Online Appendix to "Estimating a conditional density ratio model for asset returns and option demand"

Jeroen Dalderop* Oliver Linton†

December 31, 2025

B Supplementary material

B.1 Simulation implementation details

For the simulations in Section 3.5 with volatility as covariate, we use plug-in versions of the bandwidths that minimize the asymptotic weighted IMSE of $\hat{m}(y,x)$, with $h_x = c_x h_y$:

$$h_y = \left(\frac{\iint \sigma_m^2 f dy dx}{T\mu_2^2(K) \iint (m_{yy}^2 + c_x^2 m_{yy}\tilde{m}_{xx}) f dy dx} \right)^{\frac{1}{6}} \tag{B.1}$$
$$c_x = \left(\frac{\iint m_{yy}^2 f dy dx}{\iint \tilde{m}_{xx}^2 f dy dx} \right)^{\frac{1}{4}},$$

where we integrate over the middle 95% of the marginal return and volatility distributions. For the pilot estimator we use the Nadaraya-Watson estimator (12) with Silverman's rule-of-thumb bandwidths. The bandwidths are fixed in the return dimension, but are multiplied in the volatility dimension by the factor $\hat{f}_h^V(v)^{-1/6} / \int \hat{f}_h^V(u)^{5/6} du$, which moves inversely to the estimated marginal density $\hat{f}_h^V$ of the volatility factor at the MSE-optimal rate, while preserving the fixed bandwidths as their means. Both nonparametric estimators use boundary-corrected trimming at the (0.01,0.99)-quantiles of the risk-neutral distributions. We consider only the two largest sample sizes from the simulations without covariates, as localization in the conditioning variable effectively implies using only a subset of the observations at a time.

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. Email: jdalderop@nd.edu.

†Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue Cambridge, CB3 9DD, UK. Email: obl20@cam.ac.uk.

B.2 Risk-neutral density estimation

For the empirical results in Section 4, we estimate the risk-neutral densities nonparametrically estimated based on the Breeden and Litzenberger (1978) result

$$q_t(\kappa) = e^{r_t^f} \frac{\partial^2}{\partial^2 \kappa} E_t \left(C_{it} | \kappa_{it} = \kappa \right),$$

where $(C_{it}, \kappa_{it})_{i=1}^{n_t}$ is the cross-section of end-of-day call option mid prices (possibly obtained by put-call parity) with corresponding forward moneyness levels, and the risk-free rate r_t^f ensures $q_t(\cdot)$ integrates to one. When the number of recorded strike prices n_t goes to infinity, the call pricing function and its second derivative can be consistently estimated using nonparametric methods such as kernel smoothing or series approximation. In particular, we estimate the risk-neutral densities using the local cubic method in Dalderop (2020), which uses plug-in bandwidths obtained by fitting an initial Bates (2000) stochastic volatility model. Before smoothing, we apply the constrained least squares method of Aït-Sahalia and Duarte (2003) to ensure each cross section of option prices is monotone and convex in the strike price. To limit the effect of density estimation error, we only estimate each risk-neutral density on a grid $[\kappa_{tl}, \kappa_{tu}]$, where the thresholds are set as $\kappa_{tl} = \hat{Q}_t^{-1}(0.01)$ and $\kappa_{tu} = \hat{Q}_t^{-1}(0.99)$ based on an initial local-linear estimator $\hat{Q}_t(\cdot)$ of the risk-neutral cumulative distribution function. For applications that require entire densities, we 'paste' the tails of the Bates (2000) model by matching its spot volatility at the lower and upper moneyness thresholds. Figure 1 shows that the resulting risk-neutral densities for both horizons are smooth, unimodal, and left-skewed.

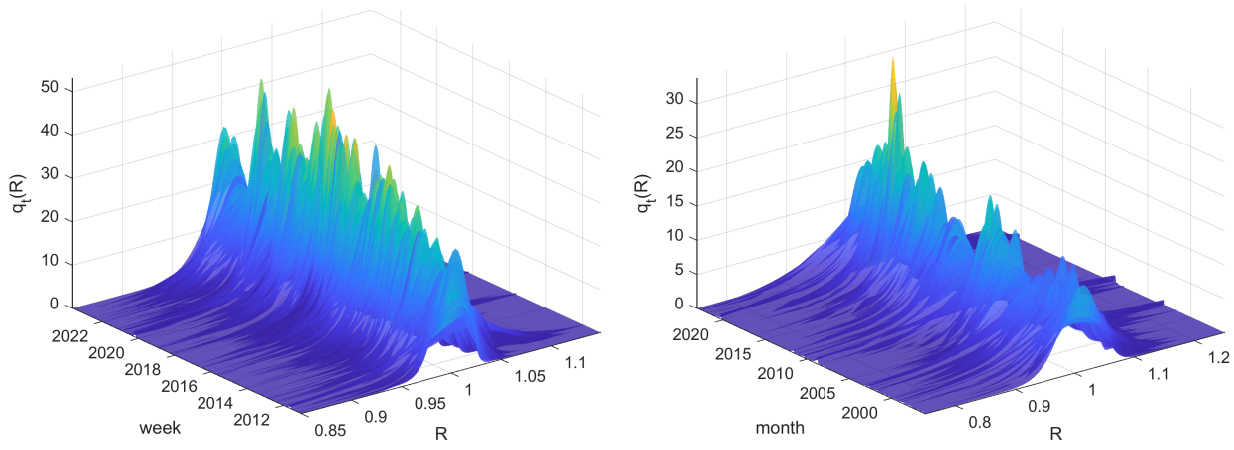
B.3 Heterogeneous investor model

This section develops a structural model to interpret the demand-based pricing kernel estimates in Section 4.4. Consider $i = 1, \ldots, N$ investors with utility U_i over wealth and subjective belief densities f_{it} for the return R_{t+1} on a stock with price S_t. Following Carr and Madan (2001), investors trade positions in the stock, a bond with risk-free return R_t^f, and a continuum of option contracts on R_{t+1}, so that the payoff function $\phi_{it}(R)$ maximizes their subjective expected utility $\int_0^\infty U_i(\phi_{it}(R)) f_{it}(R) dR$ subject to their budget constraint.

Under exponential utility $U_i(w) = -\frac{1}{\gamma_i} \exp(-\gamma_i w)$, the optimal payoff is given by

$$\phi_{it}(R) = \mu_{it} + \frac{1}{\gamma_i} \log \frac{f_{it}(R)}{q_t(R)}, \qquad i = 1, \ldots, N, \tag{B.2}$$

where μ_{it} is an intercept. The equilibrium stock market clearing condition $\sum_{i=1}^N \phi_{it}(R) = S_t R$



(a) Weekly horizon.

(b) Monthly horizon.

Figure 1: Weekly and monthly risk-neutral probability densities estimated from daily option prices from January 1996 to August 2023 using local cubic method with plug-in bandwidth from Dalderop (2020). Outside threshold points $(\kappa_{tl}, \kappa_{tu}) = (\hat{Q}_t^{-1}(0.01), \hat{Q}_t^{-1}(0.99))$, we smoothly paste tails from Bates (2000) model.

implies that the Arrow-Debreu state-price density equals

$$q_t(R) = c_t \exp\left(-\tilde{\gamma} S_t R + \sum_i \frac{\tilde{\gamma}}{\gamma_i} \log f_{it}(R)\right), \tag{B.3}$$

where $\tilde{\gamma}^{-1} = \sum_{i=1}^N \frac{1}{\gamma_i}$ sums the risk tolerances, and c_t is a normalizing constant.

Furthermore, suppose the subjective densities f_{it} are of the form

$$\log f_{it}(R) \propto \sum_{i=1}^K s_{itk} g_k(R), \tag{B.4}$$

for some basis functions $(g_k)_{k=1}^K$ and stochastic loadings s_{itk}. For example, $K = 2$ with $g_k(R) = R^k$ describes Normal densities with possible disagreement about their means and variances. Plugging (B.3) and (B.4) into (B.2) yields the equilibrium net derivative payoff functions $\omega_{it}(R)$ as

$$\omega_{it}(R) = \text{const}_{it} + \left(\frac{\tilde{\gamma}}{\gamma_i} - \omega_{it}^S\right) S_t R + \frac{1}{\gamma_i} \sum_{k=1}^K \tilde{s}_{itk} g_k(R), \tag{B.5}$$

where ω_{it}^S are the equilibrium stock positions, and $\tilde{s}_{itk} = s_{itk} - \sum_i \frac{\tilde{\gamma}}{\gamma_i} s_{itk}$ center the subjective loadings by their risk-aversion weighted means. The optimal option portfolio thus combines a finite number of claims. Given data on $\omega_{it}(R)$, we can identify the scaled loadings $x_{it} = \left(\frac{1}{\gamma_i} \tilde{s}_{itk}\right)_{k=1}^K$

as the last K coefficients in a functional regression of $\omega_{it}(R)$ on $(1, R, g_1(R), \ldots, g_K(R))$.[1]

Characterizing the pricing kernel requires connecting the subjective densities to the true conditional density. Suppose f_t has the same form as (B.4), but the true loadings are weighted averages of the subjective ones:

$$\log f_t(R) \propto \sum_k s_{tk} g_k(R), \quad s_{tk} = \sum_i \theta_{ik} s_{itk}, \quad \sum_i \theta_{ik} = 1 \text{ for all } k. \tag{B.6}$$

Combining (B.3), (B.4), and (B.6) yields the density ratio model

$$\log f_t(R) \propto \log q_t(R) + \tilde{\gamma} S_t R + \sum_{i,k} \theta_{ik} \tilde{s}_{itk} g_k(R), \tag{B.7}$$

which is parametric in terms of observed covariates. Since $\sum_{i=1}^N x_{it} = 0$ by construction, we avoid multicollinearity by standardizing the covariates as $\tilde{x}_{it} = x_{it} - x_{1t}$ for $i = 2, \ldots, N$. The resulting inverse pricing kernel can be written in terms of the parameters $\beta_{ik} = \theta_{ik}\gamma_i - \theta_{1k}\gamma_1$ as

$$m(R, S_t, \tilde{x}_t) = \exp\left(\gamma S_t R + \sum_{i=2}^{N} \sum_{k=1}^{K} \beta_{ik} \tilde{x}_{it} g_K(R)\right). \tag{B.8}$$

The informativeness (θ_{ik}) and risk aversion parameters (γ_i) are not separately identified from (β_{ik}). Still, $\beta_{ik} > (<) 0$ implies that investor i is more (less) risk averse than the first investor, which we take to be the market-maker, or more (less) informed about loading k, or both.

Pricing kernel (B.8) is exponential-linear in the covariates, so that the local parametric form (13) holds globally, and the optimal bandwidth h_x becomes infinite. However, the model relies on parametric restrictions on the utility function and the subjective and true densities. Fortunately, the local estimator is consistent even under misspecification, and has a small asymptotic bias when the true pricing kernel is approximately exponential-linear in x. The latter may be achieved by a judicious choice of x_t, for which functional form restrictions such as those discussed can serve as a guide.[2]

B.4 Conditional moment specification test

We test the null hypothesis in Section 4.5 using the distance statistic

$$I(g) = E\left[\eta_{t+1}(g) E\left(\eta_{t+1}(g) \mid x_t, z_t\right) f_{X,Z}(x_t, z_t)\right],$$

[1] See Bossu et al. (2021) for related applications of functional analysis for the static replication of payoff functions.

[2] Almeida and Freire (2022) express the pricing kernel in terms of the entire option portfolio of the marginal investor. Our heterogeneous belief formulation allows summarizing this portfolio with a low-dimensional proxy.

where the weighted error $\eta_{t+1}(g) = \varepsilon_{t+1}(g)c_t^{-1}f_X(x_t)$ serves to avoid a small denominator problem. For its empirical analog we consider the leave-one-out version given by

$$I_T(g) = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t} \hat{\eta}_{t+1}(g)\hat{\eta}_{s+1}(g) K_{h'_x}(x_t - x_s) K_{h'_z}(z_t - z_s),$$

where $\hat{\eta}_{t+1}(g) = \hat{\varepsilon}_{t+1}(g)\hat{c}_t^{-1}\hat{f}_X(x_t)$, with $\hat{\varepsilon}_{t+1}(g)$ and $\hat{c}_t$ based on the estimator $\hat{m}_{-t}(y, x_t)$ which leaves out observation t, and $\hat{f}_{X,-t}(x_t)$ the leave-one-out kernel density of x_t using the same bandwidth h_x. Li (1999, Thm 3.1.) shows that under similar conditions as ours, under H_0,

$$J_T(g) = T\sqrt{h'_x h'_z}\frac{I_T(g)}{\hat{\sigma}_I(g)} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}_I^2(g) = \frac{2h'_x h'_z}{T(T-1)} \sum_t \sum_{s \neq t} \hat{\eta}_{t+1}^2(g)\hat{\eta}_{s+1}^2(g) K_{h'_x}^2(x_t - x_s) K_{h'_z}^2(z_t - z_s)$, provided the bandwidths h'_x and h'_z vanish fast enough relative to the bandwidth h_x of the restricted model. Meanwhile, under the alternative, $J_T(g)$ will exceed any sequence of order $o(T\sqrt{h'_x h'_z})$ with probability approaching one. Therefore we reject the null for large, positive values of $J_T(g)$.

As the test statistic $J_T(g)$ converges at a slower than $\sqrt{T}$ rate to its asymptotic Normal distribution, we approximate its finite-sample distribution using the following bootstrap algorithm. First, obtain the conditional densities $\hat{f}_t$ after estimating $\hat{m}(y, x_t)$. Then, for each draw $s = 1, \ldots, S$, simulate the path of returns as $R_{t+1}^s = \hat{F}_t^{-1}(U)$, with U the standard uniform, and compute the simulated test statistic $J_T^s(g)$ based on $(R_{t+1}^s, q_t, x_t)_{t=1}^T$. As the deep tails of the pricing kernel cannot be accurately estimated, we trim the test functions as $g^{\text{trim}}(R) = g(R)1(\kappa_l \leq R \leq \kappa_u)$, with the thresholds set as the unconditional 1% and 99% percentiles of returns. The test statistic bandwidths are chosen as $(h'_x, h'_z) = 2(\hat{\sigma}(x_t), \hat{\sigma}(z_t))T^{-1/4}$ to meet the conditions in Li (1999). The bootstrapped distributions then have standard deviations close to unity, though are slightly left skewed.

B.5 Additional proofs

Proof of Theorem 2. First, we show $\hat{\beta} \xrightarrow{p} \bar{\beta}$, where $\bar{\beta} = \bar{\beta}(y, x)$. The F.O.C. of the local least squares criterion (13) are

$$0 = \frac{\partial}{\partial \beta} Q_T(y, x, \hat{\beta}) = -\frac{2}{T} \sum_{t=1}^{T} \left(\frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} - \exp\left(\hat{\beta}^T \tilde{x}_t\right) \right) \exp\left(\hat{\beta}^T \tilde{x}_t\right) \tilde{x}_t^T K_{h_x}(x_t - x),$$

where $\tilde{x}_t = (1, x_t - x)^T$. Write

$$E\left(\frac{\partial}{\partial\beta}Q_T(y,x,\beta)\right) = D(x,y,h_x,\beta) + O(h^2),$$

where

$$D(x,y,h_x,\beta) = f(x)\int\sum_{j=0}^{1}\left(\frac{\partial^j}{\partial^j x}m(y,x) - \exp(\beta_0)\right)(\beta_1 h_x u)^j \exp(\beta_0)(1, h_x u)K(u)du.$$

Then $D(x,y,h_x,\bar{\beta}) = 0$, while $D(x,y,h_x,\beta) \neq 0$ for any $\beta \neq \bar{\beta}$. Therefore $\hat{\beta}$ is consistent for $\bar{\beta}$ if the following uniform convergence holds:

$$\sup_{\beta\in\Theta}\left\|\frac{\partial}{\partial\beta}Q_T(y,x,\beta) - D(x,y,h_x,\beta)\right\| \xrightarrow{p} 0, \tag{B.9}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix. By the mean-value theorem and Cauchy-Schwartz inequality, for $j = 0, 1$, and any β and $\tilde{\beta}$

$$\left\|\frac{\partial}{\partial\beta_j}Q_T(y,x,\beta) - \frac{\partial}{\partial\beta_j}Q_T(y,x,\tilde{\beta})\right\| \leq \|\beta - \tilde{\beta}\|\sup_{\beta\in\Theta}\left\|\frac{\partial^2 Q_T(x,y,\beta)}{\partial\beta_j\partial\beta^T}\right\|.$$

The Hessian is given by

$$\frac{\partial^2 Q_T(x,y,\beta)}{\partial\beta\partial\beta^T} = -\frac{2}{T}\sum_{t=1}^{T}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - 2\exp\left(\beta^T\tilde{x}_t\right)\right)\exp\left(\beta^T\tilde{x}_t\right)\tilde{x}_t\tilde{x}_t^T K_{h_x}(x_t - x).$$

Let β_j^u and β_j^l denote the largest and smallest value of β_j in Θ for $j = 0, 1$. Since

$$\begin{aligned}
E\left(\sup_{\beta\in\Theta}\left|\frac{\partial^2 Q_T(x,y,\beta)}{\partial^2\beta_0}\right|\right) &\leq 2E\left(\sup_{\beta\in\Theta}\left|\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - 2\exp\left(\beta^T\tilde{x}_t\right)\right|\exp\left(\beta^T\tilde{x}_t\right)K_{h_x}(x_t - x)\right) \\
&\leq 2E\left(\sup_{\beta\in\Theta}\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\exp\left(\beta^T\tilde{x}_t\right)K_{h_x}(x_t - x)\right) \\
&\quad + 4E\left(\sup_{\beta\in\Theta}\exp\left(2\beta^T\tilde{x}_t\right)K_{h_x}(x_t - x)\right) \\
&\leq 2E\left(E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} \mid x_t\right)\exp\left(\beta_0^u + (|\beta_1^u| + |\beta_1^l|)|x_t - x|\right)K_{h_x}(x_t - x)\right) \\
&\quad + 4E\left(\exp\left(2\beta_0^u + 2(|\beta_1^u| + |\beta_1^l|)|x_t - x|\right)K_{h_x}(x_t - x)\right) \\
&= O(1),
\end{aligned}$$

it follows that $\sup_{\beta\in\Theta}\frac{\partial^2 Q_T(x,y,\beta)}{\partial^2\beta_0}=O_p(1)$. Other elements of the Hessian are of smaller stochastic order, so that $\sup_{\beta\in\Theta}\left\|\frac{\partial^2 Q_T(x,y,\beta)}{\partial\beta\partial\beta^\top}\right\|=O_p(1)$. Therefore $\frac{\partial}{\partial\beta}Q_T(y,x,\beta)$ is stochastically equicontinuous, which confirms (B.9) using Newey (1991, Corollary 2.2).

Consistency and $\bar{\beta}$ being an interior point imply that $0=\frac{\partial}{\partial\beta}Q_T(y,x,\hat{\beta})$ with probability approaching one. By the mean-value theorem,

$$0=\frac{\partial}{\partial\beta}Q_T(y,x,\hat{\beta})=\frac{\partial}{\partial\beta}Q_T(y,x,\bar{\beta})+\frac{\partial^2 Q_T(x,y,\beta^*)}{\partial\beta\partial\beta^T}(\hat{\beta}-\bar{\beta}), \tag{B.10}$$

for some β^* such that β_j^* lies in between $\hat{\beta}_j$ and $\bar{\beta}_j$ almost surely. Re-arranging (B.10) and scaling by $\sqrt{Th^2}H^{-1}$ yields

$$\sqrt{Th^2}H\left(\hat{\beta}-\bar{\beta}\right)=A_T(x,y,\beta^*)^{-1}\sqrt{Th^2}S_T(x,y,\beta_0),$$

where

$$\begin{aligned}A_T(x,y,\beta)&=H^{-1}\frac{\partial^2 Q_T(x,y,\beta)}{\partial\beta\partial\beta^T}H^{-1},\\ S_T(x,y,\beta)&=-H^{-1}\frac{\partial}{\partial\beta}Q_T(y,x,\beta).\end{aligned}$$

The proof consists of two main steps showing that (1) $A_T\left(x,y,\beta^*\right)$ converges to a positive-definite limit matrix, and (2) $\sqrt{Th^2}S_T\left(x,y,\bar{\beta}\right)$ satisfies a multivariate central limit theorem.

Step (1): Write

$$A_T\left(x,y,\beta^*\right)=R_{T1}\left(x,\bar{\beta}\right)+\left\{R_{T1}\left(x,\beta^*\right)-R_{T1}\left(x,\bar{\beta}\right)\right\}+R_{T2}\left(x,y,\beta^*\right),$$

where

$$\begin{aligned}R_{T1}\left(x,\beta\right)&=\frac{2}{T}\sum_{t=1}^{T}\exp\left(2\beta^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t-x\right),\\ R_{T2}(x,y,\beta)&=-\frac{2}{T}\sum_{t=1}^{T}\left\{\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}-\exp\left(\beta^T\tilde{x}_t\right)\right\}\exp\left(\beta^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t-x\right).\end{aligned}$$

We will show that, for a positive-definite limit matrix $A_{T0}(x)$ given below,

$$\left\|R_{T1}\left(x,\bar{\beta}\right)-A_{T0}(x)\right\|=\mathrm{o_p}(h_x), \tag{B.11a}$$

$$\left\|R_{T1}\left(x,\beta^*\right)-R_{T1}\left(x,\bar{\beta}\right)\right\|=\mathrm{o_p}(1), \tag{B.11b}$$

$$\left\|R_{T2}(x,y,\beta^*)\right\|=\mathrm{o_p}(1). \tag{B.11c}$$

Using second-order Taylor expansions around $x_t = x$, the means of the terms in $R_{T1}(x, \bar{\beta})$ are

$$\begin{aligned} E\left(\exp\left(2\bar{\beta}^T\tilde{x}_t\right)K_{h_x}\left(x_t - x\right)\right) &= \int \exp\left(2\bar{\beta}_0 + 2\bar{\beta}_1(x_t - x)\right)K_{h_x}\left(x_t - x\right)f(x_t)dx_t \\ &= \int \exp\left(2\bar{\beta}_0 + 2\bar{\beta}_1 h_x z\right)K\left(z\right)f(x + zh_x)dz \\ &= m(y,x)^2 f(x) + O(h_x^2), \end{aligned}$$

$$\begin{aligned} E\left(\exp\left(2\bar{\beta}^T\tilde{x}_t\right)\left(\frac{x_t - x}{h_x}\right)K_{h_x}\left(x_t - x\right)\right) &= \int \exp\left(2\bar{\beta}_0 + 2\bar{\beta}_1 h_x z\right)zK\left(z\right)f(x + zh_x)dz \\ &= m(y,x)^2\mu_2(K)\left(2\bar{\beta}_1 f(x) + f'(x)\right)h_x + O(h_x^3), \end{aligned}$$

and

$$\begin{aligned} E\left(\exp\left(2\bar{\beta}^T\tilde{x}_t\right)\left(\frac{x_t - x}{h_x}\right)^2 K_{h_x}\left(x_t - x\right)\right) &= \int \exp\left(2\bar{\beta}_0 + 2\bar{\beta}_1 h_x z\right)z^2K\left(z\right)f(x + zh_x)dz \\ &= m(y,x)^2\mu_2(K)f(x) + O(h_x^2). \end{aligned}$$

Therefore,

$$E\left(R_{T1}\left(x, \bar{\beta}\right)\right) = 2m(y,x)^2\begin{pmatrix} f(x) & \mu_2(K)r_{12}(x)h_x \\ \mu_2(K)r_{12}(x)h_x & \mu_2(K)f(x) \end{pmatrix} + \begin{pmatrix} O\left(h_x^2\right) & O\left(h_x^3\right) \\ O\left(h_x^3\right) & O\left(h_x^2\right) \end{pmatrix} \equiv A_{T0}(x) + O(h_x^2),$$

where $r_{12}(x) = 2\bar{\beta}_1 f(x) + f'(x)$. Furthermore, the variance of each element in $R_{T1}(x, \bar{\beta})$ is $O\left(\frac{1}{Th_x}\right)$ under the mixing condition. Therefore (B.11a) follows under bandwidth condition 2g).

By the mean-value theorem, there exists some $\tilde{\beta}$ in between β^* and $\bar{\beta}$ such that

$$R_{T1}\left(x, \beta^*\right) - R_{T1}\left(x, \bar{\beta}\right) = \frac{4}{T}\sum_{t=1}^{T}\left(\beta^* - \bar{\beta}\right)^T\tilde{x}_t\exp\left(2\tilde{\beta}^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t - x\right).$$

Therefore (B.11b) follows from

$$\begin{aligned} &E\left(\|R_{T1}\left(x, \beta^*\right) - R_{T1}\left(x, \bar{\beta}\right)\|\right) \\ &\leq 4E\left(\|\left(\beta^* - \bar{\beta}\right)^T\tilde{x}_t\exp\left(2\tilde{\beta}^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t - x\right)\|\right) \\ &\leq 4\|\beta^* - \bar{\beta}\|\int\|(1, zh_x)\|E\left(\exp\left(2\tilde{\beta}^T(1, zh_x)\right)\right)\|\tilde{z}\tilde{z}^T\|K\left(z\right)f(x + zh_x)dz \\ &\to 0, \end{aligned} \tag{B.12}$$

using the triangle inequality and stationarity in the first step, the Cauchy-Schwartz inequality, properties of K, compact support of $\tilde{\beta}$, and Fubini's Theorem in the second, and consistency of β^* in the third.

To establish (B.11c), write

$$\begin{aligned} R_{T2}(x,y,\beta) = &-\frac{2}{T}\sum_{t=1}^{T}\left\{\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - m(y,x_t)\right\}\exp\left(\beta^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t-x\right) \\ &-\frac{2}{T}\sum_{t=1}^{T}\left\{m(y,x_t)-\exp\left(\beta^T\tilde{x}_t\right)\right\}\exp\left(\beta^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t-x\right) \\ \equiv\ & T_{T1}(x,y,\beta)+T_{T2}(x,y,\beta). \end{aligned}$$

Using second-order expansions, $E\left(T_{T1}(x,y,\beta)\right) = O(h^2)$ for any β, while $E\left(T_{T2}(x,y,\bar{\beta})\right) = O(h_x^2)$. Furthermore, $\mathrm{Var}\left(T_{T1}(x,y,\beta)\right) = O\left(\frac{1}{Th^2}\right)$ and $\sup_{\beta\in\Theta}\frac{\partial}{\partial\beta}T_{T1}(x,y,\beta) = O_p\left(1\right)$, so that $T_{T1}(x,y,\beta) \xrightarrow{p} 0$ uniformly over β. Moreover, $\mathrm{Var}\left(T_{T2}(x,y,\bar{\beta})\right) = O\left(\frac{1}{T}\right)$, and $\|T_{T2}(x,y,\beta^*) - T_{T2}(x,y,\bar{\beta})\| = o_p(1)$ using similar steps as (B.12). Therefore $\|T_{Tj}(x,y,\beta^*)\| = o_p(1)$ for $j=1,2$, implying (B.11c).

Step (2): Write $S_T\left(x,y,\bar{\beta}\right)$ as

$$S_T\left(x,y,\bar{\beta}\right) = 2m(y,x)\left(S_{T1}\left(x,y,\bar{\beta}\right)+S_{T2}\left(x,y,\bar{\beta}\right)+S_{T3}\left(x,y,\bar{\beta}\right)\right),$$

where, writing $\tilde{x}_{t,h} = \left(1, \frac{x_t-x}{h_x}\right)^T$,

$$\begin{aligned} S_{T1}\left(x,y,\bar{\beta}\right) &= \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h}\tilde{x}_{t,h}K_{h_x}\left(x_t-x\right), \\ S_{T2}\left(x,y,\bar{\beta}\right) &= \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h}\left(\exp\left(\bar{\beta}_1(x_t-x)\right)-1\right)\tilde{x}_{t,h}K_{h_x}\left(x_t-x\right), \\ S_{T3}\left(x,y,\bar{\beta}\right) &= \frac{1}{T}\sum_{t=1}^{T}\left\{E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\mid x_t\right)-\exp\left(\bar{\beta}^T\tilde{x}_t\right)\right\}\exp\left(\bar{\beta}_1(x_t-x)\right)\tilde{x}_{t,h}K_{h_x}\left(x_t-x\right), \end{aligned}$$

with $u_{t+1,h} = \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\mid x_t\right)$, so that $E\left(S_{T1}\left(x,y,\bar{\beta}\right)\right) = E\left(S_{T2}\left(x,y,\bar{\beta}\right)\right) = 0$.

We will show below that

$$\begin{aligned} &\sqrt{Th^2}S_{T1}\left(x,y,\bar{\beta}\right) \xrightarrow{d} \mathrm{N}\left(0,V(x,y)\right), \\ &\left\|\sqrt{Th^2}S_{T2}\left(x,y,\bar{\beta}\right)\right\| = \mathrm{O_p}(h), \\ &\left\|S_{T3}\left(x,y,\bar{\beta}\right) - h^2Ha(x,y)\right\| = \mathrm{O_p}\left(h^4\right), \end{aligned}$$

where $V(x,y) = c_x^{-1} R_0(K) m(y,x) \text{diag}(R_0(K), R_2(K)) E\left(\frac{c_t}{q_t(y)} \mid x_t = x\right) f(x)$, with $R_j(K) = \int z^j K^2(z) dz$ for any integer j, and $a(x,y)$ is given below.

Given stationarity,

$$\begin{aligned}\text{Var}\left(\sqrt{Th^2} S_{T1}\left(x, y, \bar{\beta}\right)\right) &= h^2 \text{Var}\left(u_{t+1,h} \tilde{x}_{t,h} K_h(x_t - x)\right) \\ &\quad + 2h^2 \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \text{Cov}\left(u_{j+1,h} \tilde{x}_{j,h} K_h(x_j - x), u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right).\end{aligned}$$

The variance term equals

$$\begin{aligned}h^2 \text{Var}\left(u_{t+1,h} \tilde{x}_t K_{h_x}(x_t - x)\right) &= h^2 E\left(\text{Var}\left(\frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} \mid x_t\right) \tilde{x}_{t,h} \tilde{x}_{t,h}^T K_{h_x}^2(x_t - x)\right) \\ &= h^2 E\left(E\left(\int \left(\frac{K_h(R-y)}{q_t(R)}\right)^2 f_t(R) dR \mid x_t\right) \tilde{x}_{t,h} \tilde{x}_{t,h}^T K_{h_x}^2(x_t - x)\right) + O(h) \\ &= h^2 E\left(\int E\left(\frac{c_t}{q_t(R)} \mid x_t\right) K_h^2(R-y) m(R, x_t) dR \tilde{x}_{t,h} \tilde{x}_{t,h}^T K_{h_x}^2(x_t - x)\right) + O(h) \\ &= \iint E\left(\frac{c_t}{q_t(y + hu)} \mid x_t = x + h_x z\right) K^2(u) m(y + hu, x + h_x z) du \\ &\quad \times \tilde{z}\tilde{z}^T K^2(z) f(x + h_x z) dz + O(h) \\ &= V(x,y) + O(h).\end{aligned}$$

By the LIE,

$$\begin{aligned}&\text{Cov}\left(u_{j+1,h} \tilde{x}_{j,h} K_h(x_j - x), u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right) \\ &\quad = E\left(\left(\int \frac{K_h(R-y)}{q_j(R)} f_j(R) dR - E\left(\frac{K_h(R_{j+1} - y)}{q_j(R_{j+1})} \mid x_j\right)\right) \tilde{x}_{j,h} K_h(x_j - x) u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right) \\ &\quad = \text{Cov}\left(\int K_h(r-y) m(r, x_j) dr \left(c_j - 1\right) \tilde{x}_{j,h} K_h(x_j - x), u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right).\end{aligned}$$

By Davydov's inequality for strong mixing processes

$$\begin{aligned}&\text{Cov}\left(\left(c_j - 1\right) \tilde{x}_{j,h} K_h(x_j - x), u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right) \\ &\quad \leq 8\alpha(j)^{\frac{\delta}{2+\delta}} E\left(\left|\left(c_j - 1\right) \tilde{x}_{j,h} K_h(x_j - x)\right|^{2+\delta}\right)^{\frac{1}{2+\delta}} E\left(\left(u_{1,h} \tilde{x}_{0,h} K_h(x_0 - x)\right)^2\right)^{\frac{1}{2}} \\ &\quad \leq C\alpha(j)^{\frac{\delta}{2+\delta}} E\left(\left|c_j - 1\right|^{2+\delta} K_h^{2+\delta}(x_j - x)\right)^{\frac{1}{2+\delta}} E\left(\text{Var}\left(u_{1,h} \mid x_t\right) K_h^2(x_0 - x)\right)^{\frac{1}{2}} I_2 \\ &\quad \leq C\alpha(j)^{\frac{\delta}{2+\delta}} h_x^{-\frac{1+\delta}{2+\delta}} (h_y h_x)^{-\frac{1}{2}} I_2.\end{aligned}$$

Therefore, for some constant $C > 0$,

$$\sum_{j=1}^{T-1} \|\mathrm{Cov}\left(u_{j+1,h}\tilde{x}_{j,h}K_h(x_j - x), u_{1,h}\tilde{x}_{0,h}K_h(x_0 - x)\right)\| \leq Ch^{-\frac{3+2\delta}{2+\delta}} \sum_{j=1}^{T-1} \alpha(j)^{\frac{\delta}{2+\delta}} = O\left(h^{-\frac{3+2\delta}{2+\delta}}\right),$$

which is of smaller order than the $O(h^{-2})$ variance term.

Asymptotic normality of $S_{T1}(x, y, \bar{\beta})$ follows a similar large-small block argument as used for Theorem 1. The block sizes $l_T = \sqrt{T}h/\log T$ and $s_T = \left(\frac{\sqrt{T}}{h}\log T\right)^{\frac{\delta}{2+\delta}}$ satisfy the conditions in (A.1). In particular,

$$\frac{s_T}{l_T} = \left(\sqrt{T}h^{1+\delta}\right)^{-\frac{2}{2+\delta}} (\log T)^{\frac{2+2\delta}{2+\delta}} \to 0,$$

as Assumption 2g) implies that $T^{-\frac{1}{2}}h^{-1-\delta} = O(T^{-\varepsilon_0})$ for some $\varepsilon_0 > 0$. Therefore $k_T = O(T/l_T) = O\left(\frac{\sqrt{T}}{h}\log T\right) = O\left(s_T^{1+\frac{2}{\delta}}\right)$. The mixing condition implies $T^{1+\frac{2}{\delta}}\alpha(T) \to 0$, so that $k_T\alpha(s_T) \to 0$.

Meanwhile element-wise second-order mean-value expansions of $\exp(\bar{\beta}_1(x_t - x))\tilde{x}_t$ yield some x_t^* between $x_t - x$ and 0 such that

$$S_{T2}\left(x, y, \bar{\beta}\right) = \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h} \exp(\bar{\beta}_1 x_t^*) \begin{pmatrix} \bar{\beta}_1 \\ 2\bar{\beta}_1 + \bar{\beta}_1^2 x_t^* \end{pmatrix} (x_t - x)\, \tilde{x}_{t,h} K_{h_x}\left(x_t - x\right),$$

using $\frac{\partial}{\partial x}\exp(\beta x)x = \exp(\beta x)\left(1 + \beta x\right)$ and $\frac{\partial^2}{\partial^2 x}\exp(\beta x)x = \beta\exp(\beta x)\left(2 + x\right)$. Using similar steps as for $S_{T1}\left(x, y, \bar{\beta}\right)$, it follows that $\mathrm{Var}\left(S_{T2}\left(x, y, \bar{\beta}\right)\right) = O\left(\frac{h_x^2}{Th^2}\right) = O\left(\frac{1}{T}\right)$.

Performing second- and third-order Taylor expansions for the constant and slope term, respectively, in $S_{T3}(x, y, \bar{\beta})$ allows obtaining its leading terms as

$$\begin{aligned}
S_{T3}\left(x, y, \bar{\beta}\right) &= \frac{1}{2T}\sum_{t=1}^{T} \Big\{\mu_2(K)m_{yy}(y, x_t)h^2 + \left(m_{xx}(y, x) - \bar{\beta}_1^2 m(y, x)\right)(x_t - x)^2 + \mu_2(K)m_{yyx}(y, x)(x_t - x) \\
&\quad + \frac{1}{3}\left(m_{xxx}(y, x) - \bar{\beta}_1^3 m(y, x)\right)(x_t - x)^3\Big\}\left(1 + \bar{\beta}_1(x_t - x)\right)\tilde{x}_{t,h}K_{h_x}(x_t - x) + O_p(h^4), \\
&= \frac{1}{2}h^2 H \begin{pmatrix} \mu_2(K)f(x)\left(m_{yy}(y, x) + \left(m_{xx}(y, x) - \bar{\beta}_1^2 m(y, x)\right)c_x^2\right) + O_p(h^2) \\ \left(\mu_2^2(K)m_{yyx}(y, x)c_x + \frac{1}{3}\mu_4(K)\left(m_{xxx}(y, x) - \bar{\beta}_1^3 m(y, x)\right)c_x^3\right)f(x) + O_p(h) \\ + \left(\mu_2^2(K)m_{yy}(y, x)c_x + \mu_4(K)\left(m_{xx}(y, x) - \bar{\beta}_1^2 m(y, x)\right)c_x^3\right)\left(\bar{\beta}_1 f(x) + f'(x)\right) \end{pmatrix} \\
&\equiv h^2 H a(y, x) + O_p(h^4),
\end{aligned}$$

where the term $\bar{\beta}_1 f(x) + f'(x)$ in $a_2(y, x)$ stems from first-order expansions of $\exp(\beta_1(x_t - x))$

and $f(x)$.

Combining results from Steps (1) and (2), the asymptotic bias of $\hat{\beta}$ equals

$$\begin{aligned}
E\left(\hat{\beta}-\bar{\beta}\right) &= H^{-1}A_{T0}(x)^{-1}2m(y,x)h^2Ha(x,y)+o(h^2)\\
&= h^2m(y,x)^{-1}f(x)^{-1}\begin{pmatrix}1 & -f(x)^{-1}r_{12}(x)h_x\\ -f(x)^{-1}r_{12}(x) & \mu_2(K)^{-1}h_x^{-1}\end{pmatrix}\begin{pmatrix}a_1(x,y)\\ ha_2(x,y)\end{pmatrix}+o(h^2)\\
&= h^2m(y,x)^{-1}f(x)^{-1}\begin{pmatrix}a_1(x,y)\\ -f(x)^{-1}r_{12}(x)a_1(x,y)+c_x^{-1}\mu_2(K)^{-1}a_2(x,y)\end{pmatrix}+o(h^2)\\
&\equiv h^2b(x,y)+o(h^2),
\end{aligned}$$

while its asymptotic variance equals

$$\begin{aligned}
\text{Var}\left(\frac{1}{\sqrt{Th^2}}(\hat{\beta}-\bar{\beta})\right) &= A_{T0}(x)^{-1}4m(y,x)^2V(x,y)A_{T0}(x)^{-1}+o(1)\\
&= m(y,x)^{-1}f(x)^{-1}c_x^{-1}R_0(K)\text{diag}(R_0(K),R_2(K)/\mu_2^2(K))E\left(\frac{c_t}{q_t(y)}\mid x_t=x\right)\\
&\equiv \Omega(x,y)+o(1).
\end{aligned}$$

Therefore we have established that $\sqrt{Th^2}H\left(\hat{\beta}-\bar{\beta}-h^2b(x,y)\right)\xrightarrow{d}N\left(0,\Omega(x,y)\right)$. The limiting distribution of $\hat{m}(y,x)=\exp(\hat{\beta}_0)$ follows from the delta method. □

References

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.

Almeida, C. and Freire, G. (2022). Demand in the option market and the pricing kernel. *Available at SSRN 4314173*.

Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of econometrics*, 94(1-2):181–238.

Bossu, S., Carr, P., and Papanicolaou, A. (2021). A functional analysis approach to the static replication of european options. *Quantitative Finance*, 21(4):637–655.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.

Carr, P. and Madan, D. (2001). Optimal positioning in derivative securities. *Quantitative Finance*, 1(1):19–37.

Dalderop, J. (2020). Nonparametric filtering of conditional state-price densities. *Journal of Econometrics*, 214(2):295–325.

Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics*, 92(1):101–147.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167.