time series of the state variables. The first component is multiplied by the square of the kernel estimator bandwidth to ensure convergence and asymptotic normality, as in GR. The second component is similar to the minimum distance criterion introduced in Ai and Chen [2003] to estimate conditional moment restrictions models and used in Nagel and Singleton [2011] in an application to conditional asset pricing models. In their most general formulation, the GMM, cross-calibration and XMM estimators minimize quadratic forms defined by some weighting matrices. The estimation of a particular weighting matrix, for instance the one that minimizes the asymptotic variance of the SDF parameters estimator (see Hansen [1982] for the GMM estimator and GR for the XMM estimator), could introduce additional statistical errors and lower the finite sample properties of the estimators (see e.g. Altonji and Segal [1996] for the GMM method). For this reason and to lower the computation burden, identity weighting matrices are used for all the criteria. The no-arbitrage restrictions for the actual value $x_0$ of the conditioning state variables are included in both vectors $L$ and $U(x_T;.,.)$. GR show that with this choice the asymptotic efficiency of the XMM estimator increases.

In this paper two nonparametric estimators of the transition density of the state variables are considered: the kernel estimator $\hat{f}$ defined in Equation (3.8) and an adjusted kernel estimator that makes use of an arbitrage-free pricing model. This last estimator, called full-information estimator of the transition density of the state variables, minimizes a statistical divergence from the kernel density estimator subject to the no-arbitrage restrictions. This divergence is derived from the Kullback-Leibler divergence $d$ of the transition density $f$ from the kernel density estimator $\hat{f}$. When the conditioning value of the state variables is $\tilde{x}$, the Kullback-Leibler divergence is defined as

$$d(f, \hat{f}|\tilde{x}) = \int \log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right) f(x|\tilde{x})dx. \tag{3.12}$$

The full-information estimator $\hat{f}^\star$ is defined as

$$\begin{aligned} \hat{f}^\star &= \arg\min_f \int d(f, \hat{f}|x)\hat{f}_X(x)dx, \\ & s.t. \begin{cases} L(\hat{\theta}^\star, f) = 0, \\ U(x; \hat{\theta}^\star, f) = 0, \text{ for all } x, \end{cases} \end{aligned} \tag{3.13}$$

where $\hat{f}_X$ is the kernel estimator of the historical unconditional density of the state variables:

$$\hat{f}_X(x) = \frac{1}{Th_T^2} \sum_{t=1}^{T} K\left(\frac{x_t - x}{h_T}\right). \tag{3.14}$$

The transition density $\hat{f}^\star$ is the minimizer of a constrained criterion. This criterion is the average Kullback-Leibler divergence weighted by the kernel density estimator $\hat{f}_X$. Equivalently, it is the kernel estimator of the unconditional expected Kullback-Leibler divergence between the transition density and its kernel estimator. The constraints for the criterion are the no-arbitrage restrictions evaluated by using the estimated SDF parameters vector $\hat{\theta}^\star$ defined in Equation (3.11). Probability density estimation through minimization of a statistical divergence subject to conditional moment restrictions has become popular in the literature on model calibration (see e.g. Buchen and Kelly [1996] and Stutzer [1996]) and on information-based approaches to GMM (see Kitamura and Stutzer [1997], Kitamura, Tripathi and Ahn [2004] and Gagliardini, Gouriéroux and Renault [2011]).

The full-information estimator $\hat{f}^\star$ defined in Equation (3.13) is an adaptation of the full-information estimator introduced in GR. Following similar steps, we get this expression:

$$\hat{f}^\star(x|\tilde{x}) = \frac{\hat{f}(x|\tilde{x})\mathcal{T}(x, \tilde{x}; \hat{\theta}^\star, \hat{f}^\star)}{\int \hat{f}(x|\tilde{x})\mathcal{T}(x, \tilde{x}; \hat{\theta}^\star, \hat{f}^\star)dx}, \tag{3.15}$$

where $\mathcal{T}$ is a tilting (or twisting) factor. We can adapt the characterization given by GR of their full-information estimator to the setting considered in this paper and express the tilting factor in terms of stock returns, SDF and some option-to-share price ratios. These last ratios are obtained by using the estimator $\hat{f}^\star$ itself. Hence the representation of the full-information estimator given in Equation

(3.15) is implicit and yields a fixed point problem. In this paper, an iterative procedure to solve this fixed point problem is implemented, in a similar way as the one suggested by GR. The estimator $\hat{f}^{\star}$ is computed numerically on a grid of points.

## 3.3 Data

In this section the data are described. Section 3.3.1 explains the criteria adopted for their selection. Section 3.3.2 deals with the empirical characteristics of the state variables and the risk-free rate. Section 3.3.3 illustrates the considered options.

### 3.3.1 Data construction

The IBM stock traded on U.S. centralized markets is considered during the period from $2006$ to $2008$. IBM stock is one of the most liquid stock during the period in U.S. centralized markets. The price of an IBM share traded at the NYSE during the period from $2006/01/03$ to $2008/08/29$ ($671$ business days) is taken from the NYSE TAQ database.[30] This price, expressed in USD, is available at high frequency. The geometric return on an investment in the IBM stock over the considered period is about $16\%$, being the last trading price of the day for an IBM share USD $82.07$ on $2006/01/03$ and USD $118.19$ on $2008/08/29$. The lowest and highest trading prices of an IBM share are USD $72.84$ on $2006/07/18$ and USD $130.89$ on $2008/07/23$. In the considered period, the IBM stock price showed an overall upward trend with no noteworthy sequence of returns of the same sign in following days. In the considered period there are on average $7.9$ millions trades of IBM shares per day. The daily dollar trading volume of the IBM stock is always included between USD $190$ millions and $3.2$ billions, with a mean value of USD $985$ millions. The difference between the highest ask and the lowest bid price of the same day for a single IBM share is always less than USD $5$. The percentage bid-ask spread computed by these ask and bid prices is always lower than $7\%$.[31]

Let us denote by $S_t$ and $D_t$ the last trading price of the day of an IBM share and the dividend announced at day $t$. IBM's dividends are usually paid on the 9th or 10th of March, June, September and December and announced about a month earlier.[32] The daily cum-dividend geometric return from

[30]The NYSE was the primary market for IBM shares. They were also traded in regional markets, as the Philadelphia Stock Exchange (PHLX) or, in the early 2006, the Pacific Stock Exchange (PSE).

[31]For an asset ask price $ASK_t$ and bid price $BID_t$, the percentage bid-ask spread is defined as $100\,(ASK_t - BID_t)\,/\,[(ASK_t + BID_t)\,/2]$.

[32]IBM announced a dividend of USD $0.2$ per share on $2006/02/08$, a dividend of USD $0.3$ on $2006/05/08$, $2006/08/08$, $2006/11/08$ and $2007/02/07$, a dividend of USD $0.4$ on $2007/05/08$, $2007/08/08$, $2007/11/07$, $2008/02/06$ and a dividend of USD $0.5$ on $2008/05/07$, $2008/08/06$. IBM has a long history of quarterly dividend payments. From $1998/01/01$ to $2008/08/29$ the time distance between two subsequent dividend announcement days has always been between $87$ and $94$ calendar days. IBM did not perform any stock split in the considered period. Neither merger nor acquisition took place. In 2007 IBM spun its Printing Systems Division off.

day $t$ to day $t+1$ on an IBM share is defined as

$$r_{t+1} = \log\left(\frac{S_{t+1} + D_{t+1}}{S_t}\right). \tag{3.16}$$

A daily realized volatility (RV) is taken as a proxy of the true daily spot volatility. This proxy does not rely on any parametric specification of the stock return dynamics. The RV is defined by 193 intra-day IBM share trading prices $S_{j,t}$, for $j = 1, \ldots, 193$, at 2 minutes frequency[33] from $9:35$ a.m. to $3:59$ p.m.:

$$RV_t = \sqrt{\sum_{j=2}^{193}\left(\log\left(\frac{S_{j,t}}{S_{j-1,t}}\right)\right)^2}. \tag{3.17}$$

When the process $(S_t)$ is a square integrable semi-martingale, the realized volatility $RV_t$ converges in probability to the quadratic variation at day $t$ of the log-price process $\log(S_t)$ as the number of intra-day trading prices increases (see e.g. Protter [2004]). For instance, this is the case when the data generating process is a continuous time stochastic volatility or jump-diffusion model (see e.g. Andersen, Bollerslev, Diebold and Ebens [2001] and Andersen, Bollerslev, Diebold and Labys [2003]). The bias induced by the micro-structure effects is small. The IBM shares are traded at an extremely high frequency in the considered period. The NYSE tick-size is USD $0.1$, so that the return rounding and discreteness effects, as staleness, are present but negligible.[34] For instance, for the IBM share price USD $73.58$, that is the lowest IBM share closing price in the period, the four possible returns closest to zero are about $\pm 0.12\%$ and $\pm 0.06\%$. Even if there is a consistent rise in the absolute value of returns and RV and a significant decline in the risk-free rate in July 2007 (see Figure 3.1), no structural change is considered in the time series of the state variables. The last world financial crisis starts, or at least gets worse, in the summer of 2007 and festers in September 2008.[35] The IBM stock plunges in the fall of 2008: the IBM share price at close is USD $71.595$ on $2008/11/19$, then with a geometric return of $-21.8\%$ from the end of August 2008. The extreme events occurred after August 2008 make doubtful the assumption of the same data generating process as before.[36] The analysis is then restricted to the

[33]Two following transaction times are not always distant 2 minutes. For instance, the two transaction times just before 3 o'clock can be $2:56:39$ p.m. and $2:58:59$ p.m.. This inhomogeneity in the time spacing is not considered in the RV computation. High frequency data were not available at the following days: $2007/04/06$, $2007/07/03$, $2007/11/23$, $2007/12/24$, $2008/07/03$. At these days the share closing price is taken from the Ivy DB OptionMetrics database and the volatility is estimated by a linear interpolation between the RV volatility at the previous and following days.

[34]IBM shares are between the shares of the DJIA with the highest level, so that the return grid for IBM shares is finer than the one of most of the other members of the index. The method described in Aït-Sahalia, Mykland and Zhang [2005] is useful to determine the optimal sampling frequency in applications of the estimation method described in this paper accounting for microstructure effects.

[35]In the literature there is not consensus about the starting date of the crisis. See The Squam Lake Working Group on Financial Regulation [2010] for a review of the major episodes and Lo [2011] for a recent review on some different prospectives.

[36]In September 2008 the interbank lending froze (the TED spread, measure of tightness in the interbank market, sky-

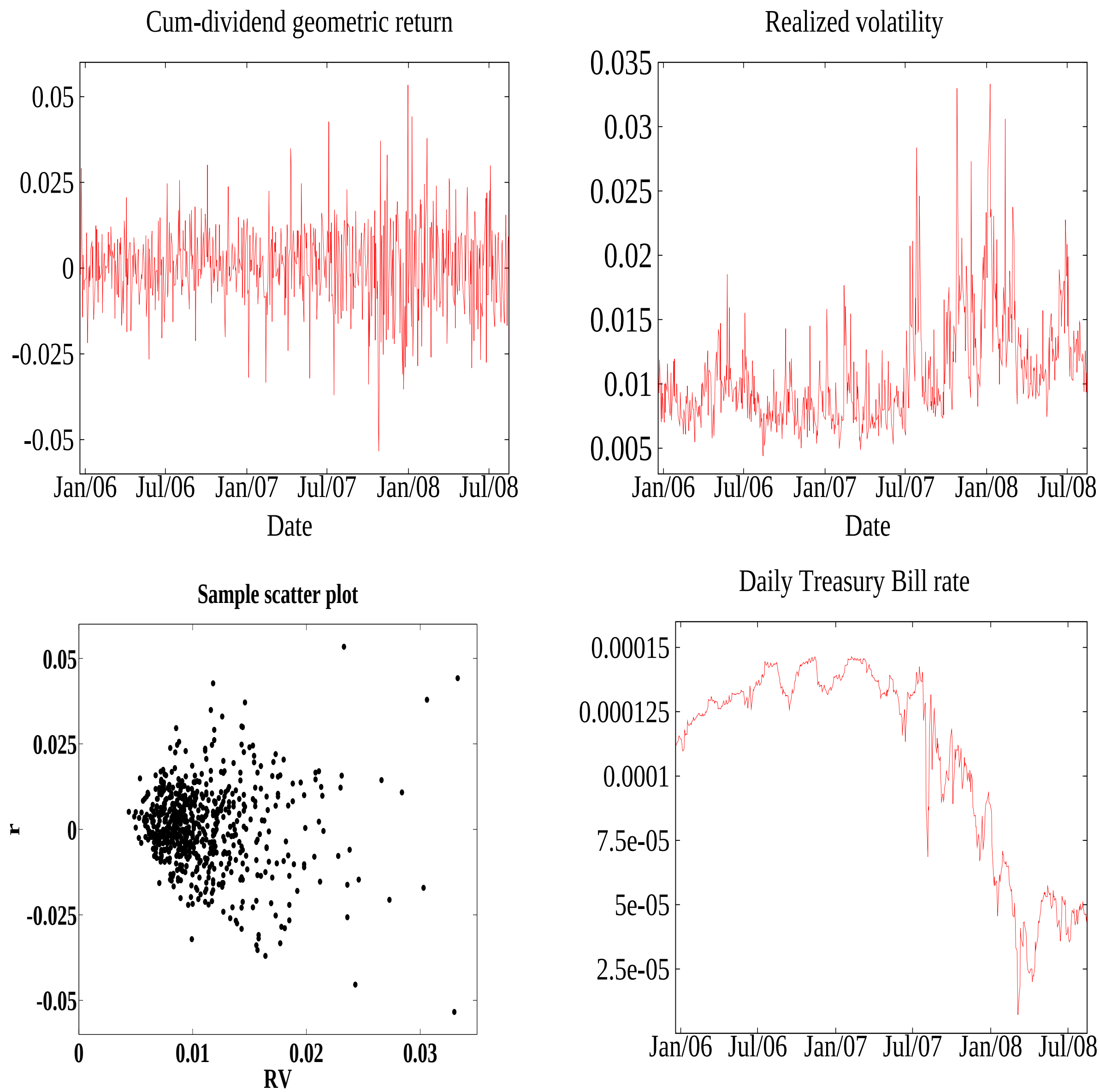


Figure 3.1: Time series of stock returns, realized volatility and risk-free rate during the period from $2006/01/03$ to $2008/08/29$. In the upper left panel we see the daily cum-dividend geometric stock return. In the upper right panel we see the stock return realized volatility from intraday geometric returns at the $2$ minutes frequency. In the lower left panel we see the scatter plot of the joint realizations of stock returns and realized volatility. In the lower right panel we see the historical realizations of the daily risk-free rate obtained by the daily $1$-month T-bill rate.

period from 2006 only up to the end of August 2008. While considering the same data generating process for the state variables on the whole period, we can distinguish between a (relatively) stable period, before July 2007 and composed by 376 business days, and a following (relatively) volatile one, composed by 295 business days. The partial sample median RV is about 0.009 and 0.013 before and after July 2007, respectively. Moreover, about two thirds of all the RV observations are in the range $[0.007 : 0.011]$ during the first period and in the range $[0.009 : 0.018]$ during the second period. The rise in IBM stock return volatility takes place in a period of growing widespread concerns about prices and ratings of financial assets, liquidity risk and counterparty risk, exacerbated in July 2007. Interbank lending sharply declines, assets prices drop, liquidity crunch spreads around the shadow banking system and a sharp contraction in real economies begins to be largely foreseen.[37]

The mid-quotes for the American call and put options written on IBM shares in the period 2008/07/01 to 2008/08/29 (43 business days), i.e. in the last two months of the considered period, are obtained by the Ivy DB OptionMetrics database. In these months options on IBM were multiple listed, traded at the CBOE, PHLX and AMEX (nowadays part of NYSE). The closing time for these exchanges was 4 : 00 p.m. New York time. The Ivy DB OptionMetrics database reports the highest ask and the lowest bid price at close across the U.S. exchanges. These values are used to compute the mid-quotes at the close of the exchanges. The unit of trading (or contract lot size) for these options is standard: only the trades of lots of options on 100 shares are allowed.[38] U.S. equity put and call options traded in centralized markets expire on the third Saturday of the month and are closed for trading the previous Friday. The markets provide at any business day the quotes of options for at least four different expiration months.[39] The two earliest expiration months are the current and the next one. The other two months are chosen on the base of some options issuing cycles. IBM belongs to the January cycle of the U.S. equity options so that the last two expiration months are the earliest between January, April, July and October. Then, at any business day in July and August 2008, 1- and 2-months options and options expiring in October and January are quoted. The average daily put option volume (that is the total amount of put contracts traded in a day) is 25490 contracts. The average daily call option volume

rocketed to over 450 basis points), Lehman Brothers failed, Fannie Mae and Freddie Mac were nationalized, Bank of America acquired Merill Lynch, the FED announced that Goldman Sachs and Morgan Stanley had been asked to turn into commercial banks, AIG was rescued by the U.S. Government, the Reserve Primary Fund "broke the buck", JPMorgan Chase acquired Washington Mutual Bank, short selling on many stocks was banned.

[37] In July 2007 the market for asset-backed commercial papers began to dry up, the asset-backed securities indices started a decline, the TED spread started to fluctuate from around 100 to around 200 basis points (see e.g. Figures 1-3 in Brunnermeier [2009] and Figure 1 in Stanton and Wallace [2011]), the rescue of IKB Deutsche Industriebank opened a series of bailouts in Europe, American Home Mortgage Investment Corporation announced its financial difficulties, the SEC relaxed the uptick rule for stocks traded at NYSE.

[38] Each of these contracts has physical settlement, that means that the delivery of 100 IBM shares must take place at exercise.

[39] The discussion does not apply to Long-term Equity AnticiPation Securities (LEAPs), that are options with time-to-maturity greater than two years when first listed. They usually expire in January.

is $12596$ contracts.

Option data are filtered on the base of several criteria. Many options are not considered: options with percentage bid-ask spread at close higher than or equal $100\%$, options with daily trading activity lower than or equal to $500$ contracts, options with time-to-maturity longer than $300$ business days, options with a moneyness strike less than $0.75$ or bigger than $1.25$. The database tick-size is USD $0.05$. The lowest option mid-quote that can survive this filter is USD $0.075$, and since the maximum value reached by an IBM share at close in the period from $2008/07/01$ to $2008/08/29$ is USD $130.03$, the lowest option mid-quote-to-share price ratio that in principle can be considered is $5.8$ E$-6$. As a consequence, this is the highest precision that can be reached in the computation of option mid-quote-to-share price ratios. The time-to-maturity filter implies to retain only options with time-to-maturity shorter than $480$ calendar days. The percentage bid-ask spread filter does it for options with an ask price at close at most three times the contemporaneous bid price. Options with null bid at close are automatically excluded.

The reference daily risk-free rate is obtained from the $1$-month Constant Maturity Treasury Rate. The U.S. Department of the Treasury provides publicly this rate, that comes from an interpolation of the daily yield curve, on an annualized basis.[40] T-bills, considered as free of default risk, are more liquid and with a broader secondary market than other assets traded in financial markets at the same time.

### 3.3.2 State variables realizations

We see in the two upper panels of Figure 3.1 the time series of the state variables at the NYSE closing time for the period from $2006/01/03$ to $2008/08/29$. In Figure 3.1 we see the related scatter plot. In Table 3 we find sample unconditional mean, standard deviation, skewness, kurtosis, minimum and maximum values and median for both stock return and RV. The $5$th, $10$th, $25$th, $75$th, $90$th, $95$th quantiles and the inter-quartile range of their distributions are also reported in this table.[41] The sample mean and median of the return are positive and lower than $0.1\%$, with standard deviation and inter-quartile range close to $1.5\%$. The empirical distribution of the returns is negatively skewed. This means that the IBM stock suffers the unconditional left tail risk that is typical in the equity markets. The sample kurtosis of the returns exceeds by about $1.91$ the kurtosis of the standard normal distribution: the unconditional return distribution is leptokurtic. The sample unconditional Sharpe ratio over a one-day horizon is $0.0443$: the expected excess return for an investment in the IBM stock normalized by its

[40]The Treasury yield curve is estimated daily using a cubic spline proprietary model. Inputs to the model are primarily bid-side yields for on-the-run Treasury securities. See more at http://www.ustreas.gov.

[41]For the descriptive statistics of IBM stock return in the period from $1970/01/02$ to $2008/12/31$ see also Tsay [2010], p. 11.

standard deviation is slightly positive.[42] The time series of the RV has mean and median close to $1\%$, with standard deviation and inter-quartile range close to $0.4\%$. The distribution of the RV is positively skewed with excess kurtosis equal to $4.60$. Also Andersen, Bollerslev, Diebold and Ebens [2001] find positive skewness and kurtosis for the RV of the $30$ DJIA stocks, among which there is the IBM stock, in the period from January $1993$ to May $1998$ (see Table $2$ in their article).

In the lower right panel of Figure 3.1 the time series of the risk-free rate is displayed. The value of this rate is quite stable at about $1.35$ E$-4$ before July $2007$ and declines afterwards, reflecting the weakening of financial markets. Some of the sample unconditional properties of this rate are reported in Table 3. The sample mean and median are close to one sixth and one fifth of the corresponding statistics of the IBM stock return, respectively.

For the purpose of data description, let us consider some contemporaneous and lagged statistical relationships between the state variables and between the return and the logarithmic RV. Let us first consider some correlation properties and then the estimation results of a linear vector autoregressive model for the return and logarithmic RV.

We see in the four panels of Figure 3.2 the values of some sample coefficients of unconditional auto- and cross-correlation of stock return, RV and logarithmic RV. The correlation coefficients are denoted by $\rho$ and are displayed as functions of the lag index. For a given lag index, the coefficients pertaining to the RV and logarithmic RV are coupled. The coefficient related to the RV is always displayed on the left, the other on the right. The $95\%$ confidence level bounds are represented by the horizontal lines. We see in the upper panels the first $20$ coefficients of auto-correlation for the returns and the first $40$ coefficients for the RV and logarithmic RV. At the $95\%$ confidence level, the autocorrelation coefficients for the return are not significant, with the exception of the $8$-th lag. The autocorrelation coefficients for the RV and the logarithmic RV are statistically significant at this level instead. We see in the lower panels the first $20$ sample coefficients of the correlation between the return and RV and the return and logarithmic RV. Few of these coefficients are marginally significant, all the others are not significant. In particular, the contemporaneous correlation coefficients between the return and RV or logarithmic RV are negative and marginally significant, indicating a contemporaneous leverage effect. The first three correlation coefficients between the return and the lagged RV or logarithmic RV are negative and statistically significant, indicating a lagged volatility feedback effect. Then, the returns are serially uncorrelated, the RV and logarithmic RV are auto-correlated and there is a cross-correlation between the return and the contemporaneous or lagged RV and logarithmic RV.

To conclude the description of the historical realizations of the state variables, let us consider a

[42] Treasury rate data are not available at $2007/10/08$ and $2007/11/12$ and a linear interpolation to get their proxies is used.

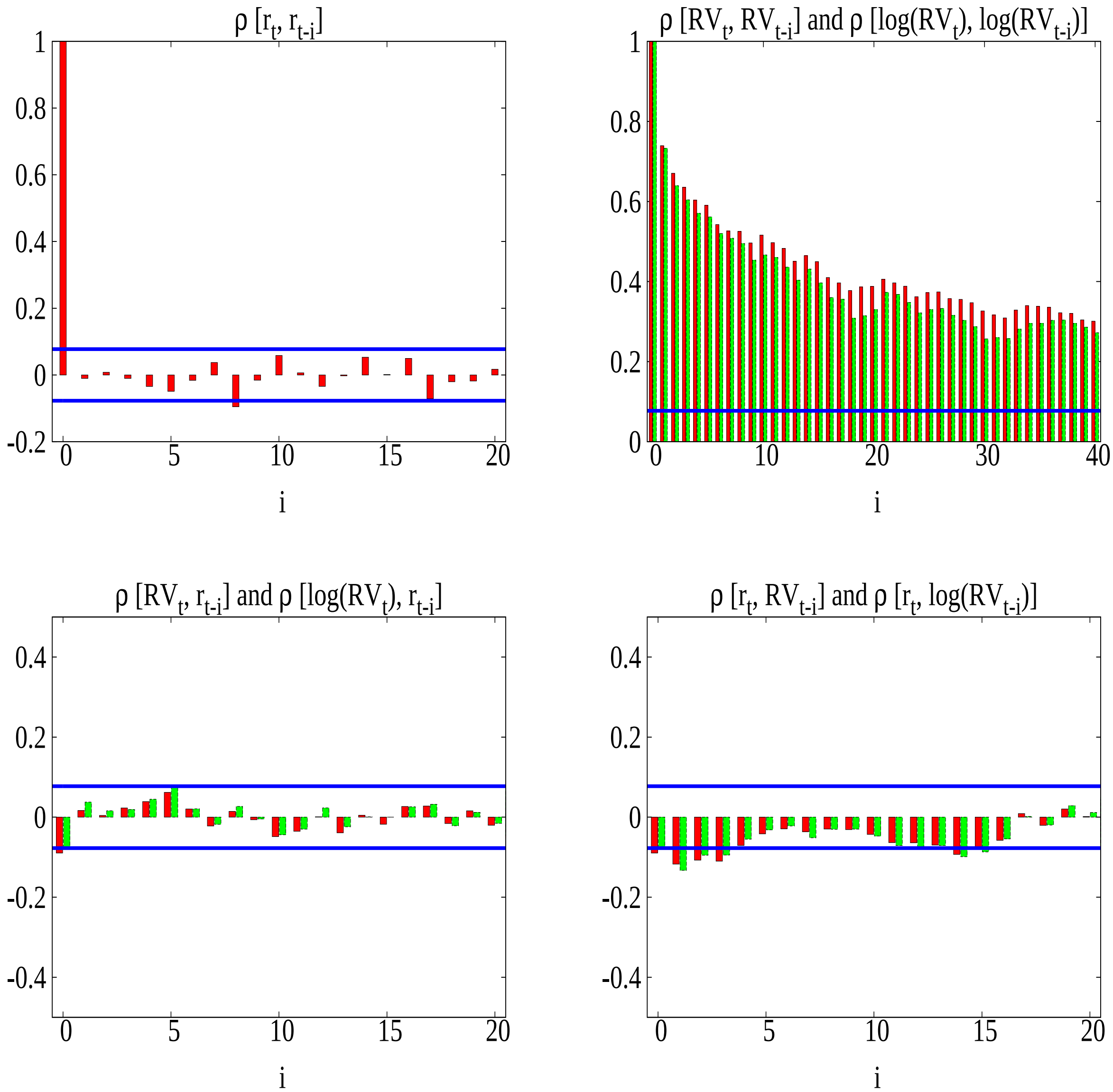


Figure 3.2: Sample auto- and cross-correlation for the daily cum-dividend geometric return, realized volatility and logarithmic realized volatility for the IBM stock price in the period $2006/01/03 - 2008/08/29$. The correlations are displayed as functions of the lag index. The coefficients concerning the RV or logarithmic RV for the same lag index are coupled, with the former displayed on the left and the latter on the right. The $95\%$ confidence interval ($2$ standard error bounds) for the coefficients is $[-0.0773 : 0.0773]$ and its borders are displayed by the straight horizontal lines.

linear vector autoregressive model of order 1 for the return and logarithmic RV:

$$\begin{cases} r_t = \mu_1 + \phi_{1,1} r_{t-1} + \phi_{1,2} log\left(RV_{t-1}\right) + \epsilon_{1,t}, \\ \\ log\left(RV_t\right) = \mu_2 + \phi_{2,1} r_{t-1} + \phi_{2,2} log\left(RV_{t-1}\right) + \epsilon_{2,t}, \end{cases} \tag{3.18}$$

where the exogenous innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ at time $t$ are $i.i.d.$ over time, have zero mean and finite variance and are possibly correlated. The coefficients $\phi_{1,1}$ and $\phi_{2,2}$ are the autoregressive coefficients that describe the impact on the current value of return and logarithmic RV of its own lagged value. The coefficient $\phi_{1,2}$ is a measure of the impact of lagged logarithmic RV on the current return and then of the one-day delayed volatility feedback effect (see e.g. French, Schwert and Stambaugh [1987] and Campbell and Hentschel [1992]). The coefficient $\phi_{2,1}$ is a measure of the impact of lagged return on the current logarithmic RV and then of a one-day delayed leverage effect (see e.g. Black [1977], Christie [1982] and Nelson [1991]). The ordinary least squares estimates of the model coefficients are reported in Table 4. Every root of the characteristic polynomial lies inside the unit circle. The coefficients of the first equation are not statistically significant at the $95\%$ confidence level and the $R^2$ of this regression equation is lower than $0.1\%$. This last finding is consistent with the idea that the returns cannot be predicted by past state variables (see e.g. Fama [1970] and Leroy [1982]). On the contrary, the coefficients of the second equation are statistically significant at the $5\%$ confidence level and the $R^2$ of this regression is $55\%$. This is consistent with the hypothesis that the logarithmic RV is predictable to some extent by lagged state variables. The estimate of the coefficient $\phi_{2,1}$ is $-1.44$. This last negative estimate is in agreement with the assumption of a delayed leverage effect.[43] The estimate of the autoregressive parameter $\phi_{2,2}$ for the logarithmic RV is $0.73$. The positive estimate is in agreement with the observed volatility persistence that we can recognize in the upper right panel of Figure 3.1 and with the sample auto-correlation coefficients displayed in the upper right panel of Figure 3.2. In Table 5 the ordinary least squares estimates of variance, covariance and correlation of the innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are reported. Every estimate is reported with its $95\%$ bias-corrected and accelerated bootstrap confidence interval (see e.g. Efron and Tibshirani [2000], ch. 14), computed by using 9999 bootstrap samples (see Andrews and Buchinsky [2003] for this number). The estimates of the variance of innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are $1.44$ E$-4$ and $0.05$, respectively. The former value is very close to the sample variance of the return, the latter is smaller than the sample variance of the logarithmic RV, that is about $0.34$. This finding is consistent with the fact that only the coefficients of the second equation of System (3.18) are statistically significant at the $95\%$ confidence level. The

[43] The scale on $log\left(RV_t\right)$ is bigger than the scale on $r$. This explains the greater magnitude of the point estimate of parameter $\phi_{2,1}$ than the magnitude of the other point estimates.

estimated covariance and correlation between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are $-4.13$ E$-4$ and $-0.15$, respectively. This negative correlation is statistically significant at the $5\%$ confidence level.

| | Mean | STD | Skewness | Kurtosis | Minimum | Maximum |
|---|---|---|---|---|---|---|
| $r_t$ | 0.0006 | 0.0120 | −0.1172 | 4.9186 | −0.0534 | 0.0534 |
| $RV_t$ | 0.0108 | 0.0042 | 1.7820 | 7.5989 | 0.0044 | 0.0333 |
| $r_{f,t}$ | 1.0759 E−4 | 0.3768 E−4 | −0.8978 | 2.3274 | 0.0722 E−4 | 1.4639 E−4 |

| | 5th quantile | 10th quantile | 25th quantile | Median | 75th quantile | 90th quantile | 95th quantile | IQR |
|---|---|---|---|---|---|---|---|---|
| $r_t$ | −0.0201 | −0.0140 | −0.0060 | 0.0006 | 0.0080 | 0.0138 | 0.0185 | 0.0140 |
| $RV_t$ | 0.0064 | 0.0069 | 0.0080 | 0.0097 | 0.0126 | 0.0160 | 0.0188 | 0.0046 |
| $r_{f,t}$ | 0.3778 E−4 | 0.4528 E−4 | 0.7944 E−4 | 1.2639 E−4 | 1.3611 E−4 | 1.4333 E−4 | 1.4500 E−4 | 0.5667 E−4 |

Table 3: Unconditional sample properties of stock returns, realized volatility and risk-free rate. The sample mean, standard deviation (STD), skewness, kurtosis, minimum and maximum values, median for the three variables are reported. The 5th, 10th, 25th, 75th, 90th, 95th quantiles and the inter-quartile range (IQR) of their sample distributions are also displayed.

| | $\mu_1$ | $\phi_{1,1}$ | $\phi_{1,2}$ | $\mu_2$ | $\phi_{2,1}$ | $\phi_{2,2}$ |
|---|---|---|---|---|---|---|
| Point estimate | 0.0032 | −0.0089 | 0.0006 | −1.2149 | −1.4410 | 0.7348 |
| Standard error | 0.0064 | 0.0390 | 0.0014 | 0.1200 | 0.7357 | 0.0261 |
| t-statistic | 0.5050 | −0.2292 | 0.4039 | −10.1210 | −1.9588 | 28.1340 |

Table 4: Ordinary least squares estimates of the coefficients of the linear vector autoregressive model in System (3.18).

| | $\mathrm{V}[\epsilon_{1,t}]$ | $\mathrm{V}[\epsilon_{2,t}]$ | $\mathrm{Cov}[\epsilon_{1,t}, \epsilon_{2,t}]$ | $\mathrm{Corr}[\epsilon_{1,t}, \epsilon_{2,t}]$ |
|---|---|---|---|---|
| Point estimate | 1.4409 E−4 | 0.0515 | −4.1313 E−4 | −0.1523 |
| Bootstrap 95% confidence interval | [1.2499 : 1.6868] E−4 | [0.0457 : 0.0592] | [−6.9079 : −1.3630] E−4 | [−0.2463 : −0.0499] |

Table 5: Ordinary least squares estimates of variance, covariance and correlation of the innovations of the linear vector autoregressive model in System (3.18). The 95% bias-corrected and accelerated bootstrap confidence intervals, computed by using 9999 bootstrap samples, are also reported.

### 3.3.3 Options

By filtering the Ivy DB OptionMetrics database as described in Section 3.3.1, the considered dataset is composed by 248 put options and 221 call options. The mean percentage bid-ask spread for the options is 11.16%, with 10.68% for the put options and 11.69% for the call options. The 90th percentile of the percentage bid-ask spread is 28.57%, with 25% for the put options and 28.57% for the call options. At each day the total number of selected options is between 4 and 23. The number of put options considered at a given day varies from 2 to 12 and its mean value is 5. For the call options the same numbers apply.

The upper panel of Figure 3.3 shows the moneyness strike for the considered options as a function of the date. The moneyness strike varies approximatively in the range $[0.71 : 1.06]$ for the put options (indicated by crosses) and in the range $[0.85 : 1.18]$ for the call options (indicated by circles). Only 1% of the put options and 1.8% of the call options are at-the-money and 22% of the put options and 26% of the call options are in-the-money. Most of the considered options in the dataset are then out-of-the-money. 24% of the put options and 32% of the call options are close to the money, with moneyness strike ranging from 0.98 to 1.02.[44] The lower panel of Figure 3.3 shows the time-to-maturity for the considered options as a function of the date. The time-to-maturity varies from 1 to 163 business days for the put options and to 137 business days for the call options, once again indicated by circles and crosses, respectively. 82% of the put options and 84% of the call options have time-to-maturity up to 70 business days. These options are the ones with the highest trading volume and lowest percentage bid-ask spread in the dataset. This characteristic of the data is explained by the fact that closer is the expiration, higher is the rate of change in option value due to time (i.e. higher is the option theta in absolute value) and higher are the potential return and leverage.

The imaginary investor able to trade at the mid-quote and without incoming in frictions would not find any arbitrage opportunity in the considered option sample. At any day, the mid-quote $C$ of an American call option written on a single share with price $S$, with strike price $K$ and time-to-maturity $h$ is not greater than $S$ and smaller than $(Se^{-\delta h} - Ke^{-r_f h})^+$, for the dividend yield $\delta$ and risk-free rate $r_f$. Similarly, the mid-quote $P$ of the put option with the same option characteristics is not greater then $Ke^{-r_f h}$ and smaller than $(Ke^{-r_f h} - Se^{-\delta h})^+$. Then, there would not be any discount arbitrage opportunity. There would not be any bull and bear spread arbitrage opportunity: for any couple of contemporaneous mid-quotes of call (put) options with the same maturity, the mid-quote of the call

[44] During July and August 2008 some deep out-of-the-money options have been traded, but with lower trading volume and bigger percentage bid-ask spread than the other traded options. More volatile is the stock, more attractive is the trading of deep out-of-the-money options because of the return and leverage opportunities. The volatility in the considered two months did not provide an incentive to (relatively) large trades in deep out-of-the-money options.

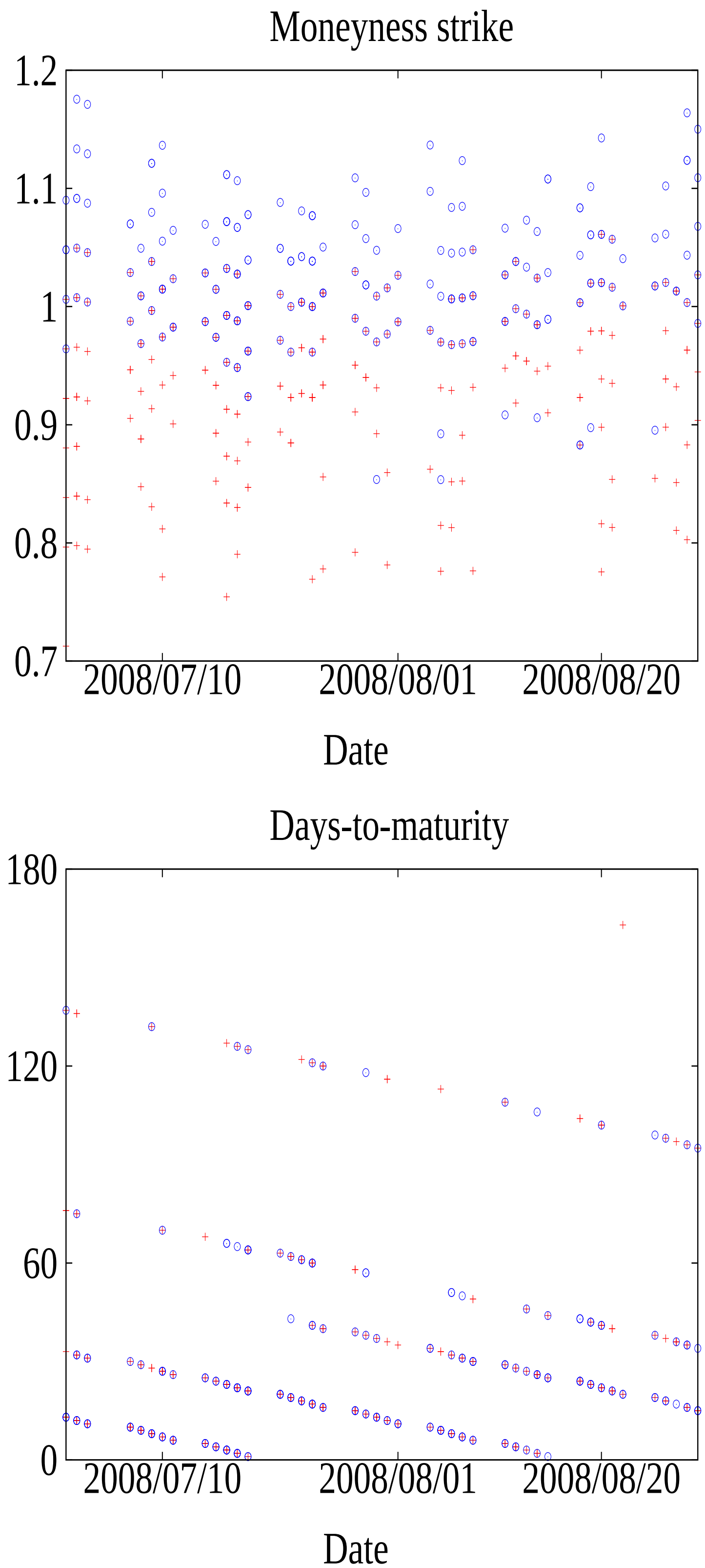


Figure 3.3: Characteristics of the considered American put and call options at any business day in July and August 2008. We see the moneyness strike of the options, in the upper panel, and their times-to-maturity, in the lower panel, as functions of the date. The time-to-maturity is expressed in business days. Crosses refer to the put options, circles to the call options.

(put) with the lower strike is not lower (higher) than the other.[45] There would not be any calendar spread arbitrage opportunity: for any couple of contemporaneous mid-quotes of call (put) options with the same strike price, the mid-quote of the call (put) option with the longer maturity is not valued less than the other.

## 3.4 Estimation results

This section discusses the estimates of the SDF and some characteristics of the historical joint dynamics of the state variables. Section 3.4.1 describes the results of the kernel estimation of the historical conditional correlation of the state variables, Sharpe ratio of an investment in the IBM stock and skewness and kurtosis of the returns. Section 3.4.2 reports the results of the XMM estimation of the SDF parameters vector. Section 3.4.3 describes the tilting factor $\mathcal{T}$ introduced in Equation (3.15) on the kernel estimator of the transition density. Finally, Section 3.4.4 describes the differences between the estimates of correlation function, Sharpe ratio, skewness and kurtosis of the returns obtained by using the kernel estimator $\hat{f}$ and the full-information estimator $\hat{f}^{\star}$.

### 3.4.1 Dynamic properties of the state variables without a no-arbitrage model

In this section fully nonparametric estimates of some properties of the historical dynamics of the state variables are described. The historical conditional correlation function between the state variables, the conditional Sharpe ratio of an investment in the IBM stock over a one-day horizon and the conditional skewness and kurtosis of the returns are considered. These quantities are estimated by using the kernel estimator $\hat{f}$ of the transition density of the state variables defined in Equation (3.8). The conditional expectations involved in the definitions of these quantities are estimated by a Nadaraya-Watson kernel regression function estimator. The matrix bandwidth for the kernel estimation is proportional to the one chosen by the multivariate generalized Scott's rule of thumb (see e.g. Hardle, Muller, Sperlich and Werwatz [2004], p. 73, and Simonoff [1996], ch. 4). The proportional constants for the four quantities are $2.5, 2.2, 2, 1.25$, respectively in the order they have been introduced in this section.[46] For these estimations, the full time series of the state variables for the period from $2006/01/03$ to $2008/08/29$ is used. The four considered estimators are asymptotically normal. The derivation of this behavior is reported in Appendix 5.1.2. The width of the $95\%$ confidence interval derived from the estimate of the asymptotic variance is smaller than the $10\%$ of the absolute value of the point estimate for most of values of the conditioning stock return and RV.

[45] For a given maturity, the call (put) option mid-quotes are convex decreasing (increasing) in the strike price.

[46] The development of a data-driven method for the selection of the optimal bandwidth matrix for the considered applications is beyond the scope of this paper.

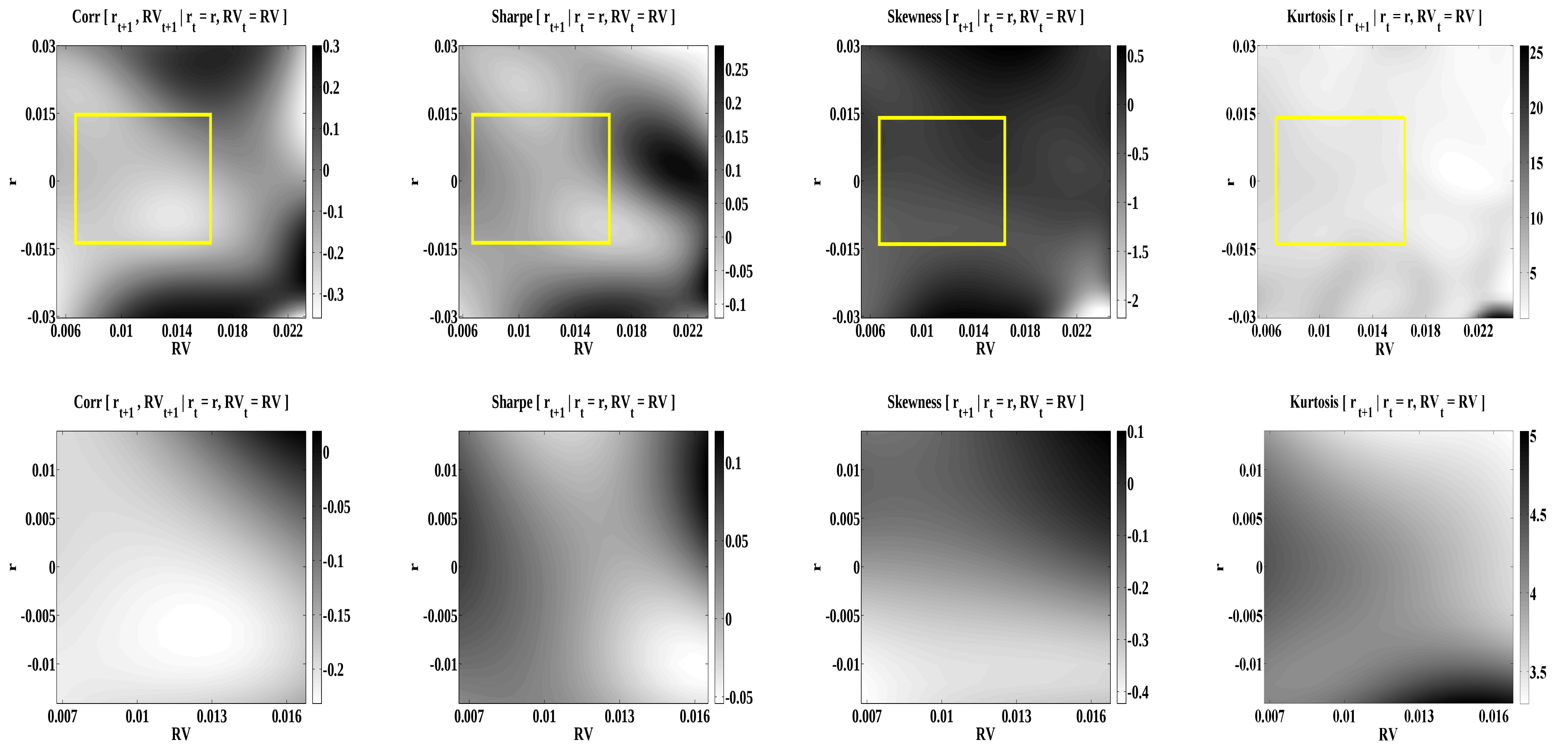


Figure 3.4: Level plots for the kernel estimates of the conditional correlation coefficient between stock returns and realized volatility, Sharpe ratio, skewness and kurtosis of the IBM stock return as functions of the conditioning state variables. The quantities are for a one-day horizon. In the upper panels, the range of the conditioning state variables on both the axes cover their sample 1st to 99th inter-quantile ranges. Each lower panel is the zoom of the area surrounded by the rectangle in the corresponding upper panel.

We see the estimates of the four quantities as functions of both the conditioning state variables in Figure 3.4. The value of each quantity is displayed by a contour plot. Each plot is presented along with its color legend on the right side. Darker (lighter) is the color of a point in the plot, higher (lower) is the value of the function. Because of space constraints, the $95\%$ confidence bounds are not reported. The figure is composed by eight panels. In the upper panels the axes cover the $1$st to $99$th inter-quantile range of the marginal sample distribution of stock return and RV, that is about $[-0.03 : 0.03]$ for the former and $[0.006 : 0.022]$ for the latter. In the lower panels the axes cover the $10$th to $90$th inter-quantile range of the marginal sample distribution of the variables, that is about $[-0.014 : 0.014]$ for the return and $[0.007 : 0.016]$ for the RV. All the estimated quantities vary over the conditioning state variable space, particularly for relatively high value of the conditioning RV. As a result, each conditional quantity can greatly differ from the values of its unconditional counterpart. As an implication for the analyzed time series of the state variables, in the (relatively) volatile period of the considered sample, i.e. after July $2007$, the quantities are more sensitive to the variations in the conditioning return than before. In each of the upper panels a rectangle surrounds the area that is zoomed in the lower panel. The joint historical realizations of the state variables outside the rectangle depicted in the upper panels are more sparse than inside. We see this in the lower left panel of Figure 3.1. Let us restrict in this section the analysis of the considered quantities for the values of the conditioning state variables considered in the lower panels. For these values the estimation is quite accurate.[47] For a positive conditioning return, the correlation between the state variables is negative, the Sharpe ratio is positive and the third and fourth moments of the distribution of the return are relatively close to those of a normal distribution. For a negative conditioning return, the correlation between the state variables is negative and stronger than for a positive conditioning return, the Sharpe ratio is almost null and the distribution of the return is leptokurtic and negatively skewed.

In the first lower panel of Figure 3.4 we see the estimated conditional correlation of the state variables. For most of the values of the conditioning state variables, the correlation is negative, consistently with a contemporaneous leverage effect. While the unconditional contemporaneous correlation between the state variables is about $-0.1$, as displayed in the two lower panels of Figure 3.2, its conditional counterpart varies approximatively in the range $[-0.2 : 0]$. The qualitative behavior of this function is increasing in the conditioning stock return. For a positive conditioning return, the function is increasing in the conditioning RV, while for a negative conditioning return the contemporaneous correlation between the state variables is quite stable at around $-0.2$. The overall behavior of this function is consistent with a contemporaneous leverage effect that gets more pronounced (i.e. more negative correlation between the state variables) for a negative conditioning return.

[47] The loss of accuracy in the estimation for values of the state variables outside the rectangle is due to boundary effects.

In the second lower panel of Figure 3.4 we see the estimated conditional Sharpe ratio over a one-day horizon. It varies approximatively in the range $[-0.05 : 0.1]$, corresponding to an annualized Sharpe ratio ranging in $[-0.79 : 1.58]$. This variation means that there is not a direct proportionality between the estimated conditional expectation of the excess return and the standard deviation of the return. The highest values of the Sharpe ratio are for very low values of the conditioning RV and for very high values of both the conditioning state variables. Its lowest values are for very high values of the conditioning RV and negative values of the conditioning return. In the former case the IBM stock is expected to outperform the T-bill, in the latter case to do the opposite. We can distinguish between two different behaviors of the Sharpe ratio for values of the conditioning RV that are lower than $0.013$, that is about the RV sample $75$th quantile, and higher values. In the former case, the Sharpe ratio varies weakly around $0$ as a function of the conditioning return. In the latter case, the Sharpe ratio is an increasing function of the conditioning return and varies in the range $[-0.05 : 0.1]$. This finding is consistent with the idea that higher is the RV, more strongly bad and good news on the return affect the expectation of the performance of the IBM stock.

In the third lower panel of Figure 3.4 we see the estimated conditional skewness of the returns. This function varies approximatively in the range $[-0.4 : 0.1]$. For most of the conditioning values of the state variables it is negative, pointing out the presence of a conditional left tail risk over a one-day horizon. The entity of this risk varies as the values of the conditioning state variables change. Higher are the values of the conditioning state variables, higher is the value of the conditional skewness, i.e. less negative return are expected to realize and lower is the conditional left tail risk. For extremely high values of the conditioning state variables, the estimated conditional skewness is positive. Only for these values of the conditioning state variables the stockholder does not face the left tail risk.

In the fourth lower panel of Figure 3.4 we see the estimated conditional kurtosis of the return. The function varies approximatively in the range $[3.5 : 5]$, showing that the sample conditional distribution of the returns is always leptokurtic. The highest (respectively lowest) values of this quantity are for high values of the conditioning RV and negative (respectively positive) values of the conditioning return. The estimated conditional kurtosis varies especially for high values of the conditioning RV. In this case, lower is the return, more extreme events are expected to occur.

### 3.4.2 Stochastic discount factor parameters

This section contains the description of the XMM estimates of the SDF parameters vector at each day of July and August $2008$. These estimates are compared with the cross-sectional calibration (CS) and GMM estimates of the same vector, i.e. the estimates obtained by a minimization of only the first or second part of the criterion in Equation (3.11).

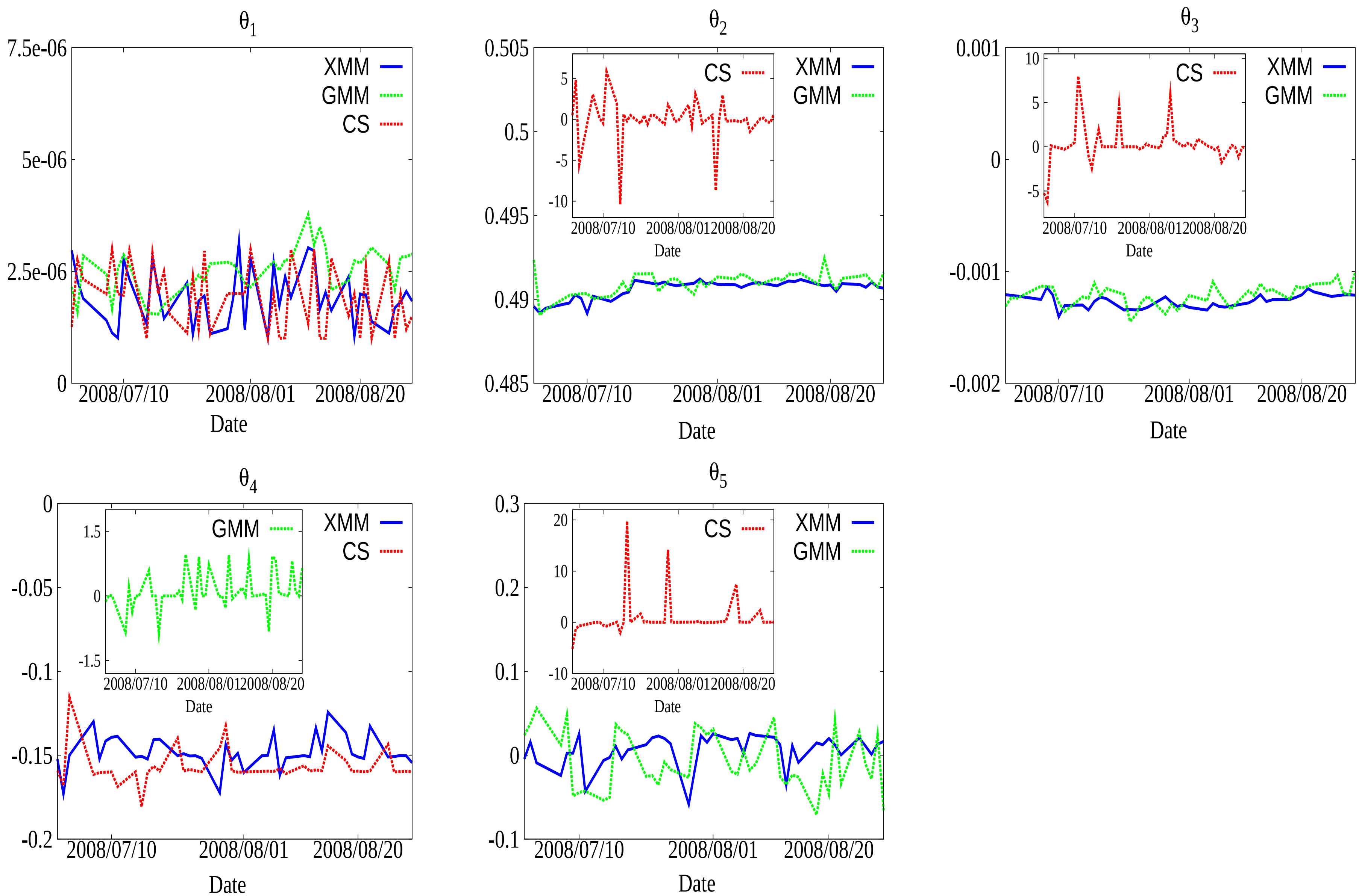


Figure 3.5: The SDF parameters estimated at each business day in July and August 2008. The solid line corresponds to the Extended Method of Moments (XMM) estimator, the two dotted line correspond to the Generalized Method of Moments (GMM) and Cross-Sectional (CS) estimators.

In Figure 3.5 we see the estimates of the SDF parameters vector as functions of the date. The XMM estimates are indicated by a solid line and the GMM and CS estimates by two different dashed lines. When estimates of the same parameter obtained by different methodologies vary in different ranges, their plots are on different axis scales. Specifically, the smaller inner graph plots the estimate that varies in the widest range. We find in Table 6 the sample mean value and standard deviation over time of the estimates of each SDF parameter obtained by using the three estimators.[48] The XMM and GMM methods give similar results in the estimation of $\theta_2$, $\theta_3$ and $\theta_5$, in terms of both mean and standard deviation over time. The XMM and CS methods do it for $\theta_4$ and all the three methodologies do it for $\theta_1$. The fact that neither the CS nor the GMM method provides the same results as the XMM supports the idea that the time series of state variables and the cross-section of option mid-quotes carry informations on the SDF parameters that are not redundant. The GMM method, that exploits the information content of the time series of share and non-defaultable zero-coupon bond prices, is able to estimate the market discount for future realizations of the excess return. It then provides a reliable estimate of parameter $\theta_2$, but it does not do the same for parameter $\theta_4$. Differently, the CS method, that exploits the information content of the cross-section of asset prices at the current date, is able to provide a reliable estimate of the market discount for future realizations of the return variance, measured through parameter $\theta_4$. It does not do the same for parameter $\theta_2$. This is consistent with the idea that the price of share and non-defaultable zero-coupon bond are sources of information mostly on the premium for the risk of a different excess return than expected and that option quotes have a similar role w.r.t. the return variance. The benefit given by the XMM method is combining the time series and cross-sectional information to get at the same time reliable estimates of parameters $\theta_2$ and $\theta_4$.

The quantity $-\theta_2 \left(r_{t+1} - r_{f,t+1}\right) - \theta_4 \sigma_{t+1}^2$ at the exponent in the RHS of Equation (3.1) depends on the realizations of the state variables at day $t+1$ and gives rise to the discount for risks. The mean over time of the XMM and GMM estimates of parameter $\theta_2$ is about $0.49$, with a standard deviation that is lower than $0.001$. For this value of parameter $\theta_2$, the discount of the asset-to-stock price ratios for the risk of a lower excess return than expected is about $1.0098$, $0.9997$ and $0.9910$ for the $5$th quantile, median value and $95$th quantile of the excess returns distribution. The mean over time of the XMM and CS estimates of parameter $\theta_4$ is about $-0.15$, with a standard deviation over time that is about $0.01$. For this value of parameter $\theta_4$, the discount of the asset-to-stock price ratios for the risk of an higher return variance than expected is about $1 + (0.61\text{ E}{-5})$, $1 + (1.41\text{ E}{-5})$ and $1 + (5.30\text{ E}{-5})$, respectively, for the $5$th quantile, median value and $95$th quantile of the RV distribution. The estimated positive sign of $\theta_2$ and negative sign of $\theta_4$ cause an overall shift of probability towards lower stock

[48] GR show that the XMM estimator of the SDF parameters vector is asymptotically normal.

| | XMM | GMM | CS |
|---|---|---|---|
| $\theta_1$ | 1.9063 E$-$6<br>(0.6255 E$-$6) | 2.4837 E$-$6<br>(0.5058 E$-$6) | 1.9007 E$-$6<br>(0.7301 E$-$6) |
| $\theta_2$ | 0.4906<br>(0.0005) | 0.4909<br>(0.0005) | 0.0531<br>(2.7723) |
| $\theta_3$ | $-$0.0013<br>(0.0001) | $-$0.0012<br>(0.0001) | 0.1719<br>(2.2241) |
| $\theta_4$ | $-$0.1483<br>(0.0095) | 0.0081<br>(0.0540) | $-$0.1571<br>(0.0100) |
| $\theta_5$ | 0.0061<br>(0.0189) | $-$0.0068<br>(0.0350) | 0.8351<br>(3.9606) |

Table 6: Estimates of the SDF parameters. The mean (upper number) and the standard deviation (lower number in brackets) over time of the estimates of each SDF parameter are reported. They are obtained by the Extended Method of Moments (XMM), Generalized Method of Moments (GMM) and Cross-Sectional (CS) estimation methodologies.

return and higher RV values when passing from the historical to the risk-neutral transition density of the state variables. Therefore, for pricing purpose more weights is put on negative stock return and high RV values. This result is consistent with the idea of risk averse investors more concerned about adverse outcomes than favorable ones. The CS estimate of parameter $\theta_2$ and the GMM estimate of the parameter $\theta_4$ vary strongly over time and do not lead to clear economic interpretations of the discount for risks.

The quantity $-r_{f,t+1} - \theta_1 - \theta_3 \left(r_t - r_{f,t}\right) - \theta_5 \sigma_t^2$ at the exponent in the RHS of Equation (3.1) depends on the values of the state variables at day $t$ and generates the fixed-discount. The estimate for parameter $\theta_5$ varies greatly at different days. The CS estimates for parameter $\theta_3$ are unstable over time. The estimate of parameter $\theta_1$ is of order E$-6$ and the GMM and XMM estimate of parameter $\theta_3$ are of order E$-3$. The risk-free rate is at least of order E$-5$ and about half of its realizations are of order E$-4$ (see Table 3). Since the excess return is at most of order E$-2$, the main contribution to the fixed-discount is given by the risk-free rate.

The dispersion in the estimates of the model parameters can be due to the statistical variability of the estimators and the model misspecification. We have model misspecification if no admissible SDF has the specification in Equation (3.1) or if some state variables are omitted. The first possibility realizes if no admissible SDF is monotonically decreasing in the state variables. The second possibility realizes if not only stock return and return volatility but also other factors affect the price of IBM shares. For instance, this is case if the IBM stock price is correlated with other financial assets or factors linked to the business IBM strategy or to the supply and demand of IBM shares. There is no statistical evidence in the data against the chosen model specification, and in particular against the presence of a discount for the risks of a lower excess return and an higher return variance than expected.

### 3.4.3 Tilting factor

This section contains the description of the tilting factor $\mathcal{T}$ introduced in Equation (3.15) at some days and for some values of the conditioning state variables. When moving from the historical to the risk-neutral transition density, we expect a shift of probability density toward a certain part of the state variables space. We expect this shift because of the hypothesis of risk-adverse investors who weight adverse outcomes more than positive ones for pricing. The findings reported in Section 3.4.2 and in particular the estimated signs of parameters $\theta_2$ and $\theta_4$ are consistent with this idea. Differently, since functions $\hat{f}$ are $\hat{f}^{\star}$ are both estimators of the historical transition density of the state variables, we do not expect any monotonic shift of probability density when passing from a function to the other. At most, if the economic model is not redundant for the estimation, we can just expect the full-information estimator $\hat{f}^{\star}$ to be a more precise estimator of the historical transition density of the state variables than the kernel estimator $\hat{f}$. In this case, we can expect that the estimates obtained by using the former estimator are more stable over time than the ones obtained by using the latter. This is exactly what it is empirically found and described in the following Section 3.4.4.

As an illustration, let us consider the estimated tilting factor just at $2008/08/06$ and $2008/08/20$ and for the contemporaneous value of the conditioning state variables. At both these days the RV has an intermediate value between its median for the entire period from January $2006$ to August $2008$ and its median for the (relatively) volatile period after July $2007$. At $2008/08/06$ the value of the return is about $0.003$ and the one of the RV is about $0.011$. The options considered at this day are $6$ put and $6$ call options. The put options have time-to-maturities up to $31$ business days and moneyness strikes included in the range $[0.81 : 1.00]$. For the call options these numbers are $50$ and $[0.96 : 1.08]$, respectively. At $2008/08/20$ the value of the return is almost null and the one of the RV is about $0.012$. The cross-section of options considered at this day is composed by $8$ put and $5$ call options. The put options have time-to-maturities up to $101$ business days and moneyness strikes included in the range $[0.77 : 1.06]$. For the call options these numbers are $101$ and $[1.02 : 1.14]$, respectively. At these two days the adjustments provided by the tilting factor to the kernel estimator of the transition density of the state variables is different. Specifically, for most of the conditioning values of the state variables, considering the option quotes leads to a correction of the point estimate that is opposite at the two days. At $2008/08/06$ the tilting factor concentrates more weight at around the mean of the kernel distribution. At $2008/08/20$ it shifts some probability weights towards the tails of the kernel distribution.

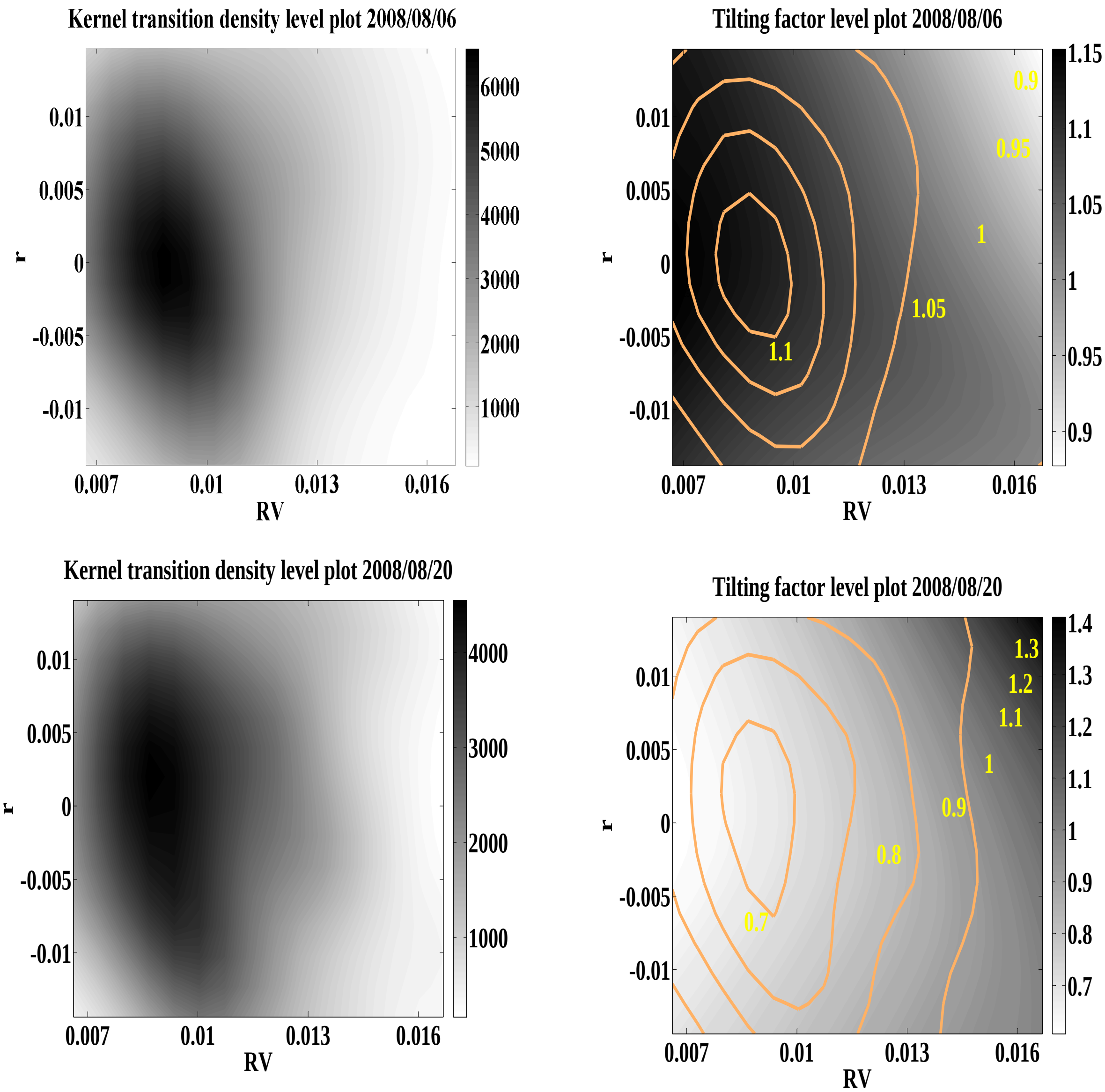


Figure 3.6: Level plots for the kernel estimator of the transition density of the state variables (left panels) and for the tilting factor normalized by its kernel expectation (right panels). The upper panels refer to $2008/08/06$, when $6$ put and $6$ call options with time-to-maturities up to $50$ business days are considered for the computation of the tilting factor, the value of the return is about $0.003$ and the value of the RV is about $0.011$. The lower panels refer to $2008/08/20$, when $8$ put and $5$ call options with time-to-maturities up to $101$ business days are considered for the computation of the tilting factor, the value of the return is almost null and the value of the RV is about $0.012$. In the right panels, the lines are level curves of the corresponding kernel estimator on the left, while the numbers indicate the value of the normalized tilting factor at the point where they are located.

We see in the left panels of Figure 3.6 the kernel estimator of the transition density of the state variables. The conditioning state variables are fixed to the value they assume on the considered day. The estimator is displayed in the form of a contour plot. The axes cover the values of the stock return and RV that correspond to their $10$th to $90$th inter-quantile ranges. The estimated kernel transition densities are both unimodal. The one that refers to $2008/08/06$ (upper panel) is more peaked than the other. The future outcomes are less uncertain at the first day, when the value of the RV is lower.

We see in the right panels of Figure 3.6 the estimator of the normalized tilting factor defined in Equation (3.15). It is the tilting factor normalized by its kernel expected value, i.e. the pointwise $\hat{f}^\star$ to $\hat{f}$ ratio. The conditioning stock return and RV are again fixed to the value they assume on the considered day. The bandwidth matrix is twice the one chosen by the Scott's rule of thumb. The normalized tilting factor is plotted in the form of a contour plot with the axes covering the same ranges as in the left panels. We see in these panels also some level curves of the kernel estimator of the transition density on the left and numbers indicating the value of the normalized tilting factor at some points. These level curves help to distinguish the areas where the option data bring a greater correction to the estimator of the transition density. In the areas with color level equal to $1$ the estimators $\hat{f}^\star$ and $\hat{f}$ coincide. The main modifications made by the tilting factor are on the tails of the transition density, i.e. close to the borders of the panels. At $2008/08/06$ the normalized tilting factor varies approximatively in the range $[0.9 : 1.15]$ (see upper right panel). It reaches its minimum for extremely high values of the state variables and its maximum for very low values of RV. At this day the tilting factor shifts the probability distribution toward lower values of RV. At $2008/08/20$ the normalized tilting factor varies approximatively in the range $[0.7 : 1.4]$ (see lower right panel). It reaches its minimum for low values of the RV and its maximum for very high values of stock return and RV. At this day the the tilting factor shifts the probability distribution toward higher values of the state variables. The contribution given by the options to a more accurate estimation of the quantities is more appreciable less accurate is the kernel estimation of the transition density of the state variables. Specifically, because of boundary effects, the kernel estimation is less accurate for extreme values of the conditioning state variables. The difference between the ranges of variation for the tilting factor at every day at July and August $2008$ is consistent with the idea that at a day with more uncertainty the adjustment provided by the options on the expected future outcomes is relatively greater.

### 3.4.4 Dynamic properties of the state variables with a no-arbitrage model

This section contains a comparison between the estimates of some dynamic properties of the state variables obtained by using the kernel estimator $\hat{f}$ and the full-information estimator $\hat{f}^\star$ at each day of July and August $2008$. As in Section 3.4.1, the considered quantities are the historical conditional correlation function between the state variables, the conditional Sharpe ratio of an investment in the IBM stock over a one-day horizon and the conditional skewness and kurtosis of the returns. In the first four panels of Figure 3.7 the estimates of the four quantities are displayed as functions of the date. They are computed for the contemporaneous value of the conditioning state variables. For each quantity, the dotted line, labeled "Kernel", indicates the estimate obtained by using $\hat{f}$, while the solid line, labeled "Tilted", indicates the estimate obtained by using $\hat{f}^{\star}$. We can see the daily stock return and RV during the two considered months in the center and right lower panels.

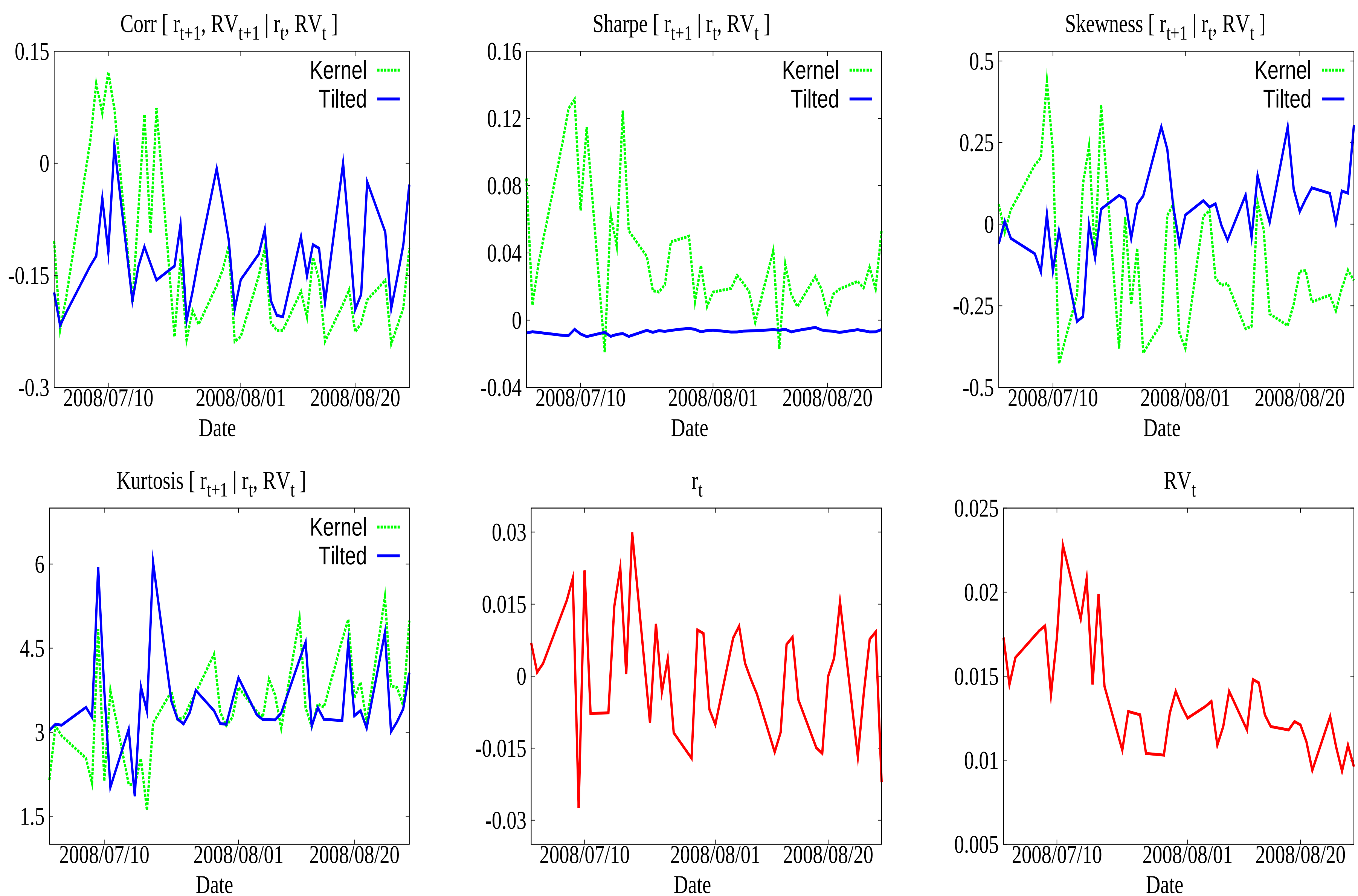


Figure 3.7: Estimates of the conditional correlation coefficient between stock returns and realized volatility, Sharpe ratio, skewness and kurtosis of the IBM stock return as functions of the date. The quantities are for a one-day horizon. We see in the first four panels the estimates obtained by using $\hat{f}$ (dotted line) and $\hat{f}^{\star}$ (solid line), with their 95% confidence intervals. We see in the center and right lower panels the values of the stock return and realized volatility.

The variation of the estimated quantities over time has potentially three causes: the change in the values of the conditioning state variables, the statistical variability in the estimation and any possible model misspecification. First, stock return and RV vary over time and the considered quantities are conditional on specific values of these variables. Second, the estimation of the model parameters is performed with different data samples. Third, the choices of the state variables and the parametrization of Equation (3.1) are assumptions. The contribution given by the two last points to the time variation of the estimated quantities is small. Concerning the different data samples, let us make two general considerations. First, the application of the kernel estimator to time series that consist of hundreds of observations that differ only for few of them most likely does not lead to statistically different results. Second, it seems reasonable to assume that the quotes of the considered options carry similar information about the data generating process. Moreover, as shown in Section 3.4.2, the XMM estimates of the SDF parameter vector are quite stable over time and this stability supports the validity of the assumptions. Then, as a whole, a major part of the variation in the time series of the estimated quantities is caused by the changing value of the conditioning state variables and only a minor part is due to the statistical variability.

The range of variation over time of the estimates obtained by $\hat{f}^{\star}$ is mainly smaller than the one of the estimates obtained by $\hat{f}$. Moreover, the range of variation of the latter estimates has two regimes: it is broader before $2008/07/17$, when the state variables have extreme values, and lower afterwards, when the state variables assume very high but not extreme values. In the first period the stock return varies approximatively in the range $[-0.03 : 0.03]$ and the RV does it in $[0.01 : 0.025]$, while in the second period the return varies approximatively in the range $[-0.015 : 0.015]$ and the RV does it in $[0.01 : 0.015]$. While we observe two regimes of variability for the estimates obtained by $\hat{f}$, no clear separation in different regimes appear when the estimation is performed by $\hat{f}^{\star}$. Adopting an arbitrage-free pricing model leads to estimates of some dynamic properties of the state variables that are more stable over time.

We see in the left upper panel of Figure 3.7 the time series of the estimates of the conditional correlation between the state variables. The estimates obtained by $\hat{f}^{\star}$ are almost always negative, while the estimates obtained by $\hat{f}$ vary approximatively in the range $[-0.25 : 0.15]$ before $2008/07/17$ and are negative afterwards. While the former estimates support the presence of a contemporaneous leverage effect, the latter does it only after $2008/07/17$.

We see in the center upper panel the time series of the estimates of the conditional conditional Sharpe ratio of an investment in the IBM stock over a one-day horizon. The estimates obtained by $\hat{f}$ and by $\hat{f}^{\star}$ give very different results. The former are almost always positive, varying approximatively in the range $[-0.02 : 0.14]$ before $2008/07/17$ and in the range $[-0.02 : 0.06]$ afterwards. The latter are extremely stable over time with value at about $-0.01$. Then, while considering only the time series of stock return and RV encourages, in terms of Sharpe ratio, almost always a long position in the IBM stock, taking into account this time series and the option data in an arbitrage-free pricing model provides an opposite suggestion.

We see in the right upper panel the time series of the estimates of the conditional skewness of the returns. The estimates obtained by $\hat{f}$ vary approximatively in the range $[-0.5 : 0.5]$ before $2008/07/17$ and are almost always negative afterwards. Differently, the estimates obtained by $\hat{f}^{\star}$ are almost always negative before $2008/07/17$ and positive afterwards. Considering both the time series of the state variables and the option data under an arbitrage-free pricing model makes the shareholder fear the left tail risk only before $2008/07/17$.

Finally, we see in the left lower panel the time series of the estimates of the conditional kurtosis of the returns. The estimates obtained by both the estimators of the transition density vary approximatively in the range $[1.5 : 6]$ before $2008/07/17$ and in the range $[3 : 5.5]$ afterwards. In the second period, the transition density, estimated in both ways, is leptokurtic. Before $2008/07/17$, the information content in the option quotes suggest to take into account a greater fat tail risk, almost at any day. The opposite happens in the second period.

# References


Y. Achdou, “An Inverse Problem for a Parabolic Variational Inequality Arising in Volatility Calibration with American Options”, *SIAM Journal on Control and Optimization*, 2005, Vol. 43 (5), pp. 1583-1615.

Y. Achdou, “An Inverse Problem for a Parabolic Variational Inequality with an Integro-Differential Operator”, *SIAM Journal on Control and Optimization*, 2008, Vol. 47 (2), pp. 733-767.

C. Ai and X. Chen, “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions”, *Econometrica*, 2003, Vol. 71 (6), pp. 1795-1843.

F. Ait-Sahlia and P. Carr, “American Options: A comparison of Numerical Methods”, in *Numerical methods in finance*, 1997, Vol. 4, Ch. 36, pp. 67-87, L. C. G. Rogers and D. Talay Editors, *Cambridge University Press*, Cambridge.

Y. Aït-Sahalia, “The Delta and Bootstrap Methods for Nonparametric Kernel Functionals”, 1992, MIT *Working paper*.

Y. Aït-Sahalia, “Nonparametric Pricing of Interest Rate Derivative Securities”, *Econometrica*, 1996, Vol. 64 (3), pp. 527-560.

Y. Aït-Sahalia, “Testing Continuous-Time Models of the Spot Interest Rate”, *The Review of Financial Studies*, 1996, Vol. 9 (2), pp. 385-426.

Y. Aït-Sahalia and J. Duarte, “Nonparametric Option Pricing under Shape Restrictions”, *Journal of Econometrics*, 2003, Vol. 116 (1-2), pp. 9-47.

Y. Aït-Sahalia and A. W. Lo, “Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices”, *The Journal of Finance*, 1998, Vol. 53 (2), pp. 499-547.

Y. Aït-Sahalia, P. Mykland and L. Zhang, “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise”, *The Review of Financial Studies*, 2005, Vol. 18 (2), pp. 351-416.

J. Alcock and D. Auerswald, “Empirical Tests of Canonical Nonparametric American Option-Pricing Methods”, *The Journal of Futures Markets*, 2010, Vol. 30 (6), pp. 509-532.

J. Alcock and T. Carmichael, “Nonparametric American Option Pricing”, *The Journal of Futures Markets*, 2008, Vol. 28 (8), pp. 717-748.

L. Alili and A. E. Kyprianou, “Some Remarks on First Passage of Lévy Processes, The American Put and Pasting Principles”, *The Annals of Applied Probability*, 2005, Vol. 15 (3), pp. 2062-2080.

K. Alpert, “Taxation and the Early Exercise of Call Options”, *Journal of Business Finance & Accounting*, 2010, Vol. 37 (5-6), pp. 715-736.

W. Allegretto, G. Barone-Adesi and R. Elliott, “Numerical Evaluation of the Critical Price and American Options”, *The European Journal of Finance*, 1995, Vol. 1 (1), pp. 69-78.

J. G.. Altonji and L. M. Segal, “Small-Sample Bias in GMM Estimation of Covariance Structures”, *Journal of Business and Economic Statistics*, 1996, Vol. 14 (3), pp. 353-366.

K. I. Amin, “Jump Diffusion Option Valuation in Discrete Time”, *The Journal of Finance*, 1993, Vol. 48 (5), pp. 1833-1863.

K. I. Amin and A. Khanna, “Convergence of American Option Values from Discrete to Continuous-Time Financial Models”, *Mathematical Finance*, 1994, Vol. 4 (4), pp. 289-304.

T. G. Andersen, T. Bollerslev, F. X. Diebold and H. Hebens, “The Distribution of Realized Stock Return Volatility”, *Journal of Financial Economics*, 2001, Vol. 61 (1), pp. 43-67.

T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys, “Modeling and Forecasting Realized Volatility”, *Econometrica*, 2003, Vol. 71 (2), pp. 579-625.

L. Andersen and M. Broadie, “A Primal-Dual Simulation Algorithm for Pricing Multi-Dimensional American Options”, *Management Science*, 2004, Vol. 50 (9), pp. 1222-1234.

D. W. K. Andrews and M. Buchinsky, “A Three-Step Method for Choosing the Number of Bootstrap Repetitions”, *Econometrica*, 2003, Vol. 68 (1), pp. 23-51.

B. Antoine, H. Bonnal and E. Renault, “On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood”, *Journal of Econometrics*, 2007, Vol. 138 (2), pp. 461-487.

B. Antoine and E. Renault, ‘Efficient Minimum Distance Estimation with Multiple Rates of Convergence”, forthcoming in *Journal of Econometrics*.

K. J. Arrow, D. Blackwell and M. A. Girshick, “Bayes and Minimax Solutions of Sequential Decision Problems”, *Econometrica*, 1949, Vol. 17 (3-4), pp. 213-244.

A. N. Avramidis and H. Matzinger, "Convergence of the Stochastic Mesh Estimator for Pricing Bermudan Options", *Journal of Computational Finance*, 2004, Vol. 7 (4), pp. 7391.

G. Bakshi, C. Cao and Z. Chen, "Empirical Performance of Alternative Option Pricing Models", *The Journal of Finance*, 1997, Vol. 52 (5), pp. 2003-2049.

G. Bakshi, C. Cao and Z. Chen, "Pricing and Hedging Long-Term Options", *Journal of Econometrics*, 2000, Vol. 94 (1-2), pp. 277-318.

G. Bakshi, N. Kapadia and D. Madan, "Stock Returns Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options", *The Review of Financial Studies*, 2003, Vol. 16 (1), pp. 101-143.

G. Bakshi and D. Madan, "A Theory of Volatility Spreads", *Mangement Science*, 2006, Vol. 52 (12), pp. 1945-1956.

V. Bally and G. Pagès, "A Quantization Algorithm for Solving Multidimensional Discrete-Time Optimal Stopping Problems", *Bernoulli*, 2003, Vol. 9 (6), pp. 1003-1049.

V. Bally, G. Pagès and J. Printems, "A Quantization Method for Pricing and Hedging Multidimensional American Options", *Mathematical Finance*, 2005, Vol. 15 (1), pp. 119-168.

R. Bansal and A. Yaron, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles", *The Journal of Finance*, 2004, Vol. 59 (4), pp. 1481-1509.

G. Barles, J. Burdeau, M. Romano and N. Sansoen, "Critical Stock Price Near Expiration", *Mathematical Finance*, 1995, Vol. 5 (2), pp. 77-95.

G. Barone-Adesi, "The Saga of the American Put", *Journal of Banking & Finance*, 2005, Vol. 29 (11), pp. 2909-2918.

G. Barone-Adesi and R. Elliott, "Approximations for the Values of American Options", *Stochastic Analysis and Applications*, 1991, Vol. 9 (2), pp. 115-131.

G. Barone-Adesi and R. E. Whaley, "Efficient Analytic Approximation of American Option Values", *The Journal of Finance*, 1987, Vol. 42 (2), pp. 301-320.

K. Barty, P. Girardeau, J. S. Roy and C. Strugarek, "Application of Kernel-Based Stochastic Gradient Algorithms to Option Pricing", *Monte Carlo Methods and Applications*, 2008, Vol. 14 (2), pp. 99-127.

D. S. Bates, “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options”, *The Review of Financial Studies*, 1996, Vol. 9 (1), pp. 69-108.

J. A. Bather, “Optimal Stopping Problems for Brownian Motion”, *Advances in Applied Probability*, 1970, Vol. 2 (2), pp. 259-286.

G. Bekaert and G. Wu, “Asymmetric Volatility and Risk in Equity Markets”, *The Review of Financial Studies*, 2000, Vol. 13 (1), pp. 1-42.

R. Bellman, “Dynamic Programming”, *Princeton University Press*, 1957, Princeton (New Jersey).

A. Bensoussan, “On the Theory of Option Pricing”, *Acta Applicandae Mathematicae*, 1984, Vol. 2 (2), pp. 139-158.

A. Bensoussan and J. L. Lions, “Applications of Variational Inequalities in Stochastic Control”, *North-Holland*, 1982, Amsterdam; translation of “Applications des Inéquations Variationnelles en Contróle Stochastique”, 1978, Bordas (Dunod), Paris.

R. Bikbov and M. Chernov, “Yield Curve and Volatility: Lessons from Eurodollars Futures and Options”, *Journal of Financial Econometrics*, 2010, Vol. 9 (1), pp. 66-105.

M. Birke and K. Pilz, “Nonparametric Option Pricing with No-Arbitrage Constraints”, *Journal of Financial Econometrics*, Vol. 7 (2), pp. 53-76.

F. Black, “Studies of Stock Price Volatility Changes”, *Proceedings of the 1976 Meetings of the American Statistical Association*, 1977, Business and Economic Statistics Section, pp. 177-181.

F. Black and M. Scholes, “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy*, 1973, Vol. 81 (3), pp. 637-654.

T. Bollerslev, M. Gibson and H. Zhou, “Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities”, *Journal of Econometrics*, 2011, Vol. 160 (1), pp. 235-245.

D. Bosq, “Nonparametric Statistics for Stochastic Processes: Estimation and Prediction”, *Springer*, New York, 1998.

P. P. Boyle, “A Lattice Framework for Option Pricing with Two State Variables”, *The Journal of Financial and Quantitative Analysis*, 1988, Vol. 23 (1), pp. 1-12.

M. J. Brennan and E. S. Schwartz, “The Valuation of American Put Options”, *The Journal of Finance*, 1977, Vol. 32 (2), pp. 449-462.

M. J. Brennan and E. S. Schwartz, “Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis”, *Journal of Financial and Quantitative Analysis*, 1978, Vol. 13 (3), pp 461-474.

M. Broadie, M. Chernov and M. Johannes, “Model Specification and Risk Premia: Evidence from Futures Options”, *The Journal of Finance*, 2007, Vol. 62 (3), pp. 1453-1490.

M. Broadie and J. B. Detemple, “American Capped Call Options on Dividend-Paying Assets”, *The Review of Financial Studies*, 1995, Vol. 8 (1), pp. 161-191.

M. Broadie and J. B. Detemple, “American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods”, *The Review of Financial Studies*, 1996, Vol. 9 (4), pp. 1211-1250.

M. Broadie and J. B. Detemple, “Option Pricing: Valuation Models and Applications”, *Management Science*, 2004, Vol. 50 (9), pp. 1145-1177.

M. Broadie, J. B. Detemple, E. Ghysels and O. Torrés, “American Options with Stochastic Dividends and Volatility: A Nonparametric Investigation”, *Journal of Econometrics*, 2000, Vol. 94 (1-2), pp. 53-92.

M. Broadie, J. B. Detemple, E. Ghysels and O. Torrés, “Nonparametric Estimation of American Options’ Exercise Boundaries and Call Prices”, *Journal of Economic Dynamics and Control*, 2000, Vol. 24 (11-12), pp. 1829-1857.

M. Broadie and P. Glasserman, “Pricing American-Style Securities Using Simulation”, *Journal of Economic Dynamics and Control*, 1997, Vol. 21 (8-9), pp. 1323-1352.

M. Broadie and P. Glasserman, “A Stochastic Mesh Method for Pricing High-Dimensional American Options”, *Journal of Computational Finance*, 2004, Vol. 7 (4), pp. 35-72.

B. W. Brown and W. Newey, “Efficient Semiparametric Estimation of Expectations”, *Econometrica*, 1998, Vol. 66 (2), pp. 453-464.

M. K. Brunnermeier, “Deciphering the Liquidity and Credit Crunch 2007-2008”, *Journal of Economic Perspectives*, 2009, Vol. 23 (1), pp. 77-100.

P. W. Buchen and M. Kelly, “The Maximum Entropy Distribution of an Asset Inferred from Option Prices”, *The Journal of Financial and Quantitative Analysis*, 1996, Vol. 31 (1), pp. 143-159.

D. S. Bunch and H. Johnson, “A Simple and Numerically Efficient Valuation Method for American Puts Using a Modified Geske-Johnson Approach”, *The Journal of Finance*, 1992, Vol. 47 (2), pp. 809-816.

Z. Cai and Y. Hong, “Some Recent Developments in Nonparametric Finance”, in *Advances in Econometrics*, 2009, Vol. 25, *Nonparametric Econometric Methods*, Ch. 12, pp. 379-432, Q. Li and J. S. Racine Editors, *Emerald*, Bingley.

J. Y. Campbell and J. H. Cochrane, “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior”, *The Journal of Political Economy*, 1999, Vol. 107 (2), pp. 205-251.

J. Y. Campbell and L. Hentschel, “No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns”, *Journal of Financial Economics*, 1992, Vol. 31, pp. 281-318.

L. Canina and S. Figlewski, “The Informational Content of Implied Volatility”, *The Review of Financial Studies*, 1993, Vol. 6 (3), pp. 659-681.

J. N. Carpenter, “The Exercise and Valuation of Executive Stock Options”, *Journal of Financial Economics*, 1998, Vol. 48 (2), pp. 127-158.

P. Carr, “Randomization and the American Put”, *The Review of Financial Studies*, 1998, Vol. 11 (3), pp. 597-626.

P. Carr and D. Faguet, “Fast Accurate Valuation of American Options”, 1994, *Cornell university working paper*.

P. Carr, R. Jarrow and R. Myneni, “Alternative Characterizations of American Put Options”, *Mathematical Finance*, 1992, Vol. 2 (2), pp. 87-106.

P. Carr and L. Wu, “Variance Risk Premiums”, *The Review of Financial Studies*, 2009, Vol. 22 (3), pp. 1311-1341.

J. F. Carriere, “Valuation of the Early-Exercise Price for Options Using Simulations and Nonparametric Regression”, *Insurance: Mathematics and Economics*, 1996, Vol. 19 (1), pp. 19-30.

M. Carrasco and X. Chen, “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models”, *Econometric Theory*, 2002, Vol. 18 (1), pp. 17-39.

G. Chamberlain, “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions”, *Journal of Econometrics*, 1987, Vol. 34 (3), pp. 305-334.

D. Chapman and N. Pearson, “Is the Short Rate Drift Actually Nonlinear?”, *The Journal of Finance*, 2000, Vol. 55 (1), pp. 355-388.

X. Chen, J. Chadam, J. Jiang and W. Zheng, “Convexity of the Exercise Boundary of the American Put Option on a Zero Dividend Asset”, *Mathematical Finance*, 2008, Vol. 18 (1), pp. 185-197.

X. Chen, L. P. Hansen and M. Carrasco, “Nonlinearity and Temporal Dependence”, *Journal of Econometrics*, 2010, Vol. 155 (2), pp. 155-169.

H. C. Chen and M. Magdon-Ismail, “NN-OPT: Neural Network for Option Pricing Using Multinomial Tree”, in *Lecture Notes in Computer Science*, 2006, Vol. 4234, *Neural Information Processing*, pp. 360-369, I. King, J. Wang, L. Chan and D. Wang Editors, *Springer-Verlag*, Berlin.

E. Chevalier, “Critical Price Near Maturity for an American Option on a Dividend-Paying Stock in a Local Volatility Model”, *Mathematical Finance*, 2005, Vol. 15 (3), pp 439-463.

C. Chiarella and A. Ziogas, “American Call Options Under Jump-Diffusion Processes - A Fourier Transform Approach”, *Applied Mathematical Finance*, 2009, Vol. 16 (1), pp. 37-79.

A. A. Christie, “The Stochastic Behavior of Common Stock Variances. Value, Leverage and Interest Rate Effects”, *Journal of Financial Economics*, 1982, Vol. 10 (4), pp. 407-432.

P. Christoffersen, S. L. Heston and K. Jacobs, “Option Valuation with Conditional Skewness”, *Journal of Econometrics*, 2006, Vol. 131 (1-2), pp. 253-284.

S. L. Chung and H. C. Chang, “Generalized Analytical Upper Bounds for American Option Prices”, *Journal of Financial and Quantitative Analysis*, 2007, Vol. 42 (1), pp. 209-228

E. Clément, D. Lamberton and P. Protter, “An Analysis of a Least Squares Regression Method for American Option Pricing”, *Finance and Stochastics*, 2002, Vol. 6 (4), pp. 449-471.

G. M. Costantidines and S. Perrakis, “Stochastic Dominance Bounds on American Option Prices in Markets with Frictions”, *Review of Finance*, 2007, Vol. 11, pp. 697-703.

J. C. Cox, J. E. Ingersoll Jr. and S. A. Ross, “A Theory of the Term Structure of Interest Rates”, *Econometrica*, 1985, Vol. 53 (2), pp. 385-407.

J. C. Cox and S. A. Ross, "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics*, 1976, Vol. 3 (1-2), pp. 145166.

J. C. Cox, S. A. Ross and M. Rubinstein, "Option Pricing: A Simplified Approach", *Journal of Financial Economics*, 1979, Vol. 7 (3), pp. 229-263.

M. G. Crandall, H. Ishii and P. L. Lions, "User's Guide to Viscosity Solutions of Second Order Partial Differential Equations", *Bulletin (New Series) of the American Mathematical Society*, 1992, Vol. 27 (2), pp. 1-67.

T. Daglish, "A Pricing and Hedging Comparison of Parametric and Nonparametric Approaches for American Index Options", *Journal of Financial Econometrics*, 2003, Vol. 1 (3), pp. 327-364.

M. H. A. Davis and I. Karatzas, "A Deterministic Approach to Optimal Stopping", in *Probability, Statistics and Optimization: a Tribute to Peter Whittle*, 1995, pp. 455-466, F. P. Kelly Editor, *Wiley*, New-York.

P. Dennis and S. Mayhew, "Risk-Neutral Skewness: Evidence from Stock Options", *Journal of Financial and Quantitative Analysis*, 2002, Vol. 37 (3), pp. 471-493.

J. B. Detemple, "American Options: Symmetry Properties", in *Handbooks in Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management*, 2001, pp. 67-104, E. Jouini, J. Cvitanić and M. Musiela Editors, *Cambridge University Press*, 2001, Cambridge.

J. B. Detemple, "American-Style Derivatives: Valuation and Computation", *Chapman & Hall/CRC Financial Mathematics Series*, 2005, London.

J. B. Detemple and S. Sundaresan, "Nontraded Asset Valuation with Portfolio Constraints: A Binomial Approach", *The Review of Financial Studies*, 2002, Vol. 12 (4), pp. 835-872.

J. B. Detemple and W. Tian, "The Valuation of American Options for a Class of Diffusion Processes", *Management Science*, 2002, Vol. 48 (7), pp. 917-937.

F. Diz and T. J. Finucane, "The Rationality of Early Exercise Decisions: Evidence from the S&P 100 Index Options Market", *The Review of Financial Studies*, 1993, Vol. 6 (4), pp. 765-797.

J. C. Duan, "Nonparametric Option Pricing by Transformation", *Working paper*, 2002.

J. C. Duan and J. Wei, "Systematic Risk and the Price Structure of Individual Equity Options", *The Review of Financial Studies*, 2009, Vol. 22 (5), pp. 1981-2006.

M. Dueker and T. W. Miller Jr., “Directly Measuring Early Exercise Premiums Using American and European S&P 500 Index Options”, *Journal of Futures Markets*, Vol. 23 (3), pp. 287-313.

D. Duffie, “Dynamic Asset Pricing Theory”, 3rd Ed., *Princeton University Press*, Princeton, 2001.

D. Duffie, D. Filipovic and W. Schachermayer, “Affine Processes and Applications in Finance”, *The Annals of Applied Probability*, 2003, Vol. 13 (3), pp. 984-1053.

D. Duffie, R. Liu and A. M. Poteshman, “How American Options Are Exercised”, *Working paper*, 2005.

D. Duffie, J. Pan and K. Singleton, “Transform Analysis and Asset Pricing for Affine Jump Diffusions”, *Econometrica*, 2000, Vol. 68 (6), pp. 1343-1376.

D. J. Duffy, “Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach”, *Wiley*, 2006, Chichester.

B. Dupire Editor, “Monte Carlo Methodologies and Applications for Pricing and Risk Management”, 1998, *Risk Publications*, London.

B. Efron and R. J. Tibshirani, “An Introduction to the Bootstrap”, Third Edition, *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*, New York, 2000.

L. G. Epstein and S. E. Zin, “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework”, *Econometrica*, 1989, Vol. 57 (4), pp. 937-969.

D. Egloff, M. Kohler, and N. Todorovic, “A Dynamic Look-Ahead Monte Carlo Algorithm for Pricing Bermudan Options”, *The Annals of Applied Probability*, 2007, Vol. 17 (4), pp. 1138-1171.

E. Ekstrom, “Properties of American Option Prices”, *Stochastic Processes and their Applications*, 2004, Vol. 114 (2), pp. 265-278.

E. Ekstrom and J. Tysk, “Properties of Option Prices in Models With Jumps”, *Mathematical Finance*, 2007, Vol. 17 (3), pp. 381-397.

N. El Karoui and I. Karatzas, “The Optimal Stopping Problem for a General American Put-Option”, in *Mathematical Finance*, 1995, Vol. 65, pp. 77-88, M. H. A. Davis, D. Duffie, W. H. Fleming and S. S. Shreve Editors, *Springer-Verlag*, New-York.

M. Engstrom, L. Nordén and A. Stromberg, “Early Exercise of American Put Options: Investor Rationality on the Swedish Equity Options Market”, *Journal of Futures Market*, 2000, Vol. 20 (2), pp. 167-188.

J. D. Evans, R. Kuske and J. B. Keller, “American Options on Assets with Dividends Near Expiry”, *Mathematical Finance*, 2002, Vol. 12 (3), pp. 219-237.

J. Evnine and A. Rudd, “Index Options: The Early Evidence”, *The Journal of Finance*, 1985, Vol. 40 (3), pp. 743-756.

E. F. Fama, “Efficient Capital Markets: A Review of Theory and Empirical Work”, *The Journal of Finance*, 1970, Vol. 30, pp. 34-105.

T. J. Finucane, “An Empirical Analysis of Common Stock Call Exercise: A Note”, *Journal of Banking and Finance*, 1997, Vol. 21 (4), pp. 563-571.

B. Flesaker, “Testing the Heath-Jarrow-Morton/Ho-Lee Model of Interest Rate Contingent Claims Pricing”, *Journal of Financial and Quantitative Analysis*, 1993, Vol. 28 (4), pp. 483-495.

K. R. French, G. W. Schwert and R. F. Stambaugh, “Expected Stock Returns and Volatility”, *Journal of Financial Economics*, 1987, Vol. 19 (1), pp. 3-29.

A. Friedman, “Free Boundary Problems for Parabolic Equations I: Melting of Solids”, *Journal of Mathematics and Mechanics*, 1959, Vol. 8, pp. 499-518.

M. C. Fu, S. B. Laprise, D. B. Madan, Y. Su and R. Wu, “Pricing American Options: A Comparison of Monte Carlo Simulation Approaches”, *Journal of Computational Finance*, 2001, Vol. 4 (3), pp. 39-88.

P. Gagliardini, C. Gouriéroux and E. Renault, “Efficient Derivative Pricing by the Extended Method of Moments”, *Econometrica*, 2011, Vol. 79 (4), pp. 1181-1232.

P. Gagliardini and D. Ronchetti, “Semi-Parametric Estimation of American Option Prices”, 2010, *Working paper*.

R. Garcia and R. Gencay, “Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Link”, *Journal of Econometrics*, 2000, Vol. 94 (1-2), pp. 93-115.

D. Gatarek and A. Świech, “Optimal Stopping in Hilbert Spaces and Pricing of American Options”, *Mathematical Methods of Operations Research*, 1999, Vol. 50 (1), pp. 135-147.

R. Geske and H. E. Johnson, “The American Put Option Valued Analytically”, *The Journal of Finance*, 1984, Vol. 39 (5), pp. 1511-1524.

P. Glasserman, “Monte Carlo Methods in Financial Engineering”, *Springer*, New York, 2004.

P. Glasserman and B. Yu, “Number of Paths Versus Number of Basis Functions in American Option Pricing”, *The Annals of Applied Probability*, 2004, Vol. 14 (4), pp 2090-2119.

C. Gouriéroux and J. Jasiak, “Autoregressive Gamma Processes”, *Journal of Forecasting*, 2006, Vol. 25 (2), pp. 129-152.

C. Gouriéroux and A. Monfort, “Econometric Specification of Stochastic Discount Factor Models”, *Journal of Econometrics*, 2007, Vol. 136 (2), pp. 509-530.

C. Gouriéroux, A. Monfort and V. Polimenis, “Affine Model for Credit Risk Analysis”, *Journal of Financial Econometrics*, 2006, Vol. 4 (3), pp. 494-530.

D. Grant, G. Vora and D. E. Weeks, “Simulation and the Early-Exercise Option Problem”, *The Journal of Financial Engineering*, 1996, Vol. 5 (3), pp. 211-227.

B. E. Hansen, “Uniform Convergence Rates for Kernel Estimation with Dependent Data”, *Econometric Theory*, 2008, Vol. 24 (3), pp. 726-748.

L. P. Hansen, “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica*, 1982, Vol. 50 (4), pp. 1029-1054.

L. P. Hansen and E. Renault, “Pricing Kernels”, in *Encyclopedia of Quantitative Finance*, 2010, Vol. 3, Ch. 19, pp. 518-552, R. Cont Editor, *Wiley*, New-York.

L. Hansen and S. Richard, “The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models”, *Econometrica*, 1987, Vol. 55, pp. 587-613.

L. P. Hansen and K. J. Singleton, “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models”, *Econometrica*, 1982, Vol. 50 (5), pp. 1269-1286.

W. Hardle, M. Muller, S. Sperlich and A. Werwatz, “Nonparametric and Semiparametric Models”, *Springer*, Berlin, 2004.

J. M. Harrison and D. M. Kreps, “Martingales and Arbitrage in Multiperiod Securities Markets”, *Journal of Economic Theory*, 1979, Vol. 20 (3), pp. 381-408.

J. M. Harrison and S. R. Pliska, "Martingales and Stochastic Integrals in the Theory of Continuous Trading", *Stochastic Processes and their Applications*, 1981, Vol. 11 (3), 215-260.

M. B. Haugh and L. Kogan, "Pricing American Options: a Duality Approach", *Operations Research*, 2004, Vol. 52 (2), pp. 258-270.

S. L. Heston, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *The Review of Financial Studies*, 1993, Vol. 6 (2), pp. 327-343.

S. L. Heston and S. Nandi, "A Closed-Form GARCH Option Valuation Model", *The Review of Financial Studies*, 2000, Vol. 13 (3), pp. 585-625.

T. S. Ho, M. G. Subrahmanyam and R. C. Stapleton, "The Valuation of American Options with Stochastic Interest Rates: A Generalization of the Geske-Johnson Technique", *The Journal of Finance*, 1997, Vol. 52 (2), pp. 827-840.

Y. C. Hon and X. Z. Mao, "A Radial Basis Function Method for Solving Options Pricing", *The Journal of Financial Engineering*, 1999, Vol. 8 (1), pp. 31-49.

Y. Hong and H. Li, "Nonparametric Specification Testing for Continuous-Time Models with Applications to Interest Rate Term Structures", *The Review of Financial Studies*, 2005, Vol. 18 (1), pp. 37-84.

Y. Hong, J. Tu and G. Zhou, "Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation", *The Review of Financial Studies*, 2007, Vol. 20 (5), pp. 1547-1581.

J. Huang and J. S. Pang, "Option Pricing and Linear Complementarity", *The Journal of Computational Finance*, 1998, Vol. 2 (1), pp. 31-60.

J. Huang, M. G. Subrahmanyam and G. G. Yu, "Pricing and Hedging American Options: A Recursive Integration Method", *The Review of Financial Studies*, 1996, Vol. 9 (1), pp. 277-300.

J. Hull and A. White, "The Pricing of Options on Assets with Stochastic Volatilities", *The Journal of Finance*, 1987, Vol. 42 (2), pp. 281-300.

J. Hutchinson, A. Lo and T. Poggio, "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks", *The Journal of Finance*, 1994, Vol. 49 (3), pp. 851-889.

S. Jacka, "Optimal Stopping and the American Put", *Mathematical Finance*, 1991, Vol. 1 (2), pp. 1-14.

J. C. Jackwerth, "Generalized Binomial Trees", *The Journal of Derivatives*, 1997, Vol. 5 (2), pp. 7-17.

J. C. Jackwerth, "Option-Implied Risk-Neutral Distributions and the Implied Binomial Trees", *The Journal of Derivatives*, 1999, Vol. 7 (2), pp. 66-82.

J. C. Jackwerth, "Recovering Risk Aversion from Option Prices and Realized Returns", *The Review of Financial Studies*, 2000, Vol. 13 (2), pp. 433-451.

J. C. Jackwerth and M. Rubinstein, "Recovering Probability Distributions from Option Prices", *The Journal of Finance*, 1996, Vol. 51 (5), pp. 1611-1631.

P. Jaillet, D. Lamberton and B. Lapeyre, "Variational Inequalities and the Pricing of American Options", *Acta Applicandae Mathematicae*, 1990, Vol. 21 (3), pp. 263-289.

F. Jamshidian, "An Analysis of American Options", *The Review of Futures Markets*, 1992, Vol. 1 (1), pp. 72-80.

G. J. Jiang and Y. S. Tian, "The Model-Free Implied Volatility and Its Information Content", *The Review of Financial Studies*, 2005, Vol. 18 (4), pp. 1305-1342.

H. E. Johnson, "An Analytical Approximation for the American Put Price", *Journal of Financial and Quantitative Analysis*, 1983, Vol. 18, pp. 141-148.

N. Ju, "Pricing an American Option by Approximating its Early Exercise Boundary as a Multipiece Exponential Function", *The Review of Financial Studies*, 1998, Vol. 11 (3), pp. 627-646.

N. Ju and R. Zhong, "An Approximate Formula for Pricing American Options", *The Journal of Derivatives*, 1999, Vol. 7 (2), pp. 31-40.

S. Kallast and A. Kivinukk, "Pricing and Hedging American Options Using Approximations by Kim Integral Equations", *European Finance Review*, 2003, Vol. 7 (3), pp. 361-383.

B. Kamrad and P. Ritchken, "Multinomial Approximating Models for Options with k State Variables", *Management Science*, 1991, Vol. 37 (12), pp. 1640-1652.

I. Karatzas, "On the Pricing of American Options", *Applied Mathematics & Optimization*, 1988, Vol. 17 (1), pp. 37-60.

I. Karatzas, "Optimization Problems in the Theory of Continuous Trading", *SIAM Journal on Control and Optimization*, 1989, Vol. 27 (6), pp. 1221-1259.

I. J. Kim, "The Analytic Valuation of American Options", *The Review of Financial Studies*, 1990, Vol. 3 (4), pp. 547-572.

Y. Kitamura and M. J. Stutzer, “An Information-Theoretic Alternative to Generalized Method of Moments Estimation”, *Econometrica*, 1997, Vol. 65 (4), pp. 861-874.

Y. Kitamura, G. Tripathi and H. Ahn, “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models”, *Econometrica*, 2004, Vol. 72 (6), pp. 1667-1714.

R. C. Klemkosky and B. G. Resnick, “Put-Call Parity and Market Efficiency”, *The Journal of Finance*, 1979, Vol. 34 (5), pp. 1141-1155.

M. Kohler, “A Review on Regression-based Monte Carlo Methods for Pricing American Options”, in *Recent Developments in Applied Probability and Statistics. Dedicated to the Memory of Jurgen Lehn.*, 2010, pp. 37-58, L. Devroye, B. Karasozen, M. Kohler and R. Korn Editors, *Springer-Verlag*, Berlin.

M. Kohler, A. Krzyżak and N. Todorovic, “Pricing of High Dimensional American Options by Neural Networks”, *Mathematical Finance*, 2010, Vol. 20 (3), pp. 383-410.

H. Ichimura and P. E.. Todd, “Implementing Nonparametric and Semiparametric Estimators”, in *Handbook of Econometrics*, 2007, Vol. 6 (2), Ch. 74, pp. 5369-5468, J. J. Heckman and E. E. Leamer Editors, *Elsevier*, North-Holland.

D. Lamberton and S. Villeneuve, “Critical Price for an American Option on a Dividend-Paying Stock”, *The Annals of Applied Probability*, 2003, Vol. 13 (2), pp. 800-815.

C. G. Lamoureux and W. D. Lastrapes, “Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities”, *The Review of Financial Studies*, 1993, Vol. 6 (2), pp. 293-326.

S. B. Laprise, M. C. Fu, S. I. Marcus, A. E. B. Lim and H. Zhang, “Pricing American-Style Derivatives with European Call Options”, *Management Science*, 2006, Vol. 52 (1), pp. 95-110.

S. F. Leroy, “Expectations Models of Asset Prices: A Survey of Theory”, *The Journal of Finance*, 1982, Vol. 37, pp. 185-217.

H. Li and F. Zhao, “Nonparametric Estimation of State-Price Densities Implicit in Interest Rate Cap Prices”, *The Review of Financial Studies*, 1993, Vol. 22 (11), pp. 4335-4376.

T. Little, V. Pant and C. Hou, “A New Integral Representation of the Early Exercise Boundary for American Put Options”, *Journal of Computational Finance*, 2000, Vol. 3 (3), pp. 73-96.

Q. Liu, “Pricing American Options by Canonical Least-Squares Monte Carlo”, *The Journal of Futures Markets*, 2010, Vol. 30 (2), pp. 175-187.

G. Liu and L. J. Hong, "Revisit of Stochastic Mesh Method for Pricing American Options", *Operations Research Letters*, 2009, Vol. 37 (6), pp. 411-414.

A. W. Lo, "Reading About the Financial Crisis: A 21-Book Review", *Working paper*.

F. A. Longstaff, P. Santa-Clara and E. S. Schwartz, "Throwing Away a Billion Dollars: the Cost of Suboptimal Exercise Strategies in the Swaptions Market", *Journal of Financial Economics*, 2001, Vol. 62 (1), pp. 39-66.

F. A. Longstaff and E. S. Schwartz, "Valuing American Options by Simulation: a Simple Least-Squares Approach", *The Review of Financial Studies*, 2001, Vol. 14 (1), pp. 113-147.

R. E. Lucas, Jr., "Asset Prices in an Exchange Economy", *Econometrica*, 1978, Vol. 46 (6), pp. 1429-1445.

L. W. MacMillan, "Analytic Approximation for the American Put Prices", *Advances in Futures and Options Research*, 1986, Vol. 1, pp. 119-139.

H. P. McKean, Jr., "Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics", *Industrial Management Review*, 1965, Vol. 6 (2), pp. 32-39.

R. Mehra and E. C. Prescott, "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 1985, Vol. 15, pp. 145-161.

R. Merton, "Theory of Rational Option Pricing", *The Bell Journal of Economics and Management Science*, 1973a, Vol. 4 (1), pp. 141-183.

R. Merton, "The Relationship between Put and Call Option Prices: Comment", *The Journal of Finance*, 1973b, Vol 28 (1), pp. 183-184.

R. Merton, "Continuous-Time Finance", *Wiley-Blackwell*, Cambridge, 1990.

E. Mordecki, "Optimal Stopping and Perpetual Options for Lévy Processes", *Finance and Stochastics*, 2002, Vol. 6 (4), pp. 473-493.

M. Musiela and M. Rutkowski, "Martingale Methods in Financial Modelling", *Springer*, 2005, New York.

R. Myneni, "The pricing of the American option", *The Annals of Applied Probability*, 1992, Vol. 2 (1), pp. 1-23.

S. Nagel and K. Singleton, "Estimation and Evaluation of Conditional Asset Pricing Models", *The Journal of Finance*, 2011, Vol. 66 (3), pp. 873-909.

D. B. Nelson, "Conditional Heteroscedasticity in Asset Returns: A New Approach", *Econometrica*, 1991, Vol. 59 (2), pp. 347-370.

J. Neveu, "Discrete-Parameter Martingales", *North-Holland*, Oxford, 1975.

W. Newey and D. McFadden, "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics*, 1999, Vol. 4, Ch. 36, pp. 2113-2245, R. Engle and D. McFadden Editors, *Elsevier*, North-Holland.

J. A. Overdahl and P. G. Martin, "The Exercise of Equity Options: Theory and Empirical Tests", *Journal of Derivatives*, 1994, Vol. 2 (1), pp. 38-51.

M. Parkinson, "Option Pricing: the American Put", *Journal of Business*, 1977, Vol. 50 (1), pp. 21-36.

G. Peskir and A. Shiryaev, "Optimal Stopping and Free-Boundary Problems", 2006, *Birkhäuser Verlag*, Basel.

H. Pham, "Optimal Stopping, Free Boundary, and American Option in a Jump-Diffusion Model", *Applied Mathematics & Optimization*, 1997, Vol. 35 (2), pp. 145-164.

A. M. Poteshman and V. Serbin, "Clearly Irrational Financial Market Behavior: Evidence from the Early Exercise of Exchange Traded Stock Options", *The Journal of Finance*, 2003, Vol. 58 (1), pp. 37-70.

J. L. Powell, "Estimation of Semiparametric Models", in *Handbook of Econometrics*, 1999, Vol. 4, Ch. 41, pp. 2443-2521, R. Engle and D. McFadden Editors, *Elsevier*, North-Holland.

F. Pressacco, Gaudenzi, Zanette and Ziani, "New Insights on Testing the Efficiency of Methods of Pricing and Hedging American Options", *European Journal of Operational Research*, 2008, Vol. 185 (1), pp. 235-254.

M. Pritsker, "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models", *The Review of Financial Studies*, 1998, Vol. 11 (3), pp. 449-487.

P. E. Protter, "Stochastic Integration and Differential Equations", *Springer*, 2004, Berlin.

R. Rendleman and B. Bartter, "Two-State Option Pricing", *The Journal of Finance*, 1979, Vol. 34 (5), pp 1093-1110.

P. Ritchken and R. Trevor, “Pricing Options under Generalized GARCH and Stochastic Volatility Processes”, *The Journal of Finance*, 1999, Vol. 54 (1), pp. 377-402.

L. C. G. Rogers, “Monte Carlo Valuation of American Options”, *Mathematical Finance*, 2002, Vol. 12 (3), pp. 271-286.

A. Roux and T. Zastawniak, “American Options under Proportional Transaction Costs: Pricing, Hedging and Stopping Algorithms for Long and Short Positions”, *Acta Applicandae Mathematicae*, Vol. 106 (2), pp. 199-228.

M. Rubinstein, “Implied Binomial Trees”, *The Journal of Finance*, 1994, Vol. 49 (3), pp. 771-818.

M. Rutkowski, “The Early Exercise Premium Representation of Foreign Market American Options”, *Mathematical Finance*, 1994, Vol. 4 (4), pp. 313-325.

P. A. Samuelson, “Rational Theory of Warrant Pricing”, *Industrial Management Review*, 1965, Vol. 6 (2), pp. 13-32.

M. Schroder, “Changes of Numeraire for Pricing Futures, Forwards, and Options”, *The Review of Financial Studies*, 1999, Vol. 12 (5), pp. 1143-1163.

W. F. Sharpe, “Investments”, *Prentice-Hall*, 1978, Englewood Cliffs (New Jersey).

A. N. Shiryaev, “Optimal Stopping Rules”, *Springer-Verlag*, 1978 (English edition), New York.

B. W. Silverman, “Density Estimation for Statistics and Data Analysis”, *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*, London, 1986.

J. S. Simonoff, “Smoothing Methods in Statistics”, *Springer*, New York, 1996.

L. J. Snell, “Applications of Martingale System Theorems”, *Transactions of the American Mathematical Society*, 1952, Vol. 73 (2), pp. 293-312.

The Squam Lake Working Group on Financial Regulation, “The Squam Lake Report: Fixing the Financial System”, *Princeton University Press*, Princeton, 2010.

R. Stanton and N. Wallace, “The Bear’s Lair: Index Credit Default Swaps and the Subprime Mortgage Crisis”, *The Review of Financial Studies*, 2011, Vol. 24 (10), pp. 3250-3280.

E. M. Stein and J. Stein, “Stock Price Distributions with Stochastic Volatility: an Analytic Approach”, *The Review of Financial Studies*, 1991, Vol. 4 (4), pp. 727-752.

L. Stentoft, “Assessing the Least Squares Monte-Carlo Approach to American Option Valuation”, *The Review of Derivatives Research*, 2004, Vol. 7 (3), pp. 129-168.

J. Stock and J. Wright, ”GMM with Weak Identification”, *Econometrica*, Vol. 68 (5), pp. 1055-1096.

H. R. Stoll, “The Relationship between Put and Call Prices”, *The Journal of Finance*, 1969, Vol. 24, pp. 810-822.

M. Stutzer, “A Simple Nonparametric Approach to Derivative Security Valuation”, *The Journal of Finance*, 1996, Vol. 51 (5), pp. 1633-1652.

M. A. Sullivan, “Valuing American Put Options Using Gaussian Quadrature”, *The Review of Financial Studies*, 2000, Vol. 13 (1), pp. 75-94.

Y. S. Tian, “Extracting Risk-Neutral Density and Its Moments from American Option Prices”, *The Journal of Derivatives*, 2011, Vol. 18 (3), pp. 17-34.

J. A. Tilley, “Valuing American Options in a Path Simulation Model”, *Transactions of the Society of Actuaries*, 1993, Vol. 45, pp. 83-104.

J. Topper, “Financial Engineering with Finite Elements”, *Wiley*, 2005, Chichester.

R. S. Tsay, “Analysis of Financial Time Series”, Third Edition, *Wiley*, Hoboken, New Jersey, 2010.

J. N. Tsitsiklis, B. Van Roy, “Regression Methods for Pricing Complex American-Style Options”, *IEEE Transactions on Neural Networks*, 2001, Vol. 12 (4), pp. 694-703.

P. Van Moerbeke, “On Optimal Stopping and Free Boundary Problems”, *Advances in Applied Probability*, 1973, Vol. 5 (1), pp. 33-35.

A. Wald and J. Wolfowitz, “Optimum Character of the Sequential Probability Ratio Tests”, *Annals of Mathematical Statistics*, 1948, Vol. 19 (3), pp. 326-339.

P. Wilmott, S. Howison and J. Dewynne, “Option Pricing: Mathematical Models and Computation”, *Oxford Financial Press*, 1993, Oxford.

A. Yatchew and W. Härdle, “Nonparametric State Price Density Estimation Using Constrained Least Squares and the Bootstrap”, *Journal of Econometrics*, 2006, Vol. 133 (2), pp. 579-599.

X. L. Zhang, “Numerical Analysis of American Option Pricing in a Jump-Diffusion Model”, *Mathematics of Operations Research*, 1997, Vol. 22 (3), pp. 668-690.

S. P. Zhu, “An Exact and Implicit Solution for the Valuation of American Put Options”, *Quantitative Finance*, Vol. 6 (3), pp. 229-242.

T. L. Zivney, “The Value of Early Exercise in Option Prices: An Empirical Investigation”, *The Journal of Financial and Quantitative Analysis*, 1991, Vol. 26 (1), pp. 129-138.

# 4 Appendix to Chapter 2

## 4.1 Regularity Assumptions

In this appendix we list the additional regularity assumptions used to derive the theoretical results of the paper.

**Assumption A 1.** *The support $\mathcal{X} = \mathcal{R} \times \mathcal{S} \subset \mathbb{R}^d$ of process $(X_t)$ is compact.*

**Assumption A 2.** *The stationary pdf $f_Z$ of the vector $Z_t := [X_t' \; X_{t-1}']'$ is of differentiability class $C^\rho(\mathbb{R}^{2d})$, for integer $\rho \geq 2$, with uniformly continuous $\rho$-th order derivatives, and such that $f_Z > 0$ in the interior of the support $\mathcal{X} \times \mathcal{X}$. The same conditions are satisfied by the stationary pdf $f_X$ of $X_t$.*

**Assumption A 3.** *The stationary pdf's $f_Z$ and $f_X$ are such that $\int_{\mathcal{X}} \int_{\mathcal{X}} \left[ \frac{f_Z(x,\tilde{x})}{f_X(x) f_X(\tilde{x})} \right]^q f_Z(x,\tilde{x}) dx d\tilde{x} < \infty$ for real $q > 1$.*

**Assumption A 4.** *There exists a growing sequence of sets $\mathcal{X}_T := \mathcal{R}_T \times \mathcal{S}_T \subset \mathcal{X}$, for $T \in \mathbb{N}$, and real constants $c_1, c_2 > 0$ such that*

$$\sup_{x \in \mathcal{X}_T} \mathrm{P}\left[ X_{t+1} \in \mathcal{X}_T^C | X_t = x \right] \to 0, \text{ for } T \to \infty,$$

$$\inf_{x,\tilde{x} \in \mathcal{X}_T} f_Z(x,\tilde{x}) \geq \frac{c_1}{\log(T)^{c_2}}, \; \inf_{x \in \mathcal{X}_T} f_X(x) \geq \frac{c_1}{\log(T)^{c_2}}.$$

**Assumption A 5.** *The kernel function $K$ is a bounded and Lipschitz function on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} \|x\|^\rho K(x) dx < \infty$, where $\rho$ is defined in Assumption A 2, and $\int_{\mathbb{R}^d} x^j K(x) dx = 0$, for all multi-indices $j \in \mathbb{N}^d$ such that $|j| \leq \rho - 1$.*

**Assumption A 6.** *The bandwidth $h_T = o(1)$ is such that $\frac{\log(T)^2}{T h_T^{3d}} = o(1)$, $T h_T^{2\rho} = o(1)$.*

**Assumption A 7.** *The parameter $\theta_0$ is in the interior of compact set $\Theta \subset \mathbb{R}^p$.*

**Assumption A 8.** *The SDF $m(x;\theta)$ is of differentiability class $C^1(\Theta)$ w.r.t. $\theta \in \Theta$, for all $x \in \mathcal{X}$.*

**Assumption A 9.** *The SDF $m(x;\theta)$ satisfies: (i) $E\left[|m(X_{t+1};\theta_0)|^{2p}\right] < \infty$ for real $p > 1$ such that $1/p + 1/q = 1$, where $q > 1$ is defined in Assumption A 3; (ii) $\sup_{\substack{\theta \in \Theta \\ x \in \mathcal{X}}} \mathrm{E}\left[ |m(X_{t+1};\theta)|^{2+\delta} \middle| X_t = x \right] < \infty$, for real $\delta > 0$.*

**Assumption A 10.** *The matrix $\Omega_T$ converges in probability to the positive-definite matrix $\Omega_0$.*

**Assumption A 11.** *The matrix $\tilde{\Omega}_T(x)$ converges in probability to the positive-definite matrix $\tilde{\Omega}_0(x)$, uniformly in $x \in \mathcal{X}$.*

**Assumption A 12.** *The weight $\omega_T$ converges in probability to the non-negative scalar $\omega$.*

Assumptions A 1-4 concern the distribution of process $(X_t)$. In particular, the condition of compact support in Assumption A 1 simplifies the proofs and can be relaxed at the expense of additional technical burden. Assumption A 2 is standard for kernel estimation. Assumption A 3 restricts the dependence between $X_t$ and $X_{t-1}$ at the boundaries of the support. It is used to prove that the American put pricing operator $\mathcal{A}$ maps $L_2(\mathcal{Y})$ into $L_2(\mathcal{Y})$ in Appendix 4.3. Assumption A 4 constraints the decay behavior of the stationary densities of $X_t$ and $[X_t' \; X_{t-1}']'$ at the boundary of their supports. The sequence of sets $\mathcal{X}_T$, $T \in \mathbb{N}$, is such that these densities are bounded away from zero from below on $\mathcal{X}_T$ and $\mathcal{X}_T \times \mathcal{X}_T$, respectively, at an inverse logarithmic rate as $T$ increases. This sequence of sets is introduced to define trimmed versions of the kernel regression estimators (see the proof of Proposition 4 in Appendix 4.6.1) and control for boundary effects.

Assumptions A 5-6 concern the kernel and the bandwidth. Function $K$ is a kernel of order $\rho$, that is the same as the differentiability order of the densities in Assumption A 2. The bandwidth conditions in Assumption A 6 are stronger than the standard ones used for $d$-dimensional kernel estimation. The first condition ensures that the second-order terms in the Fréchet expansions are negligible asymptotically (see the proof of Proposition 5 in Appendix 4.6.2). The second condition is used to show that the bias of estimators constructed by averaging kernel regression estimators over the conditioning value is asymptotically negligible (see the proof of Proposition 7 in Appendix 4.6.3). When $h_T = cT^{-\eta}$ for real constants $c, \eta > 0$, Assumption A 6 is satisfied if $\dfrac{1}{2\rho} < \eta < \dfrac{1}{3d}$.

Assumption A 7 is standard for parametric estimation. Assumptions A 8-9 concern the SDF. They involve a differentiability condition w.r.t. parameter $\theta$, as well as a uniform boundedness condition for higher-order conditional moments of the SDF. Finally, Assumptions A 10-12 concern the weighting matrices in the criteria to estimate vector $\theta$, and the scalar weight in the criterion of the estimator of density $f$ in Definition 4. These assumptions ensure well-defined large sample limits for these criteria and are used to prove uniqueness of the extrema.

## 4.2 Proof of Proposition 1

At maturity, i.e. for $h = 0$, the American put option price is $V_t(0, K) = (K - S_t)^+ = S_t(k_t - 1)^+ = S_t v(0, y_t)$. The proof proceeds by induction w.r.t $h$. Let us assume that the homogeneity property holds at time-to-maturity $h-1$, i.e., $V_{t+1}(h-1, K) = S_{t+1} v(h-1, Y_{t+1})$. From Equation (1.9), the definition of the moneyness strike and the Markov property of $Y_t$ under $\mathscr{Q}$ we get

$$
\begin{aligned}
V_t(h,K) &= \max\left[(K-S_t)^+, \mathrm{E}^{\mathscr{Q}}\left[V_{t+1}(h-1,K)|Y_t\right]\right] \\
&= \max\left[(K-S_t)^+, \mathrm{E}^{\mathscr{Q}}\left[\frac{K}{k_{t+1}}v(h-1,Y_{t+1})\middle| Y_t\right]\right] \\
&= S_t \max\left[(k_t-1)^+, \mathrm{E}^{\mathscr{Q}}\left[\frac{k_t}{k_{t+1}}v(h-1,Y_{t+1})\middle| Y_t\right]\right] = S_t v(h,Y_t).
\end{aligned}
$$

### 4.3 Domain and range of the American put pricing operator

Let $\varphi \in L_2(\mathcal{Y})$ and define the operator $\mathcal{E}$ by

$$
\mathcal{E}[\varphi](y_t) := \mathrm{E}^{\mathscr{Q}}\left[\frac{k_t}{k_{t+1}}\varphi(Y_{t+1})\middle| Y_t = y_t\right]. \tag{4.1}
$$

From Equation (2.5) and by the Cauchy-Schwarz inequality we get

$$
|\mathcal{E}[\varphi](\tilde{y})| \leq \left(\int_{\mathcal{X}} e^r \varphi(\tilde{k}e^{-r},x)^2 f_X(x)dx\right)^{1/2}\left(\int_{\mathcal{X}} m(x;\theta_0)^2 e^r \frac{f(x|\tilde{x})^2}{f_X(x)}dx\right)^{1/2}, \tag{4.2}
$$

for any $\tilde{y} = [\tilde{k}\ \tilde{x}']' \in \mathcal{Y}$. Then we have

$$
\begin{aligned}
&\int_{\mathcal{Y}} |\mathcal{E}[\varphi](\tilde{y})|^2 \frac{f_X(\tilde{x})}{\tilde{k}^2}d\tilde{y} \\
&\leq \left(\int_{\mathbb{R}_+}\int_{\mathcal{X}} e^r \varphi(\tilde{k}e^{-r},x)^2 f_X(x)\frac{1}{\tilde{k}^2}d\tilde{k}dx\right)\left(\int_{\mathcal{X}}\int_{\mathcal{X}} m(x;\theta_0)^2 e^r \frac{f(x|\tilde{x})^2}{f_X(x)} f_X(\tilde{x})dxd\tilde{x}\right) \\
&= \left(\int_{\mathcal{Y}} \varphi(y)^2 \frac{f_X(x)}{k^2}dy\right)\left(\int_{\mathcal{X}}\int_{\mathcal{X}} m(x;\theta_0)^2 e^r \frac{f_Z(x,\tilde{x})^2}{f_X(x)f_X(\tilde{x})}dxd\tilde{x}\right) < \infty,
\end{aligned} \tag{4.3}
$$

where we use the change of variable from $\tilde{k}$ to $k = \tilde{k}e^{-r}$ and that the double integral in the RHS of (4.3) is finite from Assumptions A 1, A 3 and A 9 (i) and the Hölder inequality. Thus, $\mathcal{E}[\varphi] \in L_2(\mathcal{Y})$. Since $v(0,.) \in L_2(\mathcal{Y})$, it follows $\mathcal{A}[\varphi] = \max[v(0,.),\mathcal{E}[\varphi]] \in L_2(\mathcal{Y})$. Thus, operator $\mathcal{A}$ maps $L_2(\mathcal{Y})$ into $L_2(\mathcal{Y})$.

### 4.4 Proof of Proposition 2

In this appendix we use the simplified notation $\mathcal{A} = \mathcal{A}_{\theta,f}$, $m(.) = m(.;\theta)$, $g = g(\theta,f)$, $\mathrm{E}^{\mathscr{Q}} = \mathrm{E}^{\mathscr{Q}}_{\theta,f}$, $\mathcal{E} = \mathcal{E}_{\theta,f}$ and $f^{\mathscr{Q}}_{l-1} = f^{\mathscr{Q}}_{\theta,l-1}$. Moreover, we denote by $F^{\mathscr{Q}}_Y(\cdot|y)$ the conditional cdf of $Y_{t+1}$ given $Y_t = y$

under the risk-neutral probability measure.

### 4.4.1 Differentiability of $g$ almost everywhere

Let us first consider the differentiability of $g$ w.r.t. $\theta$. The holding-to-stock price $u(h,.)$ and the American put option-to-stock price $v(h,.)$ depend on the SDF parameter $\theta$ for any $h > 0$. For expository purpose, we omit this dependence in the notation. By Definition 1, Equations (2.6) and (4.1) and the linearity of operator $\mathcal{E}$, we can write the holding-to-stock price ratio as

$$\begin{aligned} u(h,y) &= \mathcal{E}\left[v(h-1,.)\right](y) = \mathcal{E}\left[\max\left[v(0,.), u(h-1,.)\right]\right](y) \\ &= \mathcal{E}\left[\max\left[0, u(h-1,.) - v(0,.)\right]\right](y) + \mathcal{E}\left[v(0,.)\right](y). \end{aligned} \tag{4.4}$$

We know that $u(h-1,y) - v(0,y) \geq 0$ if and only if $k \leq k^*(h-1,x)$, where the critical moneyness strike $k^*(h-1,x)$ is the solution of the equation

$$k - 1 = u(h-1,k,x) \tag{4.5}$$

in $k \in \mathbb{R}_+$. Thus from Equations (2.5) and (4.4)

$$\begin{aligned} u(h,y) &= \int_{\mathcal{X}} m(\tilde{x}) e^{\tilde{r}} \mathbf{1}\{ke^{-\tilde{r}} \leq k^*(h-1,\tilde{x})\}[u(h-1,ke^{-\tilde{r}},\tilde{x}) - v(0,ke^{-\tilde{r}},\tilde{x})] f(\tilde{x}|x) d\tilde{x} \\ &\quad + \int_{\mathcal{X}} m(\tilde{x}) e^{\tilde{r}} v(0,ke^{-\tilde{r}},\tilde{x}) f(\tilde{x}|x) d\tilde{x}, \end{aligned}$$

for $y = [k\ x']'$ and the indicator function $\mathbf{1}\{.\}$. For expository purpose, let us assume that $ke^{-\tilde{r}} \leq k^*(h-1,\tilde{x})$ if and only if $\tilde{r} \geq r^*(h-1,k,\tilde{\sigma})$, where $r^*(h-1,k,\tilde{\sigma})$ is the solution of the equation

$$ke^{-\tilde{r}} = k^*(h-1,\tilde{r},\tilde{\sigma}) \tag{4.6}$$

in $\tilde{r} \in \mathcal{R}$, for given $[k\ \tilde{\sigma}]' \in \mathbb{R}_+ \times \mathcal{S}$.[49] Then we have:

$$\begin{aligned} u(h,y) &= \int_{\mathcal{S}} \int_{r^*(h-1,k,\tilde{\sigma})}^{b} m(\tilde{x}) e^{\tilde{r}} [u(h-1,ke^{-\tilde{r}},\tilde{x}) - v(0,ke^{-\tilde{r}},\tilde{x})] f(\tilde{x}|x) d\tilde{r} d\tilde{\sigma} \\ &\quad + \int_{\mathcal{X}} m(\tilde{x}) e^{\tilde{r}} v(0,ke^{-\tilde{r}},\tilde{x}) f(\tilde{x}|x) d\tilde{x}, \end{aligned} \tag{4.7}$$

[49] This holds for instance when the transition density of $X_t$ given $X_{t-1}$ does not depend on $r_{t-1}$. The argument of the proof extends easily when the set $\{\tilde{r} : ke^{-\tilde{r}} \leq k^*(h-1,\tilde{x})\}$ can be written as the union of a finite number of intervals, but the notation is more cumbersome.

where $b$ is the upper boundary of $\mathcal{R}$. Let us now show that $u$ is continuous and differentiable w.r.t. $\theta$ by induction. This is true for $h = 0$. Now, let us assume that $u(h-1,.)$ is continuous and differentiable w.r.t. $\theta$. From Equations (4.5) and (4.6) and the implicit function theorem, it follows that $r^*(h-1,k,\tilde{\sigma})$ is differentiable w.r.t. $\theta$. Then, by the Leibniz integral rule for differentiation of a definite integral applied to Equation (4.7) and Assumption A 8, $u(h,.)$ is differentiable w.r.t. $\theta$. By using $\mathcal{A}^h[v(0,.)](y) = \max[v(0,y), u(h,y)]$, we get that $\mathcal{A}^h[v(0,.)](y)$ is continuous for all $\theta$ and differentiable for all $\theta$ apart from the values such that $v(0,y) = u(h,y)$. By replacing the differentiability w.r.t. $\theta$ with the Fréchet differentiability w.r.t. $f$, and by following a similar argument, we can show that $\mathcal{A}^h[v(0,.)](y)$ is Fréchet-differentiable w.r.t. $f$, for all $f$, apart from the values such that $v(0,y) = u(h,y)$.

### 4.4.2 Total differential of $g$ w.r.t. the parameters

Let us consider a generic payoff-to-stock price ratio $\varphi \in L_2(\mathcal{Y})$ and the mapping $(\theta, f) \mapsto \mathcal{E}[\varphi]$. The differential of $\mathcal{E}[\varphi]$ w.r.t. $(\theta, f)$ is given by

$$\delta\mathcal{E}[\varphi](y) \;=\; \int_{\mathcal{X}} m(\tilde{x})e^{\tilde{r}}\varphi(ke^{-\tilde{r}},\tilde{x})\delta f(\tilde{x}|x)d\tilde{x} + \int_{\mathcal{X}} \nabla_{\theta'} m(\tilde{x})e^{\tilde{r}}\varphi(ke^{-\tilde{r}},\tilde{x})f(\tilde{x}|x)d\tilde{x}\delta\theta, \quad (4.8)$$

where $\delta f$ and $\delta\theta$ denote infinitesimal variations of parameters $f$ and $\theta$, respectively. Let us now consider the mapping $(\theta, f) \mapsto \mathcal{A}^h[v(0,.)]$, for a given integer $h \geq 1$, and compute its differential w.r.t. $(\theta, f)$ in terms of the differential of $\mathcal{E}$ given in Equation (4.8). We write

$$\mathcal{A}^h[v(0,.)](y) = \left(\mathcal{E} \circ \mathcal{A}^{h-1}[v(0,.)](y) - v(0,y)\right)^+ + v(0,y),$$

where $\circ$ denotes operator composition. The right derivative of function $(.)^+$ is the indicator $\mathbf{1}\{. \geq 0\}$. Then, by the chain rule and the product rule for differentiation and the total differential, we get

$$\delta\mathcal{A}^h[v(0,.)](y) = \mathbf{1}_{\mathcal{C}(h)}(y)\Big(\delta\mathcal{E}\left[v(h-1,.)\right](y) + \mathcal{E} \circ \delta\mathcal{A}^{h-1}[v(0,.)](y)\Big), \quad (4.9)$$

where we make use of the definitions of the continuation region in Equation (2.4) and the American put option-to-stock price ratio in Equation (2.7). We can iterate Equation (4.9) to get

$$\begin{aligned}\delta\mathcal{A}^h[v(0,.)](y) &= \mathbf{1}_{\mathcal{C}(h)}\delta\mathcal{E}\left[v(h-1,.)\right](y) + \mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\delta\mathcal{E}\left[v(h-2,.)\right](y)\\ &\quad +\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-2)}\delta\mathcal{E}\left[v(h-3,.)\right](y)+\dots\\ &\quad +\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\dots\circ\mathbf{1}_{\mathcal{C}(2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(1)}\delta\mathcal{E}\left[v(0,.)\right](y), \end{aligned}\tag{4.10}$$

where operator $\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}$ is such that $\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\left[\varphi\right](y) = \mathbf{1}_{\mathcal{C}(h)}(y)\mathcal{E}\left[\varphi\right](y)$. By using $v(h-l,.) = \mathcal{A}^{h-l}\left[v(0,.)\right]$, for $1\leq l\leq h$, we rewrite Equation (4.10) as:

$$\delta\mathcal{A}^h[v(0,.)](y) = \sum_{l=1}^{h}\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\dots\circ\mathbf{1}_{\mathcal{C}(h-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h-l}\left[v(0,.)\right](y).\tag{4.11}$$

Thus, the total differential of vector $g$ w.r.t. $f$ and $\theta$ is given by

$$\delta g_j = \sum_{l=1}^{h_j}\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-1)}\mathcal{E}\circ\dots\circ\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_j),\ \text{if } j=1,\dots,N.\tag{4.12}$$

### 4.4.3 Fréchet derivative of $g$ w.r.t. the historical transition density

To compute the Fréchet derivative of the vector $g$ w.r.t. $f$, we replace $\delta\mathcal{E}$ in Equation (4.12) from Equation (4.8) with $\delta f(\tilde{x}|x) = \Delta f(\tilde{x}|x)$ and $\delta\theta = 0$. Let us focus on the quantity

$$\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\dots\circ\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_j),$$

for some integers $l$ and $h_j$ such that $1\leq l\leq h_j$, and let us write it explicitly. For $l=1$ this quantity is equal to

$$\begin{aligned}&\mathbf{1}_{\mathcal{C}(h_j)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-1}\left[v(0,.)\right](y_j)\\ &= \mathbf{1}_{\mathcal{C}(h_j)}(y_j)\int_{\mathcal{X}}\mathcal{A}^{h_j-1}[v(0,.)](k_je^{-r_{t+1}},x_{t+1})m(x_{t+1})e^{r_{t+1}}\Delta f(x_{t+1}|x_0)dx_{t+1}.\end{aligned}\tag{4.13}$$

Let us now consider the case $l\geq 2$. First, the operator $\mathcal{A}$ is applied $h_j-l$ times to discount the payoff-to-stock price ratio $v(0,.)$ from date $t+h_j$ back to date $t+l$. Second, $\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}$ is applied to discount from date $t+l$ back to date $t+l-1$:

$$\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_{t+l-1})$$

$$=\mathbf{1}_{\mathcal{C}(h_j-l+1)}(y_{t+l-1})\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-1}e^{-r_{t+l}},x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}.$$

Third, $\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}$ is applied to discount from date $t+l-1$ back to date $t+l-2$:

$$\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_{t+l-2})$$

$$=\mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})\int_{\mathcal{X}}m(x_{t+l-1})e^{r_{t+l-1}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-r_{t+l-1}},x_{t+l-1})\left(\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\right.$$

$$\left.\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-2}e^{-r_{t+l-1}-r_{t+l}},x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}\right)f(x_{t+l-1}|x_{t+l-2})dx_{t+l-1}.$$

Fourth, operators $\mathbf{1}_{\mathcal{C}(h_j-l+3)}\mathcal{E},\ldots,\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}$ are applied successively to discount from date $t+l-2$ back to date $t$ to get

$$\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\ldots\circ\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_j)$$

$$=\int_{\mathcal{Y}}\ldots\int_{\mathcal{Y}}\mathbf{1}_{\mathcal{C}(h_j)}(y_j)\ldots\mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})\frac{k_j}{k_{t+l-2}}$$

$$\cdot\int_{\mathcal{X}}m(x_{t+l-1})e^{r_{t+l-1}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-r_{t+l-1}},x_{t+l-1})$$

$$\cdot\left(\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-2}e^{-r_{t+l-1}-r_{t+l}},x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}\right)$$

$$\cdot f(x_{t+l-1}|x_{t+l-2})dx_{t+l-1}dF^{\mathcal{Q}}(y_{t+l-2}|y_{t+l-3})\ldots dF^{\mathcal{Q}}(y_{t+1}|y_j).$$

By rearranging the terms, the RHS of the previous equation is equal to

$$\mathbf{1}_{\mathcal{C}(h_j)}(y_j)\int_{\mathcal{X}}\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\zeta(h_j,l,x_{t+l},x_{t+l-1};y_j)\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}dx_{t+l-1},$$

where

$$\zeta(h_j,l,x,\tilde{x};y_j) := m(\tilde{x})e^{\tilde{r}}\int_{\mathcal{Y}}\dots\int_{\mathcal{Y}}\mathbf{1}_{\mathcal{C}(h_j-1)}(y_{t+1})\dots\mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})\frac{k_j}{k_{t+l-2}}$$
$$\cdot\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-\tilde{r}},\tilde{x})\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-2}e^{-\tilde{r}-r},x)f(\tilde{x}|x_{t+l-2})$$
$$\cdot dF^{\mathcal{Q}}(y_{t+l-2}|y_{t+l-3})\dots dF^{\mathcal{Q}}(y_{t+1}|y_j). \tag{4.14}$$

Thus, we get

$$\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\dots\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}\left[v(0,.)\right](y_j)$$
$$=\mathbf{1}_{\mathcal{C}(h_j)}(y_j)\int_{\mathcal{X}}\int_{\mathcal{X}}m(x)e^{r}\zeta(h_j,l,x,\tilde{x};y_j)\Delta f(x|\tilde{x})dxd\tilde{x}, \tag{4.15}$$

for $l\geq 2$. From Equations (4.13) and (4.15) we deduce the Fréchet derivative w.r.t. $f$:

$$\langle Dg_j,\Delta f\rangle = \mathbf{1}_{\mathcal{C}(h_j)}(y_j)\int_{\mathcal{X}}m(x)e^{r}\mathcal{A}^{h_j-1}[v(0,.)](k_je^{-r},x)\Delta f(x|x_0)dx$$
$$+\mathbf{1}_{\mathcal{C}(h_j)}(y_j)\sum_{l=2}^{h_j}\int_{\mathcal{X}}\int_{\mathcal{X}}m(x)e^{r}\zeta(h_j,l,x,\tilde{x};y_j)\Delta f(x|\tilde{x})dxd\tilde{x}, \tag{4.16}$$

for $j=1,\dots,N$. To conclude the proof, we rewrite function $\zeta$ in terms of a risk-neutral expectation using

$$e^{\tilde{r}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-\tilde{r}},\tilde{x})\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-2}e^{-\tilde{r}-r},x)$$
$$=\mathrm{E}^{\mathcal{Q}}\left[\frac{k_{t+l-2}}{k_{t+l-1}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(Y_{t+l-1})\mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-1}e^{-r},x)\middle|X_{t+l-1}=\tilde{x},Y_{t+l-2}=y_{t+l-2}\right].$$

Moreover, by the Markov property of $Y_t$ and $X_t$ under $\mathcal{Q}$, and Assumption 2, we have the following equalities:

$$f^{\mathcal{Q}}(y_{t+l-2},\dots,y_{t+1}|x_{t+l-1},y_t)=\frac{f^{\mathcal{Q}}(x_{t+l-1},y_{t+l-2},\dots,y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l-1}|y_t)}$$
$$=\frac{f^{\mathcal{Q}}(x_{t+l-1}|x_{t+l-2})f^{\mathcal{Q}}(y_{t+l-2},\dots,y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l-1}|x_t)}=\frac{m(x_{t+l-1})f(x_{t+l-1}|x_{t+l-2})f^{\mathcal{Q}}(y_{t+l-2},\dots,y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l-1}|x_t)},$$

where we omit the subscripts for sake of notation. Hence:

$$m(\tilde{x})f(\tilde{x}|x_{t+l-2})dF_Y^{\mathscr{Q}}(y_{t+l-2},\ldots,y_{t+1}|y_j) = f_{l-1}^{\mathscr{Q}}(\tilde{x}|x_0)dF_Y^{\mathscr{Q}}(y_{t+l-2},\ldots,y_{t+1}|X_{t+l-1}=\tilde{x},Y_t=y_j).$$

Thus, from (4.14) and the Law of Iterated Expectations:

$$\begin{aligned}\zeta(h_j,l,x,\tilde{x};y_j) &= f_{l-1}^{\mathscr{Q}}(\tilde{x}|x_0)\mathrm{E}^{\mathscr{Q}}\Bigg[\mathbf{1}_{\mathcal{C}(h_j-1)}(Y_{t+1})\ldots\mathbf{1}_{\mathcal{C}(h_j-l+2)}(Y_{t+l-1})\frac{k_j}{k_{t+l-1}} \\ &\qquad \mathcal{A}^{h_j-l}[v(0,.)](k_{t+l-1}e^{-r},x)\Bigg|X_{t+l-1}=\tilde{x},Y_t=y_j\Bigg].\end{aligned}$$

Equation (2.15) follows.

### 4.4.4 Gradient of $g$ w.r.t. the SDF parameter

The gradient of the vector $g$ w.r.t. $\theta$ is obtained by replacing $\delta\mathcal{E}$ in Equation (4.12) with the expression in Equation (4.8) for $\delta f(\tilde{x}|x)=0$ and $\delta\theta=d\theta$. By similar arguments as in Appendix 4.4.3 we get Equation (2.16).

## 4.5 Proof of Proposition 3

The differential w.r.t. the historical transition density $f$ of the functional Lagrangian in Equation (2.26) is

$$\begin{aligned}\delta\mathcal{L} &= \delta D_T(f,\hat{f}) - \omega_T\lambda'\delta g(\hat{\theta}^*,f) - \omega_T\nu_0'\int_{\mathcal{X}}\Gamma_U(x;\hat{\theta}^*)\delta f(x|x_0)dx - \omega_T\mu_0\int_{\mathcal{X}}\delta f(x|x_0)dx \\ &\quad - \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\nu(\tilde{x})'\int_{\mathcal{X}}\Gamma_U(x;\hat{\theta}^*)\delta f(x|\tilde{x})dxd\tilde{x} - \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\mu(\tilde{x})\int_{\mathcal{X}}\delta f(x|\tilde{x})dxd\tilde{x}. \qquad (4.17)\end{aligned}$$

Let us compute explicitly the first two differential terms in the RHS of Equation (4.17). The differential of the criterion $D_T$ is

$$
\begin{aligned}
\delta D_T(f,\hat{f}) &= \int_{\mathcal{X}} \hat{f}_X(\tilde{x}) \int_{\mathcal{X}} \left(1+\log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right)\right) \delta f(x|\tilde{x}) dx d\tilde{x} \\
&\quad +\omega_T \int_{\mathcal{X}} \left(1+\log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)}\right)\right) \delta f(x|x_0) dx \\
&= \int_{\mathcal{X}} \hat{f}_X(\tilde{x}) \int_{\mathcal{X}} \log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right) \delta f(x|\tilde{x}) dx d\tilde{x} + \omega_T \int_{\mathcal{X}} \log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)}\right) \delta f(x|x_0) dx,
\end{aligned}
$$

where we use that $f \in \mathcal{F}$ satisfies the unit mass constraint and hence $\int \delta f(x|\tilde{x}) dx = 0$ for any $\tilde{x} \in \mathcal{X}$. We get the expression of the differential $\delta g(\hat{\theta}^*, f)$ from Proposition 2 by replacing $\theta$ with $\hat{\theta}^*$ and $\Delta f$ with $\delta f$ into Equation (2.15), and using the definition of vectors $\gamma_S$ and $\gamma_L$:

$$
\delta g(\hat{\theta}^*, f) = \int_{\mathcal{X}} \gamma_S(x; \hat{\theta}^*, f) \delta f(x|x_0) dx + \int_{\mathcal{X}} \int_{\mathcal{X}} \gamma_L(x, \tilde{x}; \hat{\theta}^*, f) \delta f(x|\tilde{x}) dx d\tilde{x}.
$$

Then, the differential of the functional Lagrangian $\mathcal{L}$ is

$$
\begin{aligned}
\delta \mathcal{L} &= \int_{\mathcal{X}} \omega_T \left( \log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)}\right) - \lambda' \gamma_S(x; \hat{\theta}^*, f) - \nu_0' \Gamma_U(x; \hat{\theta}^*) - \mu_0 \right) \delta f(x|\, x_0) dx \\
&\quad + \int_{\mathcal{X}} \int_{\mathcal{X}} \left( \log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right) - \omega_T \lambda' \gamma_L(x, \tilde{x}; \hat{\theta}^*, f) \Big/ \hat{f}_X(\tilde{x}) \right. \\
&\qquad \left. - \nu(\tilde{x})' \Gamma_U(x; \hat{\theta}^*) - \mu(\tilde{x}) \right) \hat{f}_X(\tilde{x})\, \delta f(x|\, \tilde{x}) dx d\tilde{x}.
\end{aligned} \tag{4.18}
$$

By the optimality condition in Equation (2.27) and the fundamental lemma of the calculus of variations we get

$$
\log\left(\frac{\hat{f}^*(x|x_0)}{\hat{f}(x|x_0)}\right) - \hat{\lambda}' \gamma_S(x; \hat{\theta}^*, \hat{f}^*) - \hat{\nu}_0' \Gamma_U(x; \hat{\theta}^*) - \hat{\mu}_0 = 0, \tag{4.19}
$$

for a.e. $x \in \mathcal{X}$, and

$$
\log\left(\frac{\hat{f}^*(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right) - \omega_T \hat{\lambda}' \gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) - \hat{\nu}(\tilde{x})' \Gamma_U(x; \hat{\theta}^*) - \hat{\mu}(\tilde{x}) = 0, \tag{4.20}
$$

for a.e. $x, \tilde{x} \in \mathcal{X}$ with $\tilde{x} \neq x_0$. From Equations (4.19) and (4.20) we get

$$\hat{f}^*(x|x_0) = \hat{f}(x|x_0) \exp\left(\hat{\mu}_0 + \hat{\lambda}'\gamma_S(x; \hat{\theta}^*, \hat{f}^*) + \hat{\nu}_0'\Gamma_U(x; \hat{\theta}^*)\right), \tag{4.21}$$

for a.e $x \in \mathcal{X}$, and

$$\hat{f}^*(x|\tilde{x}) = \hat{f}(x|\tilde{x}) \exp\left(\hat{\mu}(\tilde{x}) + \hat{\nu}(\tilde{x})'\Gamma_U(x; \hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^*)\Big/ \hat{f}_X(\tilde{x})\right), \tag{4.22}$$

for a.e. $x, \tilde{x} \in \mathcal{X}$ with $\tilde{x} \neq x_0$. By imposing the unit mass constraints, Equation (2.28) follows. Finally, by imposing that the empirical counterpart of System (2.11) holds for $(\hat{\theta}^*, \hat{f}^*)$, System (2.29) follows.

## 4.6 Large sample properties

In this section we denote by $\mathcal{A}_{\theta,f}$ and $\mathcal{E}_{\theta,f}$ the operators $\mathcal{A}$ and $\mathcal{E}$ with generic parameters $\theta$, $f$.

### 4.6.1 Proof of Proposition 4

For technical reasons, the empirical operators used to define the components of the sample counterpart $G(\theta, \hat{f})$ of the local restrictions are based on a trimmed kernel estimator of the historical transition density. More precisely, we have $G(\theta, \hat{f}) = [g(\theta, \hat{f})' \quad \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_0]']'$. Here, $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_0] = \int_{\mathcal{X}_T} \Gamma_U(x; \theta)\hat{f}(x|x_0)dx$ and the components of $g(\theta, \hat{f})$ are defined through the pricing operator $\mathcal{A}_{\theta,\hat{f}}$ such that $\mathcal{A}_{\theta,\hat{f}}[\varphi](y) = \max\left[(k-1)^+, \mathcal{E}_{\theta,\hat{f}}[\varphi](y)\right]$, where

$$\mathcal{E}_{\theta,\hat{f}}[\varphi](y) = \int_{\mathcal{X}_T} m(x_{t+1}; \theta)e^{r_{t+1}}\varphi(ke^{-r_{t+1}}, x_{t+1})\hat{f}(x_{t+1}|x)dx_{t+1} \tag{4.23}$$

and $(\mathcal{X}_T)$ is the sequence of sets defined in Assumption A 4. We prove Proposition 4 by checking the Assumptions **i-iv** of Theorem 2.1 in Newey and McFadden [1999].

**i)** Let us consider the limit criterion $\mathcal{Q}_0(\theta) = G(\theta, f_0)'\Omega_0 G(\theta, f_0)$, for $\theta \in \Theta$, that is the asymptotic limit of the criterion $\mathcal{Q}_T$ minimized by $\hat{\theta}$ (see Definition 2). This criterion is uniquely minimized at $\theta_0$ by the identification condition in Assumption 5 and since $\Omega_0$ is positive-definite (Assumption A 10).

**ii)** The set $\Theta$ is compact by Assumption A 7.

**iii)** The criterion $\mathcal{Q}_0(\theta)$ is continuous. Indeed, the mapping $\theta \to \mathrm{E}_0[\Gamma_U(X_{t+1};\theta)|X_t = x_0]$ is continuous and, as shown in Appendix 4.4.1, the functions $g_j$, for $j = 1, \dots, N$, are continuous w.r.t. $\theta$ as well.

**iv)** Let us verify that $\mathcal{Q}_T(\theta)$ converges to $\mathcal{Q}_0(\theta)$ uniformly in $\theta \in \Theta$. By uniform convergence of kernel estimators (see Hansen [2008]) and Assumptions A 1-2, A 4-6 and A 9, we can show that

$$\sup_{\theta\in\Theta} \left\| \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_0] - \mathrm{E}_0[\Gamma_U(X_{t+1};\theta)|X_t = x_0] \right\| = o_p(1). \tag{4.24}$$

Let us now consider the uniform convergence of $g(\theta, \hat{f})$. For this purpose, let us start with some definitions and a lemma. Let $a, b > 0$ be such that $k_j \in [e^{-a}, e^{a}]$, for all $j = 1, \dots, N$, and $\mathcal{R} \subset [e^{-b}, e^{b}]$ (see Assumptions 1 and A 1). We consider the sets $\mathcal{Y}_T := [e^{-a}, e^{a}] \times \mathcal{X}_T$ and $\mathcal{Y}'_T := [e^{-(a+b)}, e^{(a+b)}] \times \mathcal{X}_T$. The supremum norm of a continuous function $\varphi \in C^0(\mathbb{R}^{d+1})$ on set $\mathcal{Y}_T$ is defined as $\|\varphi\|_{\mathcal{Y}_T,\infty} := \sup_{y\in\mathcal{Y}_T} |\varphi(y)|$. The supremum norm on set $\mathcal{Y}'_T$ is defined similarly.

**Lemma 1.** *Let $\varphi_\theta \in L_2(\mathcal{Y}) \cap C^0(\mathcal{Y})$ be a function that may depend on parameter $\theta \in \Theta$ and is such that*

$$\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}_0\left[ \varphi_\theta(Y_{t+1})^2 \middle| Y_t = y \right] < \infty. \tag{4.25}$$

*Let $\hat{\varphi}_\theta$ be an estimator of $\varphi_\theta$ such that*

$$\sup_{\theta\in\Theta} \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T,\infty} = o_p(1). \tag{4.26}$$

*Then, under Assumptions A 1-2, A 4-6 and A 9, we have* $\sup_{\theta\in\Theta} \|\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta]\|_{\mathcal{Y}_T,\infty} = o_p(1)$.

*Proof.* □

We use the uniform convergence of the kernel estimator to prove Lemma 1. Let us now write the American option pricing operator as

$$\mathcal{A}_{\theta,f}[\varphi] = v(0,.) + (\mathcal{E}_{\theta,f}[\varphi] - v(0,.))^+ \tag{4.27}$$

and do similarly for its estimator $\mathcal{A}_{\theta,\hat{f}}[\varphi]$. Since $|\max[t,0] - \max[s,0]| \le |t-s|$, for all $t, s \in \mathbb{R}$, we get from Lemma 1 that for any $\varphi_\theta$ satisfying Inequality (4.25)

$$\sup_{\theta\in\Theta} \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T,\infty} = o_p(1) \quad \Rightarrow \quad \sup_{\theta\in\Theta} \|\mathcal{A}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{A}_{\theta,f_0}[\varphi_\theta]\|_{\mathcal{Y}_T,\infty} = o_p(1). \tag{4.28}$$

**Lemma 2.** *Under Assumption A 9, if*

$$\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a-b},e^{a+b}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h_{\theta,f_0}[v(0,.)](Y_{t+1})^2 \middle| Y_t = y\right] < \infty, \tag{4.29}$$

*for $h \in \mathbb{N}$, then*

$$\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^{h+1}_{\theta,f_0}[v(0,.)](Y_{t+1})^2 \middle| Y_t = y\right] < \infty. \tag{4.30}$$

*Proof.* □

By Lemma 2, we can iterate $h \geq 1$ times the Implication (4.28) starting with $\varphi_\theta = \hat{\varphi}_\theta = v(0,.)$ and a sufficiently large moneyness strike support, and get

$$\sup_{\theta\in\Theta} \|\mathcal{A}^h_{\theta,\hat{f}}[v(0,.)] - \mathcal{A}^h_{\theta,f_0}[v(0,.)]\|_{\mathcal{Y}_T,\infty} = o_p(1). \tag{4.31}$$

We deduce that vector $g(\theta,\hat{f})$ converges to $g(\theta, f_0)$ uniformly in $\theta \in \Theta$. Then, from Equation (4.24), vector $G(\theta,\hat{f})$ converges to $G(\theta, f_0)$ uniformly in $\theta \in \Theta$. By Assumption A 10, $\mathcal{Q}_T(\theta)$ converges to $\mathcal{Q}_0(\theta)$ uniformly in $\theta \in \Theta$.

### 4.6.2 Proof of Proposition 5

We prove Proposition 5 in two steps.

**a)** First, we show that there exists an open neighborhood $\Theta_0 \subset \Theta$ such that $\theta_0 \in \Theta_0$ and the criterion $\mathcal{Q}_T(\theta)$ is differentiable w.r.t. $\theta \in \Theta_0$ w.p.a. 1.

**b)** Second, by the consistency of estimator $\hat{\theta}$, we deduce that $\hat{\theta} \in \Theta_0$ w.p.a. 1. From part a), it follows that $\hat{\theta}$ satisfies the first-order condition $\nabla_\theta \mathcal{Q}_T(\hat{\theta}) = 0$ w.p.a. 1. Hence, we can follow the approach in the proof of Theorem 3.2 in Newey and McFadden [1999] to prove Equation (2.36) and conclude.

Let us first prove part a). Since $y_j \in \mathcal{C}_{\theta_0,f_0}(h_j)$ for all $j = 1,\ldots,N$, by using the consistency of estimator $\hat{f}$ and the fact that the continuation region $\mathcal{C}_{\theta,f}(h)$ depends continuously on $\theta$ and $f$, for given $h \geq 1$, we deduce that there exists an open set $\Theta_0 \subset \Theta$ such that $\theta_0 \in \Theta_0$, and $y_j \in \mathcal{C}_{\theta,\hat{f}}(h_j)$ for all $j = 1,\ldots,N$ and $\theta \in \Theta_0$, w.p.a. 1. By the argument in Appendix 4.4.1, this implies that $g_j(\theta,\hat{f})$ is differentiable w.r.t. $\theta \in \Theta_0$, for all $j = 1,\ldots,N$, w.p.a. 1. By using that $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_0]$ is differentiable w.r.t. $\theta$, part a) follows.

For part b), let us check the conditions of Theorem 3.2 in Newey and McFadden [1999].

**i)** The true parameter value $\theta_0$ is an interior point of $\Theta_0$ by part a).

**ii)** Vector $G(\theta, \hat{f})$ is differentiable w.r.t. $\theta \in \Theta_0$, w.p.a. 1, as shown in part a).

**iii)** Let us now show that $G(\theta_0, \hat{f})$ is asymptotically normal. Let us introduce the quantity $\Delta\hat{f}(x|\tilde{x}) := \hat{f}(x|\tilde{x}) - f_0(x|\tilde{x})$. From Equation (2.14) and Proposition 2 we get

$$\sqrt{Th_T^d} g(\theta_0, \hat{f}) = \sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\gamma}_S(x) \Delta\hat{f}(x|x_0) dx$$

$$+ \sqrt{Th_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \gamma_L(x, \tilde{x}; \theta_0, f_0) \Delta\hat{f}(x|\tilde{x}) dx d\tilde{x} + O_p\left(\sqrt{Th_T^d} \|\Delta\hat{f}\|_\infty^2\right). \quad (4.32)$$

Then, by using that the last two components of $G(\theta_0, \hat{f})$ are equal to $\int \bar{\Gamma}_U(x; \hat{\theta}) \Delta\hat{f}(x|x_0) dx$, we get

$$\sqrt{Th_T^d} G(\theta_0, \hat{f}) = \sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta\hat{f}(x|x_0) dx$$

$$+ \sqrt{Th_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta\hat{f}(x|\tilde{x}) dx d\tilde{x} + O_p\left(\sqrt{Th_T^d} \|\Delta\hat{f}\|_\infty^2\right). \quad (4.33)$$

From the uniform convergence of kernel density estimators (see e.g. Hansen [2008]), the supremum norm of $\Delta\hat{f}$ is such that

$$\|\Delta\hat{f}\|_\infty = O_p\left(\sqrt{\frac{\log(T)}{Th_T^{2d}}} + h_T^\rho\right). \quad (4.34)$$

Then, the remainder term in the RHS of Equation (4.33) is such that

$$O_p\left(\sqrt{Th_T^d} \|\Delta\hat{f}\|_\infty^2\right) = O_p\left(\sqrt{Th_T^d}\left(\frac{\log(T)}{Th_T^{2d}} + h_T^{2\rho}\right)\right) = o_p(1), \quad (4.35)$$

under the bandwidth conditions in Assumption A 6. Equations (4.33) and (4.35) yield Equation (2.37). Moreover, from the asymptotic normality of kernel density estimators (see Aït-Sahalia [1992]), the asymptotic distribution of the first term in the RHS of Equation (4.33) is

$$\sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta\hat{f}(x|x_0) dx \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)} \Sigma_S(x_0)\right). \quad (4.36)$$

The bias term is asymptotically vanishing under Assumption A 6 on the bandwidth. Let us now consider the second term of the RHS of Equation (4.33). The integration w.r.t. $\tilde{x} \in \mathcal{X}$ increases the rate of convergence, namely

$$\int_{\mathcal{X}}\int_{\mathcal{X}}\bar{\Gamma}_L(x,\tilde{x})f_X(\tilde{x})\Delta\hat{f}(x|\tilde{x})dxd\tilde{x} = O_p\left(\frac{\log(T)}{\sqrt{T}}+h_T^{\rho}\right) = o_p\left(1/\sqrt{Th_T^d}\right), \tag{4.37}$$

from the bandwidth conditions in Assumption A 6. Thus, the second term of the RHS of Equation (4.33) is negligible as $T\to\infty$ and

$$\sqrt{Th_T^d}G(\theta_0,\hat{f}) \xrightarrow{D} \mathcal{N}\left(0,\frac{\mathcal{K}}{f_X(x_0)}\Sigma_S(x_0)\right). \tag{4.38}$$

**iv)** By a similar argument as in Appendix 4.4.1, the function $\nabla_{\theta'}G(\theta,f_0)$ is continuous w.r.t. $\theta\in\Theta_0$ and, by a similar argument as in Appendix 4.6.1, we have $\sup_{\theta\in\Theta_0}\left\|\nabla_{\theta'}G(\theta,\hat{f})-\nabla_{\theta'}G(\theta,f_0)\right\| = o_p(1)$.

**v)** Finally, the matrix $J_0'\Omega_0J_0$ is nonsingular since $J_0=\nabla_{\theta'}G(\theta_0,f_0)$ is full column-rank (Assumption 6) and $\Omega_0$ is positive definite (Assumption A 10).

Then, the same arguments as in the proof of Theorem 3.2 in Newey and McFadden [1999] imply Equation (2.36), and by using Expression (4.38) the conclusion follows.

### 4.6.3 Proof of Proposition 7

The first order condition for estimator $\hat{\theta}^*$ is

$$\begin{aligned}&h_T^d\left[\nabla_{\theta'}G\left(\hat{\theta}^*,\hat{f}\right)\right]'\Omega_TG\left(\hat{\theta}^*,\hat{f}\right)\\&+\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big|X_t=x_t\right]'\tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big|X_t=x_t\right]=0.\end{aligned}$$

By the mean-value theorem we get

$$\begin{aligned}&h_T^d\left[\nabla_{\theta'}G\left(\hat{\theta}^*,\hat{f}\right)\right]'\Omega_TG\left(\theta_0,\hat{f}\right)\\&+\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big|X_t=x_t\right]'\tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\big|X_t=x_t\right]\\&+\left(h_T^d\left[\nabla_{\theta'}G\left(\hat{\theta}^*,\hat{f}\right)\right]'\Omega_T\left[\nabla_{\theta'}G\left(\tilde{\theta},\hat{f}\right)\right]+\right.\\&\left.+\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big|X_t=x_t\right]'\tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\tilde{\theta}\right)\Big|X_t=x_t\right]\right)\\&\cdot\left(\hat{\theta}^*-\theta_0\right)=0,\end{aligned} \tag{4.39}$$

where $\tilde{\theta}$ is between $\hat{\theta}^*$ and $\theta_0$ componentwise. Let $R_T = [T^{-1/2}R_1 \ \ \left(Th_T^d\right)^{-1/2} R_2]$. By multiplying the two sides of Equation (4.39) by $TR_T'$ and using that

$$R_T^{-1}\left(\hat{\theta}^* - \theta_0\right) = \left(\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right)' \ \ \sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right)'\right)',$$

we get

$$\begin{aligned} A_T \begin{pmatrix} \sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right) \\ \sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right) \end{pmatrix} &= -Th_T^d R_T' \left[\nabla_{\theta'} G\left(\hat{\theta}^*, \hat{f}\right)\right]' \Omega_T G\left(\theta_0, \hat{f}\right) \\ -\frac{1}{T}\sum_{t=1}^{T} TR_T' \mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big| X_t = x_t\right]' &\tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\middle| X_t = x_t\right], \end{aligned} \tag{4.40}$$

where

$$\begin{aligned} A_T &:= Th_T^d R_T'\left[\nabla_{\theta'} G\left(\hat{\theta}^*, \hat{f}\right)\right]' \Omega_T \left[\nabla_{\theta'} G\left(\tilde{\theta}, \hat{f}\right)\right] R_T \\ &+\frac{1}{T}\sum_{t=1}^{T} TR_T' \mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big| X_t = x_t\right]' \tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\tilde{\theta}\right)\Big| X_t = x_t\right] R_T. \end{aligned}$$

By using that $\tilde{J}_0(x)R_2 = 0$ for a.e. $x \in \mathcal{X}$, we get

$$A_T = \begin{pmatrix} R_1'\mathrm{E}_0\left[\tilde{J}_0(X_t)'\tilde{\Omega}_0(X_t)\tilde{J}_0(X_t)\right] R_1 & 0 \\ 0 & R_2' J_0' \Omega_0 J_0 R_2 \end{pmatrix} + o_p(1).$$

Moreover, in the RHS of Equation (4.40) we have

$$Th_T^d R_T'\left[\nabla_{\theta'} G\left(\hat{\theta}^*, \hat{f}\right)\right]' \Omega_T G\left(\theta_0, \hat{f}\right) = \begin{pmatrix} 0 \\ R_2' J_0' \Omega_0 \sqrt{Th_T^d} G\left(\theta_0, \hat{f}\right) \end{pmatrix} + o_p(1)$$

and

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T} TR_T' \mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big| X_t = x_t\right]' \tilde{\Omega}_T(x_t)\,\mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\middle| X_t = x_t\right] \\ = &\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_1' \tilde{J}_0(x_t)' \tilde{\Omega}_0(x_t)\,\mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\middle| X_t = x_t\right] \\ 0 \end{pmatrix} + o_p(1). \end{aligned}$$

Therefore, we get

$$\left(\begin{array}{c}\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right) \\ \sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right)\end{array}\right)$$

$$= - \left(\begin{array}{cc}\left(R_1' \mathrm{E}_0\left[\tilde{J}_0(X_t)'\tilde{\Omega}_0(X_t)\tilde{J}_0(X_t)\right] R_1\right)^{-1} & 0 \\ 0 & \left(R_2' J_0' \Omega_0 J_0 R_2\right)^{-1}\end{array}\right) \Psi_T + o_p(1), \quad (4.41)$$

where

$$\Psi_T := \left(\begin{array}{c}\frac{1}{\sqrt{T}} \sum_{t=1}^T R_1' \tilde{J}_0\left(x_t\right)' \tilde{\Omega}_0\left(x_t\right) \mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1}; \theta_0\right) \middle| X_t = x_t\right] \\ R_2' J_0' \Omega_0 \sqrt{Th_T^d} G\left(\theta_0, \hat{f}\right)\end{array}\right).$$

By similar arguments as in Lemma A.1 in Gagliardini, Gouriéroux and Renault [2011] we have $\Psi_T \overset{D}{\to} N\left(0, W\right)$, where

$$W = \left(\begin{array}{cc}R_1' \mathrm{E}_0\left[\tilde{J}_0\left(X_t\right)' \tilde{\Omega}_0\left(X_t\right) \Sigma_U(X_t) \tilde{\Omega}_0\left(X_t\right) \tilde{J}_0\left(X_t\right)\right] R_1 & 0 \\ 0 & R_2' J_0' \Omega_0 \Sigma_S\left(x_0\right) \Omega_0 J_0 R_2\end{array}\right).$$

The bias induced by the nonparametric estimator vanishes asymptotically since $Th_T^{2\rho} = o(1)$ in Assumption A 6. Hence, $\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right)$ and $\sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right)$ are asymptotically normal, independent, with asymptotic variances

$$AsVar\left[\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right)\right] = \left(R_1' \mathrm{E}_0\left[\tilde{J}_0(X_t)'\tilde{\Omega}_0(X_t)\tilde{J}_0(X_t)\right] R_1\right)^{-1}$$

$$\cdot \left(R_1' \mathrm{E}_0\left[\tilde{J}_0\left(X_t\right)' \tilde{\Omega}_0\left(X_t\right) \Sigma_U(X_t) \tilde{\Omega}_0\left(X_t\right) \tilde{J}_0\left(X_t\right)\right] R_1\right) \left(R_1' \mathrm{E}_0\left[\tilde{J}_0(X_t)'\tilde{\Omega}_0(X_t)\tilde{J}_0(X_t)\right] R_1\right)^{-1},$$

$$AsVar\left[\sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right)\right] = \left(R_2' J_0' \Omega_0 J_0 R_2\right)^{-1} \left(R_2' J_0' \Omega_0 \Sigma_S\left(x_0\right) \Omega_0 J_0 R_2\right) \left(R_2' J_0' \Omega_0 J_0 R_2\right)^{-1},$$

respectively. By the standard argument for the efficient GMM, these asymptotic variances are minimized by choosing $\Omega_0 = \Sigma_S\left(x_0\right)^{-1}$ and $\tilde{\Omega}_0(x) = \Sigma_U(x)^{-1}$, for any $x \in \mathcal{X}$.

### 4.6.4 Proof of Propositions 8 and 9

In this section we sketch the derivation of the asymptotic distribution for the estimators of the density $\hat{f}^*$, of the Lagrange multipliers $\hat{\lambda}$, $\hat{\nu}_0$ and $\hat{\nu}(x)$, for $x \neq x_0$, and of functional $\hat{a}^*$.

### 4.6.5 Asymptotic expansion of the density estimator

Let us consider the tilting function in Equation (2.28) and derive its first-order Taylor expansion. Since $\hat{f}$ and $\hat{f}^*$ converge in probability to $f_0$, vector $\hat{\theta}^*$ to $\theta_0$, Lagrange multipliers $\hat{\lambda}$, $\hat{\nu}_0$ and $\hat{\nu}(x)$ to 0 and weight $\omega_T$ to $\omega$, we keep only the terms of first-order in the Lagrange multipliers estimators. For $\tilde{x} = x_0$ we have

$$\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right) \simeq 1 + \hat{\nu}_0'\Gamma_U(x;\theta_0) + \hat{\lambda}'\gamma_S(x;\theta_0,f_0) = 1 + \hat{\Lambda}'\Gamma_S(x),$$

where $\hat{\Lambda} = [\hat{\lambda}' \ \ \hat{\nu}_0']'$ and $\Gamma_S$ is defined in Equations (2.34), so that

$$\int_{\mathcal{X}} \hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)dx \simeq 1 + \hat{\Lambda}'\mathrm{E}_0\left[\Gamma_S(X_{t+1})|X_t = x_0\right] = 1.$$

Similarly, for any $\tilde{x} \neq x_0$ we have

$$\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/\hat{f}_X(\tilde{x})\right) \simeq 1 + \hat{\nu}(\tilde{x})'\Gamma_U(x;\theta_0) + \omega\hat{\Lambda}'\bar{\Gamma}_L(x,\tilde{x}),$$

where $\bar{\Gamma}_L$ is defined in Equations (2.34). Then

$$\begin{aligned}
&\int_{\mathcal{X}} \hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/\hat{f}_X(\tilde{x})\right)dx \\
&\simeq 1 + \omega\hat{\Lambda}'\mathrm{E}_0\left[\bar{\Gamma}_L(X_{t+1},\tilde{x})|X_t = \tilde{x}\right],
\end{aligned}$$

where we have used that $\mathrm{E}_0\left[\Gamma_U(X_{t+1};\theta_0)|X_t = \tilde{x}\right] = 0$ for a.e. $\tilde{x} \in \mathcal{X}$. Thus, we can approximate the tilting function for the value $\tilde{x} = x_0$ of the conditioning volatility factor as

$$\frac{\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)}{\displaystyle\int_{\mathcal{X}} \hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)dx} \simeq 1 + \hat{\Lambda}'\Gamma_S(x), \tag{4.42}$$

and for any value $\tilde{x} \neq x_0$ as

$$\frac{\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/\hat{f}_X(\tilde{x})\right)}{\int_{\mathcal{X}}\hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/\hat{f}_X(\tilde{x})\right)dx}$$
$$\simeq 1 + \hat{\nu}(\tilde{x})'\Gamma_U(x;\theta_0) + \omega\hat{\Lambda}'\Gamma_L(x,\tilde{x}), \tag{4.43}$$

where $\Gamma_L$ is defined in Equations (2.34). By plugging Approximations (4.42) and (4.43) into Equation (2.28) and keeping only the first-order terms in the estimators we get Approximation (2.41).

### 4.6.6 Asymptotic expansion of the Lagrange multipliers

Let us consider the constraints in System (2.29). They can be rewritten as:

$$\begin{cases} G(\hat{\theta}^*,\hat{f}^*) = 0, \\ \int_{\mathcal{X}}\Gamma_U(x;\hat{\theta}^*)\hat{f}^*(x|\tilde{x})dx = 0, \text{ for a.e. } \tilde{x} \neq x_0. \end{cases} \tag{4.44}$$

Let us expand the LHS of the first equation in System (4.44) around $(\theta_0, f_0)$:

$$\left\langle DG(\theta_0,f_0),\Delta\hat{f}^*\right\rangle + J_0\left(\hat{\theta}^*-\theta_0\right) + O_p\left(\|\Delta\hat{f}^*\|_\infty^2\right) + O_p\left(\|\hat{\theta}^*-\theta_0\|^2\right) = 0, \tag{4.45}$$

where matrix $J_0$ is defined in Assumption 6 and is the sum of the matrices defined in Equations (2.35). Similarly, the expansion of the LHS of the second equation of System (4.44) around $(\theta_0, f_0)$ is

$$\int_{\mathcal{X}}\bar{\Gamma}_U(x)\Delta\hat{f}^*(x|\tilde{x})dx + \tilde{J}_0(\tilde{x})\left(\hat{\theta}^*-\theta_0\right) + O_p\left(\|\hat{\theta}^*-\theta_0\|^2\right) = 0, \tag{4.46}$$

for a.e. $\tilde{x} \neq x_0$, where the $2\times p$ Jacobian matrix $\tilde{J}_0(\tilde{x})$ is defined in Proposition 7 and is such that $\tilde{J}_0(\tilde{x}) = \mathrm{E}_0\left[\bar{\Gamma}_U(X_{t+1})\nabla_{\theta'}\log\left(m(X_{t+1};\theta_0)\right)|X_t=\tilde{x}\right]$. We use Proposition 2 and Approximation

(2.41) and keep only the leading terms to approximate the first term in the LHS of Equation (4.45) as

$$
\begin{aligned}
&\left\langle DG(\theta_0, f_0), \Delta\hat{f}^*\right\rangle = \int_{\mathcal{X}} \bar{\Gamma}_S(x)\Delta\hat{f}^*(x|x_0)dx + \int_{\mathcal{X}}\int_{\mathcal{X}} \bar{\Gamma}_L(x,\tilde{x})f_X(\tilde{x})\Delta\hat{f}^*(x|\tilde{x})dxd\tilde{x} \\
&\simeq \int_{\mathcal{X}} \bar{\Gamma}_S(x)\Delta\hat{f}(x|x_0)dx + \int_{\mathcal{X}}\int_{\mathcal{X}} \bar{\Gamma}_L(x,\tilde{x})f_X(\tilde{x})\Delta\hat{f}(x|\tilde{x})dxd\tilde{x} \\
&\quad + \left[\int_{\mathcal{X}} \bar{\Gamma}_S(x)\Gamma_S(x)' f_0(x|x_0)dx + \omega\int_{\mathcal{X}}\int_{\mathcal{X}} \bar{\Gamma}_L(x,\tilde{x})\Gamma_L(x,\tilde{x})' f_0(x|\tilde{x})dx f_X(\tilde{x})d\tilde{x}\right]\hat{\Lambda} \\
&\quad + \int_{\mathcal{X}}\int_{\mathcal{X}} \bar{\Gamma}_L(x,\tilde{x})\Gamma_U(x;\theta_0)' f_0(x|\tilde{x})dx\hat{\nu}(\tilde{x})f_X(\tilde{x})d\tilde{x} \\
&= \left\langle DG(\theta_0, f_0), \Delta\hat{f}\right\rangle + \left(\Sigma_S(x_0) + \omega\int_{\mathcal{X}} \Sigma_L(x)f_X(x)dx\right)\hat{\Lambda} + \int_{\mathcal{X}} \Sigma_{L,U}(x)\hat{\nu}(x)f_X(x)dx. \qquad (4.47)
\end{aligned}
$$

Similarly, we use Approximation (2.41) to approximate the first term in the LHS of Equation (4.46) as

$$
\int_{\mathcal{X}} \bar{\Gamma}_U(x)\Delta\hat{f}^*(x|\tilde{x})dx \simeq \int_{\mathcal{X}} \bar{\Gamma}_U(x)\Delta\hat{f}(x|\tilde{x})dx + \omega\Sigma_{U,L}(\tilde{x})\hat{\Lambda} + \Sigma_U(\tilde{x})\hat{\nu}(\tilde{x}), \qquad (4.48)
$$

for $\tilde{x} \neq x_0$. We use then Equation (2.14), Approximation (4.47) in Equation (4.45) and Approximation (4.48) in Equation (4.46) to get a linearization of the constraints in System (4.44):

$$
\left\{
\begin{aligned}
G(\theta_0, \hat{f}) + \left(\Sigma_S(x_0) + \omega\int_{\mathcal{X}} \Sigma_L(x)f_X(x)dx\right)\hat{\Lambda} + \int_{\mathcal{X}} \Sigma_{L,U}(x)\hat{\nu}(x)f_X(x)dx + J_0\left(\hat{\theta}^* - \theta_0\right) &\simeq 0, \\
\int_{\mathcal{X}} \bar{\Gamma}_U(x)\Delta\hat{f}(x|\tilde{x})dx + \omega\Sigma_{U,L}(\tilde{x})\hat{\Lambda} + \Sigma_U(\tilde{x})\hat{\nu}(\tilde{x}) + \tilde{J}_0(\tilde{x})\left(\hat{\theta}^* - \theta_0\right) &\simeq 0,
\end{aligned}
\right. \qquad (4.49)
$$

for $\tilde{x} \neq x_0$. We now solve System (4.49) w.r.t. the Lagrange multipliers. Since matrix $\Sigma_U(\tilde{x})$ is invertible for any $\tilde{x}$, we can solve the second approximation of System (4.49) w.r.t. $\hat{\nu}(\tilde{x})$:

$$
\hat{\nu}(\tilde{x}) \simeq -\Sigma_U(\tilde{x})^{-1}\left(\int_{\mathcal{X}} \bar{\Gamma}_U(x)\Delta\hat{f}(x|\tilde{x})dx + \tilde{J}_0(\tilde{x})\left(\hat{\theta}^* - \theta_0\right) + \omega\Sigma_{U,L}(\tilde{x})\hat{\Lambda}\right), \qquad (4.50)
$$

for $\tilde{x} \neq x_0$. We plug Approximation (4.50) into the first approximation of System (4.49) and omit the negligible terms:

$$
G(\theta_0, \hat{f}) + \left(\Sigma_S(x_0) + \omega\int_{\mathcal{X}} \Sigma_{L\perp U}(x)f_X(x)dx\right)\hat{\Lambda} + \left(J_0 - J_{L\|U}\right)\left(\hat{\theta}^* - \theta_0\right) \simeq 0, \qquad (4.51)
$$

for the $(N+2)\times p$ matrix $J_{L\|U} := \mathrm{E}_0\left[\Sigma_{L,U}(X_t)\Sigma_U(X_t)^{-1}\bar{\Gamma}_U(X_{t+1})\nabla_{\theta'}\log\left(m(X_{t+1};\theta_0)\right)\right]$. By inverting Equation (2.40), i.e. $\theta = R\eta$, and using the Equation (4.41) with $\Omega_0 = \Sigma_S(x_0)^{-1}$ and

$\tilde{\Omega}_0(x) = \Sigma_U(x)^{-1}$, for any $x \in \mathcal{X}$, we get

$$\begin{aligned}\hat{\theta}^* - \theta_0 &= R_1\left(\hat{\eta}_1^* - \eta_{1,0}\right) + R_2\left(\hat{\eta}_2^* - \eta_{2,0}\right) \\ &\simeq -R_2\left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2\right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1} G(\theta_0, \hat{f}) \\ &\simeq -P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx, \end{aligned} \tag{4.52}$$

for the $p \times (N+2)$ matrix $P := R_2\left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2\right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1}$. Thus, Approximation (4.51) yields

$$\hat{\Lambda} \simeq -A \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx, \tag{4.53}$$

for the $(N+2) \times (N+2)$ matrix $A$ defined as

$$A := \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L \perp U}(x) f_X(x) dx\right)^{-1} \left(I_{N+2} - \left(J_0 - J_{L \| U}\right) P\right). \tag{4.54}$$

Finally, we use Approximations (4.50) and (4.52) to approximate $\hat{\nu}(\tilde{x})$, for any $\tilde{x} \neq x_0$, as

$$\hat{\nu}(\tilde{x}) \simeq \Sigma_U(\tilde{x})^{-1}\left(\left(\tilde{J}_0(\tilde{x}) P + \omega \Sigma_{U,L}(\tilde{x}) A\right) \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx - \int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|\tilde{x}) dx\right). \tag{4.55}$$

### 4.6.7 Asymptotic distribution of the Lagrange multipliers

Let us first derive the asymptotic distribution of $\hat{\Lambda}$. From Approximation (4.53) and Expression (4.36), we get

$$\sqrt{T h_T^d} \hat{\Lambda} \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)} A \Sigma_S(x_0) A'\right). \tag{4.56}$$

Let us now consider estimator $\hat{\nu}(x)$, for any $x \neq x_0$. By a similar argument as for Expression (4.36), we deduce that the two integrals in Approximation (4.55), standardized by the appropriate rate of convergence, are asymptotically normal and independent, since they involve different conditioning values in $\Delta \hat{f}$. Then we get

$$\sqrt{T h_T^d} \hat{\nu}(x) \xrightarrow{D} \mathcal{N}\left(0, \Sigma_\nu(x)\right), \tag{4.57}$$

for any $x \neq x_0$, where the $2 \times 2$ matrix $\Sigma_\nu$ is defined as

$$\begin{aligned}\Sigma_\nu(x) \;=\; & \frac{\mathcal{K}}{f_X(x_0)}\Sigma_U(x)^{-1}\big(\tilde{J}_0(x)P + \omega\Sigma_{U,L}(x)A\big)\Sigma_S(x_0)\big(\tilde{J}_0(x)P + \omega\Sigma_{U,L}(x)A\big)'\Sigma_U(x)^{-1} \\ & + \frac{\mathcal{K}}{f_X(x)}\Sigma_U(x)^{-1}.\end{aligned}$$

### 4.6.8 Pointwise asymptotic normality of the estimator of the historical transition density

From Approximation (2.41) and the asymptotic distribution of the Lagrange multipliers is Section 4.6.7, Equation (2.42) follows. Then, we deduce Proposition 8 by standard results on the pointwise asymptotic normality of the kernel density estimator (see e.g. Bosq [1998]).

### 4.6.9 Asymptotic distribution of the functionals of the historical transition density

From Equation (2.44) and Approximation (4.52) we get

$$\hat{a}^* - a_0 \simeq -\nabla_{\theta'} a(\theta_0, f_0)P\int_{\mathcal{X}} \Gamma_S(x)\Delta\hat{f}(x|x_0)dx + \left\langle Da(\theta_0, f_0), \Delta\hat{f}^*\right\rangle. \tag{4.58}$$

Let us focus on the last term of the RHS of Approximation (4.58) and proceed in a similar way as done in Section 4.6.6. From Equation (2.32) for direction $\Delta\hat{f}^*$ and state variables vector $x^\star = x_0$, Approximation (2.41) and a similar argument as for Equations (4.37), we get the following approximation of the Fréchet derivative:

$$\begin{aligned}\left\langle Da(\theta_0, f_0), \Delta\hat{f}^*\right\rangle \simeq & \int_{\mathcal{X}} \alpha_S(x)\Delta\hat{f}(x|x_0)dx + \int_{\mathcal{X}} \Sigma_{\alpha_L,U}(x)\hat{\nu}(x)f_X(x)dx \\ & + \left(\Sigma_{\alpha_S,S}(x_0) + \omega\int_{\mathcal{X}} \Sigma_{\alpha_L,L}(x)f_X(x)dx\right)\hat{\Lambda}.\end{aligned} \tag{4.59}$$

Moreover, from Approximation (4.55) and a similar argument as for Equations (4.37) we have

$$\begin{aligned}\int_{\mathcal{X}} \Sigma_{\alpha_L,U}(x)\,\hat{\nu}(x)\,f_X(x)dx \;\simeq\; & \left(\omega\int_{\mathcal{X}} \Sigma_{\alpha_L,U}(x)\,\Sigma_U(x)^{-1}\,\Sigma_{U,L}(x)\,f_X(x)dxA\right. \\ & \left. + J_{\alpha_L\|U}P\right)\int_{\mathcal{X}} \Gamma_S(x)\Delta\hat{f}(x|x_0)dx.\end{aligned}$$

Thus, by using Approximation (4.53) we get

$$\begin{aligned}
\left\langle Da(\theta_0, f_0), \Delta\hat{f}^*\right\rangle &\simeq \int_{\mathcal{X}} \alpha_S(x)\Delta\hat{f}(x|x_0)dx + \Bigg(J_{\alpha_L\|U}P - \Sigma_{\alpha_S,S}\left(x_0\right)A \\
&\quad -\omega\int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}\left(x\right)f_X(x)dxA\Bigg)\int_{\mathcal{X}} \Gamma_S(x)\Delta\hat{f}(x|x_0)dx.
\end{aligned}$$

By using that $(B_1+B_2)^{-1} = B_1^{-1} - (B_1+B_2)^{-1}B_2B_1^{-1}$ for invertible matrices $B_1$ and $B_2$, the matrix $A$ defined in Equation (4.54) can be written as

$$\begin{aligned}
A &= \left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}} \Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1} - \left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}} \Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1} \\
&\quad \cdot\left(J_0 - J_{L\|U}\right)P \\
&= \Sigma_S\left(x_0\right)^{-1} - \left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}} \Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\left(J_0 - J_{L\|U}\right)P \\
&\quad -\omega\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}} \Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\int_{\mathcal{X}} \Sigma_{L\perp U}\left(x\right)f_X(x)dx\Sigma_S\left(x_0\right)^{-1}.
\end{aligned}$$

Thus, we get

$$
\begin{aligned}
&\left\langle Da(\theta_0, f_0), \Delta\hat{f}^{*}\right\rangle \simeq \int_{\mathcal{X}} \Big(\alpha_S(x) - \Sigma_{\alpha_S,S}\left(x_0\right)\Sigma_S\left(x_0\right)^{-1}\Gamma_S(x)\Big)\Delta\hat{f}(x|x_0)dx \\
+ \quad & \omega\Sigma_{\alpha_S,S}\left(x_0\right)\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx \\
& \cdot\Sigma_S\left(x_0\right)^{-1}\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx \\
+ \quad & \Sigma_{\alpha_S,S}\left(x_0\right)\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\left(J_0 - J_{L\|U}\right)P\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx \\
- \quad & \omega\int_{\mathcal{X}}\Sigma_{\alpha_L,L\perp U}\left(x\right)f_X(x)dx\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx \\
+ \quad & \omega\int_{\mathcal{X}}\Sigma_{\alpha_L,L\perp U}\left(x\right)f_X(x)dx \\
& \cdot\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\left(J_0 - J_{L\|U}\right)P\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx \\
+ \quad & J_{\alpha_L\|U}P\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx.
\end{aligned}
$$

Then, from Approximation (4.58) we get

$$
\begin{aligned}
\hat{a}^{*} - a_0 \quad \simeq \quad & \int_{\mathcal{X}}\Big(\alpha_S(x) - \Sigma_{\alpha_S,S}\left(x_0\right)\Sigma_S\left(x_0\right)^{-1}\Gamma_S(x)\Big)\Delta\hat{f}(x|x_0)dx \\
& + \Big(\omega B(\omega) + C(\omega)P\Big)\int_{\mathcal{X}}\Gamma_S(x)\Delta\hat{f}(x|x_0)dx, \qquad (4.60)
\end{aligned}
$$

where the matrix $B(\omega)$ is defined as

$$
\begin{aligned}
B(\omega) \quad := \quad & \Sigma_{\alpha_S,S}\left(x_0\right)\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\Sigma_S\left(x_0\right)^{-1} \\
& - \int_{\mathcal{X}}\Sigma_{\alpha_L,L\perp U}\left(x\right)f_X(x)dx\left(\Sigma_S\left(x_0\right) + \omega\int_{\mathcal{X}}\Sigma_{L\perp U}\left(x\right)f_X(x)dx\right)^{-1}
\end{aligned}
$$

and the matrix $C(\omega)$ as

$$
\begin{aligned}
C(\omega) \quad := \quad & \left(\Sigma_{\alpha_S,S} + \omega \int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}(x)\, f_X(x)dx\right)\left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x)\, f_X(x)dx\right)^{-1} \\
& \cdot \left(J_0 - J_{L\|U}\right) + J_{\alpha_L\|U} - \nabla_{\theta'} a(\theta_0, f_0),
\end{aligned}
$$

for any non-negative scalar $\omega$. The integrand $\alpha_S - \Sigma_{\alpha_S,S}\Sigma_S(x_0)^{-1}\Gamma_S$ in the first term in the RHS of Approximation (4.60) is the residual of the projection of $\alpha_S$ onto $\Gamma_S$, and hence orthogonal to $\Gamma_S$. Then, by a similar argument as for Expression (4.36) and using that $J_0 - J_{L\|U} = J_S + J_{L\perp U}$, we deduce that the difference $\hat{a}^* - a_0$, standardized by the appropriate rate of convergence, is asymptotically normal with variance given in Equation (2.48).

### 4.6.10 Proof of Lemma 1

By the triangular inequality we get

$$
\left\|\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty} \leq \left\|\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty} + \left\|\mathcal{E}_{\theta,\hat{f}}[\varphi_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty}. \tag{4.61}
$$

The first term in the RHS of Inequality (4.61) is the supremum norm on set $\mathcal{Y}_T$ of the function $\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]$, which can be written as

$$
\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta](\tilde{y}) - \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta](\tilde{y}) \quad = \quad \int_{\mathcal{X}_T} m(x;\theta)e^r \left[\hat{\varphi}_\theta - \varphi_\theta\right](\tilde{k}e^{-r}, x)\hat{f}(x|\tilde{x})dx,
$$

for any $\tilde{y} := [\tilde{k} \;\; \tilde{x}']'$ in $\mathcal{Y}$. Then we have

$$
\begin{aligned}
&\left\|\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty} \leq \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T,\infty} \sup_{\tilde{x}\in\mathcal{X}_T} \int_{\mathcal{X}_T} \left| m(x;\theta)e^r \left[\frac{\Delta\hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})} + 1\right] f_0(x|\tilde{x})\right| dx \\
&\leq \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T,\infty} \left[\sup_{x,\tilde{x}\in\mathcal{X}_T}\left|\frac{\Delta\hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})}\right| + 1\right] e^b \sup_{x\in\mathcal{X}} \mathrm{E}\left[\left|m(X_{t+1};\theta)\right| \middle| X_t = x\right],
\end{aligned}
$$

where $b > 0$ is defined at point **iv)** of Section 4.6.1 and $\Delta\hat{f} := \hat{f} - f_0$.

**Lemma 3.** *Under Assumptions A 1-2 and A 4-6,* $\sup_{x,\tilde{x}\in\mathcal{X}_T}\left|\frac{\Delta\hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})}\right| = o_p(1)$.

*Proof.* See Section 4.6.12. □

From the Cauchy-Schwarz inequality, Assumption A 9, Lemma 3 and Equation (4.26) we get

$$\sup_{\theta\in\Theta}\left\|\mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta]-\mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty}=O_p\left(\sup_{\theta\in\Theta}\|\hat{\varphi}_\theta-\varphi_\theta\|_{\mathcal{Y}'_T,\infty}\right)=o_p(1). \tag{4.62}$$

The second term in the RHS of Inequality (4.61) is the supremum norm on set $\mathcal{Y}_T$ of the function $\mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]-\mathcal{E}_{\theta,f_0}[\varphi_\theta]$ given by

$$\begin{aligned}&\mathcal{E}_{\theta,\hat{f}}[\varphi_\theta](\tilde{y})-\mathcal{E}_{\theta,f_0}[\varphi_\theta](\tilde{y})\\&=\int_{\mathcal{X}_T}m(x;\theta)e^r\varphi_\theta(\tilde{k}e^{-r},x)\Delta\hat{f}(x|\tilde{x})dx-\int_{\mathcal{X}_T^C}m(x;\theta)e^r\varphi_\theta(\tilde{k}e^{-r},x)f_0(x|\tilde{x})dx.\end{aligned}$$

Then by the triangle inequality we have

$$\begin{aligned}\left\|\mathcal{E}_{\theta,\hat{f}}[\varphi_\theta]-\mathcal{E}_{\theta,f_0}[\varphi_\theta]\right\|_{\mathcal{Y}_T,\infty}\ \leq\ &\sup_{x,\tilde{x}\in\mathcal{X}_T}\left|\frac{\Delta\hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})}\right|\sup_{\tilde{y}\in\mathcal{Y}}\int_{\mathcal{X}_T}m(x;\theta)e^r\left|\varphi_\theta(\tilde{k}e^{-r},x)\right|f_0(x|\tilde{x})dx\\&+\sup_{\tilde{y}\in\mathcal{Y}}\int_{\mathcal{X}_T^C}m(x;\theta)e^r\left|\varphi_\theta(\tilde{k}e^{-r},x)\right|f_0(x|\tilde{x})dx. \end{aligned}\tag{4.63}$$

By the Cauchy-Schwarz inequality we get

$$\begin{aligned}&\int_{\mathcal{X}_T}m(x;\theta)e^r\left|\varphi_\theta(\tilde{k}e^{-r},x)\right|f_0(x|\tilde{x})dx\\&\leq e^b\left(\mathrm{E}\left[m(X_{t+1};\theta)^2\middle|X_t=\tilde{x}\right]\right)^{\frac{1}{2}}\left(\mathrm{E}\left[|\varphi_\theta(\tilde{k}e^{-r_{t+1}},X_{t+1})|^2\middle|X_t=\tilde{x}\right]\right)^{\frac{1}{2}}.\end{aligned}$$

Similarly, we get the counterpart of this inequality for the set-theoretical complement $\mathcal{X}_T^C$ of the domain of integration:

$$\begin{aligned}&\int_{\mathcal{X}_T^C}m(x;\theta)e^r\left|\varphi_\theta(\tilde{k}e^{-r},x)\right|f_0(x|\tilde{x})dx\\&\leq e^b\left(\mathrm{E}\left[m(X_{t+1};\theta)^2\mathbf{1}_{\mathcal{X}_T^C}(X_{t+1})\middle|X_t=\tilde{x}\right]\right)^{\frac{1}{2}}\left(\mathrm{E}\left[|\varphi_\theta(\tilde{k}e^{-r_{t+1}},X_{t+1})|^2\middle|X_t=\tilde{x}\right]\right)^{\frac{1}{2}}.\end{aligned}$$

Let us focus on the square of the second term in the RHS of the previous inequality. By the Hölder inequality we get

$$\mathrm{E}\left[ m(X_{t+1};\theta)^2 \mathbf{1}_{\mathcal{X}_T^C}(X_{t+1}) \middle| X_t = \tilde{x} \right] \leq \left(\mathrm{E}\left[ |m(X_{t+1};\theta)|^{2\bar{p}} \middle| X_t = \tilde{x}\right]\right)^{\frac{1}{\bar{p}}} \underbrace{\left(\mathrm{E}\left[ \mathbf{1}_{\mathcal{X}_T^C}(X_{t+1}) \middle| X_t = \tilde{x}\right]\right)^{\frac{1}{\bar{q}}}}_{=\left(\mathbb{P}\left[ X_{t+1} \in \mathcal{X}_T^C \middle| X_t = \tilde{x}\right]\right)^{\frac{1}{\bar{q}}}},$$

where $\bar{p}, \bar{q} > 1$ are such that $\frac{1}{\bar{p}} + \frac{1}{\bar{q}} = 1$. Thus from Inequality (4.63) we get

$$\left\| \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta] \right\|_{\mathcal{Y}_T,\infty} \leq e^b \left\{ \sup_{x,\tilde{x}\in\mathcal{X}_T} \left| \frac{\Delta \hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})} \right| \sup_{x\in\mathcal{X}} \left(\mathrm{E}\left[ m(X_{t+1};\theta)^2 \middle| X_t = x\right]\right)^{\frac{1}{2}} \right.$$

$$\left. + \sup_{x\in\mathcal{X}_T} \left(\mathrm{E}\left[ |m(X_{t+1};\theta)|^{2\bar{p}} \middle| X_t = x\right]\right)^{\frac{1}{\bar{p}}} \sup_{x\in\mathcal{X}_T} \left(\mathbb{P}\left[ X_{t+1} \in \mathcal{X}_T^C \middle| X_t = x\right]\right)^{\frac{1}{\bar{q}}} \right\}$$

$$\cdot \sup_{y\in[e^{-a},e^a]\times\mathcal{X}} \left(\mathrm{E}\left[ |\varphi_\theta(Y_{t+1})|^2 \middle| Y_t = y\right]\right)^{\frac{1}{2}}.$$

Let us choose $\bar{p}$ such that $2\bar{p} = 2 + \delta$, where $\delta > 0$ is defined in Assumption A 9. From Lemma 3, Assumptions A 4 and A 9 and Inequality (4.25) we get

$$\sup_{\theta\in\Theta} \left\| \mathcal{E}_{\theta,\hat{f}}[\varphi_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta] \right\|_{\mathcal{Y}_T,\infty} = o_p(1). \tag{4.64}$$

Thus, from Inequality (4.61) and Equations (4.62) and (4.64), we get $\sup_{\theta\in\Theta} \left\| \mathcal{E}_{\theta,\hat{f}}[\hat{\varphi}_\theta] - \mathcal{E}_{\theta,f_0}[\varphi_\theta] \right\|_{\mathcal{Y}_T,\infty} = o_p(1)$.

### 4.6.11 Proof of Lemma 2

We use the notation $\mathcal{E} = \mathcal{E}_{\theta,f_0}$ and $\mathcal{A} = \mathcal{A}_{\theta,f_0}$. Let $h \in \mathbb{N}$ and assume that

$$\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a-b},e^{a+b}]\times\mathcal{X}}} \mathrm{E}\left[ \mathcal{A}^h[v(0,.)](Y_{t+1})^2 \middle| Y_t = y\right] < \infty. \tag{4.65}$$

Let us now prove Inequality (4.30). We write the $(h+1)$-fold application of operator $\mathcal{A}$ as in Equation (4.27):

$$\mathcal{A}^{h+1}[v(0,.)] = v(0,.) + \left(\mathcal{E} \circ \mathcal{A}^h[v(0,.)] - v(0,.)\right)^+. \tag{4.66}$$

Since $(t-s)^+ \le |t| + |s|$, for all $t, s \in \mathbb{R}$, we get

$$\mathcal{A}^{h+1}[v(0,.)]^2 \le \left(2v(0,.) + \mathcal{E} \circ \mathcal{A}^h[v(0,.)]\right)^2. \tag{4.67}$$

From Inequality (4.67) we get

$$\begin{aligned}
&\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^{h+1}[v(0,.)](Y_{t+1})^2 \middle| Y_t = y\right] \\
&\le 4A + \sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\left(\mathcal{E} \circ \mathcal{A}^h[v(0,.)](Y_{t+1})\right)^2 \middle| Y_t = y\right] \\
&\quad + 4 \sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[v(0,Y_{t+1})\mathcal{E} \circ \mathcal{A}^h[v(0,.)](Y_{t+1}) \middle| Y_t = y\right]. \qquad (4.68)
\end{aligned}$$

We apply the Cauchy-Schwarz inequality and the Law of Iterated Expectations to the second term in the RHS of Inequality (4.68) to get

$$\begin{aligned}
&\sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\left(\mathcal{E} \circ \mathcal{A}^h[v(0,.)](Y_{t+1})\right)^2 \middle| Y_t = y\right] \\
&= \sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\left(\int_{\mathcal{X}} m(\tilde{x};\theta)e^{\tilde{r}}\mathcal{A}^h[v(0,.)](k_{t+1}e^{-\tilde{r}},\tilde{x})f(\tilde{x}|X_{t+1})d\tilde{x}\right)^2 \middle| Y_t = y\right] \\
&\le e^{2b} \sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\mathrm{E}\left[m(X_{t+2};\theta)^2 \middle| X_{t+1}\right]\mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2 \middle| Y_{t+1}\right] \middle| Y_t = y\right] \\
&\le e^{2b}\, C_2 \sup_{\substack{\theta\in\Theta \\ y\in[e^{-a},e^a]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2 \middle| Y_t = y\right], \qquad (4.69)
\end{aligned}$$

where $C_2 := \sup_{\substack{\theta\in\Theta \\ x\in\mathcal{X}}} \mathrm{E}\left[|m(X_{t+1};\theta)|^2 \middle| X_t = x\right]$ is finite from Assumption A 9 (ii). We apply the Cauchy-Schwarz inequality to the last term in the RHS of Inequality (4.68) and we make use of Inequality

(4.69):

$$
\begin{aligned}
&\sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[v(0,Y_{t+1})\mathcal{E}\circ\mathcal{A}^h[v(0,.)](Y_{t+1})\middle|\, Y_t=y\right]\\
&\le \left(A \sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\left(\mathcal{E}\circ\mathcal{A}^h[v(0,.)](Y_{t+1})\right)^2\middle|\, Y_t=y\right]\right)^{\frac{1}{2}}\\
&\le \left(e^{2b}\,C_2 A \sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2\middle|\, Y_t=y\right]\right)^{\frac{1}{2}}.
\end{aligned}
\tag{4.70}
$$

By grouping Inequalities (4.65) and (4.68)-(4.70) we get

$$
\begin{aligned}
&\sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^{h+1}[v(0,.)](Y_{t+1})^2\middle|\, Y_t=y\right]\\
&\le 4A + e^{2b}\,C_2 \sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2\middle|\, Y_t=y\right]\\
&+ \left(e^{2b}\,C_2 A \sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2\middle|\, Y_t=y\right]\right)^{\frac{1}{2}}.
\end{aligned}
\tag{4.71}
$$

By using the Law of Iterated Expectations and $k_{t+1}=k_t e^{-r_{t+1}}$ with $|r_{t+1}|\le b$, we have:

$$
\sup_{\substack{\theta\in\Theta\\ y\in[e^{-a},e^{a}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2\middle|\, Y_t=y\right] \le \sup_{\substack{\theta\in\Theta\\ y\in[e^{-a-b},e^{a+b}]\times\mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h[v(0,.)](Y_{t+2})^2\middle|\, Y_{t+1}=y\right] < \infty,
$$

from Inequality (4.65). The conclusion follows.

### 4.6.12 Proof of Lemma 3

Let us consider the kernel estimator $\hat{f}_X$ of the stationary pdf $f_X$ of $X_t$ defined in Equation (2.20) and the kernel estimator $\hat{f}_Z$ of the stationary pdf $f_Z$ of $[X_t'\ X_{t-1}']'$ defined by

$$
\hat{f}_Z(x,\tilde{x}) = \frac{1}{Th_T^{2d}}\sum_{t=2}^{T} K\left(\frac{X_t-x}{h_T}\right) K\left(\frac{X_{t-1}-\tilde{x}}{h_T}\right).
$$

Let us define $\Delta \hat{f}_Z(x,\tilde{x}) := \hat{f}_Z(x,\tilde{x}) - f_Z(x,\tilde{x})$ and $\Delta \hat{f}_X(x) := \hat{f}_X(x) - f_X(x)$. From the uniform convergence of the kernel density estimation (see Hansen [2008]) and Assumptions A 1-2 and A 5-6 we have

$$\sup_{x,\tilde{x}\in\mathcal{X}_T} \left|\Delta \hat{f}_Z(x,\tilde{x})\right| = O_p\left(\sqrt{\frac{\log(T)}{Th_T^{2d}}} + h_T^{\rho}\right), \quad \sup_{x\in\mathcal{X}_T} \left|\Delta \hat{f}_X(x)\right| = O_p\left(\sqrt{\frac{\log(T)}{Th_T^{d}}} + h_T^{\rho}\right). \tag{4.72}$$

From Assumptions A 4 and A 6 and Equations (4.72), we have

$$\begin{aligned}
\sup_{x,\tilde{x}\in\mathcal{X}_T} \left|\frac{\Delta \hat{f}_Z(x,\tilde{x})}{f_Z(x,\tilde{x})}\right| &= O_p\left((\log(T))^{c_1}\left(\sqrt{\frac{\log(T)}{Th_T^{2d}}} + h_T^{\rho}\right)\right) = o_p(1), \\
\sup_{x\in\mathcal{X}_T} \left|\frac{\Delta \hat{f}_X(x)}{f_X(x)}\right| &= O_p\left((\log(T))^{c_2}\left(\sqrt{\frac{\log(T)}{Th_T^{d}}} + h_T^{\rho}\right)\right) = o_p(1).
\end{aligned} \tag{4.73}$$

Since $f_0(x|\tilde{x}) = f_Z(x,\tilde{x})/f_X(\tilde{x})$ and $\hat{f}(x|\tilde{x}) = \hat{f}_Z(x,\tilde{x})/\hat{f}_X(\tilde{x})$ we get

$$\begin{aligned}
\frac{\Delta \hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})} &= \frac{\hat{f}(x|\tilde{x})}{f_0(x|\tilde{x})} - 1 = \frac{\hat{f}_Z(x,\tilde{x})}{\hat{f}_X(\tilde{x})f_0(x|\tilde{x})} - 1 = \frac{f_Z(x,\tilde{x}) + \Delta \hat{f}_Z(x,\tilde{x})}{\left[f_X(\tilde{x}) + \Delta \hat{f}_X(\tilde{x})\right]\dfrac{f_Z(x,\tilde{x})}{f_X(\tilde{x})}} - 1 \\
&= \frac{1 + \dfrac{\Delta \hat{f}_Z(x,\tilde{x})}{f_Z(x,\tilde{x})}}{1 + \dfrac{\Delta \hat{f}_X(\tilde{x})}{f_X(\tilde{x})}} - 1 = \frac{\dfrac{\Delta \hat{f}_Z(x,\tilde{x})}{f_Z(x,\tilde{x})} - \dfrac{\Delta \hat{f}_X(\tilde{x})}{f_X(\tilde{x})}}{1 + \dfrac{\Delta \hat{f}_X(\tilde{x})}{f_X(\tilde{x})}},
\end{aligned} \tag{4.74}$$

for any $x, \tilde{x} \in \mathcal{X}_T$. From Equations (4.73)-(4.74) the conclusion follows.

# 5 Appendix to Chapter 3

## 5.1 Kernel estimation

This section describes the kernel estimators. Section 5.1.1 deals with their implementation. Section 5.1.2 with their large sample properties.

### 5.1.1 Implementation

The multivariate kernel density estimators defined in Equations (3.8) and (3.14) and any Nadaraya-Watson estimator of the regression function is computed by using the bandwidth matrix $H = T^{-\frac{1}{6}} V^{\frac{1}{2}}$, where $V$ is the sample unconditional variance-covariance matrix of the state variables and $T$ is the sample size. This bandwidth matrix is chosen for its ease of computation (see e.g. Hardle, Muller, Sperlich and Werwatz [2004], p. 73, and Simonoff [1996], ch. 4). The estimation of conditional and unconditional densities with this bandwidth matrix is equivalent to a three steps estimation procedure: standardizing the data, applying a linear transformation to make them uncorrelated and finally transforming the density back to the original scale.[50]

The estimates considered in this paper are obtained trough the estimation of several regression functions. Any conditional expectation for the transition density $\hat{f}$ of Equation (2.19) of a generic stochastic variable $Z_t$ conditional to the value $x$ of the state variables is estimated by a Nadaraya-Watson estimator:

$$\mathrm{E}_{\hat{f}} \left[ Z_{t+1} \middle| X_t = x \right] \simeq \sum_{t=1}^{T-1} Z_{t+1} K \left( H^{-1} \left( x_t - x \right) \right) \Bigg/ \sum_{t=1}^{T-1} K \left( H^{-1} \left( x_t - x \right) \right). \tag{5.1}$$

For any given $\theta$, the American put option-to-stock price ratio is estimated in an iterative way, using Equations (3.3)-(3.4) evaluated at the kernel estimator $\hat{f}$ of the transition density. The estimation of the ratio at time-to-maturity $h > 0$ passes through the estimation of the discounted conditional expectation of the ratio at time-to-maturity $h - 1$. In particular, this estimation requires the value of the ratio at time-to-maturity $h - 1$ for any historical realization of the state variables and for any value of the moneyness strike. Therefore, at any day the value of the American put option mid-quote-to-share price ratio is computed recursively on a grid on the state variables and moneyness strike domain. The American call option mid-quote-to-share price ratio is computed in a similar way. For any computation of the American option-to-share price ratio, when the ratio on a point outside the grid is necessary, the nearest grid point is selected. The lowest and highest returns on the grid are $1.5$ times the most negative

[50]The first two steps are known in statistics as Mahalanobis transformation.

and positive return on the return time series. The extremes for the realized volatility grid are the $1\%$ and $99\%$ of the realized volatility time series. The extremes of the moneyness strike grid are $0.75$ and $1.25$. Both the return and moneyness strike ranges are divided in $100$ equally spaced points and the RV range is divided in $30$ equally spaced points. The option mid-quote-to-share price ratio when the considered moneyness strike is higher than $1.25$, for a put option, or lower than $0.75$, for a call option, is obtained by a linear extrapolation procedure. When the considered moneyness strike is lower than $0.75$, for a put option, or greater than $1.25$, for a call option, the option-to-share price ratio is set to $0$.

### 5.1.2 Large sample properties

This appendix provides a derivation of the large sample properties of the estimators introduced in Section 3.4.1. The properties are first obtained for a generic quantity that depends on the transition density of the state variables and then adapted to the considered quantities.

Let us indicate the true value of a generic function that depends on the transition density of the state variables by $\mathrm{Q}_0$ and let us consider a real scalar function $g$ and a real stochastic vector $Z_t$ such that this generic function can be written in the form

$$\mathrm{Q}_0(x) = g\left(\mathrm{E}_{f_0}\left[Z_{t+1}\middle| X_t = x\right]\right).$$

The kernel estimator $\mathrm{Q}_{\hat{f}}$ of this function is defined by considering the kernel regression estimator $\mathrm{E}_{\hat{f}}\left[Z_{t+1}\middle| X_t = x\right]$ in place of the true conditional expected value. Each quantity considered in Section 3.4.1 can be written in the form of function $\mathrm{Q}_0$ for an appropriate choice of $g$ and $Z_t$. Let us rescale the state variables such that a common bandwidth $h_T$ can be used. From the theory of kernel estimators, for any value $x$ of the state variables, the kernel regression estimator is pointwise asymptotically normal with $\sqrt{Th_T^2}$-rate of convergence:

$$\sqrt{Th_T^2}\left(\mathrm{E}_{\hat{f}}\left[Z_{t+1}\middle| X_t = x\right] - \mathrm{E}_{f_0}\left[Z_{t+1}\middle| X_t = x\right]\right) \xrightarrow{D} \mathcal{N}\left(0, V(x)\right),$$

where $V(x)$ is defined as

$$V(x) = \mathrm{V}_{f_0}\left[Z_{t+1}\middle| X_t = x\right]\left(\int K^2(x)dx \Big/ f_X(x)\right),$$

for the conditional variance operator $\mathrm{V}_{f_0}\left[.\middle| X_t = x\right]$ under the true historical probability measure (see e.g. Bosq [1998]). The asymptotic distribution of estimator $\mathrm{Q}_{\hat{f}}(x)$ can be derived by the delta method:

$$\sqrt{Th_T^2}\left(\mathrm{Q}_{\hat{f}}(x) - \mathrm{Q}_0(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \gamma(x)' V(x) \gamma(x)\right),$$

where the vector $\gamma(x)$ is defined as

$$\gamma(x) = \frac{\partial g}{\partial b}\left(\mathrm{E}_0\left[Z_{t+1} \middle| X_t = x\right]\right).$$

for the real vector $b$ with the same dimension of $Z_t$.

Let us adapt the expressions to the quantities considered in Section 3.4.1.

**i) Conditional correlation between the state variables**

Let us take the vector $Z_t = [r_t\ \sigma_t\ r_t^2\ \sigma_t^2\ (r_t\sigma_t)]'$ and the scalar function $g$ that depends on the real 5-dimensional vector $b = [b_1\ b_2\ b_3\ b_4\ b_5]'$:

$$g(b) = (b_5 - b_1 b_2)\left(b_3 - b_1^2\right)^{-0.5}\left(b_4 - b_2^2\right)^{-0.5}.$$

The derivative of $g$ w.r.t. $b$ at $a = [a_1\ a_2\ a_3\ a_4\ a_5]'$ is

$$\frac{\partial g}{\partial b}(a) = g(a)\begin{bmatrix} a_1\left(a_3 - a_1^2\right)^{-1} - a_2\left(a_5 - a_1 a_2\right)^{-1} \\ a_2\left(a_4 - a_2^2\right)^{-1} - a_1\left(a_5 - a_1 a_2\right)^{-1} \\ -0.5\left(a_3 - a_1^2\right)^{-1} \\ -0.5\left(a_4 - a_2^2\right)^{-1} \\ \left(a_5 - a_1 a_2\right)^{-1} \end{bmatrix}.$$

**ii) Conditional Sharpe ratio**

Let us take the vector $Z_t = [(r_t - r_{f,t})\ \ (r_t - r_{f,t})^2]'$ and the real scalar function $g$ that depends on the real 2-dimensional real vector $b = [b_1\ b_2]'$:

$$g(b) = b_1\left(b_2 - b_1^2\right)^{-0.5}.$$

The derivative of $g$ w.r.t. $b$ at $a = [a_1\ a_2]'$ is

$$\frac{\partial g}{\partial b}(a) = \begin{bmatrix} g(a)\left(1 + g^2(a)\right) a_1^{-1} \\ -0.5\ g^3(a) a_1^{-2} \end{bmatrix}.$$

**iii) Conditional skewness**

Let us take the vector $Z_t = [r_t \;\; r_t^2 \; r_t^3]'$ and the real scalar function $g$ that depends on the real 3-dimensional real vector $b = [b_1 \; b_2 \; b_3]'$:

$$g(b) = \left(b_3 - 3b_1b_2 + 2b_1^3\right)\left(b_2 - b_1^2\right)^{-1.5}.$$

The derivative of $g$ w.r.t. $b$ at $a = [a_1 \; a_2 \; a_3]'$ is

$$\frac{\partial g}{\partial b}(a) = \begin{bmatrix} \left(-3a_2 + 6a_1^2 + 3a_1\left(a_2 - a_1^2\right)^{0.5} g(a)\right)\left(a_2 - a_1^2\right)^{-1.5} \\ \left(-3a_1 - 1.5\left(a_2 - a_1^2\right)^{0.5} g(a)\right)\left(a_2 - a_1^2\right)^{-1.5} \\ \left(a_2 - a_1^2\right)^{-1.5} \end{bmatrix}.$$

**iv) Conditional kurtosis**

Let us take the vector $Z_t = [r_t \;\; r_t^2 \; r_t^3 \; r_t^4]'$ and the real scalar function $g$ that depends on the real 4-dimensional real vector $b = [b_1 \; b_2 \; b_3 \; b_4]'$:

$$g(b) = \left(b_4 - 4b_1b_3 + 6b_1^2b_2 - 3b_1^4\right)\left(b_2 - b_1^2\right)^{-2}$$

The derivative of $g$ w.r.t. $b$ at $a = [a_1 \; a_2 \; a_3 \; a_4]'$ is

$$\frac{\partial g}{\partial b}(a) = \begin{bmatrix} \left(-4a_3 + 12a_1a_2 - 12a_1^3 + 4a_1\left(a_2 - a_1^2\right) g(a)\right)\left(a_2 - a_1^2\right)^{-2} \\ \left(6a_1^2 - 2\left(a_2 - a_1^2\right) g(a)\right)\left(a_2 - a_1^2\right)^{-2} \\ -4a_1\left(a_2 - a_1^2\right)^{-2} \\ \left(a_2 - a_1^2\right)^{-2} \end{bmatrix}.$$