SEMI-PARAMETRIC ESTIMATION

OF AMERICAN OPTION PRICES

Patrick Gagliardini

University of Lugano and *Swiss Finance Institute*

patrick.gagliardini@usi.ch

Diego Ronchetti

University of Lugano and *Columbia Business School*

diego.ronchetti@usi.ch

First version: April 2010
This version: May 2012[*]

Semi-Parametric Estimation of American Option Prices

Abstract

We introduce a novel semi-parametric estimator of the price of American options in a discrete time, Markovian framework. The estimator is based on a parametric specification of the stochastic discount factor and is nonparametric w.r.t. the historical dynamics of the state variables. The estimation method exploits the no-arbitrage conditions for a non-defaultable short term bond, the underlying asset and a cross-section of observed prices of American options written on it. We use the dynamic programming representation of American option prices to make explicit the nonlinear restrictions on the Euclidean and functional parameters coming from option data. We obtain an estimator of the transition density of the state variables process by minimizing a statistical measure based on the Kullback-Leibler divergence from a kernel-based transition density, subject to the no-arbitrage restrictions. We use the estimator to compute the price of American options not traded in the market by recursive valuation. Other functionals of the transition density interesting for financial applications can be estimated in a similar way.

Keywords: American option, kernel estimator, semi-parametric estimation, dynamic programming, recursive valuation, Fréchet derivative, nonlinear functional.

JEL Codes: C14, C60, G13.

1 Introduction

This paper deals with the estimation of American option prices in a discrete time, incomplete market, Markovian framework. The state variables vector includes the return on the fundamental asset and 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]).[1] 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 one-period-ahead option value. 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 [1997], 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.

---

[1]Alternative characterizations of the American option pricing problem for special parameterizations 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]).

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 to be in a parametric family indexed by the finite-dimensional parameter θ. The goal is to estimate the true values f_0 and θ_0 of the model parameters by the information in a time-series of state variables observations and a cross-section of observed American option prices. The estimates of θ_0 and f_0 are used to estimate the prices of American options that are not actively traded on the market at the current time. We also propose new semi-parametric estimators for a class of linear or nonlinear functionals of θ 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 (e.g., Bakshi, Kapadia and Madan [2003]).

The semi-parametric setting introduced in this paper 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 ex-

ploited through the associated no-arbitrage restrictions. In our framework these restrictions are multi-period and involve the recursive valuation operator for American options. The resulting constraints on θ_0 and f_0 are nonlinear w.r.t. both parameters and do not correspond to standard 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 (e.g., Ai and Chen [2003]; see also Powell [1994] and Ichimura and Todd [2007] for reviews). This difference explains the methodological novelty of our paper. To get numerically tractable estimators, we consider a two-step approach. First, the SDF parameter θ_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 vector θ, 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 we describe the discrete time Markovian framework and define the American option pricing operator for recursive valuation. In Section 3 we introduce the semi-parametric specification with historical transition density f of the state variables and SDF parameter θ. 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. θ and their Fréchet derivative w.r.t. f. In Section 4 we introduce the semi-parametric estimators of the true SDF parameter θ_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 5. The asymptotics is for a long time-series of state variables observations and a fixed number of cross-sectionally observed option prices. We link the asymptotic properties of the proposed estimators to 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.

5

In Section 6 we present the results of a Monte Carlo experiment to study the finite-sample properties of the estimators. Section 7 concludes. In Appendix A we list the set of regularity assumptions for the validity of the asymptotic properties. Proofs of the propositions are gathered in Appendices B-F and proofs of technical lemmas in supplementary materials available on our web-pages.

2 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. We then state an homogeneity property w.r.t. the underlying asset price for a class of American options in Section 2.2. Finally in Section 2.3 we introduce an operator formulation for the American option price useful for the derivation of the theoretical results.

2.1 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 σ_t of relevant pricing factors, with $d \geq 2$. The vector σ_t can include the daily volatility of the stock return, the stock dividend yield and the discount rate. We refer generically to σ_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 return volatility is included in vector σ_t, Assumption 1 is compatible with the usual discrete time stochastic volatility models and multivariate volatility factor models.[2] Assumption 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

---

[2]In a standard discrete time one-factor stochastic volatility model σ_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 $[\varepsilon_t \ u_t]' \sim IIN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. This model allows for a leverage effect through the contemporaneous correlation ρ 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 σ_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 $d = m + 1$.

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 4 and 5.

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 random vectors 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 2.2-5 we consider null risk-free rate and dividend yield on the stock.[3] The results can be extended to stochastic risk-free rate and dividend yield by including them in vector σ_t and considering cum-dividend stock returns. We use a constant non-zero risk-free rate in Section 6 for our Monte Carlo experiment.

2.2 The American put options

Let us consider an American put stock option with strike price $K > 0$. Its payoff at exercise is $(K - S)^+ := \max[K - S, 0]$, if the stock price is S.[4] By the principle of dynamic programming and Assumption 1, the price $V(h, K, S, x)$ of the American put option with time-to-maturity h and strike price K at a date with underlying asset price S and state vector x is such that

$$
V(h, K, S, x) = \begin{cases} \max \left[(K - S)^+, \mathrm{E}^{\mathscr{Q}} \left[V(h - 1, K, S_{t+1}, X_{t+1}) | S_t = S, X_t = x \right] \right], & \text{for } h > 0, \\[2em] (K - S)^+, & \text{for } h = 0, \end{cases}
$$

(2.1)

where $\mathrm{E}^{\mathscr{Q}}[\cdot | S_t = S, X_t = x]$ denotes the conditional expectation operator under the risk-neutral probability measure $\mathscr{Q}$ given S_t and X_t. The quantities $\mathrm{E}^{\mathscr{Q}}[V(h - 1, K, S_{t+1}, X_{t+1}) | S_t = S, X_t = x]$ and

---

[3]In such a case, the prices of some American options (like calls) can be equal to the prices of European options written on the same underlying and with the same contract characteristics. We do not use this equivalence to derive our results.

[4]The results in the paper extend to options with payoff at exercise $\varphi(S, K)$ that is linearly homogeneous w.r.t. the stock price, i.e., $\varphi(S, K) = S\varphi(1, K/S)$. For instance, an American chooser option has payoff at exercise $\varphi(S, K) = \max \left[(K - S)^+, (S - K)^+ \right]$. When the homogeneity property is not satisfied, the approach in the paper 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 the paper.

$(K - S)^+$ are the continuation (or holding) value and the early exercise payoff (or intrinsic value) of the option when $S_t = S$, $X_t = x$, respectively. The latter is the value of the option if it is exercised, the former if it is not. The American option price is the maximum between them. The option price depends on the available information at date t by means of S_t and X_t only, since process (S_t, X_t) is Markov under $\mathcal{Q}$, and it is therefore time-homogeneous. Equation (2.1) corresponds to the value iteration algorithm (see Carriere [1996] and Tsitsiklis and Van Roy [2001]).

Let us show that the dimensionality of the option valuation problem can be reduced by exploiting an homogeneity property of the American option price function. 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

$$Y_t := [k_t \ X_t']'$$ (2.2)

in $\mathcal{Y} := \mathbb{R}_+ \times \mathcal{X}$ is time-homogeneous and Markov of order 1 under both $\mathscr{P}$ and $\mathcal{Q}$. Its transition law is independent of the strike K under both $\mathscr{P}$ and $\mathcal{Q}$. By the Markovianity of process (Y_t) under $\mathcal{Q}$, we deduce the next Proposition 1, which states an homogeneity property of the American option price w.r.t. the underlying asset price similar to Merton [1973] and Merton [1990].[5]

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

$$V(h, K, S, x) = Sv(h, y),$$

where $y = [k \ x']'$, $k = K/S$, the American put option-to-stock price ratio function v is such that

$$v(h, y) = \begin{cases} \max\left[(k - 1)^+, \mathrm{E}^{\mathcal{Q}}\left[e^{r_{t+1}}v(h - 1, Y_{t+1})\,|\,Y_t = y\right]\right], & \textit{for } h > 0, \\ (k - 1)^+, & \textit{for } h = 0, \end{cases}$$ (2.3)

for any $y \in \mathcal{Y}$, and $\mathrm{E}^{\mathcal{Q}}\left[\cdot\,|\,Y_t = y\right]$ denotes the conditional expectation under the risk-neutral probability measure $\mathcal{Q}$ given $Y_t = y$.

Proof. See Appendix B. □

From Proposition 1, the American put option-to-stock price ratio $V(h, K, S, x)/S$ is a function of only the time-to-maturity h, the current moneyness strike $k = K/S$ and the current state variables

---

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

vector x. Since the risk-neutral transition law of the Markov process (Y_t) is independent of strike K, the option-to-stock price ratio is independent of K when the moneyness strike k is given. Thus, the homogeneity property in Proposition 1 reduces the dimensionality of the valuation problem, since function $v(h, \cdot)$ gives the option-to-stock price ratio at time-to-maturity h for any strike K, stock price S and state variables vector x. The daily stock gross return $e^{r_{t+1}} = S_{t+1}/S_t$ in the conditional expectation in Equation (2.3) accounts for the fact that we consider option-to-stock price ratios. Since $\mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}} | Y_t = y\right] = 1$, for almost every (a.e.) $y \in \mathcal{Y}$, by the martingale property of the stock price under $\mathscr{Q}$, the conditional expectation $\mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}} v(h-1, Y_{t+1}) | Y_t = y\right]$ in Equation (2.3) can be written as the expectation $\mathrm{E}^{\tilde{\mathscr{Q}}}\left[v(h-1, Y_{t+1}) | Y_t = y\right]$ under an equivalent probability measure $\tilde{\mathscr{Q}}$.[6] Hence, Equation (2.3) admits the standard formulation of a backward dynamic programming iteration.

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

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

The set-theoretical complement of $\mathcal{C}(h)$ in $\mathcal{Y}$ is the exercise (or stopping) region. The frontier between the two regions is the exercise boundary, and the values of y on this frontier are called critical.

2.3 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 φ on $\mathcal{Y}$ such that $\int_{\mathcal{Y}} \varphi(y)^2 \dfrac{f_X(x)}{k^2} dy < \infty$, where f_X denotes the stationary density of X_t.[7]

Definition 1. *The American put pricing operator* $\mathcal{A} : L_2(\mathcal{Y}) \to L_2(\mathcal{Y})$ *maps a payoff-to-stock price ratio* $\varphi \in L_2(\mathcal{Y})$ *into the put option-to-stock price ratio* $\mathcal{A}[\varphi] \in L_2(\mathcal{Y})$, *that is defined as*

$$\mathcal{A}[\varphi](y) := \max \left[(k-1)^+, \mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}} \varphi(Y_{t+1}) | Y_t = y\right] \right], \qquad \textit{for all } y = [k \; x']' \in \mathcal{Y}.$$

The linear operator that maps $\varphi \in L_2(\mathcal{Y})$ into $\mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}} \varphi(Y_{t+1}) | Y_t = \cdot\right] \in L_2(\mathcal{Y})$ is the conditional expectation operator for Markov process (Y_t) under the probability measure $\tilde{\mathscr{Q}}$. This operator acts on a payoff at date $t+1$ and returns its price at date t taking the stock price as numéraire. By a change of

---

[6]This equivalent (to $\mathscr{P}$ and $\mathscr{Q}$) probability measure, under which the process $(1/S_t)$ is a martingale, is sometimes called the dual (to $\mathscr{Q}$) martingale measure (see, e.g., Shiryaev, Kabanov, Kramkov and Melnikov [1994]).

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

9

variable and Assumption 2, we can rewrite this operator through the historical transition density of X_t and the SDF:

$$\mathrm{E}^{\mathcal{Q}}\left[e^{r_{t+1}}\varphi(Y_{t+1})\middle|\, Y_t = y\right] = \int_{\mathcal{X}} m(x_{t+1})e^{r_{t+1}}\varphi(ke^{-r_{t+1}}, x_{t+1})f(x_{t+1}|x)dx_{t+1}, \qquad y \in \mathcal{Y}. \quad (2.5)$$

From Proposition 1 the option-to-stock price ratio function satisfies the backward recursion $v(h,y) = \mathcal{A}[v(h-1,\cdot)](y)$, with value at maturity $v(0,y) = (k-1)^+$. Thus, we get

$$v(h,y) = \mathcal{A}^h[v(0,\cdot)](y), \qquad \text{for all } h \in \mathbb{N} \text{ and } y \in \mathcal{Y}, \quad (2.6)$$

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

3   A semi-parametric option pricing model

Building on the framework of Section 2, we now introduce a semi-parametric option pricing model. We consider the parameterization of the SDF in Section 3.1 and describe the restrictions on the parameters induced by the no-arbitrage assumption in Section 3.2. Finally in Section 3.3 we derive the sensitivity of the American option-to-stock price ratios to a change in the model parameters.

3.1   The historical and risk neutral parameters

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

Assumption 3. *The one-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 θ 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 θ in Section 5 (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 θ and the infinite-dimensional parameter f. We denote by $\mathcal{A}_{\theta,f}$ the pricing operator $\mathcal{A}$ defined for generic parameters θ and f. This operator yields a semi-parametric pricing model for American put options through Equation (2.6). The goal is to estimate the true SDF parameter θ_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 $\mathcal{A}_{\theta_0,f_0}^{h^\star}[v(0,\cdot)](k^\star, x_{t_0})$ at the current date t_0 for any given moneyness strike $k^\star$ and time-to-maturity $h^\star$, as well as other functionals of interest that depend on the true parameters (θ_0, f_0).

3.2 The no-arbitrage restrictions

The true values θ_0 and f_0 of the model parameters are estimated from the information contained in the no-arbitrage restrictions implied by the market prices. The data consist of two sets of observations. First, we have a sample of N cross-sectionally observed prices of American put options with times-to-maturity h_j and moneyness strikes k_j, where $j = 1, \ldots, N$, traded at the current date t_0. The corresponding option-to-stock price ratios are denoted by v_j, for $j = 1, \ldots, N$. The N options are in the continuation region, i.e. $v_j > (k_j - 1)^+$, for $j = 1, \ldots, N$. Second, we have a sample of T historical observations x_t, where $t = t_0 - T + 1, \ldots, t_0$, for the state variables vector before date t_0.

The observational design for the options reflects the common practice of cross-sectional calibration. This practice accounts for the fact that the set of actively traded options changes from one trading day to the next one. The results of the paper 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 option prices at every trading day in the sample is more difficult because of the time-varying random number and characteristics (time-to-maturity and moneyness strike) of the actively traded options and is beyond the scope of this paper. Furthermore, we do not include in the sample options which are exercised at date t_0 since the corresponding no-arbitrage restrictions imply inequality constraints on (θ_0, f_0), which make the econometric analysis considerably more complex.

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

$$
\begin{cases}
\mathrm{E}_0\left[m(X_{t+1};\theta_0)e^{r_{t+1}}|\, X_t = x\right] = 1, \\[2mm]
\mathrm{E}_0\left[m(X_{t+1};\theta_0)|\, X_t = x\right] = 1,
\end{cases}
\qquad \text{for a.e. } x \in \mathcal{X}, \tag{3.1}
$$

respectively, where $\mathrm{E}_0\left[\cdot\,|\, X_t = x\right]$ denotes the conditional expectation under the true historical probability measure given $X_t = x$. The conditional moment restrictions (3.1) 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

$$
g(\theta_0, f_0) = 0, \tag{3.2}
$$

11

where the vector functional $g = [g_1 \ \ldots \ g_N]'$ with argument (θ, f) is defined by $g_j(\theta, f) := \mathcal{A}^{h_j}_{\theta,f}[v(0, \cdot)](y_j) - v_j$, with $y_j := [k_j \ x'_0]'$ and $x_0 := x_{t_0}$, for $j = 1, \cdots, N$. The derivative market restrictions (3.2) are not parametric moment restrictions, since we cannot write them as an expectation under f_0 of a known function of the unknown parameter θ_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 (3.2) 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 set of no-arbitrage restrictions is given by System (3.1) and Equation (3.2). For the definition and interpretation of the estimators in Section 4, we rewrite these restrictions in an equivalent form. The restrictions (3.1) can be written as $\mathrm{E}_0\left[\Gamma_U(X_{t+1}; \theta_0)|X_t = x\right] = 0$, for a.e. $x \in \mathcal{X}$, where

$$\Gamma_U(x; \theta) := m(x; \theta)[e^r \ 1]' - [1 \ 1]' \tag{3.3}$$

is the moment function for the capital market restrictions. Moreover, since the N traded options at date t_0 are in the continuation region, their prices equal the holding values. Thus, by using Definition 1 and Equation (2.5), the restriction (3.2) can be rewritten as $\mathrm{E}_0\left[\gamma_S(X_{t+1}; \theta_0, f_0)|X_t = x_0\right] = 0$, where the vector function $\gamma_S = [\gamma_{S,1} \ \ldots \ \gamma_{S,N}]'$ is defined as

$$\gamma_{S,j}(x; \theta, f) := m(x; \theta)\gamma_{1,j}(x; \theta, f) - v_j, \qquad \gamma_{1,j}(x; \theta, f) := e^r \mathcal{A}^{h_j-1}_{\theta,f}[v(0, \cdot)](k_j e^{-r}, x), \tag{3.4}$$

for $j = 1, \cdots, N$ and any $x \in \mathcal{X}$. We gather the restrictions (3.1) and (3.2) into system

$$\begin{cases} \mathrm{E}_0\left[\Gamma_U(X_{t+1}; \theta_0)|X_t = x\right] = 0, & \text{for a.e. } x \in \mathcal{X}, \\ \\ \mathrm{E}_0\left[\gamma_S(X_{t+1}; \theta_0, f_0)|X_t = x_0\right] = 0. \end{cases} \tag{3.5}$$

Vector γ_S defines a short-term quasi moment function for the derivative market restrictions. Vector γ_S is not a feasible parametric 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.[8]

---

[8]We could consider γ_S as a moment function involving both a finite-dimensional parameter θ 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 of the state variables vector. Moreover, the estimation approach has to account for parameter f being the transition density of the observed state variables.

3.3 Sensitivity of the derivative market constraints to the model parameters

The informational content of the derivative market restrictions (3.2) depends on the sensitivity of vector functional g to an infinitesimal change in parameters θ and f. In Proposition 2 below we compute the gradient $\nabla_{\theta'} g_j$ of function $g_j(\cdot, f)$ w.r.t. the finite-dimensional parameter θ, and the Fréchet derivative of functional $g_j(\theta, \cdot)$ w.r.t. the infinite-dimensional parameter f, for $j = 1, \ldots, N$. The Fréchet derivative at f in the direction Δf, 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 θ fixed. Hence

$$g_j(\theta, f + \Delta f) = g_j(\theta, f) + \langle Dg_j(\theta, f), \Delta f \rangle + O\left(\|\Delta f\|_\infty^2\right), \tag{3.6}$$

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

Proposition 2. *Let parameters (θ, 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 (θ, f), for all $j = 1, \ldots, N$. Then, under Assumptions 1-3, and A 2 and A 8 in Appendix A, the Fréchet derivative of $g_j(\theta, \cdot)$ at f in the direction Δf is*

$$\langle Dg_j(\theta, f), \Delta f \rangle = \int_{\mathcal{X}} m(x; \theta) \gamma_{1,j}(x; \theta, f) \Delta f(x | x_0) dx$$

$$+ \int_{\mathcal{X}} \int_{\mathcal{X}} m(x; \theta) \gamma_{2,j}(x, \tilde{x}; \theta, f) \Delta f(x | \tilde{x}) dx d\tilde{x}, \tag{3.7}$$

and the gradient of g_j w.r.t. θ is

$$\nabla_{\theta'} g_j(\theta, f) = \mathrm{E}_f\left[\left(\nabla_{\theta'} m(X_{t+1}; \theta)\right) \gamma_{1,j}(X_{t+1}; \theta, f) | X_t = x_0\right]$$

$$+ \int_{\mathcal{X}} \mathrm{E}_f\left[\left(\nabla_{\theta'} m(X_{t+1}; \theta)\right) \gamma_{2,j}(X_{t+1}, \tilde{x}; \theta, f) | X_t = \tilde{x}\right] d\tilde{x}, \tag{3.8}$$

for $j = 1, \ldots, N$, where functions $\gamma_{1,j}(x; \theta, f)$ are given in Equations (3.4) and

$$\gamma_{2,j}(x, \tilde{x}; \theta, f) := \sum_{l=2}^{h_j} f_{\theta,l-1}^{\mathcal{Q}}(\tilde{x} | x_0) \mathrm{E}_{\theta,f}^{\mathcal{Q}}\left[\mathbf{1}_{\mathcal{C}_{\theta,f}(h_j-1)}(Y_{t+1}) \ldots \mathbf{1}_{\mathcal{C}_{\theta,f}(h_j-l+1)}(Y_{t+l-1})\right.$$

$$\left. \cdot e^{R_{t,l}} \mathcal{A}_{\theta,f}^{h_j-l}[v(0, \cdot)](Y_{t+l}) \Big| X_{t+l} = x, X_{t+l-1} = \tilde{x}, Y_t = y_j\right], \tag{3.9}$$

and where $R_{t,l} = \sum_{i=1}^{l} r_{t+i}$ is the cumulated geometric stock return between day t and day $t+l$, $\mathbf{1}_{\mathcal{C}_{\theta,f}(h)}$ is the indicator of the continuation region for time-to-maturity h and parameters (θ, f), the conditional expectation $\mathrm{E}_{\theta,f}^{\mathcal{Q}}[\cdot|\cdot]$ is taken under the risk-neutral probability measure of (Y_t) for parameters (θ, f), and $f_{\theta,l-1}^{\mathcal{Q}}$ is the $(l-1)$-day risk-neutral transition density of (X_t) for parameters (θ, f).

Proof. See Appendix D. $\square$

The Fréchet derivative in Equation (3.7) 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 $\tilde{x} \in \mathcal{X}$. 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 (θ, f), we have

$$g_j(\theta, f) = \int_{\mathcal{X}} m(x; \theta)\gamma_{1,j}(x; \theta, f)f(x|x_0)dx - v_j, \tag{3.10}$$

in a neighborhood of parameters values, for $j = 1, \dots, N$. 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 (3.7). 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 (3.7). 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 $2 \le l \le h_j$, involves a conditional expectation under the risk-neutral probability measure of the l-day-ahead option price in units of the stock price at the current date. The expectation is w.r.t. the paths of process (Y_t) that lie in the continuation region between t and $t+l-1$, and is conditional on $X_{t+l} = x$, $X_{t+l-1} = \tilde{x}$ and $Y_t = y_j$. The weight $f_{\theta,l-1}^{\mathcal{Q}}(\tilde{x}|x_0)$ accounts for the risk-neutral likelihood of a $(l-1)$-day transition of the state variables vector from x_0 to $\tilde{x}$. Function $\gamma_{2,j}$ is equal to zero if the j-th option has time-to-maturity $h_j = 1$.[9]

Finally, the gradient of the local constraint vector g w.r.t. θ in Equation (3.8) also involves two

---

[9]The $\max$ operator in $\mathcal{A}$ does not prevent differentiability of $g(\theta, \cdot)$. Indeed, the kinks induced by the exercise decisions at future dates are smoothed by a subsequent application of the conditional expectation operator (see the proof of Proposition 2 in Appendix D), while the kink for the exercise decision at the current date is irrelevant since the N options are in the continuation region.

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. θ in the RHS of Equation (3.10).

4   Semi-parametric estimation

In this section we introduce semi-parametric estimators of the model parameters 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 θ_0, and then using it to derive an estimator of the historical transition density f_0. 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 4.1), and another one that exploits the full set of no-arbitrage restrictions (Section 4.2). 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 4.3). Finally, we introduce an estimator for a class of functionals of θ_0 and f_0 that includes the prices of American options (Section 4.4).

4.1   The cross-sectional estimator of the SDF parameter

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

Assumption 4. *Under the physical probability measure $\mathscr{P}$, the process (X_t) is geometrically strong mixing, that is, the α-mixing coefficients α_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 time lag j increases. Assumption 4 is satisfied by a wide class of commonly used linear and non-linear time-series processes (see, e.g., Carrasco and Chen [2002]). When the state variables process (X_t) corresponds to discrete-time observations of an underlying continuous time diffusion process, the Markov property in Assumption 1 holds and we require that the skeleton of the process satisfies Assumption 4 (see Chen, Hansen and Carrasco [2010] for sufficient conditions on the drift and volatility functions). The kernel estimator of the historical transition density of process (X_t) is

$$\hat{f}(x|\tilde{x}) := \frac{1}{h_T^d} \sum_{t=2}^{T} K\left(\frac{x_t - x}{h_T}\right) K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right) \bigg/ \sum_{t=2}^{T} K\left(\frac{x_{t-1} - \tilde{x}}{h_T}\right) \qquad (4.1)$$

15

and the kernel estimator of the historical stationary density f_X is

$$\hat{f}_X(x) := \frac{1}{Th_T^d} \sum_{t=1}^{T} K\left(\frac{x_t - x}{h_T}\right), \tag{4.2}$$

where K is a d-dimensional kernel, h_T is the bandwidth (see, e.g., Bosq [1998]) and we have switched to the simpler notation $x_1 := x_{t_0-T+1}, \ldots, x_T := x_{t_0}$.[10]

The full set of no-arbitrage restrictions at date t_0 includes the capital market restrictions (3.1) for the state value x_0 and the derivative market restrictions (3.2). This set of local restrictions can be written as

$$G(\theta_0, f_0) = 0, \tag{4.3}$$

where the $(N + 2)$-dimensional vector functional $G(\theta, f)$ is defined by

$$G(\theta, f) = [\mathrm{E}_f[\Gamma_U(X_{t+1}; \theta)|X_t = x_0]' \ \ g(\theta, f)']'. \tag{4.4}$$

We follow the minimum distance principle and estimate parameter θ 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 (4.4). 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 $\Gamma_U(\cdot; \theta)$ w.r.t. the kernel density $\hat{f}(\cdot|x_0)$. Vector $g(\theta, \hat{f})$ involves the empirical American put pricing operator

$$\mathcal{A}_{\theta,\hat{f}}[\varphi](y) = \max\left[(k-1)^+, \mathrm{E}_{\hat{f}}[m(X_{t+1}; \theta)e^{r_{t+1}}\varphi(ke^{-r_{t+1}}, X_{t+1})|X_t = x]\right], \tag{4.5}$$

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. The practical implementation of operator $\mathcal{A}_{\theta,\hat{f}}$ is discussed in Section 6 in an example.

Definition 2. *The cross-sectional estimator of the SDF parameter θ_0 is $\hat{\theta} := \underset{\theta \in \Theta}{\arg\min} \ \mathcal{Q}_T(\theta)$, for the criterion $\mathcal{Q}_T(\theta) := G(\theta, \hat{f})'\Omega_T G(\theta, \hat{f})$, where Ω_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.

---

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

4.2 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 options developed in Gagliardini, Gouriéroux and Renault [2011] and we introduce a second estimator of the SDF parameter.

Definition 3. *The XMM estimator of the SDF parameter θ_0 is $\hat{\theta}^* := \arg\min_{\theta \in \Theta} \mathcal{Q}_T^*(\theta)$, for the criterion*

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

where $\tilde{\Omega}_T(x)$ is a positive-definite 2×2 weighting matrix for all T and $x \in \mathcal{X}$, P-a.s., and matrix Ω_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 time-series average of quadratic forms in the vectors $E_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_t]$ with weighting matrices $\tilde{\Omega}_T(x_t)$, for $t = 1, \ldots, T$. The average is over the state variables observations. The vector $E_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_t]$ 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 [2010] for the use of optimal instruments in estimating conditional asset pricing models). In the cross-sectional component of criterion $\mathcal{Q}_T^*(\theta)$ we single out the factor h_T^d that shrinks to zero with the sample size. This factor introduces a down-weighting of the cross-sectional component of the criterion $\mathcal{Q}_T^*$ to ensure a suitable convergence rate for estimator $\hat{\theta}^*$ (see Section 5.2).

4.3 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 (4.1) 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 between the transition density f and the kernel

transition density estimator $\hat{f}$ for a given conditioning value $\tilde{x} \in \mathcal{X}$, that is defined as

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

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

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

where

$$D_T(f, \hat{f}) := \int_{\mathcal{X}} d_{KL}(f, \hat{f}|x)\hat{f}_X(x)dx + \omega_T d_{KL}(f, \hat{f}|x_0), \tag{4.6}$$

estimators $\hat{f}$, $\hat{f}_X$ and $\hat{\theta}^*$ are defined in Equations (4.1) and (4.2) and Definition 3, set $\mathcal{F}$ is the set of d-dimensional Markov transition densities and the weight ω_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 ω_T. This local component ensures that the minimization admits a unique solution for $\hat{f}^*(\cdot|x_0)$. The constraints involve both the uniform and the local 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:

$$\mathcal{L} := D_T(f, \hat{f}) - \omega_T \lambda' g(\hat{\theta}^*, f) - \omega_T \nu_0' \mathrm{E}_f[\Gamma_U(X_{t+1}; \hat{\theta}^*)|X_t = x_0] - \omega_T \mu_0 \int_{\mathcal{X}} f(x|x_0)dx$$

$$- \int_{\mathcal{X}} \hat{f}_X(\tilde{x})\nu(\tilde{x})' \mathrm{E}_f[\Gamma_U(X_{t+1}; \hat{\theta}^*)|X_t = \tilde{x}]d\tilde{x} - \int_{\mathcal{X}} \hat{f}_X(\tilde{x})\mu(\tilde{x}) \int_{\mathcal{X}} f(x|\tilde{x})dxd\tilde{x}. \tag{4.7}$$

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 $\nu(\cdot) := [\nu_1(\cdot) \ \nu_2(\cdot)]'$ is a bivariate functional Lagrange multiplier vector for the uniform no-arbitrage restrictions. The scalar μ_0 is the Lagrange multiplier for the local unit mass constraint $\int_{\mathcal{X}} f(x|x_0)dx = 1$ and the Lagrange multiplier scalar function μ 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 Lagrange multipliers λ, ν_0 and μ_0 in Equation (4.7) are multiplied by the weight ω_T, and functions ν and μ 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 $\hat{f}^*$:

$$\delta\mathcal{L}|_{f=\hat{f}^*} = 0. \tag{4.8}$$

The differential of the functional Lagrangian is derived in Appendix E by using Proposition 2. By solving the first-order condition in Equation (4.8), 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 $\hat{\nu}(\cdot)$ of the Lagrange multiplier vectors are such that*

$$\hat{f}^*(x|\tilde{x}) = \begin{cases} \dfrac{\hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)}{\displaystyle\int_{\mathcal{X}} \hat{f}(x|x_0)\exp\left(\hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*)\right)dx}, & \text{if } \tilde{x} = x_0, \\[4mm] \dfrac{\hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/ \hat{f}_X(\tilde{x})\right)}{\displaystyle\int_{\mathcal{X}} \hat{f}(x|\tilde{x})\exp\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*)\Big/ \hat{f}_X(\tilde{x})\right)dx}, & \text{if } \tilde{x} \neq x_0, \end{cases} \tag{4.9}$$

and

$$\begin{cases} \mathrm{E}_{\hat{f}^*}\left[\Gamma_U(X_{t+1};\hat{\theta}^*)\Big| X_t = x\right] = 0, & \text{for a.e. } x \in \mathcal{X}, \\[3mm] \mathrm{E}_{\hat{f}^*}\left[\gamma_S(X_{t+1};\hat{\theta}^*,\hat{f}^*)\Big| X_t = x_0\right] = 0, \end{cases} \tag{4.10}$$

where the vector function γ_L is defined by

$$\gamma_L(x,\tilde{x};\theta,f) := m(x;\theta) \cdot [\gamma_{2,1}(x,\tilde{x};\theta,f) \ \ldots \ \gamma_{2,N}(x,\tilde{x};\theta,f)]', \tag{4.11}$$

for functions $\gamma_{2,j}$ defined in Equation (3.9).

Proof. See Appendix E. □

The estimator $\hat{f}^*$ of the historical transition density in Proposition 3 is an exponential tilting transformation of the kernel estimator $\hat{f}$, i.e. $\hat{f}^* = \hat{\mathcal{T}}^*\hat{f}$, say, where $\hat{\mathcal{T}}^*$ is the exponential tilting factor. When the conditioning value for the historical transition density is x_0, the tilting in Equation (4.9) involves the moment function Γ_U of the uniform capital market restrictions as well as the vector γ_S with the short-term components of the Frechét derivatives of the constraints for the options. Otherwise, the tilting involves moment vector Γ_U and vector γ_L, which is the analogue of vector γ_S for the long-term components of the Frechét derivatives of the constraints for the options. The constraints in System (4.10) are empirical counterparts of the constraints in System (3.5). Moreover, the estimator $\hat{f}^*$ is defined implicitly by Equation (4.9) and System (4.10). Indeed, the vector functions γ_S and γ_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 the initial consistent estimator $\hat{f}^{*(0)} = \hat{f}$ for 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

$$
\left[\hat{\lambda}^{(1)\prime} \ \hat{\nu}_0^{(1)\prime} \right]' = \underset{\lambda \in \mathbb{R}^N, \ \nu_0 \in \mathbb{R}^2}{\arg\min} \ \log \mathrm{E}_{\hat{f}} \left[\exp\left(\nu_0' \Gamma_U(X_{t+1}; \hat{\theta}^*) + \lambda' \gamma_S(X_{t+1}; \hat{\theta}^*, \hat{f}^{*(0)}) \right) \Big| X_t = x_0 \right].
$$

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

$$
\hat{\nu}^{(1)}(\tilde{x}) = \underset{\nu \in \mathbb{R}^2}{\arg\min} \ \log \mathrm{E}_{\hat{f}} \left[\exp\left(\nu' \Gamma_U(X_{t+1}; \hat{\theta}^*) + \frac{\omega_T}{\hat{f}_X(\tilde{x})} \hat{\lambda}^{(1)\prime} \gamma_L(X_{t+1}, X_t; \hat{\theta}^*, \hat{f}^{*(0)}) \right) \Big| X_t = \tilde{x} \right].
$$

v) We derive an updated estimator $\hat{f}^{*(1)}$ for f from Equation (4.9) 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 the paper. In the Monte Carlo experiment in Section 6 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 (4.6) gets a zero weight. In such a case, the estimator in Equation (4.9) and System (4.10) admits a simple interpretation. Estimate $\hat{f}^*(\cdot|\tilde{x})$ is the conditional density that 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.[11]

---

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

The computation of the estimated conditional densities at different conditioning points $\tilde{x}$ can be done separately. While our two-step approach may yield asymptotically inefficient estimates, the joint optimization w.r.t. θ and f combined with the grid method used to evaluate the constraint vector (see Section 6.2) is numerically challenging.

4.4   The estimators of functionals of the model parameters

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 of the SDF parameter θ and the historical transition density f. A functional a in this class is characterized by the first-order expansion around the true parameters value (θ_0, f_0):

$$a(\theta, f) = a(\theta_0, f_0) + \nabla_{\theta'} a(\theta_0, f_0) \, (\theta - \theta_0) + \langle Da(\theta_0, f_0), \Delta f \rangle + O\left(\|\Delta f\|_\infty^2 + \|\theta - \theta_0\|^2\right), \quad (4.12)$$

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

$$\langle Da(\theta_0, f_0), \Delta f \rangle = \int_{\mathcal{X}} \alpha_S(x) \Delta f(x|x^\star) dx + \int_{\mathcal{X}} f_X(\tilde{x}) \int_{\mathcal{X}} \alpha_L(x, \tilde{x}) \Delta f(x|\tilde{x}) dx d\tilde{x}, \quad (4.13)$$

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

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 $\hat{\theta}^*$ and $\hat{f}^*$ are given in Definitions 3 and 4, respectively.*

We exploit Equations (4.12) and (4.13) to derive the large sample properties of estimator $\hat{a}^*$ in Section 5. The class of functionals defined by these equations contains several functionals of interest for financial applications. We provide three examples for which we characterize functions α_S and α_L.

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

From Equation (2.6) we write the American put option-to-stock price ratio for time-to-maturity $h^\star$, moneyness strike $k^\star$ and state variables vector $x^\star$ as $a(\theta, f) = \mathcal{A}_{\theta, f}^{h^\star}[v(0, \cdot)](y^\star)$, for $y^\star = [k^\star \ x^{\star\prime}]'$. Proposition 2 shows that this functional satisfies Equations (4.12) and (4.13) with

$$\alpha_S(x) = m(x; \theta_0)\gamma_1^\star(x; \theta_0, f_0), \quad \alpha_L(x, \tilde{x}) = m(x; \theta_0)\gamma_2^\star(x, \tilde{x}; \theta_0, f_0)/f_X(\tilde{x}), \quad (4.14)$$

where functions $\gamma_1^\star$ and $\gamma_2^\star$ are defined as $\gamma_{1,j}$ and $\gamma_{2,j}$ in Equations (3.4) and (3.9) 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 involves integration w.r.t. the estimated 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 $k^\star_{\theta,f}$ is the solution of the equation $\mathcal{A}^{h^\star}_{\theta,f}[v(0,\cdot)](k^\star_{\theta,f}, x^\star) = \left(k^\star_{\theta,f} - 1\right)^+$ and depends on (θ, f). This defines a functional $a(\theta, f) = k^\star_{\theta,f}$, which satisfies Equations (4.12) and (4.13) with

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

where functions $\gamma_1^\star$ and $\gamma_2^\star$ are as in Equations (4.14) and $y^\star = [k^\star_{\theta_0, f_0} \ x^{\star\prime}]'$. 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 conditional expectation in the RHS involves the one-day transition density f only, due to the Markov property of process (X_t). Functional a satisfies Equations (4.12) and (4.13) 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 $\Psi(X_{t+h^\star}; \theta) = \exp\left(ur_{t+h^\star} + w'\sigma_{t+h^\star}\right)$, with $u \in \mathbb{R}$ and $w \in \mathbb{R}^{d-1}$. The historical conditional moments and cross-moments of the one-day stock return and volatility factor correspond to $\Psi(X_{t+h^\star}; \theta) = r^m_{t+h^\star}\sigma^n_{t+h^\star}$, 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 $h^\star$-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 σ_t, the conditional historical (resp. risk-neutral) cross-moments are the basis for the estimation of the conditional historical (resp. risk-neutral) leverage effects.

5 Large sample properties of the estimators

In this section we study the large sample properties of the semi-parametric estimators introduced in Section 4. 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:

$$\bar{\Gamma}_U(x) := m(x;\theta_0)[e^r \ 1]', \qquad \bar{\gamma}_S(x) := m(x;\theta_0)[\gamma_{1,1}(x;\theta_0,f_0) \ \ldots \ \gamma_{1,N}(x;\theta_0,f_0)]',$$

$$\bar{\Gamma}_S(x) := [\bar{\Gamma}_U(x)' \ \bar{\gamma}_S(x)']', \qquad \bar{\Gamma}_L(x,\tilde{x}) := [0 \ 0 \ \gamma_L(x,\tilde{x};\theta_0,f_0)']'/f_X(\tilde{x}), \qquad (5.1)$$

$$\Gamma_S(x) := [\Gamma_U(x;\theta_0)' \ \gamma_S(x;\theta_0,f_0)']', \qquad \Gamma_L(x,\tilde{x}) := \bar{\Gamma}_L(x,\tilde{x}) - \mathrm{E}_0\left[\bar{\Gamma}_L(X_{t+1},X_t)\big|\, X_t = \tilde{x}\right],$$

where functions $\Gamma_U, \gamma_S, \gamma_{1,j}$, for $j = 1, \ldots, N$, and γ_L are defined in Equations (3.3), (3.4) and (4.11).

5.1 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 A, 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 $\Omega_0 := \plim_{T\to\infty} \Omega_T$ is a symmetric $(N+2) \times (N+2)$ matrix assumed to be positive-definite. Let us assume the global identification of parameter θ_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 θ_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 A, estimator $\hat{\theta}$ is consistent, i.e. $\hat{\theta} \xrightarrow{P} \theta_0$.*

Proof. See Appendix F.1. □

Let us now prove the asymptotic normality of estimator $\hat{\theta}$. The criterion function $\mathcal{Q}_T(\theta)$ is not everywhere differentiable on Θ because of the maximum operator in $\mathcal{A}_{\theta,\hat{f}}$ (see Equation (4.5)). However, by using Proposition 2, the consistency of kernel estimator $\hat{f}$ and the fact that the N options are in the continuation region, we show in Appendix F.2 that the criterion $\mathcal{Q}_T(\theta)$ is differentiable w.r.t. any θ in an open neighborhood of θ_0, with probability approaching 1 (w.p.a. 1). 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 θ_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 (4.4) and Proposition 2 the Jacobian matrix is $J_0 = J_S + J_L$, where

$$J_S := \mathrm{E}_0 \left[\bar{\Gamma}_S(X_{t+1}) \nabla_{\theta'} \log \left(m(X_{t+1}; \theta_0) \right) | X_t = x_0 \right],$$

$$J_L := \mathrm{E}_0 \left[\bar{\Gamma}_L \left(X_{t+1}, X_t \right) \nabla_{\theta'} \log \left(m(X_{t+1}; \theta_0) \right) \right].$$

(5.2)

Moreover, in Appendix F.2 we derive the following asymptotic expansion of the estimator $\hat{\theta}$:

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

(5.3)

where

$$\sqrt{Th_T^d} G(\theta_0, \hat{f}) = \sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}(x|x_0) dx$$

$$+ \sqrt{Th_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} + o_p(1).$$

(5.4)

We use the asymptotic normality of the integrals of kernel estimators (see, e.g., Aït-Sahalia [1992]) to deduce the next Proposition 5.

Proposition 5. *Under Assumptions 1-6 and A 1-10 in Appendix A, 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 Σ_θ is defined as

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

(5.5)

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 $\mathrm{V}_0 \left[\cdot | X_t = x_0 \right]$ denoting the conditional variance under the true historical probability measure given $X_t = x_0$.

Proof. See Appendix F.2. □

The convergence rate of estimator $\hat{\theta}$ is d-dimensional nonparametric due to the conditioning on the d-dimensional vector $X_t = x_0$ in the constraints. Moreover, the bias in the asymptotic distribution is negligible under the bandwidth conditions in Assumption A 6 in Appendix A. The matrix J_0, that is the sum of the matrices defined in Equations (5.2), and the matrix Σ_θ defined in Equation (5.5) 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 $\bar{\Gamma}_S$ (and of Γ_S as well). This matrix does not involve vector function $\bar{\Gamma}_L$ since the second term in the RHS of Equation (5.4) 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* $\dfrac{\mathcal{K}}{f_X(x_0)}(J_0'\Sigma_S(x_0)^{-1}J_0)^{-1}$.

5.2 The XMM estimator of the SDF parameter

The XMM criterion in Definition 3 exploits both the uniform restrictions (3.1) and the restrictions (4.3) at x_0. The global and local identification conditions for parameter θ_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 $\nabla_{\theta'}G\left(\theta_0, f_0\right)\beta = 0$ and $\mathrm{E}_0\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1}; \theta_0\right)|X_t = x\right]\beta = 0$, for a.e. $x \in \mathcal{X}$, is $\beta = 0$.*

Local and uniform restrictions contain information on parameter θ_0 of different character. Therefore, it is important to distinguish between the linear transformations of θ_0 that are identifiable from the uniform restrictions (3.1) alone, and the linear transformations of θ_0 that are identifiable only when the local restrictions (4.3) at x_0 are also taken into account. Following Gagliardini, Gouriéroux and Renault [2011], the former are called full-information identifiable, the latter full-information unidentifiable. To characterize the two types of transformations, let us define the linear space

$$\mathcal{J} := \left\{\beta \in \mathbb{R}^p : \mathrm{E}_0\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1}; \theta_0\right)|X_t = x\right]\beta = 0, \text{ for a.e. } x \in \mathcal{X}\right\}, \tag{5.6}$$

and let $s \leq p$ be the dimension of $\mathcal{J}$. Any linear transformation $\beta'\theta_0$ with $\beta \in \mathcal{J}$ is (locally) full-information unidentifiable, since a change in $\beta'\theta$ has a vanishing first-order impact on vector $\mathrm{E}_0\left[\Gamma_U\left(X_{t+1}; \theta\right)|X_t = x\right]$ for a.e. $x \in \mathcal{X}$. Now, 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 θ to $\eta = [\eta_1'\ \eta_2']'$, defined by

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} := \begin{pmatrix} R_1'\theta \\ R_2'\theta \end{pmatrix}, \tag{5.7}$$

25

is such that the $(p - s)$-dimensional vector η_1 involves full-information identifiable parameters only, while the s-dimensional vector η_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 Ω_0, defined in Section 5.1, and $\tilde{\Omega}_0(x) := \plim_{T \to \infty} \tilde{\Omega}_T(x)$, for any $x \in \mathcal{X}$, be the limit weighting matrices. We prove in Appendix F.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 $\Sigma_U(x) := V_0[\Gamma_U(X_{t+1}; \theta_0)|X_t = x]$, 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 A, 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*

$$\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right) \xrightarrow{D} \mathcal{N}\left(0, \left(R_1' E_0\left[\tilde{J}_0(X_t)' \Sigma_U(X_t)^{-1} \tilde{J}_0(X_t)\right] R_1\right)^{-1}\right),$$

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 $\tilde{J}_0(x) := E_0\left[\nabla_{\theta'} \Gamma_U(X_{t+1}; \theta_0)|X_t = x\right]$, matrices R_1 and R_2 are defined in Equation (5.7) and vectors $\eta_{1,0}$ and $\eta_{2,0}$ denote the true values of parameters η_1 and η_2, respectively.

Proof. See Appendix F.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 as in standard GMM, and the nonparametric rate $\sqrt{T h_T^d}$ for the full-information unidentifiable components. The uniform restrictions in the time-series component of criterion $\mathcal{Q}_T^*$ in Definition 3 are non-informative for the full-information unidentifiable parameter $\eta_{2,0}$, and the kernel estimation conditional on $X_t = x_0$ in the cross-sectional component of criterion $\mathcal{Q}_T^*$ explain the nonparametric convergence rate for that parameter. The factor h_T^d that down-weights the cross-sectional component of $\mathcal{Q}_T^*$ ensures that the kernel estimation conditional on $X_t = x_0$ in the local restrictions does not impact asymptotically the parametric convergence rate of the full-information identifiable parameter $\eta_{1,0}$. 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 $\hat{\eta}_1^*$ in Proposition 7 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 (3.1) allows us to increase the convergence rate of the full-information identifiable parameters and to decrease in general the asymptotic variance of the full-information unidentifiable parameters.

5.3   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 (4.9) in a neighborhood of (θ_0, f_0). Under Assumption A 12 in Appendix A the weight ω_T converges to the non-negative scalar ω. We get (see Appendix F.4)

$$
\hat{f}^*(x|\tilde{x}) \simeq
\begin{cases}
\hat{f}(x|x_0) + f_0(x|x_0)\hat{\Lambda}'\Gamma_S(x), & \text{if } \tilde{x} = x_0, \\[2mm]
\hat{f}(x|\tilde{x}) + f_0(x|\tilde{x})\left(\hat{\nu}(\tilde{x})'\Gamma_U(x;\theta_0) + \omega\hat{\Lambda}'\Gamma_L(x,\tilde{x})\right), & \text{if } \tilde{x} \neq x_0,
\end{cases}
\tag{5.8}
$$

where $\hat{\Lambda} = [\hat{\nu}_0' \ \hat{\lambda}']'$. We prove in Appendix F.4 that the estimators of the Lagrange multipliers $\hat{\Lambda}$ and $\hat{\nu}(\tilde{x})$ for $\tilde{x} \neq x_0$ convergence to zero at rate $\sqrt{Th_T^d}$ and are asymptotically normal. Thus, we get

$$
\hat{f}^*(x|\tilde{x}) = \hat{f}(x|\tilde{x}) + O_p\left(1/\sqrt{Th_T^d}\right),
\tag{5.9}
$$

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 A, the estimator $\hat{f}^*$ is pointwise asymptotically normal with $\sqrt{Th_T^{2d}}$-rate of convergence:*

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

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

Proof. See Appendix F.4. □

The asymptotic distribution of the estimators of smooth functionals of f_0 and θ_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 (4.13). This corresponds, e.g., to American put option-to-stock price ratios and exercise boundary for the value x_0 of the state variables vector, or to conditional cross-moments of the future state variables given $X_t = x_0$ (see examples **i)**-**iii)** in Section 4.4). The derivation of this asymptotic distribution is based on the asymptotic expansion obtained from Equation (4.12):

$$\hat{a}^* - a_0 = \nabla_{\theta'} a(\theta_0, f_0) \left(\hat{\theta}^* - \theta_0 \right) + \left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle + O_p \left(\|\Delta \hat{f}^*\|_\infty^2 \right) + O_p \left(\|\hat{\theta}^* - \theta_0\|^2 \right), \quad (5.10)$$

where the Fréchet derivative $\left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle$ is given in Equation (4.13) with $\Delta f = \Delta \hat{f}^* := \hat{f}^* - f_0$. Since we expect a nonparametric convergence rate for $\hat{a}^*$, the estimation of the SDF parameter affects the asymptotic distribution of $\hat{a}^*$ only through the estimation of the full-information unidentifiable component η_2 (see Proposition 7). The asymptotic expansion is (see Appendix F.4):

$$\sqrt{Th_T^d} \left(\hat{\theta}^* - \theta_0 \right) = -R_2 \left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2 \right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1} \sqrt{Th_T^d} G(\theta_0, \hat{f}) + o_p(1). \quad (5.11)$$

We insert Expansions (5.4), (5.8) and (5.11) into Equation (5.10), and use the asymptotic normality of integral transformations of kernel estimators (see, e.g., Aït-Sahalia [1992]). To state the result we define the following conditional covariance matrices under the true historical probability measure:

$$\Sigma_{\alpha_j,i}(x) := \mathrm{Cov}_0 \left[\alpha_j(X_{t+1}), \Gamma_i(X_{t+1}, X_t) | X_t = x \right],$$

$$\Sigma_{i,l}(x) := \mathrm{Cov}_0 \left[\Gamma_i(X_{t+1}, X_t), \Gamma_l(X_{t+1}, X_t) | X_t = x \right], \quad (5.12)$$

for the subscripts $j = S, L$ and $i, l = S, L, U$ and the state variables vector $x \in \mathcal{X}$.[12] We further define the matrix $\Sigma_{i,j \perp l}(x) := \Sigma_{i,j}(x) - \Sigma_{i,l}(x) \Sigma_l(x)^{-1} \Sigma_{l,j}(x)$, 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

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

that corresponds to the unconditional cross-second moment between $\nabla_{\theta'} \log m$ and the conditional

---

[12]Even if functions Γ_S and Γ_U are independent of the lagged value of the state variables, we use Equations (5.12) for a compact notation. We also omit the dependence of Γ_U on θ_0.

orthogonal projection of α_L onto $\bar{\Gamma}_U$, and the Jacobian matrix

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

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 A, 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 Σ_a is defined as

$$\Sigma_a := \Sigma_{\alpha_S \perp S}(x_0) + M_0(\omega)\Sigma_S(x_0)M_0(\omega)', \tag{5.13}$$

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

$$\begin{aligned}
M_0(\omega) \;:=\; & \omega\left(\Sigma_{\alpha_S,S}\left(x_0\right)\left(\Sigma_S\left(x_0\right) + \omega\mathrm{E}_0\left[\Sigma_{L\perp U}\left(X_t\right)\right]\right)^{-1}\mathrm{E}_0\left[\Sigma_{L\perp U}(X_t)\right]\Sigma_S\left(x_0\right)^{-1}\right. \\[2mm]
& \left. - \mathrm{E}_0\left[\Sigma_{\alpha_L,L\perp U}(X_t)\right]\left(\Sigma_S\left(x_0\right) + \omega\mathrm{E}_0\left[\Sigma_{L\perp U}(X_t)\right]\right)^{-1}\right) \\[2mm]
& + \left(\left(\Sigma_{\alpha_S,S}\left(x_0\right) + \omega\mathrm{E}_0\left[\Sigma_{\alpha_L,L\perp U}\left(X_t\right)\right]\right)\left(\Sigma_S\left(x_0\right) + \omega\mathrm{E}_0\left[\Sigma_{L\perp U}\left(X_t\right)\right]\right)^{-1}(J_S + J_{L\perp U})\right. \\[2mm]
& \left. + J_{\alpha_L\|U} - \nabla_{\theta'}a(\theta_0, f_0)\right)R_2\left(R_2'J_0'\Sigma_S(x_0)^{-1}J_0R_2\right)^{-1}R_2'J_0'\Sigma_S(x_0)^{-1}, \tag{5.14}
\end{aligned}$$

where ω is the probability limit of weight ω_T in Definition 4.

Proof. See Appendix F.4. □

If the SDF parameter vector θ_0 is full-information identifiable, that is, the linear space $\mathcal{J}$ defined in Equation (5.6) is null and $R_2 = 0$, the term in the third and fourth line in the RHS of Equation (5.14) is zero. Then, the asymptotic variance of estimator $\hat{a}^*$ is minimized for $\omega = 0$, that is, when the criterion D_T in Equation (4.6) 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 α_S onto Γ_S given x_0. To get the intuition, suppose that functional a is the conditional expectation of function α_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 $\tilde{f}^*(\cdot|x_0) = \underset{f \in \mathcal{F}_0}{\arg\min}\ d_{KL}(f, \hat{f}|x_0)$ under the constraint $\int_{\mathcal{X}} \Gamma_S(x)f(x|x_0)dx = 0$, and $\mathcal{F}_0$ denotes the set of d-dimensional Markov 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 θ_0 has no effect on the accuracy of estimator $\hat{a}^*$.

If some components of the SDF parameter θ_0 are full-information unidentifiable and $\omega > 0$, matrix $M_0(\omega)\Sigma_S(x_0)M_0(\omega)'$ in the RHS of Equation (5.13) 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 θ_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 Γ_L onto Γ_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 ω can be selected in order to minimize the asymptotic variance matrix $\frac{\mathcal{K}}{f_X(x_0)}\Sigma_a$.[13] This optimal weight $\underset{\omega \in \mathbb{R}_+}{\arg\min}\ M_0(\omega)\Sigma_S(x_0)M_0(\omega)'$ 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 of an American put option with time-to-maturity $h^\star$ and moneyness strike $k^\star$ in state x_0. From example **i)** in Section 4.4, the asymptotic variance of the estimator $\hat{a}^*$ is obtained by using α_S and α_L defined in Equations (4.14), 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 (5.2) by replacing $\bar{\Gamma}_S$ and $\bar{\Gamma}_L$ by α_S and α_L, respectively.

6 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 σ_t (i.e. $d = 2$) representing the volatility of the stock return. We describe the Data-Generating Process (DGP) in Section 6.1, the numerical implementation in Section 6.2 and the results in Section 6.3.

---

[13] Weighting matrix Ω_0 is considered as given and equal to $\Omega_0 = \Sigma_S(x_0)^{-1}$, which is asymptotically optimal for the estimation of θ_0. The asymptotic variance $\frac{\mathcal{K}}{f_X(x_0)}\Sigma_a$ could be minimized by optimizing jointly w.r.t. ω and Ω_0, but the optimization problem becomes more difficult. We do not consider this alternative approach.

6.1 The design

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

$$r_t = r_f + \gamma\sigma_t^2 + \sigma_t\varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1), \tag{6.1}$$

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 σ_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 σ_t^2 is defined by the conditional Laplace transform

$$\mathrm{E}_0[\exp{(-u\sigma_t^2)}\big|\,\sigma_{t-1}^2] = \exp{\left(-\phi_1(u)\sigma_{t-1}^2 - \phi_2(u)\right)}, \qquad u \geq 0, \tag{6.2}$$

where the functions ϕ_1 and ϕ_2 are defined as $\phi_1(u) := \rho u / (1 + cu)$ and $\phi_2(u) := \delta \log{(1 + cu)}$ 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:

$$M_{t,t+1}(\theta) = \exp{(-r_f)} \exp{\left(-\theta_1 - \theta_2\sigma_{t+1}^2 - \theta_3\sigma_t^2 - \theta_4\left(r_{t+1} - r_f\right)\right)}. \tag{6.3}$$

Parameters θ_2 and θ_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 set of conditioning state variables for option valuation becomes smaller, as stated in the next Corollary 10 for a general volatility process (σ_t).

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 $Y_t = [k_t\ \sigma_t']'$ and $y = [k\ \sigma']'$.*

Thus, under the above DGP, the option-to-stock price ratio depends on time-to-maturity h, moneyness strike k and volatility σ only. Moreover, in Definitions 2 and 3 and in Proposition 3, the conditioning variable X_t is replaced by σ_t. Gagliardini, Gouriéroux and Renault [2011] show that the SDF in Equation (6.3) is admissible for the DGP defined in Equations (6.1) and (6.2). Specifically, the no-arbitrage conditions for the stock and the non-defaultable bond are satisfied, i.e. $\mathrm{E}_0\left[M_{t,t+1}(\theta_0)e^{r_{t+1}}|\sigma_t = \sigma\right] = 1$ and $\mathrm{E}_0\left[M_{t,t+1}(\theta_0)e^{r_f}|\sigma_t = \sigma\right] = 1$, for all $\sigma \in \mathbb{R}_+$, iff the true parameter value $\theta_0 = [\theta_1^0\ \theta_2^0\ \theta_3^0\ \theta_4^0]'$ is such that $\theta_1^0 = -\phi_2(\xi)$, $\theta_3^0 = -\phi_1(\xi)$ and $\theta_4^0 = 1/2 + \gamma$, where $\xi = \theta_2^0 + \gamma^2/2 - 1/8$. We report in

$c = 3.65 \cdot 10^{-6}$	$\rho = 9.6 \cdot 10^{-1}$	$\delta = 1.05$	$\gamma = 3.6 \cdot 10^{-1}$
$\theta_1 = 4.55 \cdot 10^{-7}$	$\theta_2 = -5.9 \cdot 10^{-2}$	$\theta_3 = 1.14 \cdot 10^{-1}$	$\theta_4 = 8.6 \cdot 10^{-1}$

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

Table 1 the values of the historical and SDF parameters. They satisfy the above constraints and are calibrated on real data for liquid assets.

Let us now describe the data we create. We generate 1000 time-series of returns and volatility with length $T = 1000$ from date $t_0 - T + 1$ to current date t_0. The volatility σ_0 at date t_0 is the same across simulations[14] and is equal to $6.5 \cdot 10^{-3}$. 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 1. We compute the price ratios 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 moneyness strike $k = 0.966, 0.976, 0.983, 0.991, 0.997, 1.007, 1.011, 1.031$. We display these price ratios by crosses in Figure 1. For each Monte Carlo replication, the data available to the econometrician are a different time-series of state variables and the same 8 American put option-to-stock price ratios. This simulation design reflects the analysis in previous sections, where the value x_0 of the conditioning state variables (that in this Monte Carlo experiment reduces to the volatility σ_0) is constant and given.

We assume that the econometrician does not know the true DGP under $\mathscr{P}$ described in Equations (6.1) and (6.2) but correctly adopts the parametric specification of the SDF in Equation (6.3) and is aware of the Granger non-causality of r_{t-1} on X_t. The transition density f of X_t given σ_{t-1} is treated as a functional parameter. We estimate the model parameters and some American put option-to-stock price ratios for each Monte Carlo replication. We start with the estimation of the SDF parameter θ. In this semi-parametric setting, not all the components of vector θ are full-information identifiable. Indeed, the linear space $\mathcal{J}$ defined in Equation (5.6) is one-dimensional and spanned by the vector $[-\delta c/(1 + c\xi) \;\; 1 \;\; -\rho/(1 + c\xi)^2 \;\; 0]'$ (see Proposition 1 in Gagliardini, Gouriéroux and Renault [2011]). By constructing matrix $[R_1 \; R_2]$ as in Equation (5.7) and inverting the parameter transformation, it is seen that the SDF parameter θ_4 is full-information identifiable, while parameters θ_1, θ_2 and θ_3 are not. We consider the cross-sectional and XMM estimators of the SDF parameter in Definitions

---

[14]We use the time-reversibility of Markov process (σ_t) to generate the paths.

2 and 3 with identity weighting matrices $\Omega_T = I_{N+2}$ and $\tilde{\Omega}_T = I_2$. The XMM estimator becomes

$$\hat{\theta}^* = \arg\min_{\theta \in \Theta} \left(h_T \left\| G(\theta, \hat{f}) \right\|^2 + \frac{1}{T} \sum_{t=1}^{T} \left\| \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|\sigma_t] \right\|^2 \right). \tag{6.4}$$

The cross-sectional estimator $\hat{\theta}$ minimizes the first component of the criterion in the RHS of Equation (6.4). For comparison purpose, we also consider the time-series GMM estimator of parameter θ, that is defined as the minimizer of the second component of the same criterion. We then pass to the estimation of the transition density of the state variables and compute the estimator $\hat{f}^*$ defined in Proposition 3. Finally, we use the estimators $\hat{\theta}^*$ and $\hat{f}^*$ to compute the American put option-to-stock price ratios $\mathcal{A}^{h^\star}_{\hat{\theta}^*, \hat{f}^*}[v(0, \cdot)](y^\star)$ for time-to-maturity $h^\star = 20$, volatility $\sigma^\star = \sigma_0$ and moneyness strikes $k^\star = 0.972, 0.986, 1, 1.03$.

6.2   The numerical implementation

The evaluation at a given θ of the criterion functions minimized by estimators $\hat{\theta}$ and $\hat{\theta}^*$ requires the computation of the conditional expectation $\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|\sigma_t]$ for any sample value σ_t (including σ_0) and of the price ratio $v_{\theta, \hat{f}}(h, y_j) := \mathcal{A}^h_{\theta, \hat{f}}[v(0, \cdot)](y_j)$ for any option $j = 1, \ldots, N$, where operator $\mathcal{A}_{\theta, \hat{f}}$ is defined as in Equation (4.5). Any integral w.r.t. the kernel conditional density $\hat{f}$ involved in the computations is replaced by an empirical weighted sum via 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 σ_t, respectively. The computation of the option price ratios involves recursive applications of the pricing operator $\mathcal{A}_{\theta, \hat{f}}$ to functions defined on the two-dimensional moneyness-volatility space $\mathcal{Y} = \mathbb{R}^2_+$. Specifically, we use the backward dynamic programming iteration $v_{\theta, \hat{f}}(h, \cdot) = \mathcal{A}_{\theta, \hat{f}}[v_{\theta, \hat{f}}(h-1, \cdot)]$, for $h = 1, \ldots, 20$, with $v_{\theta, \hat{f}}(0, y) = (k-1)^+$, and evaluate function $v_{\theta, \hat{f}}(20, \cdot)$ at y_j to get the option-to-stock price ratio of option j. To make the estimation procedure feasible, functions $\varphi = v_{\theta, \hat{f}}(h-1, \cdot)$, for $h = 1, \ldots, 20$, 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 $\mathcal{Y}$. When the computation of $\mathcal{A}_{\theta, \hat{f}}[\varphi](y)$, for a given y in the grid, requires to evaluate function φ at a point $(\tilde{k}, \tilde{\sigma})$ within $[k_{low}, k_{high}] \times [\sigma_{low}, \sigma_{high}]$ but outside the grid, the value at the nearest grid point is selected. When $\tilde{k} < k_{low}$ we set $\varphi(\tilde{k}, \tilde{\sigma}) = 0$ and when $\tilde{\sigma} < \sigma_{low}$ the value at the nearest grid point is selected. When $\tilde{k} > k_{high}$ and/or $\tilde{\sigma} > \sigma_{high}$ we use a linear extrapolation procedure. The use of a finite subset of $\mathcal{Y}$ and a finite grid introduces a numerical error, that becomes negligible as $\sigma_{low}, k_{low} \to 0$ and $\sigma_{high}, k_{high}, N_k, N_\sigma \to \infty$. In the Monte Carlo experiment, we set $N_k = 300$ and $N_\sigma = 30$ for the discretization and $k_{low} = 0.8$ and $k_{high} = 1.2$ for the moneyness strike domain. The bounds σ_{low} and

σ_{high} of the volatility domain are set equal to the 1% and 99% quantiles of the volatility realizations in the Monte Carlo repetition. The grid spacing is homogeneous, with an adjustment at the border such that the volatility σ_0 coincides with one of the N_σ points that discretize $[\sigma_{low}, \ \sigma_{high}]$. For our choice of domain and fineness of the grid, the absolute percentage numerical error in price ratios[15] at time-to-maturity $h = 20$ and volatility state σ_0 is monotonically decreasing in the moneyness strike and less than 3% for $k \geq 0.9$. 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 (6.4) is feasible in less than a minute in most of the repetitions.

We compute the estimator $\hat{f}^*$ by the iterative algorithm described in Section 4.3 with $\omega_T = 0$. This choice simplifies the estimation procedure, since we can dispense with computing function γ_L. The computation of the function $\gamma_S(\cdot; \hat{\theta}^*, \hat{f}^{*(i-1)})$ in the tilting factor $\hat{\mathcal{T}}^{*(i)}$ at the i-th iteration uses option price ratios evaluated with operator $\mathcal{A}_{\hat{\theta}^*, \hat{f}^{*(i-1)}}$ implemented as above. The computation of the integral of a function w.r.t. the conditional density $\hat{f}^{*(i-1)}$ is implemented via a kernel regression of the function multiplied by the tilting factor $\hat{\mathcal{T}}^{*(i-1)}$. Then, at the i-th iteration we use $\hat{f}^{*(i)}$ for the estimation of the four option-to-stock price ratios of interest. This requires evaluation of the tilting factor $\hat{\mathcal{T}}^{*(i)}$ only for conditional values $\tilde{\sigma}$ corresponding to volatility values of grid points. We take as convergence criterion the stability of price 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.

6.3 The results

We show in Figure 2 the kernel smoothed density functions of the XMM and cross-sectional estimators of the SDF parameters. The XMM estimators of parameters θ_1, θ_2 and θ_3 feature small bias and their distributions are slightly skewed. The skew is more pronounced for parameter θ_3. The estimator of parameter θ_4 is downward biased. The estimated values of the parameters have the same sign as the true parameter values in most of the Monte Carlo repetitions. The cross-sectional estimates feature larger standard deviations than the XMM estimates. Hence, accounting for the uniform restrictions (3.1) 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 θ_4. The two estimators have similar biases, but the distribution of the cross-sectional estimator features larger variance and is more skewed and leptokurtic.[16] These findings are compatible with the different rate of convergence of the XMM and cross-sectional estimators of θ_4, that is parametric for the former

---

[15]The error is computed w.r.t. the results obtained with a grid 10 times finer in the moneyness and volatility dimensions.

[16]The bias of the XMM estimator of θ_4 is $-3.45 \cdot 10^{-1}$, that of the cross-sectional estimator is $-5.69 \cdot 10^{-1}$.

and nonparametric for the latter (see Propositions 5 and 7). Figure 3 displays the kernel smoothed density functions of the time-series GMM estimators of the SDF parameters. The distributions of the GMM estimators for parameters θ_1, θ_2 and θ_3 are more disperse than the distributions of the XMM estimators by orders of magnitude (see Figure 2). This is the finite-sample manifestation of the lack of full-information identifiability for these SDF parameters discussed in Section 6.1. The distribution of the GMM and the XMM estimates of the full-information identifiable parameter θ_4 are similar. Overall, Figures 2 and 3 show that the XMM estimator of the SDF parameter is preferable compared to both the cross-sectional and the GMM estimators in this Monte Carlo experiment.

We show in Figure 4 the kernel smoothed density functions of the estimates of the American option-to-stock price ratio for the four moneyness strikes $k^\star$ of interest (solid line). 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 $k^\star = 1.03$, the distribution of the estimated option-to-stock price ratio has 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 since they are relatively far from the critical moneyness strike. For comparison purposes, in Figure 4 we display by a dashed curve the smoothed density functions of the estimates of the American option-to-stock price ratios obtained using the kernel density $\hat{f}$ as 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 estimators based on $\hat{f}^*$ and $\hat{f}$ are similar. However, for each considered moneyness strike, the option-to-stock price ratio estimator based on $\hat{f}^*$ features a smaller variance than the one 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 that are not currently observed on the market.

In practice, the number N of options to use in the estimation procedure depends on selection criteria reflecting the assumed absence of arbitrage opportunities, as thresholds in daily trading activity and ranges in moneyness strike and time-to-maturity.[17] The larger is N, the larger is the number of restrictions and therefore the more efficient the estimators are expected to be. However, the computational burden is also increasing in the number N and in the maximal time-to-maturity $\bar{h} := \max_{j=1,\ldots,N} h_j$ of the options. In our Monte Carlo experiment, the elapsed CPU time for the computation of estimate $\hat{\theta}^*$ is rather stable w.r.t. N and linearly increasing w.r.t. $\bar{h}$. For instance, for $\bar{h} = 100$ days-to-maturity and

---

[17]For example, Ronchetti [2011] reports that at any day in July and August 2008 only between 4 and 23 American put and call options on the IBM stock traded on U.S. centralized markets have daily trading activity larger than 500 contracts, time-to-maturity shorter than 300 business days, moneyness strike inside the range $[0.75 : 1.25]$ and bid-ask spread smaller than 100%.

$N = 8$ options (with moneyness strikes as in Section 6.1), the computation of $\hat{\theta}^*$ requires on average about 2 minutes, and less than 5 minutes in most repetitions. The elapsed CPU time for computation of estimate $\hat{f}^*$ is linearly increasing w.r.t. the number N of options. For instance, with $N = 40$ options (with moneyness strike ranging from 0.832 to 1.031) and $\bar{h} = 20$ days, the computation of $\hat{f}^*$ requires on average about 5 minutes, and up to 10 minutes in some repetitions.

7   Concluding remarks

In this paper we introduce a novel semi-parametric estimator of American option prices. The framework involves a parametric specification of the SDF and is nonparametric w.r.t. the transition density of the Markov state process. We introduce estimators of the SDF parameter θ_0 and transition density f_0 by combining time-series and cross-sectional information from the relevant state variables and observed American option prices, respectively. These estimators are used to get an estimator of the price of an American option for a maturity and strike of interest by a dynamic programming approach. When the number T of time-series observations diverges to infinity and the number N of cross-sectionally observed option prices is fixed, the estimators are consistent and asymptotically normal. In a Monte Carlo experiment we show that estimators combining time-series and cross-sectional information feature a superior performance compared to pure cross-sectional, or time-series, estimators.

If the parametric SDF model is misspecified, the estimators of the SDF parameter θ_0 in Definitions 2 and 3, the estimator $\hat{f}^*$ of the transition density f_0 in Definition 4 and the estimators of functionals of (θ_0, f_0) in Definition 5, are typically inconsistent. Indeed, in such a case, the imposed no-arbitrage restrictions might be invalid. There exists an extensive literature on testing the correct specification of unconditional moment restrictions derived from an asset pricing model, possibly after introducing a given set of instruments. The tests are based on the Hansen statistic (Hansen [1982]), the Hansen-Jagannathan (HJ) statistic (Hansen, Heaton and Luttmer [1995] and Hansen and Jagannathan [1997]), or HJ statistics relying on information-theoretic discrepancy measures (see, e.g., Kitamura and Stutzer [2002] and Almeida and Garcia [2012]). Specification tests for conditional moment restrictions in a parametric framework are also available. For instance, under the maintained assumption that vector θ_0 is full-information identifiable, the results in Tripathi and Kitamura [2003] imply that a test of correct specification of the uniform capital market restrictions can be based on a suitably rescaled and recentered version of the minimized time-series component of criterion $\mathcal{Q}_T^*$ in Definition 3. In the same setting, Nagel and Singleton [2010] test conditional asset pricing models via the Hansen statistic with optimal instruments. When some components of vector θ_0 are full-information unidentifiable, and specification testing concerns also the local derivative market restrictions, the testing problem is

more difficult. If the derivative market restrictions can be written as local conditional moment restrictions (for example for a cross-section of prices of European options), the results in Antoine and Renault [2010] suggest the validity of standard overidentification tests. However, when the available data include the prices of American options, the derivative market restrictions are not parametric moment restrictions. We conjecture that a test statistic can be obtained by combining the values of the cross-sectional and time-series components of criterion $\mathcal{Q}_T^*$ evaluated at the XMM estimator $\hat{\theta}^*$. The derivation of such a test and of its asymptotic properties is left for future research.

References

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

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

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

Almeida, C. and R. Garcia, Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators, forthcoming in Journal of Econometrics.

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

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

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

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

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

Birke, M. and K. Pilz, 2009, Nonparametric Option Pricing with No-Arbitrage Constraints, Journal of Financial Econometrics 7 (2), 53-76.

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

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

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

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

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

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

Broadie, M., J. Detemple, E. Ghysels and O. Torrés, 2000b, Nonparametric Estimation of American Options' Exercise Boundaries and Call Prices, Journal of Economic Dynamics and Control 24 (11-12), 1829-1857.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hansen, L. P., J. Heaton and E. G. J. Luttmer, 1995, Econometric Evaluation of Asset Pricing Models, The Review of Financial Studies 8 (2), 237-274.

Hansen, L. P. and R. Jagannathan, 1997, Assessing Specification Errors in Stochastic Discount Factor Models, The Journal of Finance 52 (2), 557-590.

Hansen, L. P. and S. F. Richard, 1987, The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models, Econometrica 55 (3), 587-613.

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

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

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

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

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

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

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

Kitamura, Y. and M. J. Stutzer, 2002, Connections Between Entropic and Linear Projections in Asset Pricing Estimation, Journal of Econometrics 107 (1-2), 159-174.

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

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

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

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

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

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

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

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

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

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

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

Ronchetti, D., 2011, An Empirical Study of Stock and American Option Prices, Working paper, University of Lugano.

Shiryaev, A. N., Yu. M. Kabanov, D. O. Kramkov, A. V. Melnikov, 1994, Toward the Theory of Pricing of Options of Both European and American Types. II. Continuous Time, Theory of Probability and its Applications 39 (1), 61-102.

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

Stock, J. and J. Wright, 2000, GMM with Weak Identification, Econometrica 68 (5), 1055-1096.

Tripathi, G. and Y. Kitamura, 2003, Testing Conditional Moment Restrictions, Annals of Statistics 31 (6), 2059-2095.

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

Yatchew, A. and W. Härdle, 2006, Nonparametric State Price Density Estimation Using Constrained Least Squares and the Bootstrap, Journal of Econometrics 133 (2), 579-599.
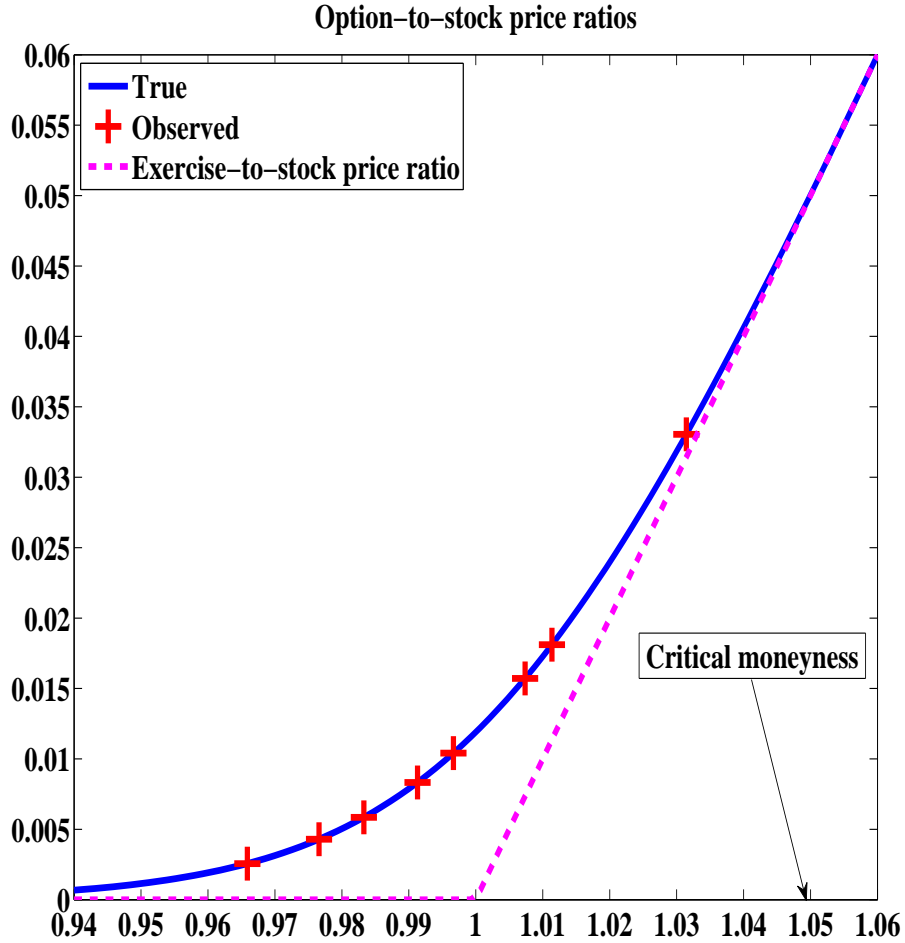
Figure 1: The cross-section of American option-to-stock price ratios at the current date t_0 as a function of the moneyness strike, for time-to-maturity $h = 20$ days. The DGP is the stochastic volatility model with exponential-affine SDF defined in Equations (6.1)-(6.3). The values of the historical and SDF parameters are given in Table 1. The value of the volatility of the stock return at the current date is $\sigma_0 = 6.5 \cdot 10^{-3}$. The solid line is the American option-to-stock price ratio function. The dashed line is the early exercise-to-stock price ratio function. The crosses are the American option-to-stock price ratios that are observed by the econometrician in each Monte Carlo repetition.
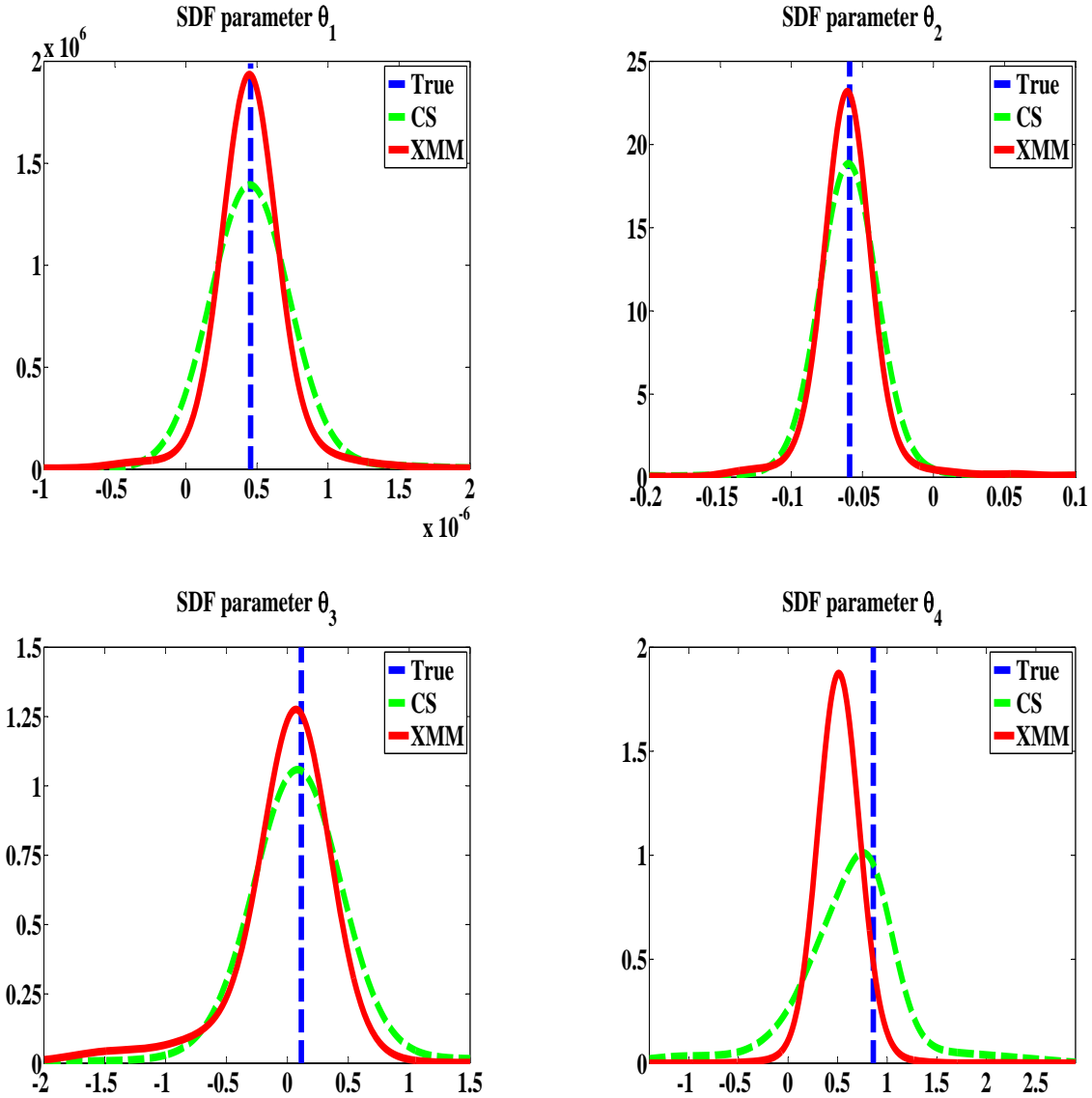
Figure 2: The distributions of the estimated SDF parameters. In each panel, the solid line corresponds to the XMM estimator $\hat{\theta}^*$ in Definition 3 with weighting matrices $\Omega_T = I_{N+2}$ and $\tilde{\Omega}_T = I_2$, and the dashed line to the cross-sectional (CS) estimator $\hat{\theta}$ in Definition 2 with weighting matrix $\Omega_T = I_{N+2}$. The true parameter values are displayed by the dashed vertical lines.
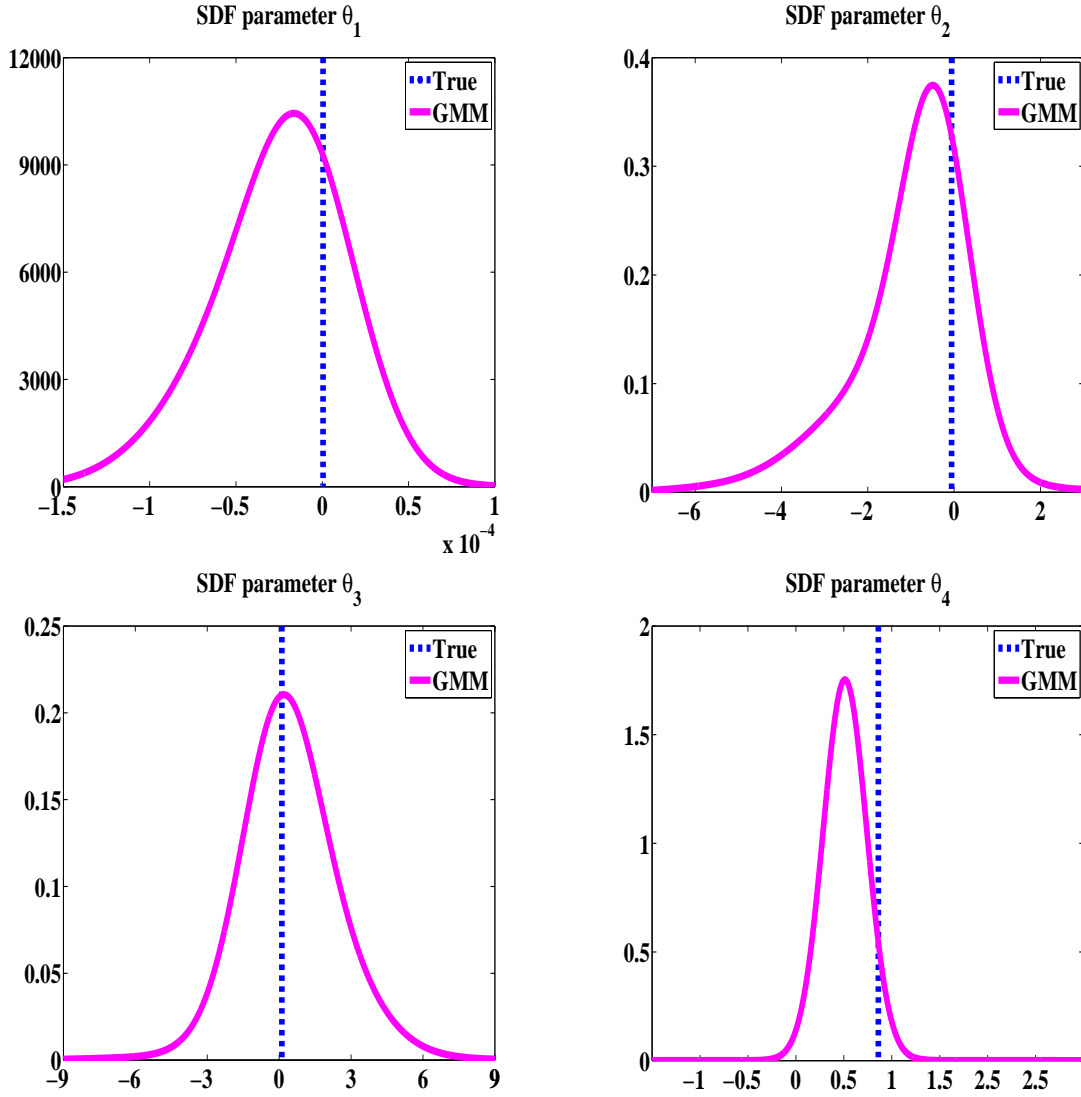
Figure 3: The distributions of the estimated SDF parameters by time-series GMM. The GMM estimator of the SDF parameter vector corresponds to $\hat{\theta}^*$ in Definition 3 with weighting matrices $\Omega_T = 0_{N+2}$ and $\tilde{\Omega}_T = I_2$. The true parameter values are displayed by the dashed vertical lines.
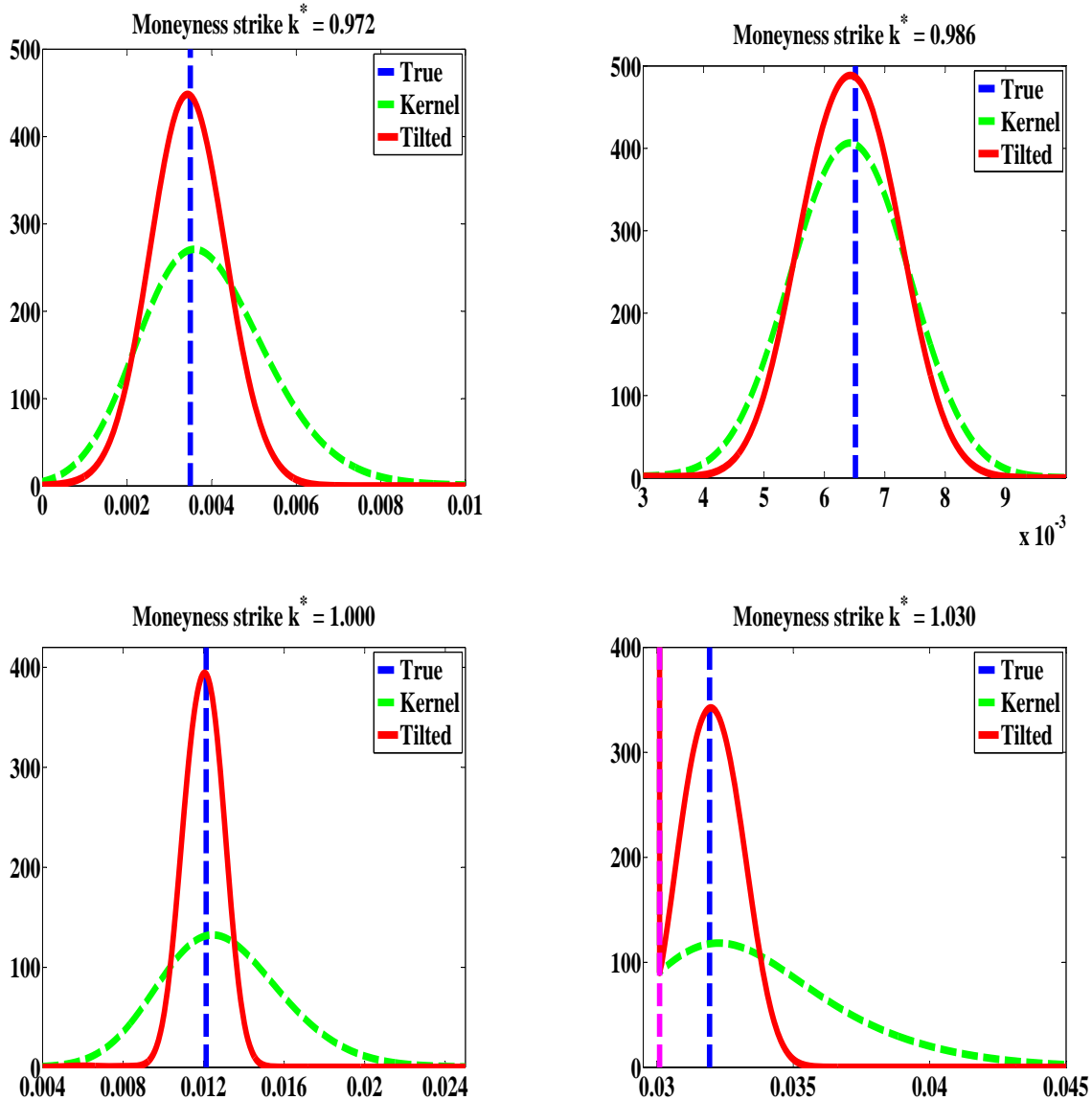
Figure 4: The distribution of the estimated American option-to-stock price ratios at the current date t_0 for time-to-maturity $h^\star = 20$ days and four different moneyness strikes $k^\star$. In each panel, the solid line is the distribution of the estimates when we use $\hat{f}^*$ defined in Equations (4.9) and (4.10) 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 (4.1). For $k^\star = 0.972, 0.986, 1$ the dashed vertical line indicates the true value of the price ratios. For $k^\star = 1.03$ the dashed vertical line on the left indicates the exercise value and the dashed vertical line on the right the true value of the price ratio. The peaks of the distributions at the left vertical line correspond to estimated option-to-stock price ratios equal to the exercise value.

45

APPENDIX

A Regularity assumptions

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

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 process of variable $Z_t := [X_t'\, X_{t-1}']'$ is of differentiability class $C^\rho(\mathbb{R}^{2d})$, for integer $\rho \geq 2$, with uniformly continuous ρ-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 $\int_{\mathcal{X}} \int_{\mathcal{X}} \left[\dfrac{f_Z(\tilde{x}, x)}{f_X(\tilde{x}) f_X(x)} \right]^q f_Z(\tilde{x}, x) d\tilde{x} dx < \infty$ 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 $\sup\limits_{x \in \mathcal{X}_T} \mathbb{P}\left[X_{t+1} \in \mathcal{X}_T^C | X_t = x \right] \to 0$, for $T \to \infty$, $\inf\limits_{x, \tilde{x} \in \mathcal{X}_T} f_Z(x, \tilde{x}) \geq \dfrac{c_1}{\log{(T)}^{c_2}}$ and $\inf\limits_{x \in \mathcal{X}_T} f_X(x) \geq \dfrac{c_1}{\log{(T)}^{c_2}}$.*

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 ρ is defined in Assumption A 2, and $\int_{\mathbb{R}^d} x^j K(x) dx = 0$, for any multi-index $j \in \mathbb{N}^d$ such that $|j| \leq \rho - 1$.*

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

Assumption A 7. *The parameter θ_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) $\mathrm{E}\left[|m(X_{t+1}; \theta_0)|^{2p} \right] < \infty$ for real $p > 1$ such that $1/p + 1/q = 1$, where $q > 1$ is defined in Assumption A 3; (ii) $\sup\limits_{\substack{\theta \in \Theta \\ x \in \mathcal{X}}} \mathrm{E}\left[|m(X_{t+1}; \theta)|^{2+\delta} \Big| X_t = x \right] < \infty$, for real $\delta > 0$.*

Assumption A 10. *The matrix Ω_T converges in probability to the positive-definite matrix Ω_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 ω_T converges in probability to the non-negative scalar ω.*

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 C. Assumption A 4 constrains 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 F.1) and controls for boundary effects.

Assumptions A 5-6 concern the kernel and the bandwidth. Function K is a kernel of order ρ, 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 F.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 F.3). When $h_T = cT^{-\bar{\eta}}$ for real constants $c, \bar{\eta} > 0$, Assumption A 6 is satisfied if $\dfrac{1}{2\rho} < \bar{\eta} < \dfrac{1}{3d}$.

Assumption A 7 is standard in parametric estimation. Assumptions A 8-9 concern the SDF. They involve a differentiability condition w.r.t. parameter θ, 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 θ, and the scalar weight in the criterion to estimate the density f in Definition 4. These assumptions ensure well-defined large sample limits for these criteria and are used to prove uniqueness of the minima of the large sample criteria.

B Proof of Proposition 1

At maturity, i.e. for $h = 0$, the American put option price is $V(0, K, S, x) = (K-S)^+ = S(k-1)^+ = Sv(0, y)$. 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., $V(h - 1, K, S, x) = Sv(h - 1, y)$, with $y = [k \ x']'$. From Equations (2.1) and (2.3), the definition of moneyness strike and the Markov property of Y_t under $\mathscr{Q}$ we get

$$
\begin{aligned}
V(h, K, S, x) &= \max\left[(K - S)^+, \mathrm{E}^{\mathscr{Q}}\left[V(h - 1, K, S_{t+1}, X_{t+1})|S_t = S, X_t = x\right]\right] \\
&= \max\left[(K - S)^+, \mathrm{E}^{\mathscr{Q}}\left[S_{t+1}v(h - 1, Y_{t+1})|\, S_t = S, X_t = x\right]\right] \\
&= S\max\left[(k - 1)^+, \mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}}v(h - 1, Y_{t+1})|\, Y_t = y\right]\right] = Sv(h, y).
\end{aligned}
$$

C Domain and range of the American put pricing operator

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

$$
\mathcal{E}[\varphi](y) := \mathrm{E}^{\mathscr{Q}}\left[e^{r_{t+1}}\varphi(Y_{t+1})|\, Y_t = y\right] = \int_{\mathcal{X}} m\,(\tilde{x}; \theta_0)\, e^{\tilde{r}}\varphi(ke^{-\tilde{r}}, \tilde{x})f(\tilde{x}|x)d\tilde{x}. \tag{C.1}
$$

By the Cauchy-Schwarz inequality we get

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

for any $y = [k \; x']' \in \mathcal{Y}$. Then we have

$$
\begin{aligned}
\int_{\mathcal{Y}} |\mathcal{E}[\varphi](y)|^2 \frac{f_X(x)}{k^2} dy &\leq \left(\int_{\mathbb{R}_+} \int_{\mathcal{X}} e^{\tilde{r}} \varphi(ke^{-\tilde{r}}, \tilde{x})^2 f_X(\tilde{x}) \frac{1}{k^2} dk d\tilde{x} \right) \left(\int_{\mathcal{X}} \int_{\mathcal{X}} m(\tilde{x}; \theta_0)^2 e^{\tilde{r}} \frac{f(\tilde{x}|x)^2}{f_X(\tilde{x})} f_X(x) d\tilde{x} dx \right) \\
&= \left(\int_{\mathcal{Y}} \varphi(\tilde{y})^2 \frac{f_X(\tilde{x})}{\tilde{k}^2} d\tilde{y} \right) \left(\int_{\mathcal{X}} \int_{\mathcal{X}} m(\tilde{x}; \theta_0)^2 e^{\tilde{r}} \frac{f_Z(\tilde{x}, x)^2}{f_X(\tilde{x}) f_X(x)} d\tilde{x} dx \right) < \infty, \quad \text{(C.2)}
\end{aligned}
$$

where we use the change of variable from k to $\tilde{k} = ke^{-\tilde{r}}$ and that the double integral over $\mathcal{X} \times \mathcal{X}$ 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 $v(0, \cdot) \in L_2(\mathcal{Y})$, it follows that $\mathcal{A}[\varphi] = \max[v(0, \cdot), \mathcal{E}[\varphi]] \in L_2(\mathcal{Y})$. Thus, operator $\mathcal{A}$ maps $L_2(\mathcal{Y})$ into $L_2(\mathcal{Y})$.

D   Proof of Proposition 2

In this appendix we use the simplified notation $\mathcal{A} = \mathcal{A}_{\theta,f}$, $m(\cdot) = m(\cdot; \theta)$, $g = g(\theta, f)$, $\mathrm{E}^{\mathscr{Q}} = \mathrm{E}^{\mathscr{Q}}_{\theta,f}$, $\mathcal{E} = \mathcal{E}_{\theta,f}$ 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.

D.1   Differentiability of g almost everywhere

Let us first consider the differentiability of g w.r.t. θ. The American put option-to-stock price ratio $v(h, y)$ and the holding-to-stock price ratio $u(h, y) := \mathrm{E}^{\mathscr{Q}}[e^{r_{t+1}} v(h-1, Y_{t+1})| Y_t = y]$ depend on the SDF parameter θ for any $h > 0$. For expository purpose, we omit this dependence in the notation. By Definition 1 and the linearity of operator $\mathcal{E}$ defined in Equation (C.1), we can write the holding-to-stock price ratio as

$$
\begin{aligned}
u(h, y) &= \mathcal{E}[v(h-1, \cdot)](y) = \mathcal{E}[\max[v(0, \cdot), u(h-1, \cdot)]](y) \\
&= \mathcal{E}[\max[0, u(h-1, \cdot) - v(0, \cdot)]](y) + \mathcal{E}[v(0, \cdot)](y). \quad \text{(D.1)}
\end{aligned}
$$

We know that $u(h-1, y) - v(0, y) \geq 0$ iff $k \leq k^\star(h-1, x)$, where the critical moneyness strike $k^\star(h-1, x)$ is the solution of the equation $k - 1 = u(h-1, k, x)$ in $k \in \mathbb{R}_+$. From the implicit function theorem, $k^\star(h-1, x)$ is differentiable w.r.t. θ. From Equations (C.1) and (D.1), we get

$$
\begin{aligned}
u(h, y) &= \int_{\mathcal{X}} m(\tilde{x}) e^{\tilde{r}} \mathbf{1}\{ke^{-\tilde{r}} \leq k^\star(h-1, \tilde{x})\}[u(h-1, ke^{-\tilde{r}}, \tilde{x}) - v(0, ke^{-\tilde{r}}, \tilde{x})] f(\tilde{x}|x) d\tilde{x} \\
&\quad + \int_{\mathcal{X}} m(\tilde{x}) e^{\tilde{r}} v(0, ke^{-\tilde{r}}, \tilde{x}) f(\tilde{x}|x) d\tilde{x},
\end{aligned}
$$

for $y = [k\ x']'$ and the indicator function $\mathbf{1}\{\cdot\}$. For expository purpose, let us assume that $ke^{-\tilde{r}} \leq k^{\star}(h-1,\tilde{x})$ iff $\tilde{r} \geq r^{\star}(h-1,k,\tilde{\sigma})$, where $r^{\star}(h-1,k,\tilde{\sigma})$ is the solution of the equation

$$ke^{-\tilde{r}} = k^{\star}(h-1,\tilde{r},\tilde{\sigma}) \tag{D.2}$$

in $\tilde{r} \in \mathcal{R}$, for given $[k\ \tilde{\sigma}]' \in \mathbb{R}_+ \times \mathcal{S}$.[18] Then we have:

$$
\begin{aligned}
u(h,y) &= \int_{\mathcal{S}} \int_{r^{\star}(h-1,k,\tilde{\sigma})}^{b} m(\tilde{x})e^{\tilde{r}}[u(h-1,ke^{-\tilde{r}},\tilde{x}) - v(0,ke^{-\tilde{r}},\tilde{x})]f(\tilde{x}|x)d\tilde{r}d\tilde{\sigma} \\
&+ \int_{\mathcal{X}} m(\tilde{x})e^{\tilde{r}}v(0,ke^{-\tilde{r}},\tilde{x})f(\tilde{x}|x)d\tilde{x},
\end{aligned}
\tag{D.3}
$$

where b is the upper boundary of $\mathcal{R}$. Let us now show that u is differentiable w.r.t. θ by induction. This is true for $h = 0$. Now, let us assume that $u(h-1,\cdot)$ is differentiable w.r.t. θ. From Equation (D.2) and the implicit function theorem, it follows that $r^{\star}(h-1,k,\tilde{\sigma})$ is differentiable w.r.t. θ. Then, by the Leibniz integral rule for differentiation of a definite integral applied to Equation (D.3) and Assumption A 8, $u(h,\cdot)$ is differentiable w.r.t. θ. By using $\mathcal{A}^h[v(0,\cdot)](y) = \max[v(0,y),u(h,y)]$, we get that $\mathcal{A}^h[v(0,\cdot)](y)$ is continuous for all θ and differentiable for all θ apart from the values such that $v(0,y) = u(h,y)$. By replacing the differentiability w.r.t. θ with the Fréchet differentiability w.r.t. f, and by following a similar argument, we can show that $\mathcal{A}^h[v(0,\cdot)](y)$ is Fréchet-differentiable w.r.t. f, for all f, apart from the values such that $v(0,y) = u(h,y)$.

D.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 $(\theta,f) \to \mathcal{E}[\varphi]$. The differential of $\mathcal{E}[\varphi]$ w.r.t. (θ,f) is given by

$$
\delta\mathcal{E}[\varphi](y) = \int_{\mathcal{X}} m(\tilde{x})e^{\tilde{r}}\varphi(ke^{-\tilde{r}},\tilde{x})\delta f(\tilde{x}|x)d\tilde{x} + \int_{\mathcal{X}} \nabla_{\theta'}m(\tilde{x})e^{\tilde{r}}\varphi(ke^{-\tilde{r}},\tilde{x})f(\tilde{x}|x)d\tilde{x}\delta\theta, \tag{D.4}
$$

where δf and $\delta\theta$ denote infinitesimal variations of parameters f and θ, respectively. Let us now consider the mapping $(\theta,f) \to \mathcal{A}^h[v(0,\cdot)]$, for a given integer $h \geq 1$, and compute its differential w.r.t. (θ,f) in terms of the differential of $\mathcal{E}$ given in Equation (D.4). We write $\mathcal{A}^h[v(0,\cdot)](y) = \left(\mathcal{E} \circ \mathcal{A}^{h-1}[v(0,\cdot)](y) - v(0,y)\right)^+ + v(0,y)$, where $\circ$ denotes operator composition. The right derivative of function $(\cdot)^+$ is the indicator $\mathbf{1}\{\cdot \geq 0\}$. Then, by the chain and product rules for differentiation and the total differential, we get

$$
\delta\mathcal{A}^h[v(0,\cdot)](y) = \mathbf{1}_{\mathcal{C}(h)}(y)\left(\delta\mathcal{E}[v(h-1,\cdot)](y) + \mathcal{E} \circ \delta\mathcal{A}^{h-1}[v(0,\cdot)](y)\right), \tag{D.5}
$$

---

[18]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 $\{\tilde{r} \in \mathcal{R} : ke^{-\tilde{r}} \leq k^{\star}(h-1,\tilde{x})\}$ can be written as the union of a finite number of intervals, but the notation is more cumbersome.

where we make use of the definition of the continuation region in Equation (2.4) and the expression of the American put option-to-stock price ratio in Equation (2.6). We can iterate Equation (D.5) to get

$$
\begin{aligned}
\delta \mathcal{A}^h[v(0,\cdot)](y) \;=\; & \mathbf{1}_{\mathcal{C}(h)}\delta\mathcal{E}\left[v(h-1,\cdot)\right](y) + \mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\delta\mathcal{E}\left[v(h-2,\cdot)\right](y) \\
& + \mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-2)}\delta\mathcal{E}\left[v(h-3,\cdot)\right](y) + \ldots \\
& + \mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\ldots\circ\mathbf{1}_{\mathcal{C}(2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(1)}\delta\mathcal{E}\left[v(0,\cdot)\right](y), \quad\quad \text{(D.6)}
\end{aligned}
$$

where operator $\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}$ is such that $\left(\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\right)[\varphi](y) = \mathbf{1}_{\mathcal{C}(h)}(y)\mathcal{E}[\varphi](y)$. By using $v(h-l,\cdot) = \mathcal{A}^{h-l}[v(0,\cdot)]$, for $1 \le l \le h$, we rewrite Equation (D.6) as

$$
\delta\mathcal{A}^h[v(0,\cdot)](y) = \sum_{l=1}^{h}\mathbf{1}_{\mathcal{C}(h)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-1)}\mathcal{E}\circ\ldots\circ\mathbf{1}_{\mathcal{C}(h-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h-l}[v(0,\cdot)](y).
$$

Thus, the total differential of vector g w.r.t. f and θ is given by

$$
\delta g_j = \sum_{l=1}^{h_j}\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-1)}\mathcal{E}\circ\ldots\circ\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}[v(0,\cdot)](y_j), \text{ if } j=1,\ldots,N. \quad \text{(D.7)}
$$

D.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 (D.7) from Equation (D.4) with $\delta f(\tilde{x}|x) = \Delta f(\tilde{x}|x)$ and $\delta\theta = 0$. Let us focus on the quantity $\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E}\circ\ldots\circ\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}[v(0,\cdot)](y_j)$, for some integers l and h_j such that $1 \le l \le h_j$, and let us write it explicitly. For $l=1$ this quantity is equal to

$$
\begin{aligned}
& \mathbf{1}_{\mathcal{C}(h_j)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-1}[v(0,\cdot)](y_j) \\
& = \mathbf{1}_{\mathcal{C}(h_j)}(y_j)\int_{\mathcal{X}}\mathcal{A}^{h_j-1}[v(0,\cdot)](k_j e^{-r_{t+1}}, x_{t+1})m(x_{t+1})e^{r_{t+1}}\Delta f(x_{t+1}|x_0)dx_{t+1}. \quad\quad \text{(D.8)}
\end{aligned}
$$

Let us now consider the case $l \ge 2$. First, operator $\mathcal{A}$ is applied $h_j - l$ times to discount the payoff-to-stock price ratio $v(0,\cdot)$ from $t+h_j$ back to $t+l$. Second, $\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}$ is applied to discount from $t+l$ back to $t+l-1$:

$$
\begin{aligned}
& \mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}[v(0,\cdot)](y_{t+l-1}) \\
& = \mathbf{1}_{\mathcal{C}(h_j-l+1)}(y_{t+l-1})\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\mathcal{A}^{h_j-l}[v(0,\cdot)](k_{t+l-1}e^{-r_{t+l}}, x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}.
\end{aligned}
$$

Third, $\mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}$ is applied to discount from date $t+l-1$ back to date $t+l-2$:

$$
\begin{aligned}
& \mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E}\circ\mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E}\circ\mathcal{A}^{h_j-l}[v(0,\cdot)](y_{t+l-2}) \\
& = \mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})\int_{\mathcal{X}}m(x_{t+l-1})e^{r_{t+l-1}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-r_{t+l-1}}, x_{t+l-1})\left(\int_{\mathcal{X}}m(x_{t+l})e^{r_{t+l}}\right. \\
& \left. \quad \cdot\mathcal{A}^{h_j-l}[v(0,\cdot)](k_{t+l-2}e^{-r_{t+l-1}-r_{t+l}}, x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}\right)f(x_{t+l-1}|x_{t+l-2})dx_{t+l-1}.
\end{aligned}
$$

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 $t+l-2$ back to t:

$$\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E} \circ \ldots \circ \mathbf{1}_{\mathcal{C}(h_j-l+2)}\mathcal{E} \circ \mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E} \circ \mathcal{A}^{h_j-l}[v(0,\cdot)](y_j)$$

$$= \int_{\mathcal{Y}} \cdots \int_{\mathcal{Y}} \mathbf{1}_{\mathcal{C}(h_j)}(y_j) \ldots \mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})e^{R_{t,l-2}} \int_{\mathcal{X}} m(x_{t+l-1})e^{r_{t+l-1}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-r_{t+l-1}}, x_{t+l-1})$$

$$\cdot \left(\int_{\mathcal{X}} m(x_{t+l})e^{r_{t+l}}\mathcal{A}^{h_j-l}[v(0,\cdot)](k_{t+l-2}e^{-r_{t+l-1}-r_{t+l}}, x_{t+l})\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l} \right)$$

$$\cdot f(x_{t+l-1}|x_{t+l-2})dx_{t+l-1}dF^{\mathcal{Q}}(y_{t+l-2}|y_{t+l-3}) \ldots dF^{\mathcal{Q}}(y_{t+1}|y_j).$$

By rearranging the terms, the RHS of the previous equation is equal to

$$\mathbf{1}_{\mathcal{C}(h_j)}(y_j) \int_{\mathcal{X}} \int_{\mathcal{X}} m(x_{t+l})\zeta(h_j, l, x_{t+l}, x_{t+l-1}; y_j)\Delta f(x_{t+l}|x_{t+l-1})dx_{t+l}dx_{t+l-1},$$

where function ζ is defined as

$$\zeta(h_j, l, x, \tilde{x}; y_j) := m(\tilde{x})e^{r+\tilde{r}} \int_{\mathcal{Y}} \cdots \int_{\mathcal{Y}} \mathbf{1}_{\mathcal{C}(h_j-1)}(y_{t+1}) \ldots \mathbf{1}_{\mathcal{C}(h_j-l+2)}(y_{t+l-2})\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-\tilde{r}}, \tilde{x})$$

$$\cdot e^{R_{t,l-2}}\mathcal{A}^{h_j-l}[v(0,\cdot)](k_{t+l-2}e^{-\tilde{r}-r}, x)f(\tilde{x}|x_{t+l-2})dF^{\mathcal{Q}}(y_{t+l-2}|y_{t+l-3}) \ldots dF^{\mathcal{Q}}(y_{t+1}|y_j). \qquad \text{(D.9)}$$

Thus, we get

$$\mathbf{1}_{\mathcal{C}(h_j)}\mathcal{E} \circ \ldots \circ \mathbf{1}_{\mathcal{C}(h_j-l+1)}\delta\mathcal{E} \circ \mathcal{A}^{h_j-l}[v(0,\cdot)](y_j) = \mathbf{1}_{\mathcal{C}(h_j)}(y_j) \int_{\mathcal{X}} \int_{\mathcal{X}} m(x)\zeta(h_j, l, x, \tilde{x}; y_j)\Delta f(x|\tilde{x})dxd\tilde{x}, \qquad \text{(D.10)}$$

for $l \geq 2$. From Equations (D.8) and (D.10) we deduce the Fréchet derivative of g_j w.r.t. f:

$$\langle Dg_j, \Delta f \rangle = \mathbf{1}_{\mathcal{C}(h_j)}(y_j) \int_{\mathcal{X}} m(x)e^r \mathcal{A}^{h_j-1}[v(0,\cdot)](k_j e^{-r}, x)\Delta f(x|x_0)dx$$

$$+ \mathbf{1}_{\mathcal{C}(h_j)}(y_j) \sum_{l=2}^{h_j} \int_{\mathcal{X}} \int_{\mathcal{X}} m(x)\zeta(h_j, l, x, \tilde{x}; y_j)\Delta f(x|\tilde{x})dxd\tilde{x},$$

for $j = 1, \ldots, N$. To conclude the proof, we rewrite function ζ in terms of a risk-neutral expectation using

$$e^{r+\tilde{r}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(k_{t+l-2}e^{-\tilde{r}}, \tilde{x})\mathcal{A}^{h_j-l}[v(0,\cdot)](k_{t+l-2}e^{-\tilde{r}-r}, x)$$

$$= \mathrm{E}^{\mathcal{Q}}\left[e^{r_{t+l-1}+r_{t+l}}\mathbf{1}_{\mathcal{C}(h_j-l+1)}(Y_{t+l-1})\mathcal{A}^{h_j-l}[v(0,\cdot)](Y_{t+l}) \,\Big|\, X_{t+l} = x, X_{t+l-1} = \tilde{x}, Y_{t+l-2} = y_{t+l-2} \right].$$

Moreover, by the Markov property of Y_t and X_t under $\mathcal{Q}$, and Assumption 2, we have the following equalities:

$$f^{\mathcal{Q}}(y_{t+l-2}, \ldots, y_{t+1}|x_{t+l}, x_{t+l-1}, y_t) = \frac{f^{\mathcal{Q}}(x_{t+l}, x_{t+l-1}, y_{t+l-2}, \ldots, y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l}, x_{t+l-1}|y_t)}$$

$$= \frac{f^{\mathcal{Q}}(x_{t+l-1}|x_{t+l-2})f^{\mathcal{Q}}(y_{t+l-2}, \ldots, y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l-1}|x_t)} = \frac{m(x_{t+l-1})f(x_{t+l-1}|x_{t+l-2})f^{\mathcal{Q}}(y_{t+l-2}, \ldots, y_{t+1}|y_t)}{f^{\mathcal{Q}}(x_{t+l-1}|x_t)},$$

where we use the same symbol for different probability densities and omit the subscript to simplify the notation. Hence:

$$m(\tilde{x})f(\tilde{x}|x_{t+l-2})dF_Y^{\mathcal{Q}}(y_{t+l-2},\ldots,y_{t+1}|y_j) = f_{l-1}^{\mathcal{Q}}(\tilde{x}|x_0)dF_Y^{\mathcal{Q}}(y_{t+l-2},\ldots,y_{t+1}|X_{t+l}=x, X_{t+l-1}=\tilde{x}, Y_t=y_j).$$

Thus, from Equation (D.9) and the Law of Iterated Expectations, we get

$$
\begin{aligned}
\zeta(h_j, l, x, \tilde{x}; y_j) &= f_{l-1}^{\mathcal{Q}}(\tilde{x}|x_0)\mathrm{E}^{\mathcal{Q}}\Big[\mathbf{1}_{\mathcal{C}(h_j-1)}(Y_{t+1})\ldots\mathbf{1}_{\mathcal{C}(h_j-l+1)}(Y_{t+l-1})e^{R_{t,l}} \\
&\quad \cdot \mathcal{A}^{h_j-l}[v(0,\cdot)](Y_{t+l})\Big|X_{t+l}=x, X_{t+l-1}=\tilde{x}, Y_t=y_j\Big].
\end{aligned}
$$

Equation (3.7) follows.

D.4 Gradient of g w.r.t. the SDF parameter

The gradient of the vector g w.r.t. θ is obtained by replacing $\delta\mathcal{E}$ in Equation (D.7) with the expression in Equation (D.4) for $\delta f(\tilde{x}|x) = 0$ and $\delta\theta = d\theta$. By similar arguments as in Appendix D.3 we get Equation (3.8).

E Proof of Proposition 3

The differential w.r.t. the historical transition density f of the functional Lagrangian in Equation (4.7) is

$$
\begin{aligned}
\delta\mathcal{L} &= \delta D_T(f,\hat{f}) - \omega_T\lambda'\delta g(\hat{\theta}^*, f) - \omega_T\nu_0'\int_{\mathcal{X}}\Gamma_U(x;\hat{\theta}^*)\delta f(x|x_0)dx - \omega_T\mu_0\int_{\mathcal{X}}\delta f(x|x_0)dx \\
&\quad - \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\nu(\tilde{x})'\int_{\mathcal{X}}\Gamma_U(x;\hat{\theta}^*)\delta f(x|\tilde{x})dxd\tilde{x} - \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\mu(\tilde{x})\int_{\mathcal{X}}\delta f(x|\tilde{x})dxd\tilde{x}.
\end{aligned}
\tag{E.1}
$$

Let us compute explicitly the first two differential terms in the RHS of Equation (E.1). The differential of the criterion D_T is

$$
\begin{aligned}
\delta D_T(f,\hat{f}) &= \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\int_{\mathcal{X}}\left(1+\log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right)\right)\delta f(x|\tilde{x})dxd\tilde{x} + \omega_T\int_{\mathcal{X}}\left(1+\log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)}\right)\right)\delta f(x|x_0)dx \\
&= \int_{\mathcal{X}}\hat{f}_X(\tilde{x})\int_{\mathcal{X}}\log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})}\right)\delta f(x|\tilde{x})dxd\tilde{x} + \omega_T\int_{\mathcal{X}}\log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)}\right)\delta f(x|x_0)dx,
\end{aligned}
$$

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 θ with $\hat{\theta}^*$ and Δf with δf into Equation (3.7), and using the definition of vectors γ_S and γ_L:

$$\delta g(\hat{\theta}^*, f) = \int_{\mathcal{X}}\gamma_S(x;\hat{\theta}^*, f)\delta f(x|x_0)dx + \int_{\mathcal{X}}\int_{\mathcal{X}}\gamma_L(x,\tilde{x};\hat{\theta}^*, f)\delta f(x|\tilde{x})dxd\tilde{x}.$$

Then, the differential of the functional Lagrangian $\mathcal{L}$ is

$$\delta\mathcal{L} = \int_{\mathcal{X}} \omega_T \left(\log\left(\frac{f(x|x_0)}{\hat{f}(x|x_0)} \right) - \lambda'\gamma_S(x;\hat{\theta}^*,f) - \nu_0'\Gamma_U(x;\hat{\theta}^*) - \mu_0 \right) \delta f(x|x_0)dx$$

$$+ \int_{\mathcal{X}}\int_{\mathcal{X}} \left(\log\left(\frac{f(x|\tilde{x})}{\hat{f}(x|\tilde{x})} \right) - \frac{\omega_T}{\hat{f}_X(\tilde{x})}\lambda'\gamma_L(x,\tilde{x};\hat{\theta}^*,f) - \nu(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) - \mu(\tilde{x}) \right) \hat{f}_X(\tilde{x})\, \delta f(x|\tilde{x})dxd\tilde{x}.$$

By the optimality condition in Equation (4.8) and the fundamental lemma of the calculus of variations we get

$$\log\left(\frac{\hat{f}^*(x|x_0)}{\hat{f}(x|x_0)} \right) - \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*) - \hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) - \hat{\mu}_0 = 0, \tag{E.2}$$

for a.e. $x \in \mathcal{X}$, and

$$\log\left(\frac{\hat{f}^*(x|\tilde{x})}{\hat{f}(x|\tilde{x})} \right) - \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) - \hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) - \hat{\mu}(\tilde{x}) = 0, \tag{E.3}$$

for a.e. $x, \tilde{x} \in \mathcal{X}$ with $\tilde{x} \neq x_0$. From Equations (E.2) and (E.3) we get

$$\hat{f}^*(x|x_0) = \hat{f}(x|x_0) \exp\left(\hat{\mu}_0 + \hat{\lambda}'\gamma_S(x;\hat{\theta}^*,\hat{f}^*) + \hat{\nu}_0'\Gamma_U(x;\hat{\theta}^*) \right),$$

for a.e $x \in \mathcal{X}$, and

$$\hat{f}^*(x|\tilde{x}) = \hat{f}(x|\tilde{x}) \exp\left(\hat{\mu}(\tilde{x}) + \hat{\nu}(\tilde{x})'\Gamma_U(x;\hat{\theta}^*) + \omega_T\hat{\lambda}'\gamma_L(x,\tilde{x};\hat{\theta}^*,\hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) \right),$$

for a.e. $x, \tilde{x} \in \mathcal{X}$ with $\tilde{x} \neq x_0$. By imposing the unit mass constraints, Equation (4.9) follows. Finally, by imposing that the empirical counterpart of System (3.5) holds for $(\hat{\theta}^*, \hat{f}^*)$, System (4.10) follows.

F 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 θ, f.

F.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 $G(\theta, \hat{f}) = [\mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1};\theta)|X_t = x_0]'\; g(\theta,\hat{f})']'$. 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

$$\mathcal{E}_{\theta,\hat{f}}[\varphi](y) = \int_{\mathcal{X}_T} m(x_{t+1};\theta)e^{r_{t+1}}\varphi(ke^{-r_{t+1}},x_{t+1})\hat{f}(x_{t+1}|x)dx_{t+1}$$

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 θ_0 by the identification condition in Assumption 5 and since Ω_0 is positive-definite (Assumption A 10).

ii) The set Θ 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 D.1, the functions g_j, for $j = 1, \ldots, N$, are continuous w.r.t. θ 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, e.g., Hansen [2008]) and Assumptions A 1, A 2, A 4-6 and A 9, we can show that

$$\sup_{\theta \in \Theta} \left\| \mathrm{E}_{\hat{f}}[\Gamma_U(X_{t+1}; \theta)|X_t = x_0] - \mathrm{E}_0[\Gamma_U(X_{t+1}; \theta)|X_t = x_0] \right\| = o_p(1). \tag{F.1}$$

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 $j = 1, \ldots, N$, 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 scalar 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 scalar function that may depend on parameter $\theta \in \Theta$ and is such that*

$$\sup_{\substack{\theta \in \Theta \\ y \in [e^{-a}, e^a] \times \mathcal{X}}} \mathrm{E}_0 \left[\varphi_\theta(Y_{t+1})^2 \,\middle|\, Y_t = y \right] < \infty. \tag{F.2}$$

Let $\hat{\varphi}_\theta$ be an estimator of φ_θ such that $\sup_{\theta \in \Theta} \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T, \infty} = o_p(1)$. Then, under Assumptions A 1, A 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. See Section H.1 of the supplementary materials. $\square$

We use the uniform convergence of the kernel estimator to prove Lemma 1. Let us now write the American put pricing operator as

$$\mathcal{A}_{\theta, f}[\varphi] = v(0, \cdot) + (\mathcal{E}_{\theta, f}[\varphi] - v(0, \cdot))^+ \tag{F.3}$$

and do similarly for its estimator $\mathcal{A}_{\theta, \hat{f}}[\varphi]$. Since $|\max[t, 0] - \max[s, 0]| \leq |t - s|$, for all $t, s \in \mathbb{R}$, we get from Lemma 1 that for any φ_θ satisfying Inequality (F.2)

$$\sup_{\theta \in \Theta} \|\hat{\varphi}_\theta - \varphi_\theta\|_{\mathcal{Y}'_T, \infty} = o_p(1) \quad \Rightarrow \quad \sup_{\theta \in \Theta} \|\mathcal{A}_{\theta, \hat{f}}[\hat{\varphi}_\theta] - \mathcal{A}_{\theta, f_0}[\varphi_\theta]\|_{\mathcal{Y}_T, \infty} = o_p(1). \tag{F.4}$$

Lemma 2. *Under Assumption A 9, if*

$$\sup_{\substack{\theta \in \Theta \\ y \in [e^{-a-b}, e^{a+b}] \times \mathcal{X}}} \mathrm{E}\left[\mathcal{A}^h_{\theta, f_0}[v(0, \cdot)](Y_{t+1})^2 \,\Big|\, Y_t = y \right] < \infty,$$

for $h \in \mathbb{N}$, then

$$\sup_{\substack{\theta \in \Theta \\ y \in [e^{-a}, e^{a}] \times \mathcal{X}}} \mathrm{E}\left[\mathcal{A}^{h+1}_{\theta, f_0}[v(0, \cdot)](Y_{t+1})^2 \,\Big|\, Y_t = y \right] < \infty. \tag{F.5}$$

Proof. See Section H.2 of the supplementary materials. ☐

By Lemma 2, we can iterate $h \geq 1$ times the Implication (F.4) starting with $\varphi_\theta = \hat{\varphi}_\theta = v(0, \cdot)$ and a sufficiently large moneyness strike support, and get

$$\sup_{\theta \in \Theta} \| \mathcal{A}^h_{\theta, \hat{f}}[v(0, \cdot)] - \mathcal{A}^h_{\theta, f_0}[v(0, \cdot)] \|_{\mathcal{Y}_T, \infty} = o_p(1).$$

We deduce that $\sup_{\theta \in \Theta} \left\| g(\theta, \hat{f}) - g(\theta, f_0) \right\| = o_p(1)$. Then, from Equation (F.1), 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 $\theta \in \Theta$.

F.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 (5.3) 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 θ 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 D.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. θ, 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 θ_0 is an interior point of Θ_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 (3.6) and Proposition 2 we get

$$g(\theta_0, \hat{f}) = \int_{\mathcal{X}} \bar{\gamma}_S(x) \Delta \hat{f}(x|x_0) dx + \int_{\mathcal{X}} \int_{\mathcal{X}} \gamma_L(x, \tilde{x}; \theta_0, f_0) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} + O_p\left(\|\Delta \hat{f}\|^2_\infty \right).$$

Then, by using that the first two components of $G(\theta_0, \hat{f})$ are equal to $\int \bar{\Gamma}_U(x; \theta_0) \Delta \hat{f}(x|x_0) dx$, we get

$$\sqrt{Th_T^d} G(\theta_0, \hat{f}) = \sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}(x|x_0) dx$$
$$+ \sqrt{Th_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} + O_p\left(\sqrt{Th_T^d} \|\Delta \hat{f}\|_\infty^2\right). \tag{F.6}$$

From the uniform convergence of kernel density estimators (see, e.g., Hansen [2008]), the supremum norm of $\Delta \hat{f}$ is such that $\|\Delta \hat{f}\|_\infty = O_p\left(\sqrt{\dfrac{\log(T)}{Th_T^{2d}}} + h_T^\rho\right)$, for integer ρ defined in Assumption A 2. Then, the remainder term in the RHS of Equation (F.6) is such that

$$O_p\left(\sqrt{Th_T^d} \|\Delta \hat{f}\|_\infty^2\right) = O_p\left(\sqrt{Th_T^d}\left(\frac{\log(T)}{Th_T^{2d}} + h_T^{2\rho}\right)\right) = o_p(1), \tag{F.7}$$

under the bandwidth conditions in Assumption A 6. Equations (F.6) and (F.7) yield Equation (5.4). Moreover, from the asymptotic normality of kernel density estimators (see, e.g., Aït-Sahalia [1992]), the asymptotic distribution of the first term in the RHS of Equation (F.6) is

$$\sqrt{Th_T^d} \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}(x|x_0) dx \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)} \Sigma_S(x_0)\right), \tag{F.8}$$

where the bias term vanishes asymptotically under Assumption A 6 on the bandwidth. Let us now consider the second term of the RHS of Equation (F.6). The integration w.r.t. $\tilde{x} \in \mathcal{X}$ increases the convergence rate, i.e.

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} = O_p\left(\frac{\log(T)}{\sqrt{T}} + h_T^\rho\right) = o_p\left(1/\sqrt{Th_T^d}\right), \tag{F.9}$$

from the bandwidth conditions in Assumption A 6. Thus, the second term of the RHS of Equation (F.6) is negligible as $T \to \infty$ and

$$\sqrt{Th_T^d} G(\theta_0, \hat{f}) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathcal{K}}{f_X(x_0)} \Sigma_S(x_0)\right). \tag{F.10}$$

iv) By a similar argument as in Appendix D.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 F.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 $J_0' \Omega_0 J_0$ is nonsingular since $J_0 = \nabla_{\theta'} G(\theta_0, f_0)$ is full column-rank (Assumption 6) and Ω_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 (5.3), and by using Expression (F.10) the conclusion follows.

F.3   Proof of Proposition 7

The first order condition for estimator $\hat{\theta}^*$ is

$$
h_T^d \left[\nabla_{\theta'} G \left(\hat{\theta}^*, \hat{f} \right) \right]' \Omega_T G \left(\hat{\theta}^*, \hat{f} \right)
$$
$$
+ \frac{1}{T} \sum_{t=1}^T \mathrm{E}_{\hat{f}} \left[\nabla_{\theta'} \Gamma_U \left(X_{t+1}; \hat{\theta}^* \right) \middle| X_t = x_t \right]' \tilde{\Omega}_T (x_t) \, \mathrm{E}_{\hat{f}} \left[\Gamma_U \left(X_{t+1}; \hat{\theta}^* \right) \middle| X_t = x_t \right] = 0.
$$

By the mean-value theorem, there exists $\tilde{\theta}$ between $\hat{\theta}^*$ and θ_0 (componentwise) such that

$$
h_T^d \left[\nabla_{\theta'} G \left(\hat{\theta}^*, \hat{f} \right) \right]' \Omega_T G \left(\theta_0, \hat{f} \right)
$$
$$
+ \frac{1}{T} \sum_{t=1}^T \mathrm{E}_{\hat{f}} \left[\nabla_{\theta'} \Gamma_U \left(X_{t+1}; \hat{\theta}^* \right) \middle| X_t = x_t \right]' \tilde{\Omega}_T (x_t) \, \mathrm{E}_{\hat{f}} \left[\Gamma_U \left(X_{t+1}; \theta_0 \right) \middle| X_t = x_t \right]
$$
$$
+ \left(h_T^d \left[\nabla_{\theta'} G \left(\hat{\theta}^*, \hat{f} \right) \right]' \Omega_T \left[\nabla_{\theta'} G \left(\tilde{\theta}, \hat{f} \right) \right] + \right.
$$
$$
\left. + \frac{1}{T} \sum_{t=1}^T \mathrm{E}_{\hat{f}} \left[\nabla_{\theta'} \Gamma_U \left(X_{t+1}; \hat{\theta}^* \right) \middle| X_t = x_t \right]' \tilde{\Omega}_T (x_t) \, \mathrm{E}_{\hat{f}} \left[\nabla_{\theta'} \Gamma_U \left(X_{t+1}; \tilde{\theta} \right) \middle| X_t = x_t \right] \right) \left(\hat{\theta}^* - \theta_0 \right) = 0.
$$

By multiplying the two sides of the last equation by $T R_T'$, where $R_T := [T^{-1/2} R_1 \ \ (T h_T^d)^{-1/2} R_2]$, and using that $R_T^{-1} \left(\hat{\theta}^* - \theta_0 \right) = \left(\sqrt{T} \left(\hat{\eta}_1^* - \eta_{1,0} \right)' \ \sqrt{T h_T^d} \left(\hat{\eta}_2^* - \eta_{2,0} \right)' \right)'$, we get

$$
H_T \left(\begin{array}{c} \sqrt{T} \left(\hat{\eta}_1^* - \eta_{1,0} \right) \\ \sqrt{T h_T^d} \left(\hat{\eta}_2^* - \eta_{2,0} \right) \end{array} \right) = -T h_T^d R_T' \left[\nabla_{\theta'} G \left(\hat{\theta}^*, \hat{f} \right) \right]' \Omega_T G \left(\theta_0, \hat{f} \right)
$$
$$
- \frac{1}{T} \sum_{t=1}^T T R_T' \mathrm{E}_{\hat{f}} \left[\nabla_{\theta'} \Gamma_U \left(X_{t+1}; \hat{\theta}^* \right) \middle| X_t = x_t \right]' \tilde{\Omega}_T (x_t) \, \mathrm{E}_{\hat{f}} \left[\Gamma_U \left(X_{t+1}; \theta_0 \right) \middle| X_t = x_t \right], \quad \text{(F.11)}
$$

where

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

By using that $\tilde{J}_0(x) R_2 = 0$ for a.e. $x \in \mathcal{X}$, we get $H_T = H + o_p(1)$, where matrix H is given by:

$$
H = \left(\begin{array}{cc} H_{1,1} & 0 \\ 0 & H_{2,2} \end{array} \right) := \left(\begin{array}{cc} R_1' \mathrm{E}_0 \left[\tilde{J}_0(X_t)' \tilde{\Omega}_0(X_t) \tilde{J}_0(X_t) \right] R_1 & 0 \\ 0 & R_2' J_0' \Omega_0 J_0 R_2 \end{array} \right).
$$

Moreover, in the RHS of Equation (F.11) we have

$$\frac{1}{T}\sum_{t=1}^{T} TR_T' \mathrm{E}_{\hat{f}}\left[\nabla_{\theta'}\Gamma_U\left(X_{t+1};\hat{\theta}^*\right)\Big| X_t = x_t\right]' \tilde{\Omega}_T\left(x_t\right) \mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\big| X_t = x_t\right] = \begin{pmatrix} \Psi_{1,T} \\ 0 \end{pmatrix} + o_p(1)$$

and

$$Th_T^d R_T' \left[\nabla_{\theta'} G\left(\hat{\theta}^*,\hat{f}\right)\right]' \Omega_T G\left(\theta_0,\hat{f}\right) = \begin{pmatrix} 0 \\ \Psi_{2,T} \end{pmatrix} + o_p(1),$$

where

$$\Psi_T = \begin{pmatrix} \Psi_{1,T} \\ \Psi_{2,T} \end{pmatrix} := \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_1'\tilde{J}_0\left(x_t\right)' \tilde{\Omega}_0\left(x_t\right) \mathrm{E}_{\hat{f}}\left[\Gamma_U\left(X_{t+1};\theta_0\right)\big| X_t = x_t\right] \\ R_2'J_0'\Omega_0\sqrt{Th_T^d}G\left(\theta_0,\hat{f}\right) \end{pmatrix}.$$

Therefore, we get

$$\begin{pmatrix} \sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right) \\ \sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right) \end{pmatrix} = -H^{-1}\Psi_T + o_p(1), \tag{F.12}$$

As in Lemma A.1 in Gagliardini, Gouriéroux and Renault [2011] we have $\Psi_T \overset{D}{\to} \mathcal{N}\left(0, W\right)$, where

$$W = \begin{pmatrix} W_{1,1} & 0 \\ 0 & W_{2,2} \end{pmatrix} := \begin{pmatrix} R_1'\mathrm{E}_0\left[\tilde{J}_0\left(X_t\right)'\tilde{\Omega}_0\left(X_t\right)\Sigma_U(X_t)\tilde{\Omega}_0\left(X_t\right)\tilde{J}_0\left(X_t\right)\right]R_1 & 0 \\ 0 & R_2'J_0'\Omega_0\Sigma_S\left(x_0\right)\Omega_0J_0R_2 \end{pmatrix},$$

and the bias 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

$$\mathrm{AsVar}\left[\sqrt{T}\left(\hat{\eta}_1^* - \eta_{1,0}\right)\right] = H_{1,1}^{-1}W_{1,1}H_{1,1}^{-1} \qquad \text{and} \qquad \mathrm{AsVar}\left[\sqrt{Th_T^d}\left(\hat{\eta}_2^* - \eta_{2,0}\right)\right] = H_{2,2}^{-1}W_{2,2}H_{2,2}^{-1},$$

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 $x \in \mathcal{X}$. Proposition 7 follows.

F.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 $\hat{a}^*$.

F.4.1 Asymptotic expansion of the density estimator

Let us consider the tilting function in Equation (4.9) and derive its first-order Taylor expansion. Since $\hat{f}$ and $\hat{f}^*$ converge in probability to f_0, vector $\hat{\theta}^*$ to θ_0, Lagrange multipliers $\hat{\lambda}$, $\hat{\nu}_0$ and $\hat{\nu}(x)$ to 0 and weight ω_T to ω, 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 $\hat{\Lambda} = [\hat{\nu}_0' \ \hat{\lambda}']'$ and Γ_S is defined in Equations (5.1), 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 (5.1). Then, since $\mathrm{E}_0 \left[\Gamma_U(X_{t+1}; \theta_0) | X_t = \tilde{x} \right] = 0$ for a.e. $\tilde{x} \in \mathcal{X}$, we have

$$\int_{\mathcal{X}} \hat{f}(x|\tilde{x}) \exp \left(\hat{\nu}(\tilde{x})' \Gamma_U(x; \hat{\theta}^*) + \omega_T \hat{\lambda}' \gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) \right) dx \simeq 1 + \omega \hat{\Lambda}' \mathrm{E}_0 \left[\bar{\Gamma}_L(X_{t+1}, \tilde{x}) | X_t = \tilde{x} \right].$$

Thus, we can approximate the tilting function for the value $\tilde{x} = x_0$ of the conditioning volatility factor as

$$\frac{\exp \left(\hat{\nu}_0' \Gamma_U(x; \hat{\theta}^*) + \hat{\lambda}' \gamma_S(x; \hat{\theta}^*, \hat{f}^*) \right)}{\displaystyle\int_{\mathcal{X}} \hat{f}(x|x_0) \exp \left(\hat{\nu}_0' \Gamma_U(x; \hat{\theta}^*) + \hat{\lambda}' \gamma_S(x; \hat{\theta}^*, \hat{f}^*) \right) dx} \simeq 1 + \hat{\Lambda}' \Gamma_S(x), \tag{F.13}$$

and for any other value $\tilde{x} \neq x_0$ as

$$\frac{\exp \left(\hat{\nu}(\tilde{x})' \Gamma_U(x; \hat{\theta}^*) + \omega_T \hat{\lambda}' \gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) \right)}{\displaystyle\int_{\mathcal{X}} \hat{f}(x|\tilde{x}) \exp \left(\hat{\nu}(\tilde{x})' \Gamma_U(x; \hat{\theta}^*) + \omega_T \hat{\lambda}' \gamma_L(x, \tilde{x}; \hat{\theta}^*, \hat{f}^*) \Big/ \hat{f}_X(\tilde{x}) \right) dx} \simeq 1 + \hat{\nu}(\tilde{x})' \Gamma_U(x; \theta_0) + \omega \hat{\Lambda}' \Gamma_L(x, \tilde{x}),$$

$$\tag{F.14}$$

where Γ_L is defined in Equations (5.1). By inserting Approximations (F.13) and (F.14) into Equation (4.9) and keeping only the first-order terms in the estimators we get Approximation (5.8).

F.4.2 Asymptotic expansion of the Lagrange multipliers

Let us consider the constraints in System (4.10). They can be rewritten as:

$$\begin{cases} \displaystyle\int_{\mathcal{X}} \Gamma_U(x; \hat{\theta}^*) \hat{f}^*(x|\tilde{x}) dx = 0, & \text{for a.e. } \tilde{x} \neq x_0, \\ G(\hat{\theta}^*, \hat{f}^*) = 0. \end{cases} \tag{F.15}$$

The expansion of the LHS of the first equation of System (F.15) around (θ_0, f_0) is:

$$\int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}^*(x|\tilde{x}) dx + \tilde{J}_0(\tilde{x}) \left(\hat{\theta}^* - \theta_0 \right) + O_p \left(\|\hat{\theta}^* - \theta_0\|^2 \right) = 0, \tag{F.16}$$

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 $\tilde{J}_0(\tilde{x}) = \mathrm{E}_0 \left[\bar{\Gamma}_U(X_{t+1}) \nabla_{\theta'} \log \left(m(X_{t+1}; \theta_0) \right) | X_t = \tilde{x} \right]$. Similarly, the expansion of the LHS of the second equation in

System (F.15) around (θ_0, f_0) is:

$$\left\langle DG(\theta_0, f_0), \Delta \hat{f}^* \right\rangle + J_0 \left(\hat{\theta}^* - \theta_0 \right) + O_p \left(\| \Delta \hat{f}^* \|_\infty^2 \right) + O_p \left(\| \hat{\theta}^* - \theta_0 \|^2 \right) = 0, \qquad \text{(F.17)}$$

where matrix J_0 is defined in Assumption 6 and is the sum of the matrices defined in Equations (5.2).

We use Proposition 2 and Approximation (5.8) and keep only the leading terms to approximate the first term in the LHS of Equation (F.17) as

$$
\begin{aligned}
\left\langle DG(\theta_0, f_0), \Delta \hat{f}^* \right\rangle &= \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}^*(x|x_0) dx + \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}^*(x|\tilde{x}) dx d\tilde{x} \\
&\simeq \int_{\mathcal{X}} \bar{\Gamma}_S(x) \Delta \hat{f}(x|x_0) dx + \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) f_X(\tilde{x}) \Delta \hat{f}(x|\tilde{x}) dx d\tilde{x} \\
&\quad + \left[\int_{\mathcal{X}} \bar{\Gamma}_S(x) \Gamma_S(x)' f_0(x|x_0) dx + \omega \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) \Gamma_L(x, \tilde{x})' f_0(x|\tilde{x}) dx f_X(\tilde{x}) d\tilde{x} \right] \hat{\Lambda} \\
&\quad + \int_{\mathcal{X}} \int_{\mathcal{X}} \bar{\Gamma}_L(x, \tilde{x}) \Gamma_U(x; \theta_0)' f_0(x|\tilde{x}) dx \hat{\nu}(\tilde{x}) f_X(\tilde{x}) d\tilde{x} \\
&= \left\langle DG(\theta_0, f_0), \Delta \hat{f} \right\rangle + \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_L(x) f_X(x) dx \right) \hat{\Lambda} + \int_{\mathcal{X}} \Sigma_{L,U}(x) \hat{\nu}(x) f_X(x) dx. \qquad \text{(F.18)}
\end{aligned}
$$

Similarly, we use Approximation (5.8) to approximate the first term in the LHS of Equation (F.16) as

$$\int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}^*(x|\tilde{x}) dx \simeq \int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|\tilde{x}) dx + \omega \Sigma_{U,L}(\tilde{x}) \hat{\Lambda} + \Sigma_U(\tilde{x}) \hat{\nu}(\tilde{x}), \qquad \text{(F.19)}$$

for $\tilde{x} \neq x_0$. Then, we use Equation (3.6) and replace Approximation (F.18) in Equation (F.17), and Approximation (F.19) in Equation (F.16), to get a linearization of the constraints in System (F.15):

$$
\begin{cases}
\displaystyle \int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|\tilde{x}) dx + \omega \Sigma_{U,L}(\tilde{x}) \hat{\Lambda} + \Sigma_U(\tilde{x}) \hat{\nu}(\tilde{x}) + \tilde{J}_0(\tilde{x}) \left(\hat{\theta}^* - \theta_0 \right) \simeq 0, \\
\displaystyle G(\theta_0, \hat{f}) + \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_L(x) f_X(x) dx \right) \hat{\Lambda} + \int_{\mathcal{X}} \Sigma_{L,U}(x) \hat{\nu}(x) f_X(x) dx + J_0 \left(\hat{\theta}^* - \theta_0 \right) \simeq 0,
\end{cases}
$$
$$\qquad \text{(F.20)}$$

for $\tilde{x} \neq x_0$. We now solve System (F.20) w.r.t. the Lagrange multipliers. Since matrix $\Sigma_U(\tilde{x})$ is invertible for any $\tilde{x}$, we can solve the first approximation of System (F.20) w.r.t. $\hat{\nu}(\tilde{x})$:

$$\hat{\nu}(\tilde{x}) \simeq -\Sigma_U(\tilde{x})^{-1} \left(\int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|\tilde{x}) dx + \tilde{J}_0(\tilde{x}) \left(\hat{\theta}^* - \theta_0 \right) + \omega \Sigma_{U,L}(\tilde{x}) \hat{\Lambda} \right), \qquad \text{(F.21)}$$

for $\tilde{x} \neq x_0$. We insert Approximation (F.21) into the second approximation of System (F.20) and omit the negligible terms:

$$G(\theta_0, \hat{f}) + \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L \perp U}(x) f_X(x) dx \right) \hat{\Lambda} + \left(J_0 - J_{L\|U} \right) \left(\hat{\theta}^* - \theta_0 \right) \simeq 0, \qquad \text{(F.22)}$$

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 (5.7), i.e. $\theta = R\eta$, and using Equation (F.12) 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

$$
\begin{aligned}
\hat{\theta}^* - \theta_0 &= R_1 \left(\hat{\eta}_1^* - \eta_{1,0} \right) + R_2 \left(\hat{\eta}_2^* - \eta_{2,0} \right) \\
&\simeq -R_2 \left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2 \right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1} G(\theta_0, \hat{f}) \simeq -P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx, \quad \text{(F.23)}
\end{aligned}
$$

for the $p \times (N+2)$ matrix $P := R_2 \left(R_2' J_0' \Sigma_S(x_0)^{-1} J_0 R_2 \right)^{-1} R_2' J_0' \Sigma_S(x_0)^{-1}$. Approximation (F.22) yields

$$
\hat{\Lambda} \simeq -A \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx, \tag{F.24}
$$

for the $(N+2) \times (N+2)$ matrix A defined as

$$
A := \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L \perp U}(x) f_X(x) dx \right)^{-1} \left(I_{N+2} - \left(J_0 - J_{L\|U} \right) P \right). \tag{F.25}
$$

Finally, we use Approximations (F.21) and (F.23) to approximate $\hat{\nu}(\tilde{x})$, for any $\tilde{x} \neq x_0$, as

$$
\hat{\nu}(\tilde{x}) \simeq \Sigma_U(\tilde{x})^{-1} \left(\left(\tilde{J}_0(\tilde{x}) P + \omega \Sigma_{U,L}(\tilde{x}) A \right) \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx - \int_{\mathcal{X}} \bar{\Gamma}_U(x) \Delta \hat{f}(x|\tilde{x}) dx \right). \tag{F.26}
$$

F.4.3 Asymptotic distribution of the Lagrange multipliers

Let us first derive the asymptotic distribution of $\hat{\Lambda}$. From Approximation (F.24), by using Expression (F.8) we get

$$
\sqrt{Th_T^d} \hat{\Lambda} \xrightarrow{D} \mathcal{N} \left(0, \frac{\mathcal{K}}{f_X(x_0)} A \Sigma_S(x_0) A' \right).
$$

Let us now consider estimator $\hat{\nu}(x)$, for any $x \neq x_0$. By a similar argument as for Expression (F.8), we deduce that the two integrals in Approximation (F.26), 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

$$
\sqrt{Th_T^d} \hat{\nu}(x) \xrightarrow{D} \mathcal{N} \left(0, \Sigma_\nu(x) \right),
$$

for any $x \neq x_0$, where the 2×2 matrix Σ_ν is defined as

$$
\Sigma_\nu(x) = \frac{\mathcal{K}}{f_X(x_0)} \Sigma_U(x)^{-1} \left(\tilde{J}_0(x) P + \omega \Sigma_{U,L}(x) A \right) \Sigma_S(x_0) \left(\tilde{J}_0(x) P + \omega \Sigma_{U,L}(x) A \right)' \Sigma_U(x)^{-1} + \frac{\mathcal{K}}{f_X(x)} \Sigma_U(x)^{-1}.
$$

F.4.4 Pointwise asymptotic normality of the estimator of the historical transition density

From Approximation (5.8) and the asymptotic distribution of the Lagrange multipliers is Section F.4.3, Equation (5.9) follows. Then, we deduce Proposition 8 by standard results on the pointwise asymptotic normality of the kernel density estimator (see, e.g., Bosq [1998]).

F.4.5 Asymptotic distribution of the functionals of the historical transition density

From Equation (5.10) and Approximation (F.23) we get

$$\hat{a}^* - a_0 \simeq -\nabla_{\theta'} a(\theta_0, f_0) P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx + \left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle. \tag{F.27}$$

Let us focus on the last term of the RHS of Approximation (F.27) and proceed in a similar way as done in Section F.4.2. From Equation (4.13) for direction $\Delta \hat{f}^*$ and state variables vector $x^\star = x_0$, Approximation (5.8) and a similar argument as for Equations (F.9), we get the following approximation of the Fréchet derivative:

$$\left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle \simeq \int_{\mathcal{X}} \alpha_S(x) \Delta \hat{f}(x|x_0) dx + \int_{\mathcal{X}} \Sigma_{\alpha_L, U}(x) \hat{\nu}(x) f_X(x) dx$$
$$+ \left(\Sigma_{\alpha_S, S}(x_0) + \omega \int_{\mathcal{X}} \Sigma_{\alpha_L, L}(x) f_X(x) dx \right) \hat{\Lambda}.$$

Moreover, from Approximation (F.26) and a similar argument as for Equations (F.9) we have

$$\int_{\mathcal{X}} \Sigma_{\alpha_L, U}(x) \hat{\nu}(x) f_X(x) dx \simeq \left(\omega \int_{\mathcal{X}} \Sigma_{\alpha_L, U}(x) \Sigma_U(x)^{-1} \Sigma_{U,L}(x) f_X(x) dx A \right.$$
$$\left. + J_{\alpha_L \| U} P \right) \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx.$$

Thus, by using Approximation (F.24) we get

$$\left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle \simeq \int_{\mathcal{X}} \alpha_S(x) \Delta \hat{f}(x|x_0) dx + \left(J_{\alpha_L \| U} P - \Sigma_{\alpha_S, S}(x_0) A \right.$$
$$\left. - \omega \int_{\mathcal{X}} \Sigma_{\alpha_L, L\perp U}(x) f_X(x) dx A \right) \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx.$$

By using that $(B_1 + B_2)^{-1} = B_1^{-1} - (B_1 + B_2)^{-1} B_2 B_1^{-1}$ for any invertible matrices B_1 and B_2, the matrix A defined in Equation (F.25) can be written as

$$
\begin{aligned}
A &= \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} - \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \left(J_0 - J_{L\|U} \right) P \\
&= \Sigma_S(x_0)^{-1} - \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \left(J_0 - J_{L\|U} \right) P \\
&\quad - \omega \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \Sigma_S(x_0)^{-1}.
\end{aligned}
$$

Thus, we get

$$
\begin{aligned}
\left\langle Da(\theta_0, f_0), \Delta \hat{f}^* \right\rangle &\simeq \int_{\mathcal{X}} \left(\alpha_S(x) - \Sigma_{\alpha_S, S}(x_0) \Sigma_S(x_0)^{-1} \Gamma_S(x) \right) \Delta \hat{f}(x|x_0) dx \\
&\quad + \omega \Sigma_{\alpha_S, S}(x_0) \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \\
&\quad \cdot \Sigma_S(x_0)^{-1} \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx + J_{\alpha_L \| U} P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx
\end{aligned}
$$

$$+ \quad \Sigma_{\alpha_S,S}(x_0) \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} (J_0 - J_{L\|U}) P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx$$

$$- \quad \omega \int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}(x) f_X(x) dx \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx$$

$$+ \quad \omega \int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}(x) f_X(x) dx$$

$$\cdot \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} (J_0 - J_{L\|U}) P \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx.$$

Then, from Approximation (F.27) we get

$$\hat{a}^* - a_0 \quad \simeq \quad \int_{\mathcal{X}} \left(\alpha_S(x) - \Sigma_{\alpha_S,S}(x_0) \Sigma_S(x_0)^{-1} \Gamma_S(x) \right) \Delta \hat{f}(x|x_0) dx$$

$$+ \left(\omega B(\omega) + C(\omega) P \right) \int_{\mathcal{X}} \Gamma_S(x) \Delta \hat{f}(x|x_0) dx, \tag{F.28}$$

where the matrix $B(\omega)$ is defined as

$$B(\omega) \quad := \quad \Sigma_{\alpha_S,S}(x_0) \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \Sigma_S(x_0)^{-1}$$

$$- \int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}(x) f_X(x) dx \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1}$$

and the matrix $C(\omega)$ as

$$C(\omega) \quad := \quad \left(\Sigma_{\alpha_S,S} + \omega \int_{\mathcal{X}} \Sigma_{\alpha_L,L\perp U}(x) f_X(x) dx \right) \left(\Sigma_S(x_0) + \omega \int_{\mathcal{X}} \Sigma_{L\perp U}(x) f_X(x) dx \right)^{-1} (J_0 - J_{L\|U})$$

$$+ J_{\alpha_L\|U} - \nabla_{\theta'} a(\theta_0, f_0),$$

for any $\omega \geq 0$. The integrand $\alpha_S - \Sigma_{\alpha_S,S} \Sigma_S(x_0)^{-1} \Gamma_S$ in the first term in the RHS of Approximation (F.28) is the residual of the projection of α_S onto Γ_S, and hence orthogonal to Γ_S. Then, by a similar argument as for Expression (F.8) 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 (5.13).