# Estimating a conditional density ratio model for asset returns and option demand

Jeroen Dalderop* Oliver Linton†

July 1, 2025

**Abstract**

Option-implied risk-neutral densities are widely used for constructing forward-looking risk measures. Meanwhile, risk aversion introduces a multiplicative pricing kernel between the risk-neutral and true conditional densities of the underlying asset's return. This paper proposes a simple local estimator of the pricing kernel based on inverse density weighting. We characterize the asymptotic bias and variance of the estimator and its multiplicatively corrected density forecasts. A local exponential linear variant is proposed to include conditioning variables. The estimator performs well in a simulation study, even when the risk-neutral densities are noisy and/or have missing tails. We apply our estimator to a demand-based model for S&P 500 index options, and find U-shaped pricing kernels when end-users sell out-of-the-money options and volatility is high.

*Keywords:* Density Forecasting, Nonparametric Estimation, Option Pricing, Trading Data
*JEL Codes:* C14, C58, G13

## 1 Introduction

Option prices are widely used to extract forward-looking, market-implied distributions for the purpose of measuring and managing financial risk (e.g. Aït-Sahalia and Lo, 2000). Relative to

*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.

We thank seminar and conference participants at Northwestern Kellogg, University of Amsterdam, Notre Dame, Copenhagen Business School, Cambridge, QFFE Marseille, Econometric Society Meeting in Hangzhou, and SoFiE in Rio, for helpful comments. Financial support by the Keynes Fund is gratefully acknowledged.

statistical model-based risk measures, they benefit from the wide range of information available to investors beyond historical data. However, market-implied measures reflect not only rational expectations given investors' information set, but also their preferences towards risks as well as any imperfect beliefs. Together, these introduce a multiplicative deviation, known as the pricing kernel, between the risk-neutral and true conditional densities of the underlying asset's return. The shape and dynamics of the pricing kernel contain important economic insights about the nature of risk aversion and belief formation. Moreover, the pricing kernel can be used to 'correct' risk-neutral densities in order to measure the true conditional distributions of the asset return.

This paper proposes a simple local estimator of the pricing kernel, with several desirable properties. First, the estimator is nonparametric, and thus avoids imposing functional form restrictions on investor preferences or the distribution of the underlying return. This is important given the various nonlinear shapes found in empirical pricing kernels (Rosenberg and Engle, 2002). Second, the estimator does not require specifying the information set that determines the conditional return distribution. This avoids the practice of reducing investors' information set to that of the econometrician, which (Linn et al., 2017) and (Barone-Adesi et al., 2020) suggest could explain non-monotonic pricing kernel estimates. Third, it avoids global approximations of the pricing kernel, unlike recently developed methods designed to meet the preceding two points, such as conditional density integration (Linn et al., 2017; Kim, 2023) or maximum likelihood (Cuesdeanu and Jackwerth, 2018; Schreindorfer and Sichert, 2025). Such methods require numerical optimization of a growing number of coefficients, which presents challenges for computation and inference. Fourth, unlike the global approximation methods, our local estimator only requires evaluating the risk-neutral density at the realized returns, rather than over its entire support. This limits the impact of any measurement error in its tails due to sparse option trading.

In particular, our kernel-smoothing estimator is computed by taking weighted averages of the inverse of the risk-neutral density at the realized returns. Similar local estimators have been used for the multiplicative correction of initial densities obtained by kernel-smoothing (Jones et al., 1995) or a parametric start (Hjort and Glad, 1995; Hagmann and Scaillet, 2007). Our method is a dynamic extension which performs a local multiplicative correction of observed predictive densities. We show that the estimator is consistent for the multiplicative component in a general class of conditional density ratio models, not restricted to the setting of option-implied densities. Such models only require specifying which variables determine the *ratio* between the true and observed conditional densities, not which affect the densities themselves. We prove that the estimator is asymptotically Normal for a wide range of data-generating processes subject to certain moment and mixing conditions. Furthermore, we provide simple analytical expressions of its

asymptotic bias and variance, and use these to characterize asymptotically optimal bandwidths. Finally, we characterize the asymptotic distribution of the multiplicatively corrected densities.

In most economic models, the pricing kernel depends on conditioning variables that affect the stochastic discount factor, and thereby determine risk premia. To accommodate such variables, we propose a local exponential-linear variant of the estimator that guarantees a positive estimator and valid corrected densities, and establish its asymptotic properties. When the true log pricing kernel is approximately affine in the conditioning variables, this variant effectively reduces the bias and/or allows choosing a larger bandwidth to reduce the variance relative to the locally constant estimator.

A simulation study confirms the good performance of the local estimator for realistic data-generating processes and sample periods. In particular, the estimator with plug-in bandwidths outperforms or is at least competitive with the correctly specified parametric maximum likelihood estimator for small and moderate sample sizes, for both GARCH-$t$ and stochastic volatility models. The robust performance carries over to the corrected densities, and to the case of volatility-dependent pricing kernels. The heavy tails of the risk-neutral densities help to reduce the variance in the tails of the estimator. In the presence of thin tails, such as the right tail for left-skewed densities, the variance of the estimator can be effectively reduced by trimming small values of the risk-neutral densities. A simple boundary-bias correction prevents the trimming from incurring substantial biases.

In practice, risk-neutral densities are not observed directly, but estimated from large cross-sections of option prices. Therefore we investigate the robustness of our pricing kernel estimator to observational errors in the densities. Theoretically, we show that density estimation error is asymptotically negligible when our estimator trims the densities on ranges that may grow with the cross-sectional sample size. For local polynomial risk-neutral density estimators, we show that its biases have an asymptotically larger impact on the pricing kernel estimator than its noise terms, which average out over time. Our simulations confirm these biases matter for small samples of options, which could be addressed by undersmoothing. However for reasonably large sample sizes the density estimation error has little effect on our trimmed estimators.

We apply our estimator to recover the pricing kernel implicit in S&P 500 index options and returns. First, we show that the unconditional pricing kernel has a slightly upward sloping right tail, confirming this non-decreasing pattern is not just the result of misspecified conditional physical densities. When conditioning on volatility, we corroborate the finding in Schreindorfer and Sichert (2025) that lower volatility leads to steeper pricing kernels for left tail risks, but primarily at the weekly rather than the monthly horizon. We then focus on measuring time-

variation in the pricing kernel shape due to fluctuations in demand for options, whose large trading volumes suggest that they are not redundant assets. Indeed, a vast literature documents that net option demand predicts excess returns (Easley et al., 1998; Pan and Poteshman, 2006), excess implied volatility (Bollen and Whaley, 2004; Garleanu et al., 2008), and tail risk premia (Chen et al., 2019). Recent work by Almeida and Freire (2022) studies how net demand impacts option returns for varying strike prices, from which they infer the presence of non-decreasing regions in the pricing kernel. Here, we directly estimate the pricing kernel, from which other relevant quantities such as conditional expected option returns can be derived.

While our approach is nonparametric and allows for a general choice of covariates, we develop a structural parametric model to interpret our estimates and guide the choice of explanatory variables. In particular, we model net option demand by heterogeneous traders using a variant of the portfolio choice problem in Carr and Madan (2001). The model's equilibrium risk-neutral density is a risk-aversion weighted average of investor's subjective densities. By linking the latter to the true physical density, we derive a conditional density ratio model that expresses the pricing kernel in terms of investor's net positions. Any impact of net positions points to heterogeneity in either risk aversion or informativeness, although it does not identify their separate impact. Moreover, the model guides the construction of net demand proxies from traders' option portfolios, and implies the local exponential-linear approximation is exact under exponential utility. Using data on traders' positions in S&P 500 index options, we find that the monthly pricing kernel varies with inferred differences in subjective volatilities. This creates a U-shape when end-users hold net short positions in squared returns, especially when volatility is high.

Finally, we test whether the considered conditioning variables significantly change the pricing kernel using a variant of the specification test from Li (1999). Both the VIX and the net demand proxy significantly affect the right tail of the monthly pricing kernel, as inferred from call option payoffs. At the weekly horizon, net demand mainly changes the slope of the pricing kernel, and is thus associated with the equity premium. After controlling for the VIX, the net demand proxy also significantly affects the left tail of the pricing kernel, as inferred from put option payoffs.

The remainder of this paper is organized as follows. Section 2 describes the local estimator of a conditional density ratio model and its asymptotic properties. Section 3 performs the simulation study. Section 4 applies the estimator to S&P 500 index options, returns, and net positions data. Section 5 concludes. All proofs are in the Appendix.

## 2 Estimator and asymptotic properties

This section introduces a density ratio model for the relation between true and observed conditional densities, and studies the asymptotic properties of a simple nonparametric estimator with and without observed covariates.

### 2.1 Density ratio model

Suppose the econometrician observes a set of density forecasts $\{q_t(\cdot)\}_{t=1}^T$ describing the conditional distributions of each period's outcome variable $R_{t+1}$. Meanwhile, suppose that the latter's true density $f_t(y)$ given the time-$t$ information set $\mathcal{F}_t$ follows the general density ratio model

$$f_t(y) = c_t q_t(y) m(y, x_t), \tag{1}$$

where $m(y, x)$ is an unknown function whose shape may change with observed covariates $x_t$, and $c_t = \left(\int q_t(r) m(r, x_t) dr\right)^{-1}$ is the normalizing constant. Without loss of generality we assume $E(c_t \mid x_t = x) = 1$ for all $x$, so that $m(y, x) = E\left(\frac{f_t(y)}{q_t(y)} \mid x_t = x\right)$. Values of $m(y, x)$ different from one describe a density forecast bias, as they indicate $q_t(y)$ tends to under- or overstate the true density $f_t(y)$.

Density ratio models of the form (1) arise naturally in the context of pricing options on the underlying return $R_{t+1}$. In particular, no-arbitrage theory states that the price $P_t(g)$ of a contingent claim with payoff $g(R)$ equals

$$P_t(g) = E_t\left[M_{t,t+1} g(R_{t+1})\right], \tag{2}$$

for some positive stochastic discount factor $M_{t,t+1}$ with $E_t\left(M_{t,t+1}\right) = \frac{1}{R_t^f}$, where $R_t^f$ is the return on a risk-free bond. Define also the conditional pricing kernel $\pi_t(y) \equiv E\left(M_{t,t+1} \mid R_{t+1} = y, \mathcal{F}_t\right)$. By the law of iterated expectations, (2) can then be represented as

$$P_t(g) = \frac{1}{R_t^f} \int_0^\infty g(y) q_t(y) dy,$$

in terms of the risk-neutral density $q_t(y) = R_t^f f_t(y) \pi_t(y)$. Therefore, (1) holds as long as

$$\pi_t(y) = \delta_t \tilde{\pi}(y, x_t), \tag{3}$$

for some constants $\delta_t$ and covariate-dependent function $\tilde{\pi}(y, x)$, in which case $m(y, x) = \frac{1}{\tilde{\pi}(y,x)}$ and $c_t^{-1} = R_t^f \delta_t$.

A standard economic model that satisfies (3) is when investors have CRRA utility over wealth, in which case $M_{t,t+1} = \delta R_{t+1}^{-\gamma}$ for some time-discount factor $\delta$ and relative risk aversion $\gamma$ (e.g. Bliss and Panigirtzoglou, 2004). This model can be extended to allow for stochastic time preferences described by a time-varying time-discount factor $\delta_t$, as emphasized by Albuquerque et al. (2016). The stochastic discount factor $M_{t,t+1}$ may also depend on consumption or other macroeconomic variables, in which case the projection $\pi_t(y)$ depends on their joint distribution with the return. Since economic theory may not prescribe the distributional form, it is desirable not to restrict $m(y, x)$ parametrically. Furthermore, formulation (3) allows for flexible nonlinear specifications of the pricing kernel, such as those by Rosenberg and Engle (2002) and Linn et al. (2017). The constants $\delta_t$ then ensure that the implied $f_t(y)$ are valid densities. Finally, it allows for state-dependent preferences that can be described by observed covariates, such as implied measures of volatility (Song and Xiu, 2016) or tail risk (Andersen et al., 2015). Our application in Section 4.4 uses proxies of trading imbalances as covariates and measures their impact on the pricing kernel. Using a stylized equilibrium trading model, it relates $m(y, x)$ in formulation (1) to the preferences and beliefs of heterogeneous investors.

### 2.2 Estimator without covariates

First, we consider the case in which $m$ does not depend on any covariates, that is, $m(y, x) = m(y)$.[1] Let the data $\{q_t(\cdot), R_{t+1}\}_{t=1}^T$ consist of a time series of risk-neutral densities $q_t(\cdot)$ for the distribution of the return $R_{t+1}$ and the latter's realizations. The density ratio model for the conditional density $f_t(y)$ that generates $R_{t+1}$ then becomes

$$f_t(y) = \frac{q_t(y)m(y)}{\int q_t(y)m(y)dy}, \tag{4}$$

which by design integrates to unity. The multiplicative factor $m(y)$ describes the (inverse) pricing kernel, which can be interpreted as the marginal utility of investors as a function of the return outcome. Since $m$ is only identified up to scaling, we set $E(c_t) = 1$ without loss of generality. As $c_t = \left(\int q_t(y)m(y)dy\right)^{-1} = \int f_t(y)/m(y)dy$, this normalization implies $E\left(\frac{1}{m(R_{t+1})}\right) = 1$.

Based on the local multiplicative kernel density correction in Jones et al. (1995), Hjort and Glad (1995), and Gustafsson et al. (2009), we consider the following local estimator for $m(y)$:

$$\hat{m}(y) = \frac{1}{T}\sum_{t=1}^{T} \frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})}, \tag{5}$$

[1] By a change-of-variable, this case also covers models where $m(y, x) = m(G(y, x))$ for some known transformation $G$ that is monotonic in $y$ for every $x$, such as the standardized excess return or probability integral transform.

where $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$ for some bandwidth $h$ and symmetric kernel $K(\cdot)$ on $\mathbb{R}$. The asymptotic bias and variance of estimator (5) have simple analytical expressions. In particular, when $h \to 0$, its leading bias term follows from

$$\begin{aligned} E\left(\hat{m}(y)\right) &= E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right) \\ &= E\left(\int \frac{K_h(R-y)}{q_t(R)} f_t(R)dR\right) \\ &= E\left(c_t\right)\int K_h(R-y)m(R)dR \\ &= m(y) + \frac{\mu_2(K)}{2}m''(y)h^2 + o(h^2), \end{aligned} \tag{6}$$

using the Law of Iterated Expectations in the second step, where $\mu_k(K) = \int K(z)z^k dz$. As in nonparametric kernel regression, the asymptotic bias depends on the curvature of the object of interest, but not on the outcome density or its derivatives. The asymptotic variance and distribution of the estimator are established in the following result.

**Assumption 1.**

*a)* $K$ *is a symmetric bounded density with compact support*

*b)* *The function* $m(\cdot)$ *is twice differentiable on its support*

*c)* $q_t(R_{t+1}) > 0$ *a.s. and* $E\left(\frac{c_t}{q_t(y)}\right) < \infty$

*d)* $E(c_t^{2+\delta}) < \infty$ *for some* $\delta > 0$

*e)* $q_t(R) = q(R, s_t)$ *for some differentiable function* $q$ *and state variables* $s_t$ *such that* $(R_{t+1}, s_t)$ *is stationary and* $\alpha$*-mixing with* $\sum_{j=1}^{\infty}\alpha(j)^{\frac{\delta}{4+2\delta}} < \infty$

*f)* *When* $T \to \infty$, $h \to 0$, $Th^5 = O(1)$, *and* $Th^{\alpha_0} \to \infty$ *for some* $\alpha_0 > \frac{3\delta+4}{\delta+3}$

**Theorem 1.** *Suppose that Assumption 1 holds and that* $y$ *is an interior point. Then, as* $T \to \infty$

$$\sqrt{Th}\left(\hat{m}(y) - m(y) - B(y)h^2\right) \xrightarrow{d} N(0, \Omega(y)),$$

*where* $B(y) = \frac{\mu_2(K)}{2}m''(y)$ *and* $\Omega(y) = R(K)m(y)E\left(\frac{c_t}{q_t(y)}\right)$ *with* $R(K) = \int K(z)^2 dz$.

*Discussion of assumptions.* Assumption 1c) contains a moment condition on the inverse risk-neutral density, which ensures the variance of the estimator is finite. Assumption 1d) imposes a moment condition on the normalization constant that is commonly imposed on the error in nonparametric regression. It is satisfied when $m(\cdot)$ is bounded from below by a positive constant.

Otherwise, it prevents risk-neutral mass from concentrating in regions where $m(y)$ is close to zero. The mixing condition in Assumption 1e) is relatively weak, as the normalization constants are the only predictable components in $\hat{m}(y)$. The state variables $s_t$ that determine the shape of $q_t(\cdot)$ do not need to be observed or specified, and thus accommodate a wide range of dynamic latent variable models. Finally, Assumption 1f) provides upper and lower bounds on the rate at which $h$ becomes small.

*Comparison to other estimators.* Alternative nonparametric estimators involve global approximations of the function $m(y, x)$, such as orthogonal polynomials (Rosenberg and Engle, 2002) or splines (Linn et al., 2017). Such methods nest the parametric specifications in Bliss and Panigirtzoglou (2004). However, their consistency relies on increasing the approximation order with the sample size, requiring high-dimensional optimization that involves repeated numerical integration to compute the constants $c_t$. Furthermore, little is known about the asymptotic properties of estimators that depend on random densities, which complicates inference.[2]

Another advantage of the local estimator (5) is that it does not require observing the densities $q_t(\cdot)$ on their entire domain, but only at the realizations $R_{t+1}$. This matters particularly for risk-neutral densities, whose tails cannot be reliably estimated due to sparse trading of deep out-of-the-money options. Even if $q_t(\cdot)$ is only observed on a range $[\kappa_l, \kappa_u]$ that excludes some realizations $R_{t+1}$, then (5) still approximates $m(y)$ on this range as before, up to a boundary bias that could be corrected. Indeed, our simulations include the boundary-corrected estimator (18) that allows for variable ranges $[\kappa_{tl}, \kappa_{tu}]$. Moreoever, our estimator accommodates observational errors in the densities $q_t(\cdot)$, which we analyze theoretically in Section 2.4 and via simulations in Section 3.4.

*Robustness to missing conditioners.* Suppose instead the true densities are of the form $f_t(y) = q_t(y)m(y, s_t)$, where $s_t$ are state variables unobserved by the econometrician. Chabi-Yo et al. (2008) consider a general asset pricing model with latent state variables with this representation, which includes several leading models. The local estimator will then be consistent for the average inverse conditional pricing kernel $\bar{m}(y) \equiv E\left(\frac{f_t(y)}{q_t(y)}\right) = E\left(m(y, s_t)\right)$, as can be seen by adapting the bias derivation (6). While $\bar{m}(y)$ cannot be used to recover the conditional physical densities, it carries important economic insights. In particular, non-decreasing behavior in $1/\bar{m}(y)$ implies that $\pi_t(y) = 1/m(y, s_t)$ must be non-decreasing for some values of $s_t$, thus establishing the existence of non-monotonic conditional pricing kernels.

[2] Kapetanios et al. (2015) provide possibly related asymptotic theory for sieve estimators of the weighting functions in density forecast combinations, under assumptions such as bounded input densities.

### 2.3 Density forecast correction

The estimator $\hat{m}(y)$ can be plugged into (4) to estimate the normalization constants by

$$\hat{c}_t = \left( \int q_t(y)\hat{m}(y)dy \right)^{-1} = \left( \frac{1}{T}\sum_{s=1}^{T} \frac{\int q_t(y)K_h(R_{s+1}-y)dy}{q_s(R_{s+1})} \right)^{-1}, \tag{7}$$

and compute the multiplicatively corrected predictive densities as

$$\hat{f}_t(y) = q_t(y)\hat{c}_t \frac{1}{T}\sum_{s=1}^{T} \frac{K_h(R_{s+1}-y)}{q_s(R_{s+1})}, \tag{8}$$

which by construction integrates to one.

From a forecasting perspective, it is useful to perform inference on the corrected version $\hat{f}_{t^*}(y)$ of any observed density $q_{t^*}(y)$ (for any given $t^*$), which may or may not have been included in the sample used for $\hat{m}$. Thereto, for any known function $\omega(y)$ define the weighted integral $\mu_\omega = \int m(y)\omega(y)dy$. The following proposition characterizes the asymptotic distribution of its plug-in estimator $\hat{\mu}_\omega = \int \hat{m}(y)\omega(y)dy$.

**Proposition 1.** *Let Assumption 1 hold. In addition, let $\omega(y)$ be twice differentiable, $E\left(\left(\frac{\omega(R_{t+1})}{q_t(R_{t+1})}\right)^{2+\delta}\right) < \infty$, and $\text{Var}\left(\frac{\omega''(R_{t+1})}{q_t(R_{t+1})}\right) < \infty$. When $T \to \infty$ and $h \to 0$, then $\hat{\mu}_\omega - \mu_\omega = O_p(h^2 + T^{-\frac{1}{2}})$ and*

$$\sqrt{T}\left(\hat{\mu}_\omega - \mu_\omega - B_\omega h^2\right) \xrightarrow{d} N(0, \Omega_\omega),$$

*where $B_\omega = \frac{\mu_2(K)}{2}\int \omega(y)m''(y)dy$ and $\Omega_\omega = \text{Var}\left(\frac{\omega(R_{t+1})}{q_t(R_{t+1})}\right) + 2\sum_{j=1}^{\infty}\text{Cov}\left(\frac{\omega(R_1)}{q_0(R_1)}, \frac{\omega(R_{j+1})}{q_j(R_{j+1})}\right)$.*

Proposition 1 shows that the standard error of the weighted integral estimator shrinks at the faster $\sqrt{T}$ rate, although its bias is of the same order as that of $\hat{m}(y)$ pointwise. The moment conditions control the tail behavior of the function $\omega(R)$ relative to $q_t(R)$. Setting $\omega(y) = q_{t^*}(y)$ allows controlling the estimation error in the normalization constants $c_t$, which gives rise to the following corollary characterizing the relative estimation error of $\hat{f}_{t^*}(y)$.

**Corollary 1.** *Let Assumption 1 and the additional conditions in Proposition 1 hold for the fixed conditional density $q_{t^*}(y)$. Moreover, let $Th^4 \to \infty$. Then*

$$\sqrt{Th}\left(\frac{\hat{f}_{t^*}(y) - f_{t^*}(y)}{f_{t^*}(y)} - h^2\left(\frac{B(y)}{m(y)} - c_{t^*}B_c\right)\right) \xrightarrow{d} N\left(0, \frac{\Omega(y)}{m(y)^2}\right),$$

*where $B_c = \frac{\mu_2(K)}{2}\int q_{t^*}(y)m''(y)dy$.*

Corollary 1 shows that when the sample period increases, the density estimation error is of the

same order as $\hat{m}(y)$, and is not affected by any time series dependence. The assumption that $q_{t^*}(y)$ is fixed is merely for simplicity, as under the mixing conditions any dependence between $q_{t^*}(\cdot)$ and $\{q_t(\cdot)\}_{t=1}^T$ is asymptotically irrelevant. The estimation error in $\hat{c}_{t^*}$ adds an additional bias term to the density estimation error, but does not affect its variance as the normalization constants are estimated at the faster $\sqrt{T}$ rate by Proposition 1. The above CLT holds for all $y$ in the support, so that the asymptotic integrated mean squared density error IMSDE $\equiv \int E(\hat{f}_{t^*}(y) - f_{t^*})^2 dy$ follows by integrating the squared bias plus the variance over the support:

$$\text{IMSDE} = h^4 \int \left( \frac{B(y)}{m(y)} - c_{t^*} B_c \right)^2 f_{t^*} dy + \frac{1}{Th} \int \frac{\Omega(y)}{m(y)^2} f_{t^*} dy + o(1).$$

### 2.4 Estimator with densities observed with noise

Risk-neutral densities can be consistently estimated from large cross-sections of option prices, see for example Aït-Sahalia and Lo (2000). To consider the effect of density estimation error, suppose that we replace the density $q_t(\cdot)$ by an estimator $\tilde{q}_t(\cdot)$ based on a cross-sectional sample of size $n_t$, where for convenience we suppose that $n_t = n$. In practice, option contracts are only actively traded for moneyness levels in a bounded range $I_{tn} = [\kappa_{tl}, \kappa_{tu}]$. This motivates the following local inverse pricing kernel estimator that only depends on the estimated density $\tilde{q}_t(y)$ for $y \in I_{tn}$:

$$\tilde{m}^I(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(R_{t+1} - y)}{\tilde{q}_t(R_{t+1})} 1(R_{t+1} \in I_{tn}),$$

where $1(\cdot)$ is the indicator function. Similarly, define the trimmed estimator

$$\hat{m}^I(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} 1(R_{t+1} \in I_{tn}) \tag{9}$$

for when the risk-neutral density is observed without error on a finite range. When the ranges $I_{tn}$ grow with $n$, $\hat{m}^I(y)$ becomes ever closer to the untrimmed estimator $\hat{m}(y)$. The effect of density estimation error on the trimmed estimator equals

$$\tilde{m}^I(y) - \hat{m}^I(y) = -\frac{1}{T} \sum_{t=1}^{T} \frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} \frac{\tilde{q}_t(R_{t+1}) - q_t(R_{t+1})}{\tilde{q}_t(R_{t+1})} 1(R_{t+1} \in I_{tn}). \tag{10}$$

Suppose that $\tilde{q}_t(\cdot)$ is a uniformly consistent estimator of $q_t(\cdot)$ over the growing range $I_{tn}$. Commonly used kernel estimators satisfy this under general conditions, see Li et al. (2012) for uniform convergence rates and discussion. The error in the estimation of $q_t(\cdot)$ then determines the magnitude of $\tilde{m}^I(y) - \hat{m}^I(y)$:

**Proposition 2.** *Suppose that* $\max_{1\leq t\leq T}\sup_{y\in I_{tn}}|\tilde{q}_t(y)-q_t(y)|=o_p(\delta_{n,T})$ *for a sequence of positive constants* $\delta_{n,T}$ *such that* $\delta_{n,T}\to 0$ *when* $n,T\to\infty$, *and* $\max_{1\leq t\leq T}\sup_{y\in I_{tn}}\frac{|\tilde{q}_t(y)-q_t(y)|}{q_t(y)}\xrightarrow{p}0$. *Then* $\tilde{m}^I(y)-\hat{m}^I(y)=o_p(\delta_{n,T})$.

The assumptions in Proposition 2 ensure that $\tilde{q}_t(y)$ stays far enough from zero, which in practice can be controlled using the ranges $I_{tn}$. For specific estimators and error structures we can characterize (10) more explicitly. In particular, suppose we observe cross-sections of call option prices and their moneyness levels $(\tilde{C}_{it},\kappa_{it})_{i=1}^n$ according to $\tilde{C}_{it}=C_t(\kappa_{it})+\sigma_t^C(\kappa_{it})\varepsilon_{it}$, where $C_t(\kappa)$ is the true pricing function, $\sigma_t^C(\kappa)^2$ the conditional error variance function, and $\varepsilon_{it}$ a standardized error term which is *i.i.d.* across $t$ (but could be weakly dependent across $i$). Using the relation $q_t(y)=-\frac{1}{R_t^f}\frac{\partial^2 C_t(\kappa)}{\partial^2\kappa}\Big|_{\kappa=y}$, let $\tilde{q}_t(y)$ be based on the local cubic estimator of the second derivative of the call pricing function with bandwidth $h_n$. Standard conditions and results for local polynomial estimators yield, for the kernel $K_{(r,p)}$ defined in Wand and Jones (1994, p. 137),

$$\begin{aligned}\mathrm{E}\left[\tilde{q}_t(y)-q_t(y)\right]&=\frac{\mu_4(K_{(2,3)})}{4!}q_t''(y)h_n^2+o_p(h_n^2)\\ \mathrm{V}\left[\tilde{q}_t(y)-q_t(y)\right]&=\frac{R(K_{(2,3)})\sigma_t^C(y)^2}{nh_n^5}+o_p\left(\frac{1}{nh_n^5}\right).\end{aligned}$$

Plugging these into (10), and using the independence of $\varepsilon_{it}$ over time, we obtain

$$\tilde{m}^I(y)-\hat{m}^I(y)=-\frac{\mu_4(K_{(2,3)})}{4!}\frac{1}{T}\sum_{t=1}^T K_h(R_{t+1}-y)\frac{q_t''(R_{t+1})}{q_t^2(R_{t+1})}h_n^2+o_p(h_n^2)+O_p\left(\frac{1}{\sqrt{Tnh_n^5}}\right). \quad (11)$$

Thus, the contribution of the variance of $\tilde{q}_t$ to $\tilde{m}^I$ is an order $O(\frac{1}{\sqrt{T}})$ smaller than that of its bias. If the bandwidth $h_T$ for the pricing kernel estimator is chosen at its MSE-optimal rate, then $\hat{m}^I(y)-m(y)=O_p(h_T^2)$. Since $\tilde{m}^I(y)-\hat{m}^I(y)=O_p(h_n^2)$, the cross-sectional error is asymptotically irrelevant as long as $h_n/h_T\to 0$ when $n,T\to\infty$. In other words, the risk-neutral density smoothing bandwidth should be small relative to that of the pricing kernel estimator. We maintain this assumption in the sequel, though consider the effect of density estimation error in simulations.

### 2.5 Estimator with covariates

Let $x_t$ be some covariates that are thought to explain time-variation in the multiplicative bias $m(y,x_t)$ in observed predictive densities $q_t(y)$ according to formulation (1). Then, following

similar steps as for the case without covariates,

$$E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} \mid x_t = x\right) = E\left(c_t \mid x_t = x\right) \int K_h(R-y)m(R,x)dR$$
$$= m(y,x) + \frac{\mu_2(K)}{2} m_{yy}(y,x)h^2 + o(h^2).$$

Therefore, the following Nadaraya-Watson type estimator of $m(y,x)$ is asymptotically unbiased:

$$\hat{m}(y,x) = \sum_{t=1}^{T} \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} w_T(x_t,x), \tag{12}$$

when $h \to 0$ and $h_x \to 0$, where $w_T(x_t,x) = \frac{K_{h_x}(x_t-x)}{\sum_{t=1}^{T} K_{h_x}(x_t-x)}$ are kernel-smoothing weights for the covariate dimension with bandwidth $h_x$. Local linear estimators are often preferred over the locally constant estimator (12) for their reduced bias. However, the local linear estimator is not guaranteed to be positive, and thus may not produce valid corrected densities.

Instead, we propose estimating the positive multiplicative factor $m(y,x)$ as $\hat{m}(y,x) = \exp(\hat{\beta}_0)$, based on the local linear fit with an exponential link function:

$$\hat{\beta} = \arg\min_{\beta \in \Theta} \; Q_T(y,x,\beta) \tag{13}$$
$$Q_T(y,x,\beta) = \frac{1}{T}\sum_{t=1}^{T} \left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - \exp\left(\beta_0 + \beta_1(x_t-x)\right)\right)^2 K_{h_x}(x_t-x),$$

for some parameter space $\Theta \subset \mathbb{R}^2$. Related estimators have been studied by Gozalo and Linton (2000) and Hyndman and Yao (2002). Specifically, Gozalo and Linton (2000) consider local nonlinear least squares for nonparametric regression based on an initial parametric model. Meanwhile, Hyndman and Yao (2002) use a local exponential fit for conditional density estimation. Our method applies local least squares to fit $\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}$, thus combining elements of regression and density estimation.

The following result establishes the asymptotic bias, variance, and Normality of the local exponential-linear estimator. Its proof first establishes that the locally estimated coefficients $\hat{\beta}$ converge to the population coefficients $\bar{\beta}(y,x) = (\bar{\beta}_0(y,x), \bar{\beta}_1(y,x))$, which match the level and first derivative of the target and exponential link functions. It then shows that $\hat{\beta}$ is asymptotically Normal, which carries over to $\hat{m}(y,x) = \exp(\hat{\beta}_0)$ by the delta method.

**Assumption 2.**

*a)* $\bar{\beta}(y,x)$ *is an interior point of the compact parameter space* $\Theta$

*b) $K$ is a symmetric bounded density with compact support*

*c) $m(y,x)$ is twice continuously differentiable*

*d) $q_t(R_{t+1}) > 0$ a.s. and $E\left(\frac{c_t}{q_t(y)} \mid x_t = x\right) < \infty$*

*e) $E(c_t^{2+\delta} \mid x_t = x) < \infty$ for all $x$ and some $\delta > 0$*

*f) $q_t(R) = q(R, s_t, x_t)$ for some differentiable function $q$ and state variables $s_t$ such that $(R_{t+1}, s_t, x_t)$ is stationary and $\alpha$-mixing with $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta}{2+\delta}} < \infty$*

*g) When $T \to \infty$, $h \to 0$, $h_x = c_x h$ for some $c_x > 0$, $Th^6 = O(1)$, and $Th^{\alpha_0} \to \infty$ for some $\alpha_0 > \max\{2+\delta, 3\}$*

**Theorem 2.** *Under Assumption 2, when $T \to \infty$*

$$\sqrt{Th^2}\left(\hat{m}(y,x) - m(y,x) - h^2 b_m(y,x)\right) \xrightarrow{d} N\left(0, \sigma_m^2(x,y)\right),$$

*where*

$$\begin{aligned} b_m(y,x) &= \frac{\mu_2(K)}{2}\left(m_{yy}(y,x) + \left(m_{xx}(y,x) - \bar{\beta}_1^2(y,x)m(y,x)\right)c_x^2\right) \\ \sigma_m^2(y,x) &= m(y,x)f(x)^{-1}c_x^{-1}R(K)^2 E\left(\frac{c_t}{q_t(y)} \mid x_t = x\right). \end{aligned} \tag{14}$$

Assumption 2 contains mostly natural extensions to the conditions required for the case without covariates. However, introducing the covariate allows the slightly weaker mixing condition 2f), as localizing in $x_t$ diminishes any serial correlation in the summands of $\frac{\partial}{\partial\beta}Q_T$.

The asymptotic bias of $\hat{m}(y,x)$ has two components, reflecting smoothing in the return and covariate dimensions. The former is proportional to $m_{yy}(y,x)$, and does not depend on the choice of local parametric model. The latter is proportional to $\tilde{m}_{xx}(y,x) = m_{xx}(y,x) - \bar{\beta}_1^2(y,x)m(y,x)$. The closer $m(y,x)$ is to being exponential in $x$, the closer this bias component is to zero. As a result, if the true pricing kernel is approximately exponential in the covariate, the locally exponential estimator will have smaller bias than the local constant estimator (12) for a given bandwidth. Moreover, the optimal bandwidth will be larger, reducing the variance of the estimator.

### 2.6 Conditional moment specification test

The key restriction of model (1) is that $m(r, x_t, z_t) = m(r, x_t)$ for all $r$ and any variable $z_t \in \mathcal{F}_t$. It allows consistently estimating $m$ and therefore the conditional physical densities $f_t$ over time. Under correct specification, the conditional expectation of any function $g$ equals $E_t\left[g(R_{t+1})\right] =$

$c_t \int_0^\infty g(y)m(y,x_t)q_t(y)dy$. This motivates the following conditional moment-based specification test. Consider the prediction error $\varepsilon_{t+1}(g) := g(R_{t+1}) - c_t \int_0^\infty g(y)m(y,x_t)q_t(y)dy$. We can then test whether some variable $z_t \in \mathcal{F}_t$ is correctly excluded using a conditional moment restriction:

$$H_0: \quad E\left(\varepsilon_{t+1}(g)\middle|x_t, z_t\right) = 0 \quad \text{a.e.} \quad \text{versus} \quad H_a: \quad \mathbb{P}[E\left(\varepsilon_{t+1}(g)\middle|x_t, z_t\right) > 0] > 0$$

The null hypothesis is a necessary, but not sufficient, condition for the correct specification of $m$. Nonetheless, by choosing $g$ to be an interpretable function, such as a call or put option payoff, a rejection of the null reveals which part of the pricing kernel is dynamically misspecified.

Following Li (1999), we test the null hypothesis using the distance statistic

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

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 $J_T(g)$. Implementation details are provided in Section 4.5 of our empirical results.

## 3 Simulation study

We perform a simulation study into the finite-sample performance of the kernel estimator in various realistic settings, and compare its performance with the parametric maximum likelihood estimator.

### 3.1 Bandwidth choices

The asymptotic Mean Squared Error (MSE) of the estimator $\hat{m}(y)$ in (5) equals

$$\text{MSE}_T(y) = B(y)^2 h^4 + \frac{\Omega(y)}{Th}.$$

The asymptotic Integrated weighted Mean Square Error (IMSE), defined as $\int \text{MSE}_T(y) f(y) dy$ with $f(y)$ the unconditional return density, is minimized by the bandwidth

$$h^{\text{IMSE}} = \left( \frac{\int \Omega(y) f(y) dy}{4T \int B(y)^2 f(y) dy} \right)^{\frac{1}{5}}. \tag{15}$$

The integrals in (15) may not be finite over the full support, reflecting the difficulty of estimating the $m$ deep into the tails. Therefore, we truncate the IMSE over a large but finite range, and set the plug-in IMSE bandwidth based on the corresponding definite integrals of $B(y)$ and $\Omega(y)$.

We estimate the optimal bandwidth by plugging initial estimates of $\Omega(y)$ and $B(y)$ into (15). Both components depend on the unknown function $m(y)$, for which we consider two initial estimators $\hat{m}_{\text{int}}(y)$. The first fits an initial parametric model $m(y;\theta)$ by maximizing the predictive likelihood of the realized returns. The second is the nonparametric estimator (5) with a pilot bandwidth of two times Silverman's rule-of-thumb bandwidth for kernel density estimation. For both methods the local bias is estimated as $\hat{B}(y) = \frac{\mu_2(K)}{2} \hat{m}''_{\text{int}}(y)$ and the local variance as $\hat{\Omega}(y) = R(K) \hat{m}_{\text{int}}(y) \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{c}_t}{q_t(y)}$, with $\hat{c}_t$ the plug-in normalization constant (7). The plug-in bandwidth then integrates these components with respect to the kernel density estimator $\hat{f}_h(y)$.

### 3.2 Simulation design

The pseudo-code in Algorithm 1 describes our simulation design.

**Algorithm 1:** Simulation algorithm for pricing kernel estimators

```
for s = 1 : S do
    for t = 1 : T do
        - Compute model-based conditional density f_ts(y)
        - Generate return as R_{t+1,s} = F_ts^{-1}(U), where U ~ Uniform(0,1)
        - Compute risk-neutral density q_ts(y) ∝ f_ts(y) m_0(y)
    end
    Compute local estimator m̂_s(y) based on (q_ts, R_{t+1,s})_{t=1}^T
end
```

**DGPs.** We consider two data-generating process for the conditional physical densities of the returns. First, we consider a discrete-time AR(1)-GARCH(1,1) model with $t$-distributed

errors, with parameters set to their maximum likelihood estimates fitted to the S&P 500 index options and returns in our empirical application. Second, we consider the single-factor stochastic volatility model described by

$$\frac{\mathrm{d}F_t}{F_t} = r\mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t + (J-1)\mathrm{d}N(t),$$
$$\mathrm{d}V_t = \kappa(\Theta - V(t))\mathrm{d}t + \gamma\sqrt{V_t}\mathrm{d}W_t,$$

where the Brownian motions $B_t$ and $W_t$ are mutually correlated, but independent of the jump process $N(t)$ with Poisson jump times and lognormal jump sizes. The parameters are set as the estimates in Bates (2000) based on the likelihood of returns, and yield left-skewed densities.

For the true pricing kernel, we consider a exponential-quartic model, which allows capturing the main nonlinear shapes documented in the literature. Similar exponential-polynomial models are considered in Rosenberg and Engle (2002) and Schreindorfer and Sichert (2025). In particular, let

$$m(R;\gamma) = \exp\left(\sum_{l=1}^{4}\gamma_l\left(\log R\right)^l\right),$$

and let $\gamma_0$ be the maximum likelihood estimator for $\gamma$ in the sample used in our application. We then set the true pricing kernel as $m_0(R) = m(R;\gamma_0)$.

**Estimators.** Besides the local nonparametric estimators, we include the correctly specified maximum likelihood estimator of the exponential-quartic model. With $L = 4$, this yields the estimator $m(y;\hat{\gamma}_L)$, where

$$\hat{\gamma}_L = \arg\max_{\gamma}\sum_{t=1}^{T}\sum_{l=1}^{L}\gamma_l(\log R_{t+1})^l - \log\int q_t(y)\exp\left(\sum_{l=1}^{L}\gamma_l(\log y)^l\right)dy. \tag{16}$$

For the nonparametric local estimator we consider two fixed plug-in bandwidth methods that are based on either a parametric or nonparametric initial estimator.[3] The parametric initial estimator uses the slightly misspecified exponential-cubic estimator $m(y;\hat{\gamma}_3)$. We also consider two forms of trimming that aim to lower the variance of the local estimator in its tails due to small risk-neutral densities in the denominator. The first is the smoothly trimmed estimator, defined as

$$\hat{m}_{\text{sm-trim}}(y) = \frac{1}{T}\sum_{t=1}^{T}\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}S\left(\frac{q_t(R_{t+1})}{\tau_T}\right), \tag{17}$$

where $S$ is a CDF with $S(0) = 0$ and $S(1) = 1$, and $\tau_T$ a vanishing trimming parameter. We

[3]Variable bandwidths perform broadly similar to fixed bandwidths in terms of the truncated IMSE, and its results are therefore omitted. However, they allow increased bandwidths in the tails of the distribution to combat the increased variance there.

set $S$ as the Beta$(2,2)$ CDF following Linton and Xiao (2007), and $\tau_T$ as the 0.01-quantile of $q_t(R_{t+1})$. The second is the range-trimmed estimator

$$\hat{m}_{\text{corr-trim}}(y) = \frac{1}{T}\sum_{t=1}^{T} \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} \frac{1(\kappa_{tl} \leq R_{t+1} < \kappa_{tu})}{b_t(y)}, \tag{18}$$

which is the same as (9) except for the first-order boundary-bias corrections $b_t(y) = \int_{(\kappa_{tl}-y)/h}^{(\kappa_{tu}-y)/h} K(z)dz$. We set the thresholds $\kappa_{tl} = Q_t^{-1}(0.005)$ and $\kappa_{tu} = Q_t^{-1}(0.995)$ at the conditional risk-neutral quantiles. Estimator (18) only requires observing $q_t$ within user-controlled ranges, which can be used to increase robustness against estimation errors in the tails of the risk-neutral density. We choose its bandwidth based on the trimmed version of the nonparametric pilot estimator.

For all estimators, we report the simulated performance of their scaled versions, defined as

$$\hat{m}_{\text{sc}}(y) = \hat{m}(y)\frac{1}{T}\sum_{t=1}^{T}\hat{c}_t,$$

with $\hat{c}_t$ the plug-in normalization constants (7). The scaling enforces the sample version of the theoretical constraint $E(c_t) = 1$, and serves to reduce the variance.

### 3.3 Results without covariates

Table 1 reports the simulated performance of the estimators for varying sample sizes. For both models considered, all nonparametric estimators outperform the correctly specified parametric estimator at the smallest sample size ($T = 60$) and remain competitive at the medium sample size ($T = 120$). Only at the largest size ($T = 240$) does the parametric maximum likelihood estimator outperform all nonparametric estimators, due to its faster convergence rate. Thus, two decades of monthly data would be needed for the exponential-quartic estimator to outperform, even in the ideal case of correct specification. Meanwhile, the simple nonparametric estimator performs reliably at all sample sizes, and is robust against functional form misspecification. Smooth trimming benefits the local estimator by bringing down its variance substantially without incurring a similar increase in its bias. Its robust performance is remarkable given the simple 1% trimming rule used, and could be improved further by optimizing the trimming threshold and/or correcting the trimming bias (e.g. Sasaki and Ura, 2022). Boundary-corrected trimming further reduces the bias of the local estimator, and also further reduces the variance for the Bates (2000) model, by truncating some returns from its relatively thin right tail. Our simulations indicate low

sensitivity to the choice of pilot estimator $\hat{m}_{\text{int}}$ used for the plug-in bandwidth.[4] Therefore we only report results for the parametric pilot estimator, except for the range-trimmed estimator which uses the nonparametric pilot estimator, as it only requires observing the risk-neutral densities over finite ranges.

Table 1: Simulation performance of various estimators under exponential-quartic pricing kernel, based on $S = 10,000$ simulated time series of densities and returns according to two models. Nonparametric estimators use plug-in bandwidths, and either no, smooth ('sm-trim'), or boundary-corrected range-based ('corr-trim') trimming. Columns describe integrated weighted squared bias, variance, and mean squared error of scaled estimators $\hat{m}_{\text{sc}}$, truncated at the (0.005,0.995)-quantiles of the return distribution, for three numbers of months $T$.

(a) AR(1)-GARCH(1,1)-$t$-model

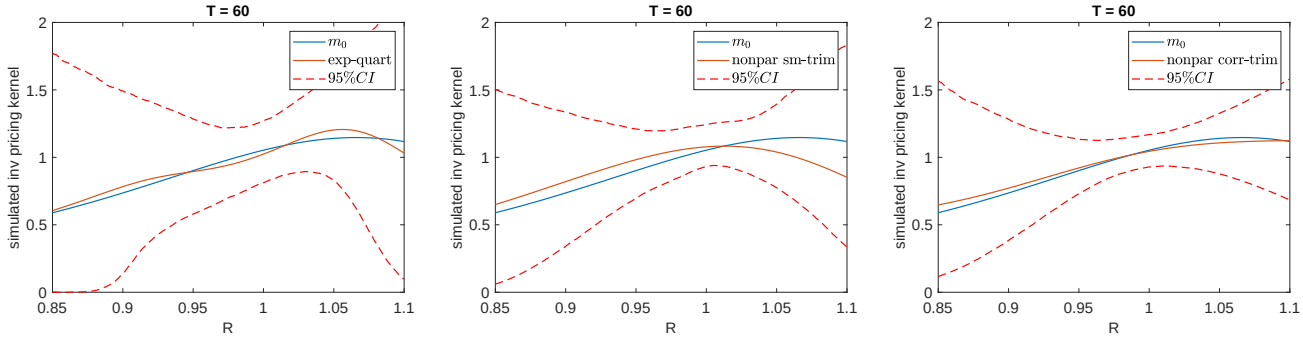
| | $T = 60$ | | | $T = 120$ | | | $T = 240$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{m}$ | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE |
| exp-quart | 0.14 | 4.08 | 4.22 | 0.07 | 1.58 | 1.64 | 0.03 | 0.79 | 0.82 |
| nonpar | 0.10 | 2.55 | 2.66 | 0.04 | 1.34 | 1.38 | 0.03 | 0.92 | 0.95 |
| nonpar sm-trim | 0.15 | 2.05 | 2.20 | 0.12 | 1.01 | 1.13 | 0.12 | 0.67 | 0.80 |
| nonpar corr-trim | 0.07 | 2.69 | 2.76 | 0.09 | 1.06 | 1.15 | 0.08 | 1.02 | 1.09 |

(b) Bates (2000) stochastic volatility-model.

| | $T = 60$ | | | $T = 120$ | | | $T = 240$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{m}$ | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE |
| exp-quart | 0.15 | 4.90 | 5.05 | 0.06 | 1.73 | 1.79 | 0.03 | 0.81 | 0.84 |
| nonpar | 0.51 | 4.26 | 4.77 | 0.28 | 2.30 | 2.58 | 0.21 | 1.72 | 1.93 |
| nonpar sm-trim | 0.73 | 2.57 | 3.30 | 0.59 | 1.23 | 1.82 | 0.77 | 0.58 | 1.35 |
| nonpar corr-trim | 0.06 | 1.75 | 1.81 | 0.01 | 1.06 | 1.07 | 0.01 | 0.58 | 0.60 |

Figure 1 shows the simulated mean and pointwise quantiles of the estimators $\hat{m}$ under the stochastic volatility model. The variance of all estimators increases substantially in the tails of the distribution due to the sparse observations there, and for the local estimators due to dividing by small $q_t(R_{t+1})$. This effect is most pronounced for the right tail, due to the left-skewed conditional densities of the Bates (2000) model. Nonetheless, with either form of trimming the local ratio estimators are less variable in the tails than the parametric estimator. While smooth trimming comes at the cost of increased biases in the tails, the boundary correction used for the range-trimmed estimator successfully reduces these. The same figure for the GARCH-$t$ model is not reported as it show similar patterns, except that its estimation errors are more equally balanced between left and right tail returns due to the model's symmetric conditional densities.

[4] A previous draft reports similar results for both pilot estimators, and also shows similar performance when using an iterative bandwidth procedure.



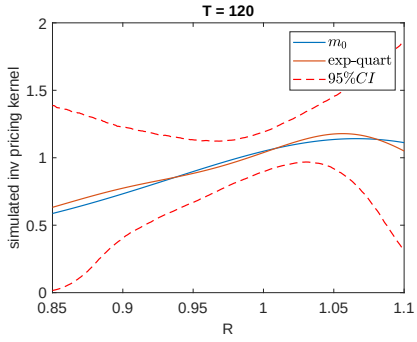

(a) Exponential-quartic estimator.



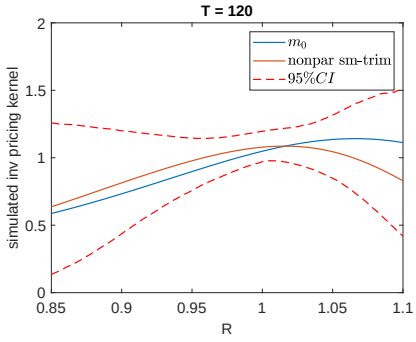

(b) Smoothly trimmed local estimator.



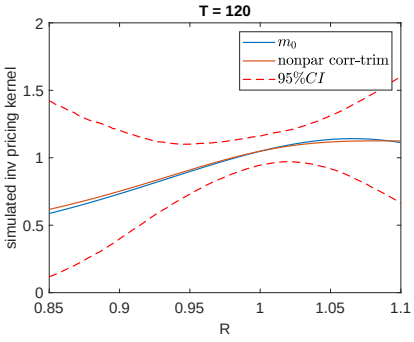

(c) Range-trimmed local estimator.



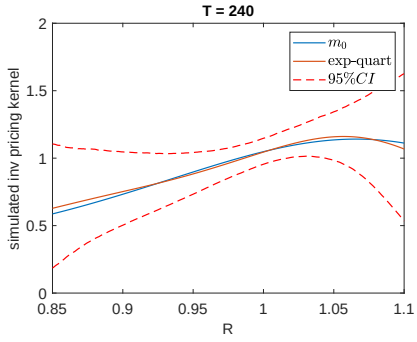

(d) Exponential-quartic estimator.



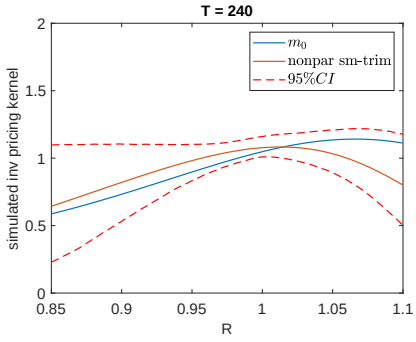

(e) Smoothly trimmed local estimator.



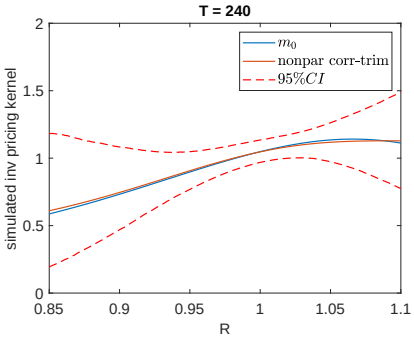

(f) Range-trimmed local estimator.




(g) Exponential-quartic estimator.




(h) Smoothly trimmed local estimator.




(i) Range-trimmed local estimator.

Figure 1: Simulated estimates $\hat{m}_{\text{sc}}$ of the inverse pricing kernel $m_0$ under the Bates (2000) stochastic volatility-model, together with simulated pointwise 95% CIs, based on $S = 10,000$ simulated time series. Local estimators use plug-in bandwidths and two types of trimming. Rows vary with $T = \{60, 120, 240\}$ months, columns with type of estimator.

Table 2 reports the simulated mean integrated squared density estimation error

$$\text{MISDE} = E\left(\int \left(\hat{f}_t(y) - f_t(y)\right)^2 dy\right)$$

of conditional density estimators $\hat{f}_t$ given by (8) for the different estimators $\hat{m}$. The integral is taken over the full support of return outcomes. The performance of the density estimators follows that of the pricing kernel estimators, with the nonparametric estimators outperforming

at $T = 60$ and being competitive at $T = 120$, and the parametric estimator outperforming at $T = 240$. Furthermore, the local estimators again favor some trimming, which particularly helps to stabilize the tails of the density estimates.

Table 2: Simulated mean integrated square density estimation error for various estimators under exponential-quartic pricing kernel, based on $S = 10,000$ simulated time series of densities and returns according to two models and sample sizes $T = \{60, 120, 240\}$ months. Nonparametric estimators use plug-in bandwidths based on the IMSE of $\hat{m}$ truncated at the (0.005,0.995)-unconditional quantiles.

(a) AR(1)-GARCH(1,1)-$t$-model.

| $\hat{m}$ | $T = 60$ | $T = 120$ | $T = 240$ |
|---|---|---|---|
| exp-quart | 13.45 | 5.85 | 2.47 |
| nonpar | 7.71 | 4.55 | 2.68 |
| nonpar sm-trim | 6.62 | 3.68 | 2.17 |
| nonpar corr-trim | 9.31 | 4.97 | 4.02 |

(b) Bates (2000) stochastic volatility-model.

| $\hat{m}$ | $T = 60$ | $T = 120$ | $T = 240$ |
|---|---|---|---|
| exp-quart | 15.17 | 6.45 | 2.85 |
| nonpar | 12.15 | 8.34 | 5.61 |
| nonpar sm-trim | 9.70 | 6.34 | 4.42 |
| nonpar corr-trim | 4.66 | 3.69 | 2.05 |

### 3.4 Results for noisy densities

Table 3 reports the simulation performance of the estimators in the presence of estimation error in the risk-neutral densities. Specifically, observed call option prices are simulated as $\tilde{C}_{it} = \max\{C_t(\kappa_{it}) + \varepsilon_{it}, 0\}$, where $C_t(\kappa)$ is the true call price at moneyness level $\kappa$ and $\varepsilon_{it}$ are i.i.d. Gaussian errors with standard deviation $\sigma_\varepsilon = 8.6 \times 10^{-5}$ equal to that of the pricing errors of the Bates (2000) model fitted to the data in the empirical application. The moneyness levels are set as $\kappa_{it} = Q_t^{-1}\left(\frac{i}{n_\kappa+1}\right)$ for $i = 1, \ldots, n_\kappa$. The densities $\tilde{q}_t$ are then obtained as the local cubic kernel estimator of the second derivative of the call pricing function using a plug-in bandwidth.

With just $n_\kappa = 50$ option prices, the density estimation error leads to starkly increased biases in all estimators $\hat{m}$, though particularly in the nonparametric estimators. This aligns with result (11) showing that the risk-neutral density estimation bias has an asymptotically larger impact than its variance. These biases only slightly weaken when the sample size increases, so that they exceed the variance when $T = 240$. Nonetheless, the range-trimmed estimator remains the best performing at all sample sizes, as it is the least affected by the noisy tails of the risk-neutral densities. With $n_\kappa = 100$ option prices, the trimmed estimators already perform similar as for the case of perfectly observed densities. Our empirical application uses on average 157 option prices to estimate the monthly risk-neutral densities, suggesting the cross-sectional smoothing error is small relative to the estimation error in $\hat{m}$. Still, even with fewer option prices, our nonparametric estimator is likely to perform well provided one uses a relatively small bandwidth

to estimate the risk-neutral densities and/or trims its tails.

Table 3: Simulation performance of various estimators under exponential-quartic pricing kernel, based on $S = 10,000$ simulated time series of *estimated* risk-neutral densities based on local cubic estimator with $n_\kappa$ option prices and returns generated according to the Bates (2000) stochastic volatility-model. Columns describe integrated weighted squared bias, variance, and mean squared error of scaled estimators $\hat{m}_{\text{sc}}$, truncated at the (0.005,0.995)-quantiles of the return distribution, for three numbers of months $T$.

(a) $n_\kappa = 50$.

| | $T = 60$ | | | $T = 120$ | | | $T = 240$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{m}$ | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE |
| exp-quart | 1.60 | 4.32 | 5.92 | 1.25 | 1.68 | 2.93 | 1.45 | 0.75 | 2.20 |
| nonpar | 2.28 | 3.84 | 6.12 | 1.69 | 2.29 | 3.98 | 1.73 | 1.70 | 3.42 |
| nonpar sm-trim | 2.45 | 2.73 | 5.18 | 1.97 | 1.51 | 3.48 | 2.16 | 0.84 | 3.00 |
| nonpar corr-trim | 1.03 | 0.20 | 1.23 | 0.50 | 0.31 | 0.81 | 0.42 | 0.21 | 0.63 |

(b) $n_\kappa = 100$.

| | $T = 60$ | | | $T = 120$ | | | $T = 240$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{m}$ | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE | IBias2 | IVar | IMSE |
| exp-quart | 0.19 | 4.64 | 4.83 | 0.10 | 1.70 | 1.80 | 0.08 | 0.79 | 0.86 |
| nonpar | 0.46 | 6.01 | 6.47 | 0.14 | 3.29 | 3.43 | 0.05 | 2.82 | 2.87 |
| nonpar sm-trim | 0.80 | 2.85 | 3.65 | 0.57 | 1.40 | 1.97 | 0.72 | 0.69 | 1.41 |
| nonpar corr-trim | 0.56 | 0.54 | 1.10 | 0.18 | 0.58 | 0.75 | 0.13 | 0.35 | 0.49 |

### 3.5 Results with volatility as covariate

This subsection analyzes the performance of the local exponential-linear estimator (13). We generate both physical and risk-neutral densities according to the single-factor stochastic volatility jump diffusion model in Bates (2000), estimated under an affine volatility risk premium constraint, and compute the true conditional pricing kernels as their ratio. While the model is affine under both measures, the pricing kernel is not an exact exponential-affine function of the conditioning volatility factor, as it is a projection on the return outcome. The conditional pricing kernels are typically downward sloping in the return outcome, but occasionally U-shaped.

We simulate the following estimators. First, we fit a exponential-polynomial model of the type (16), but now with coefficients $\gamma_l(V_t) = \sum_{k=0}^{K} a_{lk} V_t^k$ that are polynomial functions of volatility $V_t$, as considered in Schreindorfer and Sichert (2025). We consider the case with $L = K = 2$ powers of the log return and volatility, which creates a flexible model with six parameters that can reasonably approximate the various shapes of the conditional pricing kernel. The second

estimator is the static local ratio estimator (5) that simply ignores the volatility factor. The third estimator is the local exponential estimator (13) with $x_t = V_t$. We use plug-in versions of the bandwidths that minimize the asymptotic weighted IMSE of $\hat{m}(y,x)$, with $h_x = c_x h_y$:

$$h_y = \left( \frac{\iint \sigma_m^2 f dy dx}{T\mu_2^2(K) \iint (m_{yy}^2 + c_x^2 m_{yy}\tilde{m}_{xx}) f dy dx} \right)^{\frac{1}{6}} \tag{19}$$
$$c_x = \left( \frac{\iint m_{yy}^2 f dy dx}{\iint \tilde{m}_{xx}^2 f dy dx} \right)^{\frac{1}{4}},$$

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.

Table 4 reports the simulated accuracy of the pricing kernel estimators and their corresponding conditional densities. We consider only the two largest sample sizes from the previous section, as localization in the conditioning variable effectively implies using only a subset of the observations at a time. Both the parametric and nonparametric dynamic estimators outperform the static nonparametric estimator for samples of 120 and 240 months, which confirms that time-variation in the pricing kernel can be detected with realistic sample sizes. The local exponential estimator has the lowest IMSE and MISDE for both sample sizes, although the flexible parametric estimator improves more when the sample size increases due to its faster convergence. In line with Theorem 2, the bandwidths for the volatility dimension are relatively large, with their medians 30% and 15% larger than their rule-of-thumb counterparts for each sample size. This leads to slightly larger biases than the other estimators, but also the lowest variances, which help attain the lowest mean squared errors overall.

# 4 Empirical application

This section applies the local density ratio estimator to S&P 500 index returns using option price and trading data. After describing our data set, we first estimate the pricing kernel without conditioning variables, then with volatility as a covariate, and finally using net positions data based on a model with heterogeneous option traders.

Table 4: Simulation performance of various estimators of the conditional pricing kernels $m(\cdot, V_t)$ based on $S = 10,000$ simulated time series of physical and risk-neutral densities according to the Bates (2000) stochastic volatility model. Columns describe integrated weighted squared bias, variance, and mean squared error of scaled estimators $\hat{m}_{t,\text{sc}}$, and mean integrated squared density errors, for all periods with $V_t$ within the (0.025,0.975)-quantiles of its marginal distribution, for two numbers of months $T$.

| | $T = 120$ | | | | $T = 240$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\hat{m}$ | IBias2 | IVar | IMSE | MISDE | IBias2 | IVar | IMSE | MISDE |
| exp-biv-quad | 0.05 | 3.43 | 3.47 | 12.81 | 0.08 | 1.91 | 1.99 | 6.06 |
| nonpar static | 0.15 | 5.49 | 5.64 | 17.04 | 0.14 | 4.12 | 4.26 | 12.03 |
| nonpar cond exp | 0.30 | 2.15 | 2.46 | 8.91 | 0.40 | 1.38 | 1.77 | 3.98 |

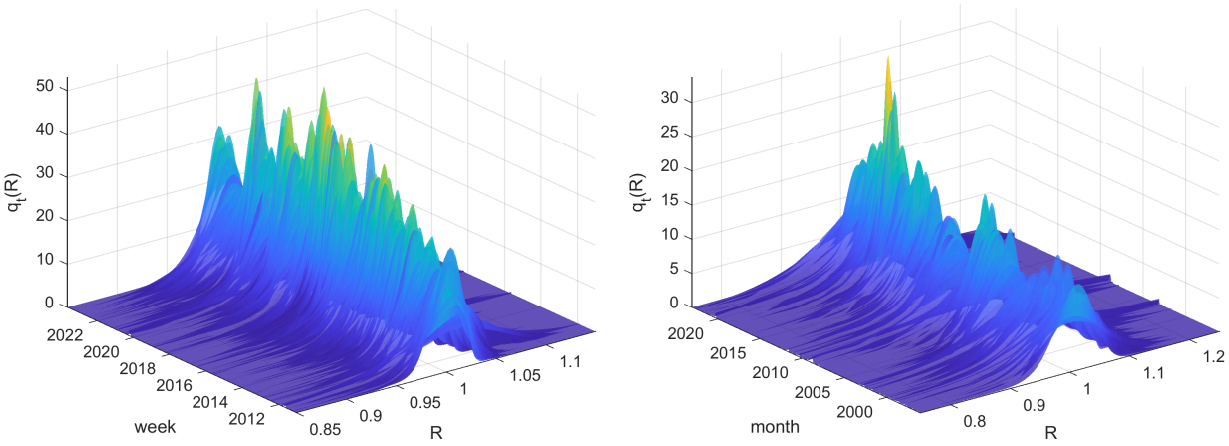
## 4.1 Data description

We construct our sample $(R_{t+1}, q_t(\cdot), x_t)_{t=1}^T$ of returns, risk-neutral densities, and conditioning variables, for both monthly and weekly frequencies, as follows. For the monthly data, we consider SPX option contracts in OptionMetrics that expire on the third Friday of each month between January 1996 and February 2023, and observe their prices on the last business day at least 28 days before expiration. For the weekly data, we consider SPX option contracts that expire on the Friday of each week in the period from January 2011 to August 2023, and observe their prices on the last business day at least seven days before expiration. We focus on contracts that are relatively close to expiration to ensure a large number of traded options per day. The resulting samples contain $T = 325$ monthly and $T = 605$ weekly non-overlapping observations. For each time $t$ in the sample we compute the realized S&P 500 index return $R_{t+1}$ over the option holding period. The risk-neutral density at time $t$ is 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 2 shows the resulting risk-neutral densities for both horizons are smooth, unimodal, and left-skewed.




(a) Weekly horizon.

(b) Monthly horizon.

Figure 2: 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.

Our application to demand-based pricing kernels in Section 4.4 uses measures of trade imbalances as the conditioning variable $x_t$. These net demand proxies are obtained from the CBOE OpenClose Volume data set, which contains daily trading positions in European option contracts on the S&P 500 index over the period 1996-2023. The trading positions are disaggregated by investor type (firm or customer) and three size categories. Observing the type of buyer or seller initiating each transaction allows tracking the direction of trades, rather than just trading volume, which helps identify the characteristics of heterogeneous investors in our model. We merge the positions data with the corresponding option prices obtained from OptionMetrics, as described in more detail in Section 4.4.2.

### 4.2 Time-invariant pricing kernels

First, we estimate a density ratio model without covariates. As an example of an economic model within this class, consider a representative investor who trades a stock with price $S_t$ and a risk-free bond with return $R_t^f$ in order to maximize his or her subjective expected utility $\tilde{E}_t\left(u(W_{t+1})\right)$ over wealth at period $t+1$. The bond is in zero net supply, so that in equilibrium $W_t = S_t$ for all $t$. Therefore, the first-order condition for investment implies that the price of a contingent claim with payoff $g(R)$ is given by (2) with stochastic discount factor

$$M_{t,t+1} = \frac{\lambda_t}{R_t^f} U'(S_{t+1}) \frac{\mathrm{d}\tilde{f}_t(S_{t+1})}{\mathrm{d}f_t(S_{t+1})},$$

with $\lambda_t$ the Lagrangian multiplier for the budget constraint. Moreover, suppose the investor has power utility with CRRA parameter $\gamma$, while the subjective and true densities are linked by the density ratio model for the stock return

$$\tilde{f}_t(R) = \frac{f_t(R)\theta(R)}{\int f_t(y)\theta(y)dy}, \tag{20}$$

where $\theta(\cdot)$ is a positive function describing incorrect beliefs. For example, Carr and Madan (2001) consider model (20) with $\theta(\cdot)$ a combination of a finite set of basis functions. Then the risk-neutral and true density are linked by density ratio model (4) with inverse pricing kernel

$$m(R) = \frac{R^\gamma}{\theta(R)}.$$

The presence of the incorrect belief term $\theta(R)$ has been offered to explain the puzzling finding of non-monotonic pricing kernels (e.g. Hens and Reichlin, 2013).[5]

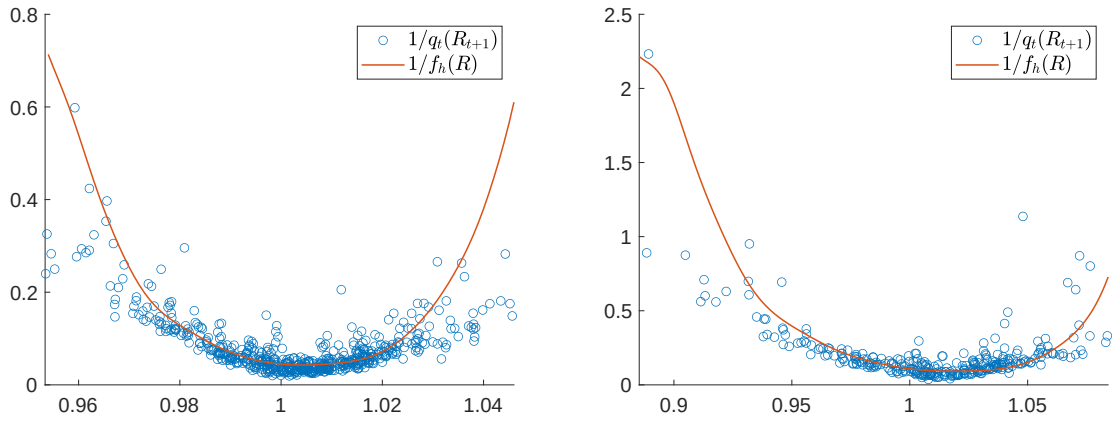
To help understand the local estimator of $m(R)$, Figure 3 plots the realized inverse densities and the inverse of the unconditional smoothed density for varying return levels. Estimator (5) takes kernel-weighted averages of $1/q_t(R_{t+1})$ locally around $R \approx y$. Meanwhile, when $q_t(R) = q(R, s_t)$ for some state variables $s_t$ with continuous density, then

$$\begin{aligned} E\left(\frac{1}{q_t(R_{t+1})} \mid R_{t+1} = y\right) &= \int \frac{1}{q(y \mid s)} f(s \mid R_{t+1} = y) ds \\ &= \int \frac{f(y \mid s)}{q(y \mid s)} \frac{f(s)}{f(y)} ds = \frac{m(y)}{f(y)}, \end{aligned}$$

using model equation (4) in the last step. Therefore, values $\frac{1}{q_t(R_{t+1})} > \frac{1}{f_h(R_{t+1})}$ for realizations of $R_{t+1}$ near $y$ indicate that $m(y) > 1$, and vice versa. Figure 3 shows no clear deviation between

[5] With rational beliefs, reference-dependent preferences provide an alternative explanation (Grith et al., 2017).

both inverse densities in the middle part of the return distribution, suggesting $m(y)$ is close to one there. Meanwhile, in the left tail $q_t(R_{t+1})$ tends to be larger than $f_h(R_{t+1})$ for both weekly and monthly horizons, suggesting $m(y)$ is below one in this region, although the variability of $1/q_t(R_{t+1})$ also increases in the tails.




(a) Weekly data.

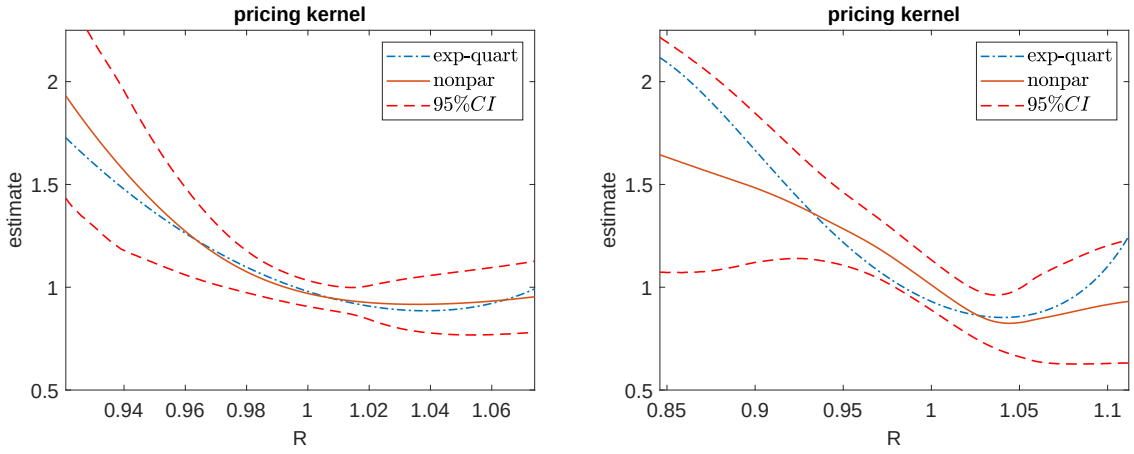
(b) Monthly data.

Figure 3: Scatterplots of $1/q_t(R_{t+1})$ against the weekly and monthly return on the S&P 500 index from January 1996 to August 2023, together with plots of the inverse unconditional smoothed return density $f_h(R)$ based on Silverman's rule-of-thumb bandwidth.

Figure 4 shows the pricing kernel estimates $1/\hat{m}_h$ based on the boundary-corrected trimmed estimator (18). Their local confidence intervals are obtained by plugging $\hat{m}_h$ into the asymptotic variance formula from Theorem 1, and using the delta-method to compute standard errors for its reciprocal. The bandwidths are chosen using the plug-in method from Section 3.1, and are typically several times larger than those used for estimating the risk-neutral densities. For both weekly and monthly horizons, the pricing kernel estimates are monotonically decreasing for negative and small positive returns. However, they flatten out and then slightly increase in the right tail of the return distribution, although we cannot confidently conclude that the pricing kernel is strictly increasing for large positive returns due to their rare occurrence. Overall, our kernel-based estimates confirm the non-decreasing right tail is a robust feature of pricing kernels, and not merely due to misspecification of the physical densities as Linn et al. (2017) suggest.[6] Even when the pricing kernel depends on latent state variables, as suggested by Chabi-Yo et al. (2008), our estimates imply that the conditional pricing kernel must be non-decreasing for some values of the state, as discussed in Section 2.2. Economically, U-shaped pricing kernels are commonly explained using variance risk aversion (Christoffersen et al., 2013). However, Heston et al.

[6]This finding aligns with the formal test results for pricing kernel non-monotonicity in Beare and Schmidt (2016).

(2024) recently estimate that variance risk aversion is not strong enough to produce a U-shape in a single-factor volatility model, which aligns with our findings of a mostly flat right tail. Finally, the plots show that the exponential-quartic fit, which we used for the simulation study, lies within our pointwise confidence bands. Nonetheless, it yields a slightly stronger U-shape for the monthly data, which may be due to parametric misspecification.




(a) Estimated pricing kernel for $\tau = 7$.

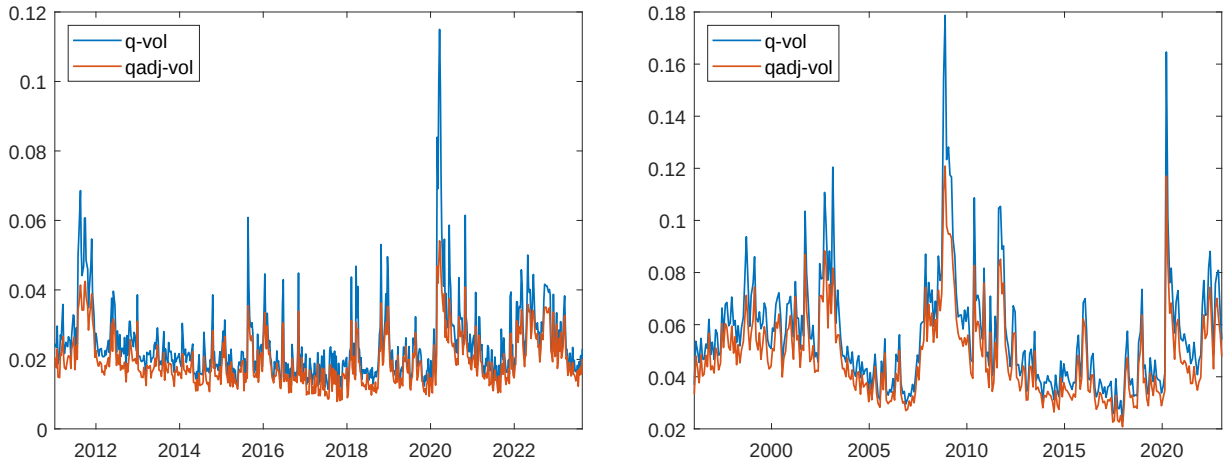
(b) Estimated pricing kernel for $\tau = 28$.

Figure 4: Local pricing kernel estimators $1/\hat{m}_h$ using weekly (left panel) and monthly (right panel) data on S&P 500 index options and returns from January 1996 to August 2023, based on plug-in IMSE optimal bandwidths based on a pilot estimator and boundary-corrected trimming with thresholds $\kappa_{tl} = Q_t^{-1}(0.01)$ and $\kappa_{tu} = Q_t^{-1}(0.99)$. Horizontal axes show 1% to 99% percentile of returns. CIs based on plug-in standard errors and delta-method.

We estimate the physical densities by multiplying the risk-neutral densities by the inverse pricing kernel, whose hump-shape dampens the probability mass in the left tails and lifts it in the center. The resulting conditional volatilities shown in Figure 5 are thus adjusted downwards, especially when risk-neutral volatility is high, and therefore fluctuate less over time. More generally, the likelihood of left-tail risks is lower and more stable under the physical than under the risk-neutral distributions. Section 4.5 tests whether the adjusted option-implied second moments are conditionally unbiased forecasts of realized squared returns.

### 4.3 Volatility-dependent pricing kernels

Volatility plays a key role in models for the risk-neutral and physical densities of returns. Moreover, several recent studies have found that the slope of their ratio, the pricing kernel, fluctuates with proxies for volatility (e.g. Kim, 2023; Schreindorfer and Sichert, 2025). Therefore we apply our local exponential estimator (13) with $x_t = \log \mathrm{VIX}_t$ as the conditioning variable, where the logarithm serves to create a closer to Gaussian conditioning variable with unbounded support.




(a) $\tau = 7$.

(b) $\tau = 28$.

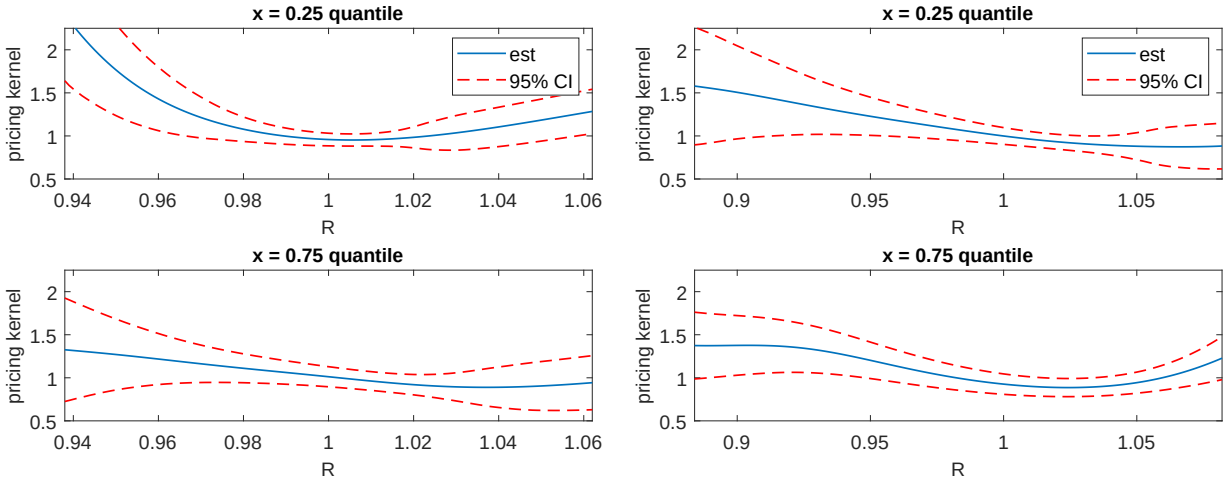
Figure 5: Original and multiplicatively corrected option-implied volatilities for the weekly and monthly return on the S&P 500 from January 1996 to August 2023. Local pricing kernel estimated with variable bandwidths $\hat{h}^{MSE}$.

While the VIX$^2$ measures the risk-neutral expected variance over the next 30 days, in most single-factor stochastic volatility models it is a monotonic transformation of spot volatility. In such models, our method produces the same pricing kernels as when conditioned on the latent spot volatility factor itself.

Figure 6 shows the estimated pricing kernels when the VIX is at its 25% and 75% quantiles. We corroborate the finding by Kim (2023) and Schreindorfer and Sichert (2025) that lower volatility is associated with steeper pricing kernels for left tail risks, especially at the weekly horizon. Therefore the positive impact of higher volatility on the equity risk premium is at least partially offset via a lower price of risk. However, quantitatively, our estimated impact of the VIX on the monthly pricing kernel is weaker than in Schreindorfer and Sichert (2025), and more in line with the moderate impacts obtained by Song and Xiu (2016), Linn et al. (2017), Barone-Adesi et al. (2020), and Kim (2023) using various nonparametric methods. Moreover, we find that volatility affects whether the pricing kernel is decreasing or U-shaped, which is not possible with the single-parameter specification of volatility-dependence in Schreindorfer and Sichert (2025).

### 4.4 Demand-based pricing kernels

This section formulates and estimates a density ratio model for the impact of net demand for options on the pricing kernel. This impact is theoretically underpinned by market-makers facing unhedgeable positions (Garleanu et al., 2008; Chen et al., 2019) and/or inventory constraints




(a) Weekly data.

(b) Monthly data.

Figure 6: Local exponential-linear estimated pricing kernels conditional on the VIX at low (25%) and high (75%) quantiles, using boundary-corrected trimming and plug-in bandwidths based on an initial bivariate Nadaray-Watson estimator. Horizontal axes show inner 98% of weekly and 95% of monthly returns. CIs based on plug-in standard errors and delta-method.

(Fournier and Jacobs, 2020), in which case option prices cannot be pinned down by no-arbitrage alone but reflect the preferences and beliefs of option traders. Therefore, we develop a stylized heterogeneous-agent equilibrium model that connects the subjective, physical, and risk-neutral densities, and allows interpreting the demand-based pricing kernels that we estimate locally.

### 4.4.1 Heterogeneous investor model

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

$$\phi_{it}(R) = \mu_{it} + \frac{1}{\gamma_i} \log \frac{f_{it}(R)}{q_t(R)}, \qquad i = 1, \ldots, N, \tag{21}$$

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

implies that the Arrow-Debreu state-price density equals

$$q_t(R) = c_t \exp\left(-\tilde{\gamma} S_t R + \sum_i \frac{\tilde{\gamma}}{\gamma_i} \log f_{it}(R)\right), \tag{22}$$

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

$$\log f_{it}(R) \propto \sum_{i=1}^K s_{itk} g_k(R), \tag{23}$$

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 (22) and (23) into (21) yields the equilibrium net derivative payoff functions $\omega_{it}(R)$ as

$$\omega_{it}(R) = \text{const}_{it} + \left(\frac{\tilde{\gamma}}{\gamma_i} - \omega_{it}^S\right) S_t R + \frac{1}{\gamma_i} \sum_{k=1}^K \tilde{s}_{itk} g_k(R), \tag{24}$$

where $\omega_{it}^S$ 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 $(1, R, g_1(R), \ldots, g_K(R))$.[7]

Characterizing the pricing kernel requires connecting the subjective densities to the true conditional density. Suppose $f_t$ has the same form as (23), but the true loadings are weighted averages of the subjective ones:

$$\log f_t(R) \propto \sum_k s_{tk} g_k(R), \quad s_{tk} = \sum_i \theta_{ik} s_{itk}, \quad \sum_i \theta_{ik} = 1 \text{ for all } k. \tag{25}$$

Combining (22), (23), and (25) yields the density ratio model

$$\log f_t(R) \propto \log q_t(R) + \tilde{\gamma} S_t R + \sum_{i,k} \theta_{ik} \tilde{s}_{itk} g_k(R), \tag{26}$$

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

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

inverse pricing kernel can be written in terms of the parameters $\beta_{ik} = \theta_{ik}\gamma_i - \theta_{1k}\gamma_1$ as

$$m(R, S_t, \tilde{x}_t) = \exp\left(\gamma S_t R + \sum_{i=2}^{N}\sum_{k=1}^{K} \beta_{ik}\tilde{x}_{it} g_K(R)\right). \tag{27}$$

The informativeness ($\theta_{ik}$) and risk aversion parameters ($\gamma_i$) are not separately identified from ($\beta_{ik}$). 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 (27) 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.[8]

### 4.4.2 Option net demand curves

We construct the empirical counterpart to $\omega_{it}$ using each investor's net number of contracts $(d^C_{ijt}, d^P_{ijt}, K_j)_{j=1}^{n_t}$ in call and put options with strike prices $K_j$ and a common maturity date. For simplicity, we suppose there are only $N = 2$ investors, namely market-makers and end-users, where the latter combines investors labeled as firms or customers (but not broker-dealers) by the CBOE. They take opposite positions, so that we drop the $i$ subscript. At current holdings, end-users' net payoff in the event $R_{t+1} = R$ is the sum of their net call and put option payoffs:

$$\omega_t(R) = \omega^C_t(R) + \omega^P_t(R), \quad \omega^C_t(R) = \sum_{j=1}^{n_t} d^C_{jt}(R - \kappa_{jt})^+, \quad \omega^P_t(R) = \sum_{j=1}^{n_t} d^P_{jt}(\kappa_{jt} - R)^+, \tag{28}$$

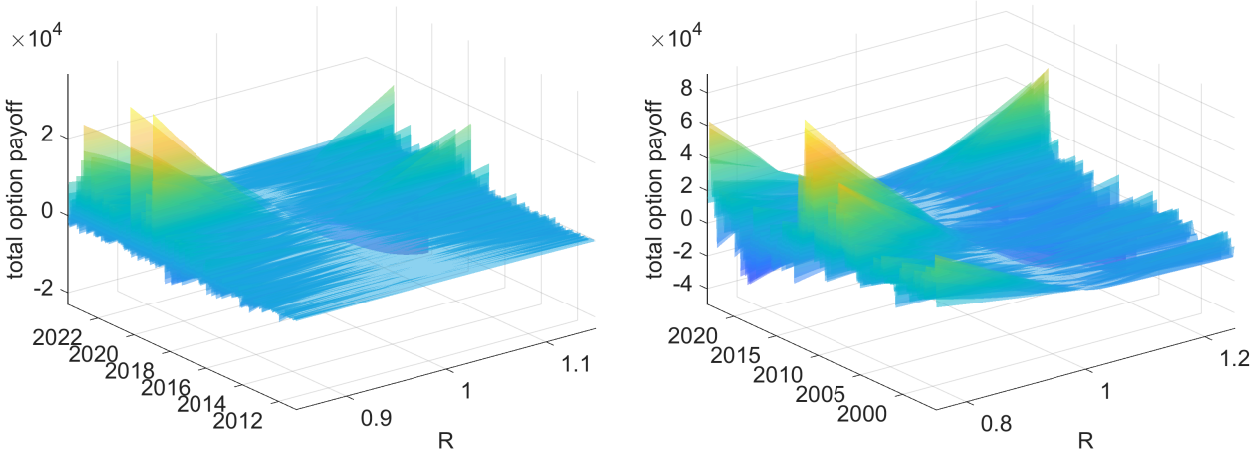
where $\kappa_{jt} = \frac{K_j}{F_t}$ are the forward moneyness levels. The payoff function $\omega_t(R)$ can be interpreted as the equilibrium net demand by end-users for Arrow-Debreu type securities that pay off in case $R_{t+1} = R$. The net number of contracts for each listed strike $K_j$ and put-call type is computed by accumulating daily net open positions $d_{js} = \text{BuyOpen}_s(K_j) - \text{SellOpen}_s(K_j)$ into $\bar{d}_{jt} = \sum_{t-l\le s\le t} d_{js}$, where $l = 21$ and $l = 60$ calendar days for weekly and monthly demand, respectively.[9] Following Chen et al. (2019) and others, we focus on open orders since close orders

[8] 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.

[9] Using a cut-off prevents 'stale' positions from dominating the demand curves. Alternative cut-offs of up to three months yield similar findings.

are mechanically influenced by existing positions and may not reflect the latest beliefs of investors. We then match the accumulated demands $\bar{d}_{jt}$ to the moneyness levels $\kappa_{jt}$ and compute the net demand curves $\omega_t(R)$ using (28).

Figure 7 shows the resulting weekly and monthly net option payoff functions. To abstract from level shifts that amount to risk-free payoffs, the plots show the net payoffs $\omega_t^c(R) = \omega_t(R) - P_t(\omega_t)$ centered by the market value of end-user positions $P_t(\omega_t) = \int_0^\infty \omega_t(y) q_t(y) dy$. By construction, the price of the payoff function $\omega_t^c(R)$ is zero. The monthly curves show end-users built up large net long positions in put options before the 2008-2009 financial crisis, but switched to net short positions during the peak and aftermath of the crisis. Similarly, weekly net end-user positions in large negative returns were highly positive around the onset of the COVID-19 pandemic in March 2020, but turned negative in the subsequent months.




(a) Net weekly positions.

(b) Net monthly positions.

Figure 7: Weekly and monthly total net end-user payoffs for the event $R_{t+1} = R$ as traded in the three weeks and two months preceding the option observation date, respectively, centered to mean zero, for the sample period 1996-2023.

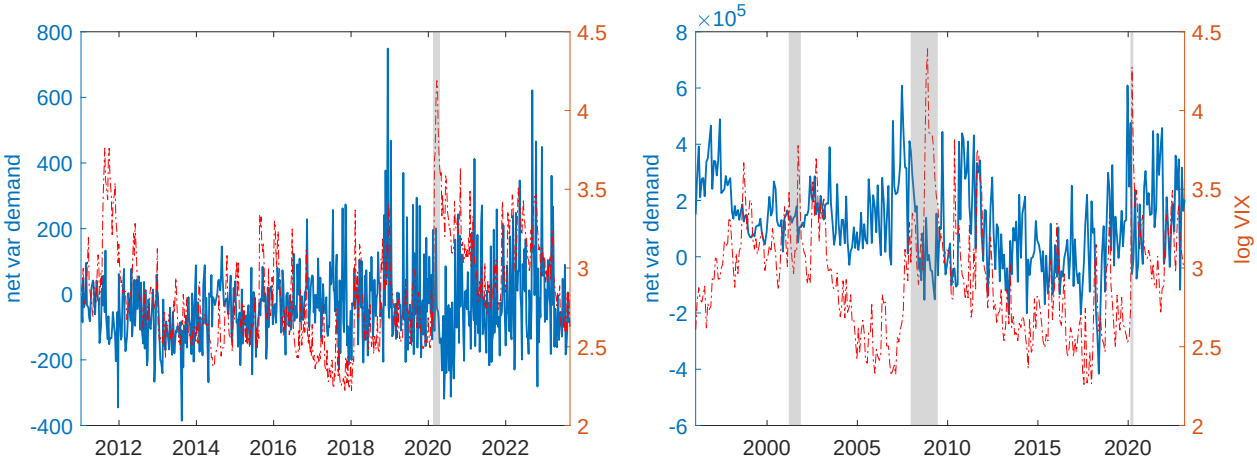
To construct the covariates in density ratio model (26), we estimate the coefficients in (24) for some basis functions $(g_k)_{k=1}^K$ by minimizing the weighted integrated least squares criterion

$$\min_b \int \left(\omega_t(y) - b^T g(y)\right)^2 q_t(y) dy,$$

which has solution

$$\hat{b}_t = \left(\int g(y) g(y)^T q_t(y) dy\right)^{-1} \int g(y)^T \omega_t(y) q_t(y) dy.$$

Setting the basis functions $g(r) = (1, r-1, (r-1)^2)$, we proxy net demand for 'directional' and 'variance risk' by their fitted coefficients $\hat{b}_t$. Since directional positions can be easier taken in the equity index directly, we focus on the net variance demand coefficient $\hat{b}_{t2}$. The demand curves in Figure 7 show that option trading volume has increased over time, in particular for short, weekly horizons. Therefore we normalize the weekly net demand variables by the square root of option trading volume during the three weeks before each observation date $t$. Figure 8 plots the resulting net variance demand proxies over time. End-users switched from buying protection against high squared returns to selling it during the peak of the financial crisis, when the price of such protection rose dramatically. However, after 2010 end-users returned to holding long variance positions. The same dynamic occured with a faster pace during the peak of the COVID-19 pandemic, as seen from both the weekly and monthly net demand variables. Interestingly, monthly net demand moved mostly opposite to the VIX during the first half of the sample, yet shared a similar downward trend during the 2010s, leading to a slightly positive overall sample correlation with the log VIX of 0.08. Weekly net demand is more erratic, and also has a rather small correlation with the log VIX of 0.07.




(a) Weekly data.

(b) Monthly data.

Figure 8: Net end-user demand (full line) for squared returns, accumulating positions over the three weeks and two months preceding the option observation date, respectively, against the log VIX (dash-dotted line), for the sample period 1996-2023. Weekly net demand is normalized by the square root of option trading volume.

### 4.4.3 Estimates

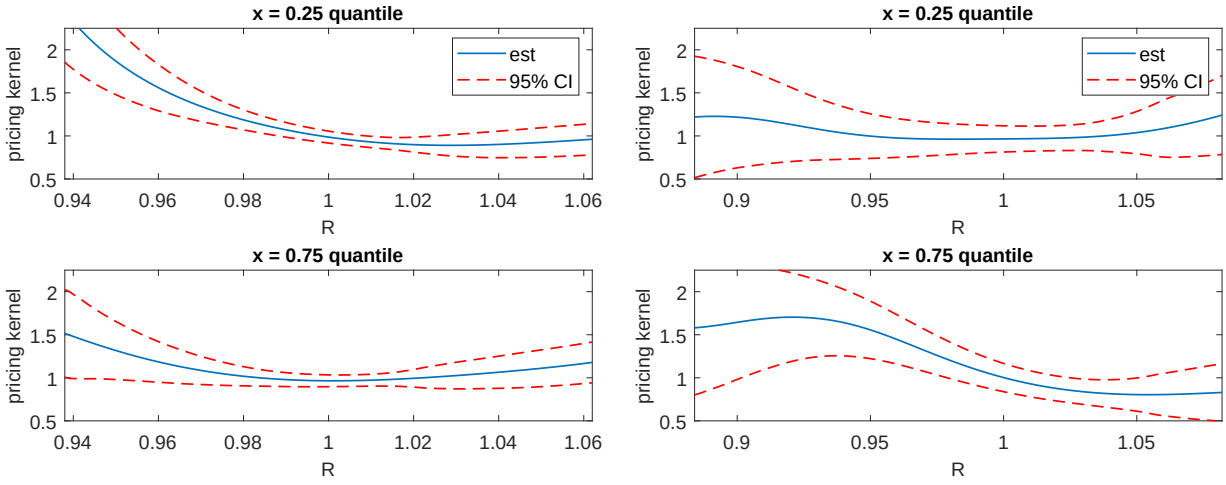
We compute the local exponential-linear estimator (13) with the net variance demand proxy $x_t = \hat{b}_{t2}$ as covariate, and plug-in bandwidths based on (19) and a Nadaraya-Watson pilot estimator.

Figure 9 shows the estimated conditional pricing kernels for low and high values of net variance demand, as defined by its 25% and 75% percentiles. For weekly data, low net variance demand leads to a significantly higher pricing kernel in the left tail than high net variance demand, though it has little effect on the right tail. For monthly data, the pricing kernel given low net variance demand is mostly flat but has a slight U-shape, which implies the overpricing of insurance of large shocks, regardless of their sign. Meanwhile for high net demand, the pricing kernel is decreasing in the middle region while flattening out in the tails. For both horizons, low net variance demand is thus associated with a higher convexity of the pricing kernel. Through the lens of model (27), this finding corresponds to a positive coefficient $\beta_{22}$, implying that end-users are more risk averse ($\gamma_2 > \gamma_1$) and/or better informed ($\theta_{22} > \theta_{12}$) about tail risks than market-makers. Intuitively, when market-makers initiated long positions in tail risks during the financial crises in order to hedge their portfolios, more risk-averse end-users demanded a large premium to take the opposite position, causing the U-shaped pricing kernel. This result aligns with the finding in Almeida and Freire (2022) that net short positions in out-of-the-money (OTM) put and call options is associated with higher respective tail risk premia. Quantitatively, the pricing kernel impact of our demand proxy appears weaker than that inferred through option returns in Almeida and Freire (2022). This might be the result of combining demand for OTM put and call options into a single proxy. For example, negative net variance demand during the low volatility mid 2010s is primarily driven by short positions in OTM puts rather than OTM calls, which affect the pricing of right tail risks more directly.[10] Therefore, the next subsection revisits the evidence for a U-shaped demand-based pricing kernel by also controlling for volatility.

### 4.5 Conditional moment test results

We now report the results of the conditional moment-based specification test described in Section 2.6. For the test function $g(R)$, we consider the linear, absolute, and squared value of the return, as well as put and call option payoffs. These functions allow detecting dynamic misspecification of the slope, V-shape, convexity, and tail values of the pricing kernel, respectively. 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 $R^s_{t+1} = \hat{F}_t^{-1}(U)$, with $U$ the standard uniform, and compute the simulated test statistic $J^s_T(g)$ based on $(R^s_{t+1}, q_t, x_t)_{t=1}^T$. As the deep tails of the pricing kernel cannot

[10] Moreover, we only consider positions initiated before the monthly option observation date, rather than up to fifteen days to expiration as in Almeida and Freire (2022).




(a) Weekly data.

(b) Monthly data.

Figure 9: Local exponential-linear estimated pricing kernels conditional on net volatility demand at low (25%) and high (75%) percentiles, using fixed plug-in bandwidths, for weekly data from 2011-2023 and monthly data from 1996-2023. Horizontal axes show inner 98% of weekly and 95% of monthly returns. CIs based on plug-in standard errors and delta-method for the pricing kernel.

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.

Table 5 shows the results for the case without covariates, which test whether the pricing kernel is constant over time. As possibly omitted variables we consider the VIX, net variance demand with two standardizations, and the risk-neutral conditional moment of each specific payoff function. At the monthly horizon, the VIX and the risk-neutral moments highly significantly predict the prediction errors in absolute and squared returns and call option payoffs, suggesting they impact the right tail of the pricing kernel in particular. The demand proxies also significantly predict the call option payoff residuals, at both horizons. Furthermore, at the weekly horizon, they significantly predict the return residuals. This suggests that net demand changes the slope of the weekly pricing kernel, which aligns with Figure 9a. While the lack of predictability of put option payoff residuals could be interpreted as a time-invariant left tail of the pricing kernel, it is partly due to the highly fat-tailed put payoffs increasing the standard error of the test statistic.

Table 6 shows the specification test results when the log VIX is included in the pricing kernel. The test statistics and their bootstrapped values use the bivariate Nadaraya-Watson estimator

Table 5: Conditional moment test results with $x_t = \{\}$. Table shows values of $J_T(g)$ with bootstrapped $p$-values based on $S = 5,000$ draws in brackets underneath. Rows show test functions $g$, columns possibly omitted variables. Test functions are trimmed at the 1% of 99% percentiles of returns. q-mom refers to corresponding risk-neutral moments. Dem-org refers to unstandardized net variance demand, dem-vol and dem-vol3 to demand standardized by the square root of volume during the last $l$ and $3l$ days, respectively, where $l = 21$ and $l = 60$ for weekly and monthly data. Bandwidth set as $h'_z = 2\hat{\sigma}(z_t)T^{-1/4}$.

(a) Weekly data.

| $\tau = 7$ | vix | q-mom | dem-vol | dem-vol3 |
|---|---|---|---|---|
| $R$ | -1.20 | -0.80 | 3.97 | 3.82 |
| | (0.98) | (0.79) | (0.02) | (0.02) |
| $\|R\|$ | 2.04 | 0.59 | -0.60 | -0.29 |
| | (0.02) | (0.14) | (0.66) | (0.45) |
| $R^2$ | 1.72 | 0.57 | -0.99 | -0.77 |
| | (0.04) | (0.17) | (0.92) | (0.80) |
| $(\kappa - R)^+$ | 0.12 | -0.35 | 0.93 | 1.04 |
| | (0.31) | (0.51) | (0.13) | (0.13) |
| $(R - \kappa)^+$ | -0.38 | -0.43 | 2.15 | 2.30 |
| | (0.50) | (0.52) | (0.03) | (0.03) |

(b) Monthly data.

| $\tau = 28$ | vix | q-mom | dem-org | dem-vol |
|---|---|---|---|---|
| $R$ | 1.21 | -0.01 | -0.18 | -0.33 |
| | (0.11) | (0.33) | (0.45) | (0.49) |
| $\|R\|$ | 2.83 | 2.66 | -0.47 | -0.43 |
| | (0.01) | (0.01) | (0.53) | (0.52) |
| $R^2$ | 4.34 | 5.52 | 0.62 | 0.70 |
| | (0.01) | (0.01) | (0.17) | (0.16) |
| $(\kappa - R)^+$ | -0.06 | -0.71 | -1.12 | -1.01 |
| | (0.39) | (0.69) | (0.95) | (0.90) |
| $(R - \kappa)^+$ | 6.84 | 10.02 | 2.31 | 2.60 |
| | (0.00) | (0.00) | (0.03) | (0.02) |

(12) for $\hat{m}(y, x_t)$, as it is faster to compute than the local exponential estimator. The risk-neutral moments are no longer significant at the monthly horizon, suggesting that the VIX is an adequate summary of the information in the risk-neutral density entering the pricing kernel. Interestingly, the net demand proxies no longer significantly predict the residuals of the call option payoff, but significantly predict those of the put option payoff, at both horizons. Moreover, net demand still predicts the return residuals at the weekly horizon.

To understand the impact of net demand after controlling for volatility, Figures 10 and 11 show estimated pricing kernels conditional on low and high net demand, for the subsamples with low and high values of the VIX. At the weekly horizon, low net demand leads to increased prices for left tail risks, for both low and high volatility periods. Meanwhile, at the monthly horizon, the impact of net demand on the left tail of the pricing kernel depends on the volatility level. In times of high volatility, low demand increases the price of left tail risk. This results in a U-shaped pricing kernel and occurs mostly during the recession dates. When volatility is low, high demand increases the price of left tail risk. This creates a monotonically decreasing pricing kernel, and occurs mostly during the optimistic 90s and mid 2000s. Meanwhile, the right tail of the pricing kernel is upward sloping given high volatility regardless of net demand, yet is slightly lifted by low net demand given low volatility. The impact of demand on the right tail of the pricing kernel

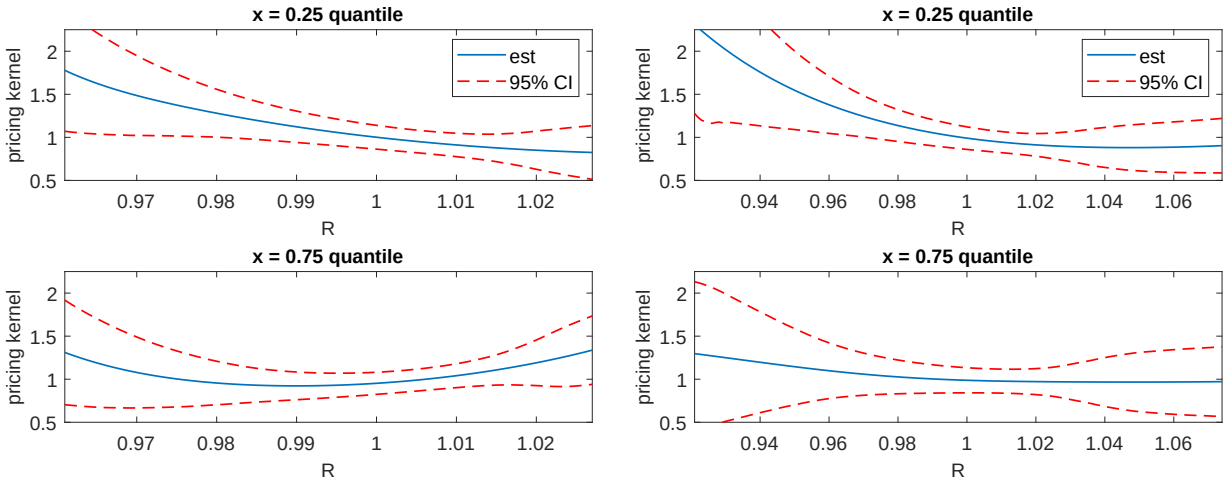
Table 6: Conditional moment test results with $x_t = \log \text{VIX}_t$ as included covariate. Table shows values of $J_T(g)$ with bootstrapped $p$-values in brackets underneath, based on leave-one-out version $\hat{m}_{-t}(y, x_t)$ of (12) with fixed plug-in bandwidths. Test bandwidths set as $(h'_x, h'_z) = 2(\hat{\sigma}(x_t), \hat{\sigma}(z_t))T^{-1/4}$. For variable descriptions see Table 5.

(a) Weekly data.

| $\tau = 7$ | q-mom | dem-vol | dem-vol3 |
|---|---|---|---|
| $R$ | -0.83 | 3.03 | 2.45 |
| | (0.68) | (0.00) | (0.01) |
| $\|R\|$ | -0.97 | -1.57 | -1.31 |
| | (0.83) | (1.00) | (0.98) |
| $R^2$ | -1.21 | -1.04 | -0.57 |
| | (0.92) | (0.86) | (0.62) |
| $(\kappa - R)^+$ | -0.92 | 1.20 | 1.69 |
| | (0.77) | (0.07) | (0.04) |
| $(R - \kappa)^+$ | 0.19 | 0.28 | -0.01 |
| | (0.16) | (0.20) | (0.29) |

(b) Monthly data.

| $\tau = 28$ | q-mom | dem-org | dem-vol |
|---|---|---|---|
| $R$ | -0.78 | -0.15 | 0.31 |
| | (0.65) | (0.32) | (0.20) |
| $\|R\|$ | -0.06 | -0.56 | -1.11 |
| | (0.29) | (0.61) | (0.88) |
| $R^2$ | -0.52 | 0.25 | -0.63 |
| | (0.52) | (0.23) | (0.64) |
| $(\kappa - R)^+$ | -0.85 | 1.50 | 1.06 |
| | (0.66) | (0.04) | (0.07) |
| $(R - \kappa)^+$ | -0.80 | -0.29 | -0.71 |
| | (0.76) | (0.44) | (0.67) |

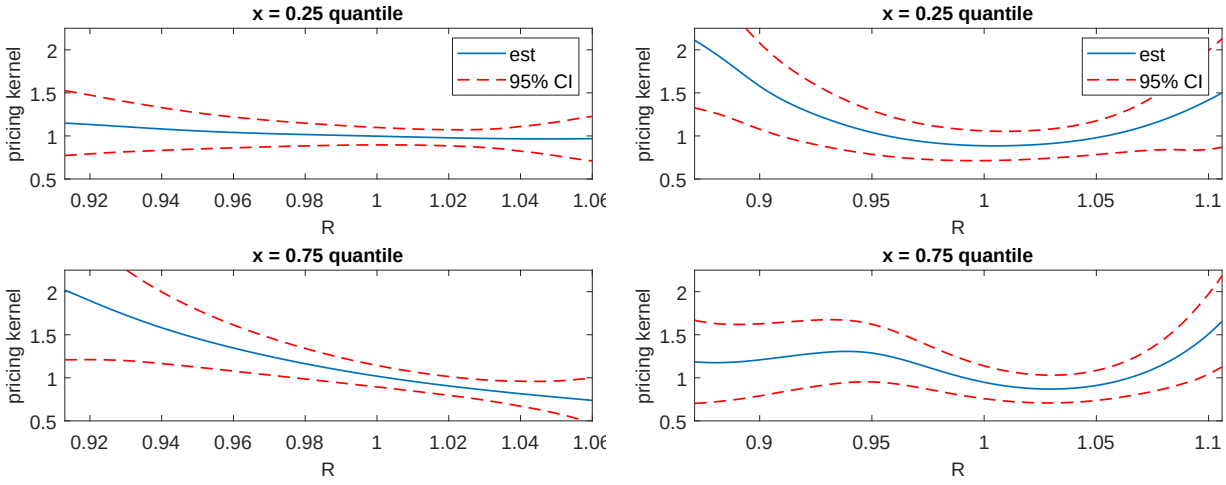



(a) Below median VIX.

(b) Above median VIX.

Figure 10: Estimated pricing kernels using weekly data conditional on net volatility demand at low (25%) and high (75%) quantiles, given the VIX below or above its median.

in Figure 9b is thus due to low volatility episodes. Together, the estimates explain why in Table 6 net demand has predictive ability for put, rather than call, option payoffs, once volatility is controlled for.




(a) Below median VIX.

(b) Above median VIX.

Figure 11: Estimated pricing kernels using monthly data conditional on net volatility demand at low (25%) and high (75%) quantiles, given the VIX below or above its median.

# 5 Conclusion

This paper propose a simple kernel-smoothing estimator, based on inverse density weighting, for the ratio between observed and true conditional densities in a general class of dynamic models. We establish its asymptotic bias, variance, and Normality, in settings with and without conditioning variables, as well as that of the multiplicatively corrected density and moment forecasts. Our simulation study reports good performance of the estimator relative to the correctly specified parametric maximum likelihood estimator, and robustness against noise in the observed densities. We apply the estimator to extract conditional pricing kernels from option-implied risk-neutral densities for index returns. We find that negative net variance demand and high volatility create a U-shaped pricing kernel for monthly returns, suggesting substantial overpricing of tail risk protection during times of financial market distress.

# A Appendix

## A.1 Proofs of results

*Proof of Theorem 1.* Given stationarity, the variance of the estimator equals

$$\operatorname{Var}\left(\widehat{m(y)}\right) = \frac{1}{T}\operatorname{Var}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right) + \frac{2}{T}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\operatorname{Cov}\left(\frac{K_h(R_{j+1}-y)}{q_j(R_{j+1})}, \frac{K_h(R_1-y)}{q_0(R_1)}\right).$$

The variance term equals

$$\begin{aligned}\frac{1}{T}\text{Var}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right) &= \frac{1}{T}E\left(\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right)^2\right) - \frac{1}{T}\left(E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right)\right)^2 \\ &= \frac{1}{T}E\left(\int\left(\frac{K_h(R-y)}{q_t(R)}\right)^2 f_t(R)dR\right) + O\left(\frac{1}{T}\right) \\ &= \frac{1}{T}\int E\left(\frac{c_t}{q_t(R)}\right)K_h^2(R-y)m(R)dR + O\left(\frac{1}{T}\right) \\ &= \frac{1}{Th}R(K)m(y)E\left(\frac{c_t}{q_t(y)}\right) + o\left(\frac{1}{Th}\right).\end{aligned}$$

The covariance terms reflect temporal dependence induced by predictability of the normalization constants $c_t$. For $t > s$ they can be written as

$$\text{Cov}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}, \frac{K_h(R_{s+1}-y)}{q_s(R_{s+1})}\right) = \int K_h(R-y)m(R)dR\text{Cov}\left(c_t, \frac{K_h(R_{s+1}-y)}{q_s(R_{s+1})}\right).$$

By Davydov's inequality for strong mixing processes

$$\text{Cov}\left(c_j, \frac{K_h(R_1-y)}{q_0(R_1)}\right) \leq 8\alpha(j)^{\frac{\delta}{4+2\delta}}E(c_t^{2+\delta})^{\frac{1}{2+\delta}}E\left(\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right)^2\right)^{\frac{1}{2}}.$$

Therefore, for some constant $C > 0$,

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|\text{Cov}\left(\frac{K_h(R_{j+1}-y)}{q_j(R_{j+1})}, \frac{K_h(R_1-y)}{q_0(R_1)}\right)\right| \leq \frac{1}{T}\sum_{j=1}^{T-1}C\alpha(j)^{\frac{\delta}{4+2\delta}}\frac{1}{\sqrt{h}} = O\left(\frac{1}{T\sqrt{h}}\right). \tag{29}$$

The total covariance term is thus a factor $\sqrt{h}$ smaller than the variance term, and vanishes asymptotically.

For asymptotic normality, we employ the large and small blocks argument (e.g. Fan and Yao, 2003, Thm 2.22). Let $Z_t = \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\right)$, $Z_{t,h} = \sqrt{h}Z_t$, and write

$$\begin{aligned}\sqrt{Th}\left(\hat{m}(y) - E\left(\hat{m}(y)\right)\right) &= \frac{1}{\sqrt{T}}\left(\sum_{i=1}^{k_T}\xi_i + \sum_{i=1}^{k_T}\eta_i + \zeta_T\right) \\ &\equiv m_T^l + m_T^s + \xi_T,\end{aligned}$$

where the large and small blocks $(\xi_i)$ and $(\eta_i)$ alternate in summing $l_T$ and $s_T$, respectively, consecutive periods of $Z_{t,h}$, $k_T = \lfloor\frac{T}{l_T+s_T}\rfloor$ is the number of blocks of each type, and $\zeta_T$ sums the

remaining periods. The block sizes $l_T$ and $s_T$ should grow to infinity such that

$$s_T/l_T \to 0, \quad l_T/T \to 0, \quad \frac{T}{l_T}\alpha(s_T) \to 0, \quad l_T/\sqrt{Th} \to 0 \tag{30}$$

It can be verified that $l_T = \sqrt{Th}/\log T$ and $s_T = \left(\sqrt{T/h}\log T\right)^{\frac{\delta}{4+2\delta}}$ satisfy these conditions. In particular,

$$\frac{s_T}{l_T} = \left(\sqrt{Th^{\frac{4+3\delta}{3+\delta}}}\right)^{-\frac{3+\delta}{4+2\delta}} (\log T)^{\frac{4+3\delta}{4+2\delta}} \to 0,$$

as Assumption 1f) implies that $T^{-1}h^{-\frac{4+3\delta}{3+\delta}} = O(T^{-\varepsilon_0})$ for some $\varepsilon_0 > 0$. Therefore $k_T = O(T/l_T) = O(\sqrt{T/h}\log T) = O(s_T^{2+\frac{4}{\delta}})$. The mixing condition implies $T^{2+\frac{4}{\delta}}\alpha(T) \to 0$, so that $k_T\alpha(s_T) \to 0$.

First, we show that the small blocks and the residual are asymptotically negligible as $T \to \infty$. Let $\Omega(y) = R(K)m(y)E\left(\frac{c_t}{q_t(y)}\right)$. Bound (29) implies that $\sum_{j=1}^{T-1}|\text{Cov}(Z_{0,h}, Z_{j,h})| \to 0$. In combination with stationarity, this implies that $\text{Var}\left(m_T^s\right) = \frac{k_T s_T}{T}\Omega(y)(1+o(1)) \to 0$ and $\text{Var}\left(\xi_T\right) = \frac{l_T+s_T}{T}\Omega(y)(1+o(1)) \to 0$, while

$$\text{Var}\left(m_T^l\right) = \frac{k_T l_T}{T}\Omega(y)(1+o(1)) \to \Omega(y). \tag{31}$$

We prove asymptotic normality using a truncation argument. For some fixed constant $\tau > 0$, let $Z_t^\tau = \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}1(q_t(R_{t+1}) \geq \tau) - E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}1(q_t(R_{t+1}) \geq \tau)\right)$, and let the superscript $\tau$ indicate that quantities sum over $Z_{t,h}^\tau = \sqrt{h}Z_t^\tau$ rather than $Z_{t,h}$. Similar to (31), it can be shown that

$$\begin{aligned}
\text{Var}(m_T^{l\tau}) &\to R(K)m(y)E\left(\frac{c_t}{q_t(y)}1(q_t(y) \geq \tau)\right) \equiv \Omega_\tau(y) \\
\text{Var}(m_T^l - m_T^{l\tau}) &\to R(K)m(y)E\left(\frac{c_t}{q_t(y)}1(q_t(y) < \tau)\right).
\end{aligned} \tag{32}$$

Consider the following bound on the difference in characteristic functions of $m_T^l$ and the Normal

distribution:

$$\begin{aligned}\left|E\left(\exp(ium_T^l\right) - \exp\left(-u^2\Omega(y)/2\right)\right| \leq & \left|E\left(\exp(ium_T^l)\left(\exp(iu(m_T^l - m_T^{l\tau})) - 1\right)\right)\right| \\ & + \left|E\left(\exp(ium_T^{l\tau})\right) - \prod_{j=1}^{k_T} E\left(\exp(iu\xi_j^\tau/\sqrt{T})\right)\right| \\ & + \left|\prod_{j=1}^{k_T} E\left(\exp(iu\xi_j^\tau/\sqrt{T})\right) - \exp\left(-u^2\Omega_\tau(y)/2\right)\right| \\ & + \left|\exp\left(-u^2\Omega_\tau(y)/2\right) - \exp\left(-u^2\Omega(y)/2\right)\right|.\end{aligned}$$

We will show that the RHS terms converge to zero when first $T \to \infty$ and then $\tau \to 0$. The first term is bounded by $E\left(\left|\exp(iu(m_T^l - m_T^{l\tau})) - 1\right|\right) = O\left(\mathrm{Var}(m_T^l - m_T^{l\tau})\right)$, which by (32) can be set arbitrary small by setting small enough $\tau$. The second term is bounded by $16(k_T - 1)\alpha(s_T)$ using the Volkonskii-Rozanov lemma, and thus converges to zero when $T \to \infty$. Since $|\xi_j^\tau| \leq Cl_T/\sqrt{h}$ as $K(\cdot)$ is bounded and has compact support,

$$\frac{1}{T}\sum_{j=1}^{k_T} E\left((\xi_j^\tau)^2 1(|\xi_j^\tau| > \varepsilon\sqrt{T})\right) \to 0,$$

for any $\varepsilon > 0$, since $\{|\xi_j^\tau| > \varepsilon\sqrt{T}\}$ becomes an empty set almost surely as $l_T/\sqrt{Th} \to 0$. Therefore $\frac{1}{\sqrt{T}}\sum_{j=1}^T \xi_j^\tau \to N(0, \Omega_\tau(y))$ by the Lindeberg-Feller central limit theorem, treating $(\xi_j)$ as independent, so that the third term vanishes when $T \to \infty$. The fourth term becomes arbitrary small when $\tau \to 0$. Therefore $m_T^l \to N(0, \Omega(y))$, completing the proof. □

*Proof of Proposition 1.* The error in estimating $\int \omega(y)m(y)dy$ equals

$$\begin{aligned}\int \omega(y)\left(\hat{m}(y) - m(y)\right)dy = & h^2 \int \omega(y)\frac{\mu_2(K)}{2}m''(y)dy + o(h^2) \\ & + \frac{1}{T}\sum_{s=1}^T \int \omega(y)\left(\frac{K_h(R_{s+1} - y)}{q_s(R_{s+1})} - E\left(\frac{K_h(R_{s+1} - y)}{q_s(R_{s+1})}\right)\right)dy.\end{aligned}$$

The first term is the leading bias term, while the variance of the estimation error equals

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{T}\sum_{s=1}^T \frac{\int \omega(y)K_h(R_{s+1} - y)dy}{q_s(R_{s+1})}\right) &= \mathrm{Var}\left(\frac{1}{T}\sum_{s=1}^T \frac{\omega(R_{s+1}) + \frac{\mu_2(K)}{2}\omega''(R_{s+1})h^2 + o_p(h^2)}{q_s(R_{s+1})}\right) \\ &= \mathrm{Var}\left(\frac{1}{T}\sum_{s=1}^T \frac{\omega(R_{s+1})}{q_s(R_{s+1})}\right) + O(h^2) \to \Omega_\omega.\end{aligned}$$

The long-run variance $\Omega_\omega$ is finite by a similar application of Davydov's inequality to $\mathrm{Cov}\left(\frac{\omega(R_{s+1})}{q_s(R_{s+1})}, \frac{\omega(R_{t+1})}{q_t(R_{t+1})}\right)$

as in the proof of Theorem 1. The result now follows from a central limit theorem for strong mixing processes, e.g. Herrndorf (1984). □

*Proof of Corollary 1.* Write

$$\begin{aligned}\hat{f}_t(y) - f_t(y) &= q_t(y)\left(\hat{c}_t\hat{m}(y) - c_t m(y)\right) \\ &= q_t(y)\left(c_t(\hat{m}(y) - m(y)) + (\hat{c}_t - c_t)m(y) - (\hat{c}_t - c_t)(\hat{m}(y) - m(y))\right).\end{aligned}$$

A second-order Taylor expansion around the true inverse normalization constant yields

$$\hat{c}_t - c_t = \frac{1}{\hat{c}_t^{-1}} - \frac{1}{c_t^{-1}} = \frac{-1}{c_t^{-2}}\left(\hat{c}_t^{-1} - c_t^{-1}\right) + O_p\left((\hat{c}_t^{-1} - c_t^{-1})^2\right).$$

The asymptotic bias of the inverse normalization constant estimator equals

$$\begin{aligned}\hat{c}_t^{-1} - c_t^{-1} &= \int q_t(y)\left(\hat{m}(y) - m(y)\right)dy \\ &= h^2\int q_t(y)\frac{\mu_2(K)}{2}m''(y)dy + o_p(h^2) + O_p\left(T^{-\frac{1}{2}}\right), \\ &= h^2 B_c + o_p(h^2),\end{aligned}$$

using Proposition 1 in the second step and the rate condition $Th^4 \to \infty$ in the last. Combining the above equations with Theorem 1, we find

$$\sqrt{Th}\left(\hat{f}_t(y) - f_t(y) - q_t(y)h^2\left(c_t B(y) - c_t^2 B_c m(y)\right)\right) \xrightarrow{d} c_t q_t(y) N(0, \Omega(y)).$$

Dividing both sides by $f_t(y) = c_t q_t(y) m(y)$ yields the stated result. □

*Proof of Proposition 2.* Equation (10) and the relation $\frac{q_t(y)}{\widetilde{q}_t(y)} = \frac{1}{1 - \frac{\widetilde{q}_t(y) - q_t(y)}{q_t(y)}}$ yield

$$\begin{aligned}\left|\widetilde{m}^I(y) - \widehat{m}^I(y)\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}\frac{K_h\left(R_{t+1} - y\right)}{q_t^2(R_{t+1})}\left(\widetilde{q}_t(R_{t+1}) - q_t(R_{t+1})\right)\frac{q_t(R_{t+1})}{\widetilde{q}_t(R_{t+1})}1\left(R_{t+1} \in I_{tn}\right)\right| \\ &\leq \frac{\max_{1\leq t\leq T}\sup_{y\in I_{tn}}\left|\widetilde{q}_t(y) - q_t(y)\right|}{1 - \max_{1\leq t\leq T}\sup_{y\in I_{tn}}\frac{\left|\widetilde{q}_t(y) - q_t(y)\right|}{q_t(y)}}\frac{1}{T}\sum_{t=1}^{T}\frac{K_h\left(R_{t+1} - y\right)}{q_t^2(R_{t+1})}1\left(R_{t+1} \in I_{tn}\right), \quad (33)\end{aligned}$$

assuming $\max_{1\leq t\leq T}\sup_{y\in I_{tn}}\frac{\left|\widetilde{q}_t(y) - q_t(y)\right|}{q_t(y)} < 1$, which occurs with probability approaching one.

The first term in (33) is $o_p(\delta_{n,T})$, and the second term is non-negative with mean bounded by

$$E\left(\frac{K_h(R_{t+1}-y)}{q_t^2(R_{t+1})}\right) = E\left(\int \frac{K_h(R-y)}{q_t^2(R)} f_t(R)dR\right)$$
$$= \int K_h(R-y)E\left(\frac{c_t}{q_t(R)}\right) m(R)dR,$$

which is bounded for all $T$. Therefore, by the Markov inequality $|\tilde{m}^I(y) - \hat{m}^I(y)| = o_p(\delta_{n,T})$. □

*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:

$$\sup_{\beta\in\Theta}\left\|\frac{\partial}{\partial\beta} Q_T(y,x,\beta) - D(x,y,h_x,\beta)\right\| \xrightarrow{p} 0, \tag{34}$$

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 $\beta$ 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 $\beta_j^u$ and $\beta_j^l$ denote the largest and smallest value of $\beta_j$ in $\Theta$ 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 (34) 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,

$$0 = \frac{\partial}{\partial\beta}Q_T(y,x,\hat{\beta}) = \frac{\partial}{\partial\beta}Q_T(y,x,\bar{\beta}) + \frac{\partial^2 Q_T(x,y,\beta^*)}{\partial\beta\partial\beta^T}(\hat{\beta}-\bar{\beta}), \tag{35}$$

for some $\beta^*$ such that $\beta_j^*$ lies in between $\hat{\beta}_j$ and $\bar{\beta}_j$ almost surely. Re-arranging (35) 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

$$R_{T1}\left(x,\beta\right)=\frac{2}{T}\sum_{t=1}^{T}\exp\left(2\beta^{T}\tilde{x}_{t}\right)H^{-1}\tilde{x}_{t}\left(H^{-1}\tilde{x}_{t}\right)^{T}K_{h_{x}}\left(x_{t}-x\right),$$

$$R_{T2}(x,y,\beta)=-\frac{2}{T}\sum_{t=1}^{T}\left\{\frac{K_{h}(R_{t+1}-y)}{q_{t}(R_{t+1})}-\exp\left(\beta^{T}\tilde{x}_{t}\right)\right\}\exp\left(\beta^{T}\tilde{x}_{t}\right)H^{-1}\tilde{x}_{t}\left(H^{-1}\tilde{x}_{t}\right)^{T}K_{h_{x}}\left(x_{t}-x\right).$$

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

$$\left\|R_{T1}\left(x,\bar{\beta}\right)-A_{T0}(x)\right\|=\mathrm{o_p}(h_x), \tag{36a}$$

$$\left\|R_{T1}\left(x,\beta^{*}\right)-R_{T1}\left(x,\bar{\beta}\right)\right\|=\mathrm{o_p}(1), \tag{36b}$$

$$\left\|R_{T2}(x,y,\beta^{*})\right\|=\mathrm{o_p}(1). \tag{36c}$$

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

$$\begin{aligned}E\left(\exp\left(2\bar{\beta}^{T}\tilde{x}_{t}\right)\left(\frac{x_{t}-x}{h_{x}}\right)^{2}K_{h_{x}}\left(x_{t}-x\right)\right)&=\int\exp\left(2\bar{\beta}_{0}+2\bar{\beta}_{1}h_{x}z\right)z^{2}K\left(z\right)f(x+zh_{x})dz\\&=m(y,x)^{2}\mu_{2}(K)f(x)+O(h_{x}^{2}).\end{aligned}$$

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 (36a) follows under bandwidth condition 2g).

By the mean-value theorem, there exists some $\tilde{\beta}$ in between $\beta^*$ 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 (36b) follows from

$$\begin{aligned}
&E\left(\|R_{T1}\left(x,\beta^*\right)-R_{T1}\left(x,\bar{\beta}\right)\|\right)\\
&\leq 4E\left(\|\left(\beta^*-\bar{\beta}\right)^T\tilde{x}_t\exp\left(2\tilde{\beta}^T\tilde{x}_t\right)H^{-1}\tilde{x}_t\left(H^{-1}\tilde{x}_t\right)^T K_{h_x}\left(x_t-x\right)\|\right) \qquad (37)\\
&\leq 4\|\beta^*-\bar{\beta}\|\int\|(1,zh_x)\|E\left(\exp\left(2\tilde{\beta}^T(1,zh_x)\right)\right)\|\tilde{z}\tilde{z}^T\|K\left(z\right)f(x+zh_x)dz\\
&\to 0,
\end{aligned}$$

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 $\beta^*$ in the third.

To establish (36c), 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 $\beta$, 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 $\beta$. 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 (37). Therefore $\|T_{Tj}(x,y,\beta^*)\|=o_p(1)$ for $j=1,2$, implying (36c).

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

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

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

$$\sqrt{Th^2}S_{T1}\left(x, y, \bar{\beta}\right) \xrightarrow{d} \mathrm{N}\left(0, V(x, y)\right), \tag{38a}$$
$$\left\|\sqrt{Th^2}S_{T2}\left(x, y, \bar{\beta}\right)\right\| = \mathrm{O_p}(h), \tag{38b}$$
$$\left\|S_{T3}\left(x, y, \bar{\beta}\right) - h^2 Ha(x, y)\right\| = \mathrm{O_p}\left(h^4\right), \tag{38c}$$

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

Given stationarity,

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

The variance term equals

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

By the LIE,

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

By Davydov's inequality for strong mixing processes

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

Therefore, for some constant $C>0$,

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

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 (30). 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

$$S_{T2}\left(x,y,\bar{\beta}\right)=\frac{1}{T}\sum_{t=1}^{T}u_{t+1,h}\exp(\bar{\beta}_1x_t^*)\begin{pmatrix}\bar{\beta}_1\\ 2\bar{\beta}_1+\bar{\beta}_1^2x_t^*\end{pmatrix}(x_t-x)\,\tilde{x}_{t,h}K_{h_x}\left(x_t-x\right),$$

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 $\text{Var}\left(S_{T2}\left(x,y,\bar{\beta}\right)\right) = O\left(\frac{h_x^2}{Th^2}\right) = O\left(\frac{1}{T}\right)$.

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

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

while its asymptotic variance equals

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

Therefore we have established that $\sqrt{Th^2}H\left(\hat{\beta}-\bar{\beta}-h^2 b(x,y)\right) \xrightarrow{d} N\left(0,\Omega(x,y)\right)$. 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.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of econometrics*, 94(1-2):9–51.

Albuquerque, R., Eichenbaum, M., Luo, V. X., and Rebelo, S. (2016). Valuation risk and asset pricing. *The Journal of Finance*, 71(6):2861–2904.

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

Andersen, T. G., Fusari, N., and Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Barone-Adesi, G., Fusari, N., Mira, A., and Sala, C. (2020). Option market trading activity and the estimation of the pricing kernel: A bayesian approach. *Journal of Econometrics*, 216(2):430–449.

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

Beare, B. K. and Schmidt, L. D. (2016). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*, 31(2):338–356.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The journal of finance*, 59(1):407–446.

Bollen, N. P. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*, 59(2):711–753.

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.

Chabi-Yo, F., Garcia, R., and Renault, E. (2008). State dependence can explain the risk aversion puzzle. *The review of Financial studies*, 21(2):973–1011.

Chen, H., Joslin, S., and Ni, S. X. (2019). Demand for crash insurance, intermediary constraints, and risk premia in financial markets. *The Review of Financial Studies*, 32(1):228–265.

Christoffersen, P., Heston, S., and Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing kernel. *The Review of Financial Studies*, 26(8):1963–2006.

Cuesdeanu, H. and Jackwerth, J. C. (2018). The pricing kernel puzzle in forward looking data. *Review of Derivatives Research*, 21(3):253–276.

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

Easley, D., O'hara, M., and Srinivas, P. S. (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2):431–465.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*, volume 20. Springer.

Fournier, M. and Jacobs, K. (2020). A tractable framework for option pricing with dynamic market maker inventory and wealth. *Journal of Financial and Quantitative Analysis*, 55(4):1117–1162.

Garleanu, N., Pedersen, L. H., and Poteshman, A. M. (2008). Demand-based option pricing. *The Review of Financial Studies*, 22(10):4259–4299.

Gozalo, P. and Linton, O. (2000). Local nonlinear least squares: Using parametric information in nonparametric regression. *Journal of econometrics*, 99(1):63–106.

Grith, M., Härdle, W. K., and Krätschmer, V. (2017). Reference-dependent preferences and the empirical pricing kernel puzzle. *Review of Finance*, 21(1):269–298.

Gustafsson, J., Hagmann, M., Nielsen, J., and Scaillet, O. (2009). Local transformation kernel density estimation of loss distributions. *Journal of Business & Economic Statistics*, 27(2):161–175.

Hagmann, M. and Scaillet, O. (2007). Local multiplicative bias correction for asymmetric kernel density estimators. *Journal of Econometrics*, 141(1):213–249.

Hens, T. and Reichlin, C. (2013). Three solutions to the pricing kernel puzzle. *Review of Finance*, 17(3):1065–1098.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, pages 141–153.

Heston, S. L., Jacobs, K., and Kim, H. J. (2024). Volatility risk and monotonic pricing kernels. *Available at SSRN 3997905*.

Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics*, pages 882–904.

Hyndman, R. J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *Journal of nonparametric statistics*, 14(3):259–278.

Jones, M., Linton, O., and Nielsen, J. (1995). A simple bias reduction method for density estimation. *Biometrika*, 82(2):327–338.

Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188(1):150–165.

Kim, H. J. (2023). Characterizing the conditional pricing kernel: A new approach. *Available at SSRN 4605072*.

Li, D., Lu, Z., and Linton, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory*, 28(5):935–958.

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

Linn, M., Shive, S., and Shumway, T. (2017). Pricing kernel monotonicity and conditional information. *The Review of Financial Studies*, 31(2):493–531.

Linton, O. and Xiao, Z. (2007). A nonparametric regression estimator that adapts to error distribution of unknown form. *Econometric Theory*, 23(3):371–413.

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

Pan, J. and Poteshman, A. M. (2006). The information in option volume for future stock prices. *The Review of Financial Studies*, 19(3):871–908.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64(3):341–372.

Sasaki, Y. and Ura, T. (2022). Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory*, 38(1):66–112.

Schreindorfer, D. and Sichert, T. (2025). Conditional risk and the pricing kernel. *Journal of Financial Economics*, 171:104106.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190(1):176–196.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC press.