UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Estimating a Density Ratio Model for Stock Market Risk and Option Demand

Jeroen Dalderop
University of Notre Dame

Oliver B. Linton
University of Cambridge

Abstract

Option-implied risk-neutral densities are widely used for constructing forward-looking risk measures. Meanwhile, investor 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, and characterizes its asymptotic bias and variance. The estimator can be used to correct biased density forecasts, and performs well in a simulation study. A local exponential linear variant of the estimator is proposed to include conditioning variables. In an application, we estimate a demand-based model for S&P 500 index options using net positions data, and attribute the U-shaped pricing kernel to heterogeneous beliefs about conditional volatility.

Reference Details

2411 Cambridge Working Papers in Economics
2405 Janeway Institute Working Paper Series

Published 5 March 2024

Keywords Density Forecasting, Nonparametric Estimation, Option Pricing, Trade Data
JEL-codes C14, G13

Websites www.econ.cam.ac.uk/cwpe
www.janeway.econ.cam.ac.uk/working-papers

Estimating a density ratio model for stock market risk and option demand

Jeroen Dalderop* Oliver Linton†

March 5, 2024

Abstract

Option-implied risk-neutral densities are widely used for constructing forward-looking risk measures. Meanwhile, investor 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, and characterizes its asymptotic bias and variance. The estimator can be used to correct biased density forecasts, and performs well in a simulation study. A local exponential linear variant of the estimator is proposed to include conditioning variables. In an application, we estimate a demand-based model for S&P 500 index options using net positions data, and attribute the U-shaped pricing kernel to heterogeneous beliefs about conditional volatility.

Keywords: Density Forecasting, Nonparametric Estimation, Option Pricing, Trade Data
JEL Codes: C14, G13

1 Introduction

Option prices can be used to extract forward-looking, market-implied distributions for the purpose of measuring and managing financial risk. However, they aggregate not only the beliefs of investors, but also their risk attitudes. 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 are of intrinsic economic

*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 participants at Northwestern Kellogg, University of Amsterdam, Notre Dame, and Copenhagen Business School for helpful comments. Financial support by the Keynes Fund is gratefully acknowledged.

interest. Moreover, they can be used to 'correct' risk-neutral densities in order to more accurately measure the conditional distribution of asset returns.

This paper proposes a simple local estimator of the pricing kernel, with several desirable properties. First, the estimator is nonparametric, which avoids imposing parametric functional form restrictions on investor preferences, as emphasized by Aït-Sahalia and Lo (2000). Second, the estimator does not require specifying the conditional density of the underlying return directly, thus avoiding Linn et al. (2017)'s critique of misspecifying the investor's information set. Third, it avoids global approximations of the pricing kernel, unlike other methods that meet the preceding points. Such methods require numerical optimization of the approximating coefficients, for example based on the probability integral transforms (Linn et al., 2017), maximum likelihood (Cuesdeanu and Jackwerth, 2018), or method-of-moments (Dalderop, 2021). These require a growing number of coefficients to be optimized, which presents challenges for computation and inference.

Instead, our kernel-smoothing estimator is easily computed by taking local averages of the *inverse* of the risk-neutral density at the realized returns. It can be interpreted as a dynamic version of the density bias correction by Jones et al. (1995), where instead of initial density estimators, it now corrects initial predictive densities. We show that the estimator is consistent for the multiplicative component in a general class of density ratio models, not restricted to the setting of option-implied predictive densities. Furthermore, we prove that the estimator is asymptotically Normal for a wide range of density-generating processes subject to certain moment and mixing conditions. We provide analytical expressions of its asymptotic bias and variance, and use these to characterize asymptotically optimal bandwidths. Moreover, we characterize the asymptotic distribution of the corrected densities.

A simulation study confirms the good performance of the local estimator for realistic data-generating processes and sample periods. In particular, the estimator outperforms or is at least competitive with the correctly specified parametric maximum likelihood estimator at moderate sample sizes, for both GARCH and stochastic volatility models. We compare four different plug-in bandwidth methods, and find the estimators are not too sensitive to the method of pilot estimator. The robust performance carries over to that of the corrected densities. Moreover, we show that the estimator can be further improved by smoothly trimming small values of the risk-neutral densities, which helps to reduce its variance in the tails of the distribution.

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 density. When the locally parametric form is approximately correct, this variant effectively reduces the bias and/or allows choosing a larger bandwidth to reduce the variance relative to the locally constant estimator.

We apply our estimator to measure the impact of net demand on the pricing kernel (Garleanu et al., 2008; Almeida and Freire, 2022). In particular, we model net demand for option contracts by heterogeneous traders using a variant of the portfolio choice problem in Carr and Madan (2001). We derive the equilibrium risk-neutral density as a risk-aversion weighted average of investor's subjective densities. By linking the latter to the true physical density, we derive a density ratio model that expresses the pricing kernel in terms of investor's net positions. Using data on trader's positions in S&P 500 index options, we find that the pricing kernel varies with differences in subjective volatilities that we infer from investor's net positions.

The remainder of this paper is organized as follows. Section 2 describes the local estimator and its asymptotic properties in the setting of a density ratio model. Section 3 performs a simulation study to assess the estimator's performance for various bandwidth choices. Section 4 applies the estimator to the pricing kernel in S&P 500 index options using 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 some true and observed conditional densities, and studies the asymptotic properties of simple nonparametric estimators with and without observed covariates.

2.1 Density ratio model

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

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

where $m(y, x)$ is an unknown multiplicative 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. Scaling $m(y, x)$ by any function of x does not change $f_t(y)$. Therefore, without loss of generality we assume $E(c_t \mid x_t = x) = 1$ for all x. Values of $m(y, x)$ different from one describe a density forecast bias, as they indicate $q_t(y)$ under- or overstates the true density at 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 $C_t(\kappa)$ of a call option on R with strike price κ equals

$$C_t(\kappa) = E_t\left(m_{t,t+1}(R_{t+1}-\kappa)^+\right), \tag{2}$$

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

$$C_t(\kappa) = \frac{1}{R_t^f}\int_0^\infty (R-\kappa)^+ q_t(y)dy,$$

in terms of a risk-neutral density $q_t(\cdot)$ that relates to the physical density as

$$q_t(y) = R_t^f f_t(y) m_t(y).$$

Therefore, (1) obtains as long as

$$m_t(y) = \delta_t \tilde{m}(y, x_t), \tag{3}$$

for some time-varying constants δ_t and covariate-dependent function $\tilde{m}(y,x)$. In particular, (1) then holds with $m(y,x) = \frac{1}{\tilde{m}(y,x)}$ and inverse normalization constant $c_t^{-1} = R_t^f \delta_t$.

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 δ and relative risk aversion γ (e.g. Bliss and Panigirtzoglou, 2004). This can be extended to allow for stochastic time preferences described by an arbitrary time-discount factor δ_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 $m_t(y)$ depends on their joint distribution with the return R. Since economic theory may not prescribe appropriate functional forms, 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). Here the constants δ_t make sure that the implied $f_t(y)$ are valid densities. Finally, it allows for state-dependence in preferences which can described by observed covariates, such as measures of volatility that may change the shape of the pricing kernel (Song and Xiu, 2016). Our application in Section 4 illustrates this using proxies of trading imbalances to measure the impact of net end-user demand on the pricing kernel. It motivates a formulation of (1) using a stylized

equilibrium trading model, which relates $m(y, x)$ to the preferences and beliefs of heterogeneous investors and the constants c_t to the Lagrangian multipliers of their budget constraints.

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. The multiplicative factor m is only identified up to scaling, so that we set $E(c_t) = 1$ without loss of generality, where $c_t = \left(\int q_t(y)m(y)dy\right)^{-1} = \int f_t(y)/m(y)dy$, which is equivalent to $E\left(\frac{1}{m(R_{t+1})}\right) = 1$.

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

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

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

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

using the Law of Iterated Expectations in the second step, where $\mu_2(K) = \int K(z)z^2dz$. As in standard nonparametric regression, the asymptotic bias is determined by the curvature of the

[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.

object of interest, and 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) $m(y)$ *is twice differentiable*

c) $q_t(R_{t+1}) > 0$ *a.s. and* $E\left(\frac{1}{q_t(y)^2} \mid R_{t+1} = y\right) < \infty$

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

e) $q_t(R) = q(R, s_t)$ *for some function* q *and state variables* s_t *such that* (R_{t+1}, s_t) *is stationary and* α*-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. *Under Assumption 1, when* $T \to \infty$

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

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, thus accommodating 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 for the constants c_t. Furthermore, little is known about the asymptotic properties of such estimators that depend on random densities, which complicates inference.[2]

Another advantage of the local estimator (5) is that it can accommodate observational errors in the density forecasts $q_t(y)$. In particular, suppose one observes $\tilde{q}_t(y) = q_t(y)z_t(y)$, where $z_t(y)$ is an i.i.d. multiplicative error with mean one. Then the bias of (5) is unchanged, even though its variance increases. Moreover, if $q_t(y)$ is only observed on a range $[y_l, y_u]$, then (5) can estimate $m(y)$ on this range as before, up to a boundary bias that could be corrected.[3]

2.3 Density forecast correction

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

$$\hat{c}_t = \left(\int q_t(y)\hat{m}(y)dy \right)^{-1} = \left(\frac{1}{T}\sum_{s=1}^{T} \frac{\int q_t(y)K_h(R_{s+1}-y)dy}{q_s(R_{s+1})} \right)^{-1}, \tag{6}$$

and compute the multiplicatively corrected predictive densities as

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

which by construction integrate to one.

From a forecasting perspective, it is useful to perform inference on the corrected version $\hat{f}_{t^*}(y)$ of any observed density $q_{t^*}(y)$ which may or may not have been included in the sample to estimate $\hat{m}$. 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 error 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 $\mathrm{Var}\left(\frac{\omega''(R_{t+1})}{q_t(R_{t+1})}\right) < \infty$. 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 $\Omega_\omega = \mathrm{Var}\left(\frac{\omega(R_{t+1})}{q_t(R_{t+1})}\right) + 2\sum_{j=1}^{\infty}\mathrm{Cov}\left(\frac{\omega(R_1)}{q_0(R_1)}, \frac{\omega(R_{j+1})}{q_j(R_{j+1})}\right)$.

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

[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.

[3] This matters for risk-neutral densities, whose tails may not be reliably estimable due to sparse option trading in corresponding strike prices.

moment conditions control the tail behavior of the function $\omega(R)$ relative to $q_t(R)$. By setting $\omega(y) = q_{t^*}(y)$, it gives rise to the following corollary characterizing the asymptotic distribution of 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 thus merely for simplicity, as under the mixing conditions any dependence between $q_{t^*}(\cdot)$ and $\{q_s(\cdot)\}_{t=1}^T$ 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 as the faster $\sqrt{T}$ rate by Proposition 1.

2.4 Adding covariates

Let x_t be some covariates that are thought to explain time-variation in the multiplicative bias in observed predictive densities $q_t(y)$. Consider the general formulation

$$f_t(y) = c_t q_t(y) m(y, x_t),$$

where $f_t(y)$ describes the conditional density of the outcome variable R_{t+1} given information available at time 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 normalize $E(c_t \mid x_t = x) = 1$ for all x. Let $K(\cdot)$ be a symmetric density function on $\mathbb{R}$ and $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$ for some bandwidth h. Then

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

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

$$\hat{m}(y,x) = \sum_{t=1}^{T} \frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} w_T(x_t,x), \tag{8}$$

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. Still, local linear estimators are typically preferred over the locally constant estimator (8) for their reduced bias. However, the local linear estimator is not guaranteed to be positive, and thus to 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:

$$\begin{aligned} \hat{\beta} &= \arg\min_{\beta \in \Theta} \; Q_T(y,x,\beta) \\ Q_T(y,x,\beta) &= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - \exp\left(\beta_0 + \beta_1(x_t-x)\right) \right)^2 K_{h_x}(x_t-x). \end{aligned} \tag{9}$$

for some parameter space Θ. Related estimators have been studied by Gozalo and Linton (2000) and Hyndman and Yao (2002). Specifically, Gozalo and Linton (2000) consider a local nonlinear least squares method to estimate a regression function nonparametrically based on an initial parametric model. Meanwhile, Hyndman and Yao (2002) use a local exponential fit to estimate the conditional densities. Our objective 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 F.O.C. of the local least squares criterion (9) 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 $\tilde{x}_t = (1, x_t - x)^T$. We prove that the locally estimated coefficients $\hat{\beta}$ consistently estimate the population coefficients β_0 that match the first two derivatives of the target and exponential link functions:

$$\begin{aligned} \beta_{00} &= \log m(y,x) \\ \beta_{01} &= \frac{\partial}{\partial x} \log m(y,x). \end{aligned}$$

Assumption 2.

a) β_0 *is an interior point of the compact parameter space* Θ

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 $E\left(\frac{1}{q_t(y)^2} \mid R_{t+1} = y, x_t = x\right) < \infty$

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 function q and state variables s_t such that (R_{t+1}, s_t, x_t) is stationary and α-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 $T \to \infty$*

$$\sqrt{Th^2} H \left(\hat{\beta} - \beta_0 - h^2 b(x,y)\right) \xrightarrow{d} N\left(0, \Omega(x,y)\right),$$

where $H = diag(1, h_x)$, and $b(x,y)$ and $\Omega(x,y)$ are given in the proof in the Appendix.

In particular, using the delta method, Theorem 2 implies the following limiting distribution of the pricing kernel estimator $\hat{m}(y,x) = \exp(\hat{\beta}_0)$:

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

where

$$\begin{aligned} b_m(y,x) &= \frac{1}{2}\mu_2(K)\left(m_{yy}(y,x) + \left(m_{xx}(y,x) - \beta_{01}^2 m(y,x)\right) c_x^2\right) \\ \sigma_m^2(y,x) &= m(y,x) f(x)^{-1} c_x^{-1} R_0(K)^2 E\left(\frac{c_t}{q_t(y)} \mid x_t = x\right). \end{aligned}$$

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 second term is proportional to $m_{xx}(y,x) - \beta_{01}^2 m(y,x)$. 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 (8) for a given bandwidth. Moreover, the optimal bandwidth will be larger, allowing to reduce the variance of the estimator. Finally, introducing the covariate allows the slightly weaker mixing condition 2f), as localizing in x_t diminishes any serial correlation in the summands of Q_T.

3 Simulation study

We perform a simulation study to assess the finite-sample performance of the kernel estimator for various bandwidth choices, and compare their performance with the correctly specified parametric estimator.

3.1 Bandwidth choices

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

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

The MSE-optimal bandwidth therefore varies with y according to

$$h^{\mathrm{MSE}}(y) = \left(\frac{\Omega(y)}{4B(y)^2 T} \right)^{\frac{1}{5}}. \tag{10}$$

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

$$h^{\mathrm{IMSE}} = \left(\frac{\int \Omega(y) f(y) dy}{4 \int B(y)^2 f(y) dy T} \right)^{\frac{1}{5}}. \tag{11}$$

The integrated variance term equals

$$\int \Omega(y) f(y) dy = R(K) \int E\left(\frac{f_t(y)}{q_t(y)^2} \right) f(y) dy = R(K) E\left(\frac{f(R_{t+1})}{q_t^2(R_{t+1})} \right),$$

using the relation $E\left(m(y) \frac{c_t}{q_t(y)} \right) = E\left(\frac{f_t(y)}{q_t(y)^2} \right)$. However, it may not be finite due to small values of $q_t^2(y)$ in the denominator. Therefore, for the simulations we consider the truncated IMSE over a large but finite range, and set the plug-in IMSE bandwidths based on the corresponding definite integrals of $B(y)$ and $\Omega(y)$.

We estimate the optimal bandwidths by plugging-in initial estimates of the components $\Omega(y)$ and $B(y)$ in (10) and (11). Both components depend on the unknown function $m(y)$, for which we consider two initial estimators $\hat{m}_{\mathrm{int}}(y)$. The first fits an initial parametric model $m(y;\theta)$, with the parameter θ estimated by maximum likelihood. The second is the nonparametric estimator (5) with a pilot bandwidth of two times Silverman's rule-of-thumb bandwidth for kernel density estimation. The local bias is then estimated as $\hat{B}(y) = \frac{\mu_2(K)}{2} \hat{m}''_{\mathrm{int}}(y)$ and the local variance as $\hat{\Omega}(y) = R(K) \hat{m}_{\mathrm{int}}(y) \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{c}_t}{q_t(y)}$, with $\hat{c}_t$ the plug-in normalization constant (6). The IMSE

optimal bandwidth then integrates these components with respect to an initial kernel density estimator $\hat{f}_h(y)$. Finally, we use the resulting estimator $\hat{m}(y)$ to compute an iterated version of the optimal plug-in bandwidth.

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 pricing kernel estimators 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

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

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.

For the true pricing kernel, we consider a exponential-quartic model, which allows capturing the main nonlinear shapes documented in the literature. In particular, we set $m_0(y) = m(y;\gamma_0)$ where

$$m(R;\gamma) = \exp\left(\sum_{l=1}^{4}\gamma_l\left(\frac{1}{2}\log R\right)^l\right),$$

and γ_0 is chosen by maximum likelihood on the sample used in our application, but without the dampening factor $\frac{1}{2}$. Using the dampened model as the DGP enhances the stability of the parametric maximum likelihood estimator, which we describe next.

Estimators. Besides the local nonparametric estimators, we include the correctly specified

exponential-quartic model and estimate it by maximum likelihood. With $L = 4$, this results in the pricing kernel estimates $m(y; \hat{\gamma}_L)$, where

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

We consider the nonparametric local estimators with four plug-in bandwidth choices. They differ by whether the optimal bandwidths are based on an initial nonparametric or parametric estimator, for which we use slightly misspecified exponential-cubic estimator $m(y; \hat{\gamma}_3)$, and whether or not they iterate the plug-in bandwidth computation.[4] We also include a smoothly trimmed version of the local estimator, defined as

$$\hat{m}_{\text{trim}} = \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} S\left(\frac{q_t(R_{t+1})}{\tau_T} \right), \tag{12}$$

where S is a CDF with $S(0) = 0$ and $S(1) = 1$, and τ_T a vanishing trimming parameter. The trimming serves to stabilize the tails of $\hat{m}(y)$ in the presence of small values of $q_t(R_{t+1})$. We set S as the Beta$(2, 2)$ CDF following Linton and Xiao (2007), and τ_T as the 0.01-quantile of $q_t(R_{t+1})$.

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 (6). The scaling enforces the sample version of the theoretical constraint $E(c_t) = 1$, and serves to reduce the variance.

3.3 Results

Table 1 reports the simulation 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. Its performance

[4] 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.

under four different plug-in bandwidths is broadly similar, indicating low sensitivity to the choice of pilot estimator $\hat{m}_{\text{int}}$. Only bandwidths based on an initial parametric fit tend to benefit from the iteration step, at least for the GARCH model, where those based on a nonparametric pilot estimator are more precise without iteration. Finally, smooth trimming benefits the local estimators by bringing down their variance substantially without incurring a similar increase in the 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). The simulation performance for the estimators without scaling are very similar, except for increased variability of the 'h-par' estimator, and are reported in Table 3 in the Appendix.

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. Optimal plug-in bandwidths chosen based on initial exponential-cubic (h-par) or local ratio estimator (h-pilot). Columns describe integrated weighted squared bias, variance, and mean squared error of scaled estimators $\hat{m}_{\text{sc}}(y)$, truncated at the (0.025,0.975)-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	0.15	4.32	4.47	0.08	1.87	1.95	0.03	0.80	0.83
h-par	0.16	3.28	3.43	0.10	2.13	2.23	0.03	1.43	1.46
h-par iter	0.12	3.36	3.48	0.05	1.89	1.94	0.02	1.06	1.08
h-par trim	0.19	2.80	2.99	0.20	1.42	1.63	0.17	0.75	0.92
h-pilot	0.05	3.57	3.62	0.02	2.07	2.09	0.01	1.05	1.07
h-pilot iter	0.06	3.89	3.94	0.02	2.23	2.25	0.01	1.14	1.15
h-pilot trim	0.10	3.56	3.66	0.12	1.96	2.08	0.13	0.95	1.08

(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	0.12	3.14	3.26	0.05	1.30	1.35	0.03	0.54	0.56
h-par	0.20	2.32	2.52	0.15	1.55	1.70	0.07	0.84	0.91
h-par iter	0.14	2.39	2.53	0.08	1.59	1.67	0.03	0.90	0.93
h-par trim	0.17	2.17	2.35	0.18	1.23	1.40	0.13	0.59	0.72
h-pilot	0.07	2.28	2.35	0.04	1.63	1.67	0.02	0.93	0.95
h-pilot iter	0.06	2.57	2.63	0.04	1.79	1.83	0.02	1.02	1.04
h-pilot trim	0.08	2.43	2.51	0.08	1.60	1.69	0.08	0.82	0.89

Figure 1 shows the simulated densities of plug-in fixed bandwidths based on the asymptotic IMSE under the Bates (2000) stochastic volatility-model. Bandwidths based on an initial exponential-cubic fit tend to centered around the infeasible optimal value, but display considerable right skewness. This pattern can be explained by the high volatility of the parametric fit, which may occasionally yield nearly flat estimates that underestimate the bias. Meanwhile, bandwidths based on the nonparametric pilot estimator appear symmetric and are less volatile, but have a downward bias at all sample sizes, suggesting these may be undersmoothing. The bandwidth densities for the GARCH model in Figure 12 in the Appendix show similar patterns. The fact that either plug-in bandwidth method has a substantial mean squared deviation from the optimal value suggests the performance of the nonparametric estimators in Table 1 could still be improved.

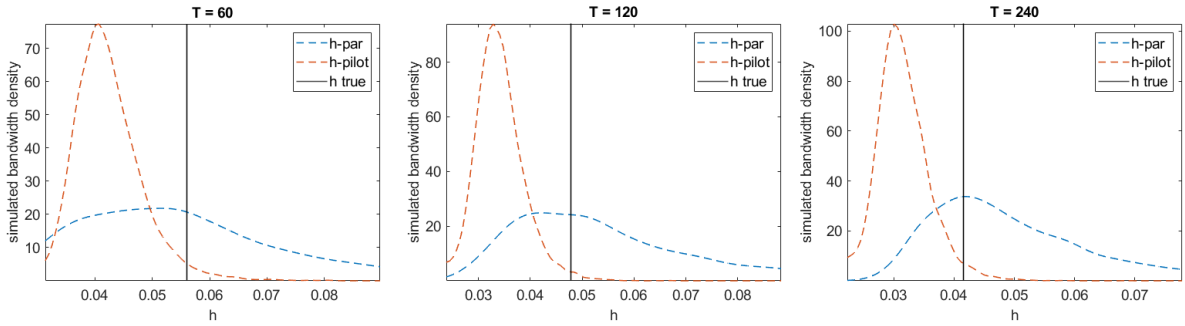
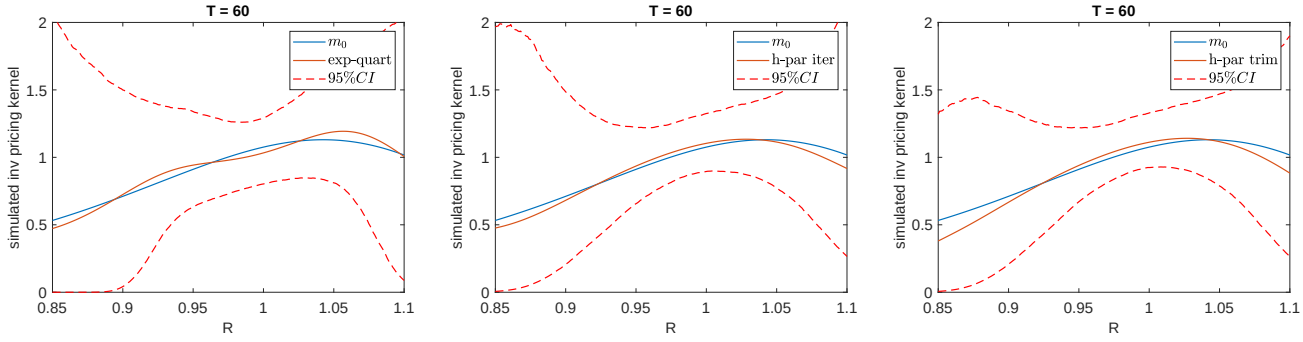


Figure 1: Simulated densities of plug-in asymptotic IMSE optimal fixed bandwidths under the Bates (2000) stochastic volatility-model, using initial exponential-cubic (h-par) and local pilot (h-pilot) estimators, for $S = 10,000$ simulated time series and sample sizes $T = \{60, 120, 240\}$ months. IMSE truncated at (0.025,0.975)-unconditional quantiles. Vertical lines show corresponding optimal bandwidths if m_0 were known.

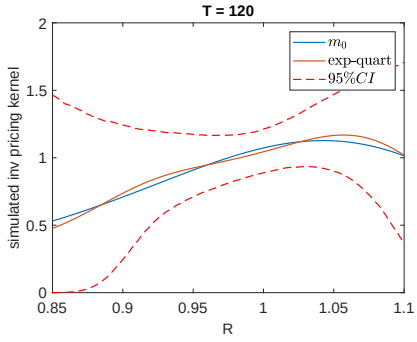
Figures 2 and 3 show the simulated mean and pointwise quantiles of the estimators $\hat{m}$ for the GARCH and stochastic volatility models, respectively. The plots highlight that the nonparametric estimators outperform the correctly specified parametric estimator in the middle part of the distribution, where the former have smaller bias but a roughly similar variance. The negative bias of the local estimators in the right tail is explained by the convexity of the true function m_0. 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 the GARCH model, which has thinner tails than the stochastic volatility model, and especially for the left tail. Still, trimming the local estimators yields more precision in the tails than the parametric estimators, without creating much larger bias.



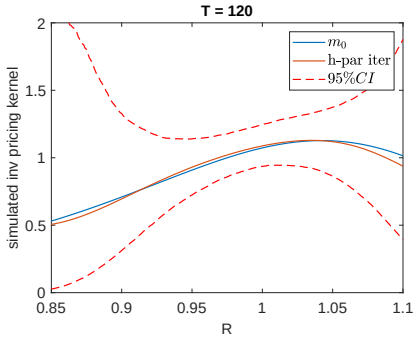
(a) Exponential-quartic estimator.

(b) Local ratio estimator.

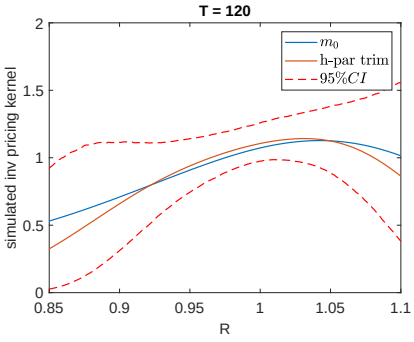
(c) Trimmed local ratio estimator.



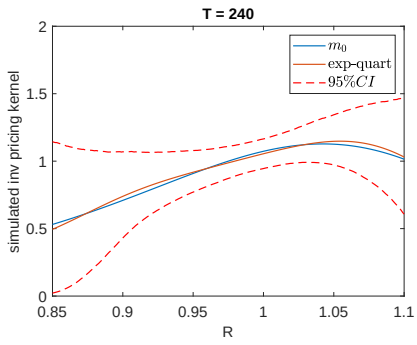
(d) Exponential-quartic estimator.



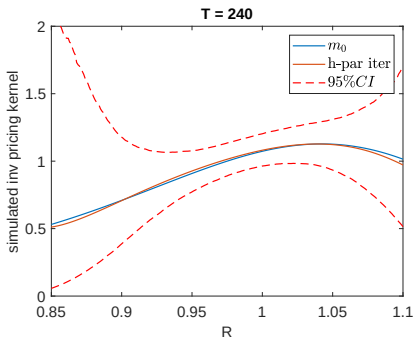
(e) Local ratio estimator.



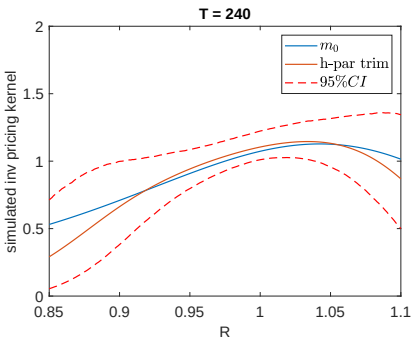
(f) Trimmed local ratio estimator.



(g) Exponential-quartic estimator.



(h) Local ratio estimator.



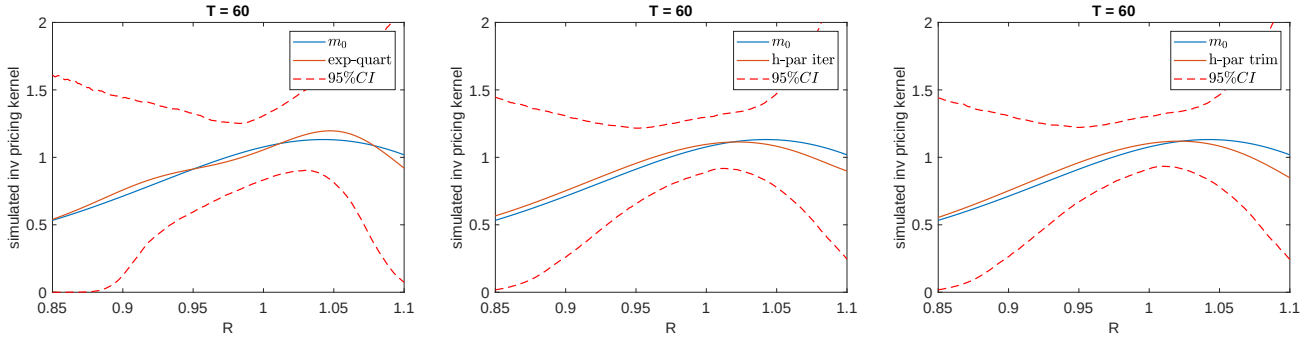
(i) Trimmed local ratio estimator.

Figure 2: Simulated estimates $\hat{m}_{\text{sc}}$ of the inverse pricing kernel m_0 under the AR(1)-GARCH(1,1)-t model, together with simulated pointwise 95% CIs, based on $S = 10,000$ simulated time series. Local estimators use plug-in bandwidths based on initial exponential-cubic fit and one iteration. Rows vary with $T = \{60, 120, 240\}$ months, columns with type of estimator.

Finally, Table 2 reports the simulated mean integrated squared density estimation error

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

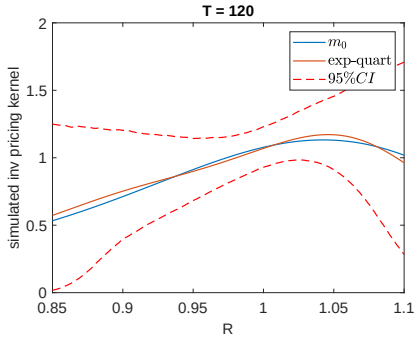
for the considered pricing kernels estimators $\hat{m}$. The density estimates are invariant to any scaling of $\hat{m}$. The performance of the corrected conditional density estimators follows that of the pricing kernel estimators, with the nonparametric estimators outperforming at $T = 60$ and being



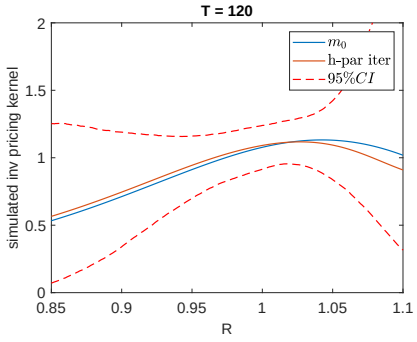
(a) Exponential-quartic estimator.

(b) Local ratio estimator.

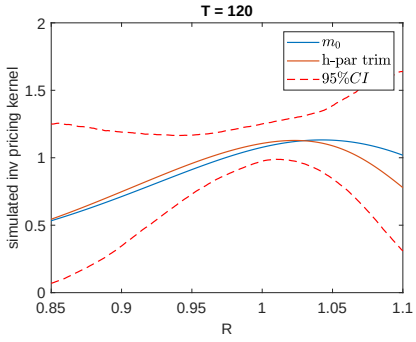
(c) Trimmed local ratio estimator.



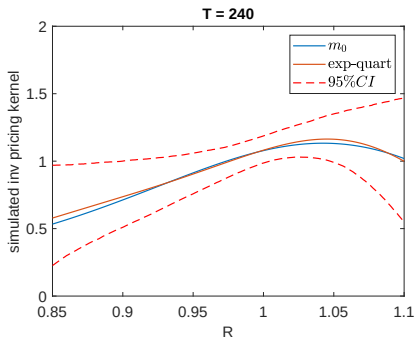
(d) Exponential-quartic estimator.



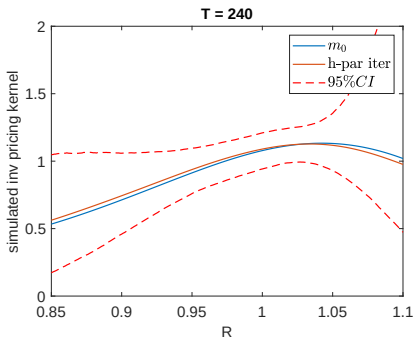
(e) Local ratio estimator.



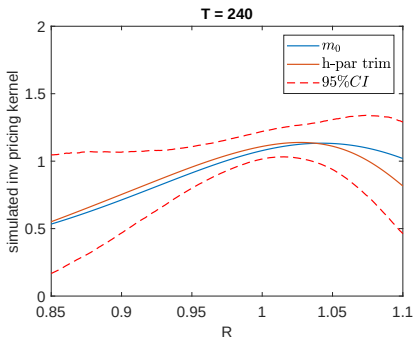
(f) Trimmed local ratio estimator.



(g) Exponential-quartic estimator.



(h) Local ratio estimator.



(i) Trimmed local ratio estimator.

Figure 3: 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 based on initial exponential-cubic fit and one iteration. Rows vary with $T = \{60, 120, 240\}$ months, columns with type of estimator.

competitive at $T = 120$, and the parametric estimator outperforming at $T = 240$. Selecting plug-in bandwidths without iteration is preferred for both types of initial estimators. Furthermore, the local estimators again favor some trimming, which particularly helps to stabilize the tails of the conditional density estimates.

Table 2: Simulated mean integrated square density estimation error for various estimators $\hat{m}$ 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. Optimal plug-in bandwidths minimize IMSE truncated at the (0.025,0.975)-unconditional quantiles using initial exponential-cubic (h-par) and local pilot (h-pilot) estimators.

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

$\hat{m}$	$T = 60$	$T = 120$	$T = 240$
exp-quart	17.17	7.60	3.43
h-par	10.90	7.18	4.93
h-par iter	12.15	7.63	5.46
h-par trim	10.24	5.77	3.58
h-pilot	12.72	8.28	5.57
h-pilot iter	14.47	9.20	6.20
h-pilot trim	12.37	7.35	4.19

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

$\hat{m}$	$T = 60$	$T = 120$	$T = 240$
exp-quart	15.29	6.47	2.77
h-par	11.31	7.71	4.49
h-par iter	11.98	8.36	4.96
h-par trim	10.77	6.59	3.51
h-pilot	11.15	8.85	5.40
h-pilot iter	12.96	10.20	6.28
h-pilot trim	11.71	8.16	4.42

4 Empirical application

This section applies the local density ratio estimator to S&P 500 index returns using option price and trade data. The first subsection describes our data set, the second estimates the pricing kernel without conditioning variables, and the third estimates a model with heterogeneous option investors using net demand data.

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, as follows. For each month from January 1996 until February 2023, we consider option prices from OptionMetrics on contracts that expire on the last expiration date with options actively traded 28 days prior (typically the last Friday).[5] Focusing on contracts with one month or shorter time-to-maturity avoids small sample sizes. The risk-neutral density at a given month 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 a cross-section or rolling window of end-of-day call option mid prices (possibly obtained by put-call parity) with corresponding moneyness levels observed at or around time t, 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

[5] We define actively traded as at least 40 strike prices being listed, which removes 60 out of 235 months.

estimated using nonparametric methods such as kernel smoothing or series approximation. In particular, we estimate monthly risk-neutral densities using the local cubic method in Dalderop (2020), based on variable plug-in bandwidths obtained by fitting an initial Bates (2000) stochastic volatility model. Furthermore, we compute the realized S&P 500 index return R_{t+1} over the option holding period.

Our application to demand-based pricing kernels in Section 4.3 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 net trading positions in European option contracts on the S&P 500 index over the period 1996-2017. The trading positions are disaggregated by investor type (firm or customer) and three size categories. This categorization of buyers and sellers of each transaction allows tracking the direction of trades, rather than just trading volume, and thereby allows identifying 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 Subsection 4.3.2.

4.2 Time-invariant pricing kernels

First, we consider the case without covariates. To help understand the local estimator, Figure 4 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 $R_{t+1} \approx y$ indicate that $m(y) > 1$, and vice versa. Figure 4 shows no clear deviation between 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 $f_h(R_{t+1})$, suggesting $m(y)$ is above one in this region, although the variability of $1/q_t(R_{t+1})$ also increases in the tails.

Figure 5 shows the estimates of $m(y)$ and its inverse, the pricing kernel, using fixed plug-in bandwidth. Their local confidence intervals are obtained by plugging $\hat{m}_h$ into the asymptotic variance in Theorem 1, and using the delta-method to compute standard errors for the pricing kernel from those of its inverse. The left panel estimate shows a similar shape as the exponential-

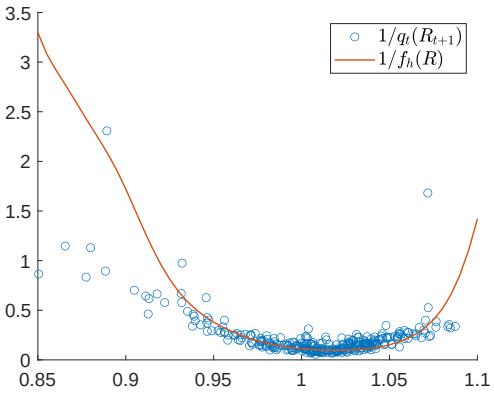
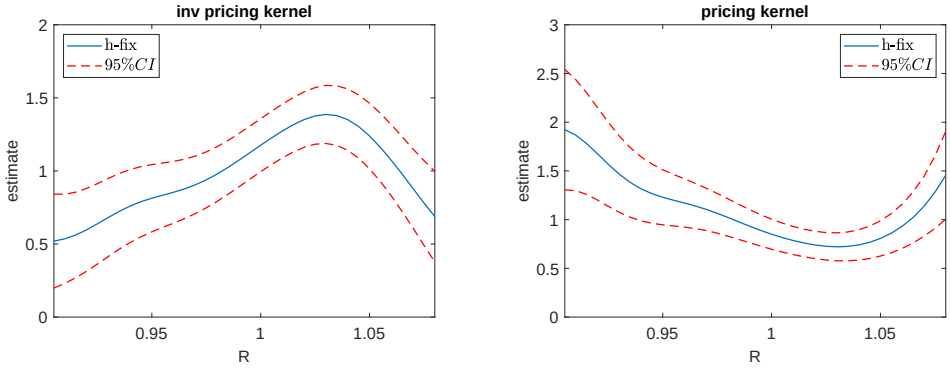


Figure 4: Scatterplot of $1/q_t(R_{t+1})$ against the monthly return on the S&P 500 from January 1996 to August 2023, together with plot of the inverse unconditional smoothed density $f_h(R)$ based on Silverman's rule-of-thumb bandwidth.

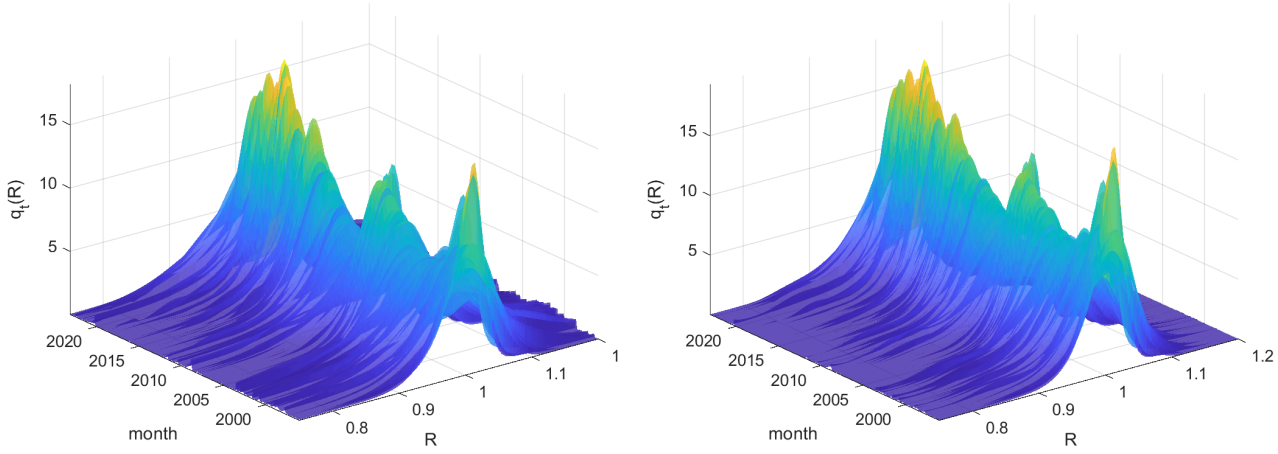
quadratic model used for the simulation study, which was fitted to the same data. In particular, $\hat{m}_h(y)$ shows a concave pattern reaching its peak for returns around one. This translates into a U-shaped pricing kernel in the right panel, which increases in both tails of the return distribution, though especially in the left.



(a) Estimated inverse pricing kernel. (b) Estimated inverse pricing kernel.

Figure 5: Local density ratio estimator $\hat{m}_h$ (left panel) with fixed plug-in bandwidth and corresponding pricing kernel (right panel) based on monthly S&P 500 index options and returns from January 1996 to August 2023. CIs based on plug-in standard errors and delta-method.

Figure 6 shows the resulting multiplicatively adjusted option-implied densities (7) for the fixed plug-in bandwidth. Multiplying the original risk-neutral densities by the hump-shaped inverse pricing kernel dampens the probability mass in the tails and lifts it in the center. As a result, conditional volatilities are typically adjusted downwards, especially when they are high, and therefore fluctuate less over time as shown in Figure 7. More generally, the likelihood of left-tail risks is lower and more stable under the physical than the risk-neutral distribution.



(a) Original densities.

(b) Corrected densities.

Figure 6: Original (left) and multiplicatively corrected (right) option-implied conditional densities, using the local estimator with bandwidth $\hat{h}^{IMSE}$, for the monthly return on the S&P 500 from January 1996 to August 2023.

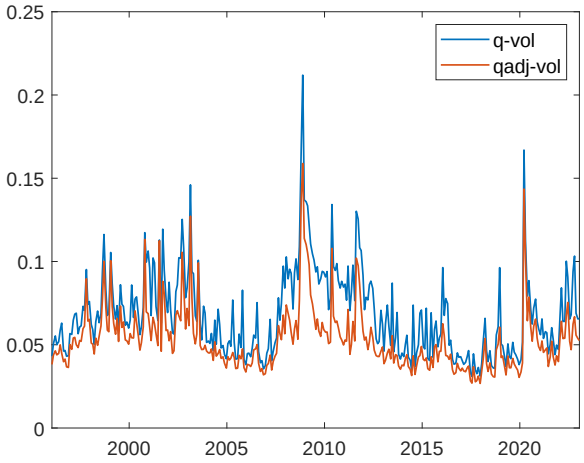


Figure 7: Original and multiplicatively corrected option-implied volatility forecasts of the monthly return on the S&P 500 from January 1996 to August 2023. Local pricing kernel estimated with fixed bandwidth $\hat{h}^{IMSE}$.

4.3 Demand-based pricing kernels

This section formulates and estimates a model for the impact of net demand for options on the ratio between the physical and risk-neutral densities, or the pricing kernel. This is motivated by empirical evidence for such impact by Easley et al. (1998), Pan and Poteshman (2006) and Bollen and Whaley (2004), among others, and theoretically underpinned by market-makers facing unhedgeable positions (Garleanu et al., 2008; Chen et al., 2019) and/or inventory constraints

(Fournier and Jacobs, 2020). Similar to recent work of Almeida and Freire (2022) based on option returns, we investigate the shape of conditional pricing kernels, which we estimate directly.

First, we discuss a stylized heterogeneous-agent equilibrium model that connects the subjective, physical, and risk-neutral densities. We then use this model to formulate and interpret a density ratio model, and study its local kernel-based estimation.

4.3.1 Heterogeneous investor model

Consider $i = 1, \ldots, N$ investors with subjective belief densities f_{it} for the stock return R_{t+1}, utility U_i over wealth, and owner share ω_i^0 of the stock. Following Carr and Madan (2001), investors trade a continuum of a contracts on R_{t+1} and choose portfolios $\phi_{it}(R_{t+1})$ according to

$$\begin{aligned} &\max_{\phi_{it}(\cdot)} \int_0^\infty U_i(\phi_{it}(R)) f_{it}(R) dR \\ &\text{s.t. } e^{-r_t^f} \int \phi_{it}(R) q_t(R) dR \leq \omega_i^0 S_{it}, \end{aligned}$$

where $q_t(R)$ is the Arrow-Debreu state-price density, and S_{it} the initial stock price. The optimal portfolio of investor i then satisfies

$$U_i'(\phi_{it}(R)) = \lambda_{it} \frac{f_{it}(R)}{q_t(R)},$$

where λ_{it} is the Lagrangian multiplier for the budget constraint. Under exponential utility $U_i(w) = -\frac{1}{\gamma_i} \exp(-\gamma_i w)$, the optimal portfolio can be solved explicitly as

$$\phi_{it}(R) = \mu_{it} + \frac{1}{\gamma_i} \log \frac{f_t^{(i)}(R)}{q_t(R)}, \qquad i = 1, \ldots, N.$$

The equilibrium stock market clearing condition $\sum_{i=1}^N \phi_{it}(R) = R$ implies

$$q_t(R) = c_t \exp\left(\beta_0 R + \sum_i \beta_{i1} \log f_{it}(R) \right),$$

where $\beta_0 = \frac{\sum_{i=1}^N \omega_i^0}{\sum_{i=1}^N \frac{1}{\gamma_i}}$, $\beta_i = \frac{\frac{1}{\gamma_i}}{\sum_{i=1}^N \frac{1}{\gamma_i}}$, and c_t is the normalizing constant.

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

$$\log f_t(R) \propto \sum_k s_{tk} g_k(R), \quad \log f_{it}(R) \propto \sum_k s_{itk} g_k(R), \quad s_{tk} = \sum_i \theta_{ik} s_{itk},$$

for some basis functions $(g_k)_{k=1}^K$ and weights such that $\sum_i \theta_{ik} = 1$ for all k. For example, $K = 2$

describes Normal densities with possible disagreement about its mean and variance. Then

$$\log f_t(y) \propto \log q_t(R) - \beta_0 R + \sum_{i,k} \theta_{ik} \tilde{s}_{itk} g_k(R), \tag{13}$$

where $\tilde{s}_{itk} = s_{itk} - \sum_i \beta_i s_{itk}$ center the subjective moments by their risk-aversion weighted means. In equilibrium, investor i holds net derivative positions ω in R according to

$$\omega_{it}(R) = \text{const}_{it} + (\beta_i - \omega_{it}^0)R + \frac{1}{\gamma_i} \sum_k \tilde{s}_{itk} g_k(R), \tag{14}$$

where ω_{it}^0 are stock positions. Therefore, given data on $\omega_{it}(R)$, we can identify coefficients $\frac{1}{\gamma_i}\tilde{s}_{itk}$ by functional regression of $\omega_{it}(R)$ on R. Plugging these into (13) yields a density ratio model that is parametric in terms of observed covariates.

4.3.2 Option net demand curves

We construct the empirical counterpart to ω_{it} using an investor's net number of contracts (d_{ijt}^C, d_{ijt}^P) in call and put options with moneyness level κ_{jt}, and payoff functions $(R - \kappa_{jt})^+$ and $(\kappa_{jt} - R)^+$, respectively. For simplicity, suppose there only $N = 2$ investors, namely market-makers and end-users, which take opposite positions so that we drop the i subscript hereafter. The non-market maker's total net payoff upon the realization $R_{t+1} = R$ equals

$$\omega_{tN}(R) = \omega_{tN}^C(R) + \omega_{tN}^P(R), \tag{15}$$

which sums up the payoffs from call and put options

$$\omega_{tN}^C(R) = \sum_{i=1}^{N_t} d_{ik}^C (R - \kappa_{it})^+, \quad \omega_{tN}^P(R) = \sum_{i=1}^{N_t} d_{ik}^P (\kappa_{it} - R)^+.$$

The payoff function $\omega_{tN}(R)$ can be interpreted as the equilibrium net demand by non-market markers for Arrow-Debreu type securities that pay off in the event $R_{t+1} = R$. Unlike the call and put option net demand curves, the net demand for such assets do not compound the demand for payoffs over a range of outcomes.

For each listed strike K_i, we accumulate net demand $d_t^c(K_i) = \sum_{t-60 \le s \le t} d_s(K_i)$ over the preceding two months from daily net positions calculated as

$$d_s(K) = \text{BuyOpen}_s(K) - \text{BuyClose}_s(K) - \text{SellOpen}_s(K) + \text{SellClose}_s(K)$$

in contracts with the same strike price, put-call type, and maturity date.[6] We then match the accumulated demand $d_t^c(K_i)$ to the forward moneyness levels $\kappa_{it} = \frac{K_i}{F_t}$, and compute the estimated net demand curves $\hat{\omega}_t(R)$ using formula (15).

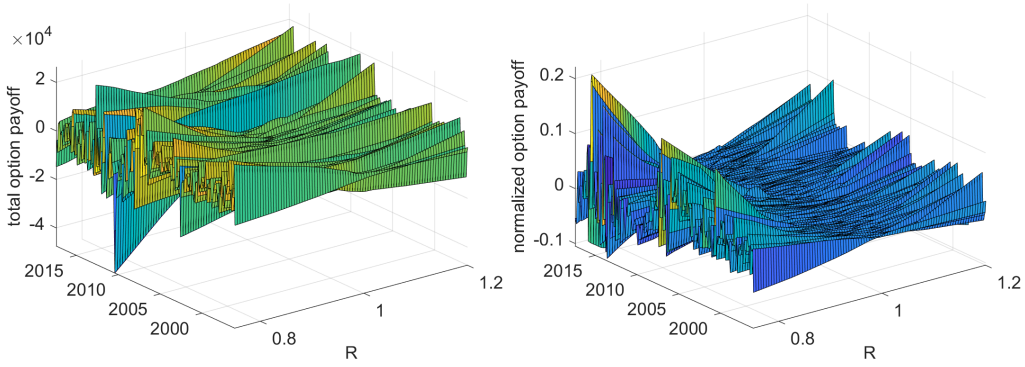
Figure 8a) shows the total monthly net option payoff $\omega_t(R)$ defined in (15) as a function of the potential end-of-month return R based on positions accumulated by the start of the month. Large imbalances occurred during the 2008-2009 financial crisis, when the sign of the net payoffs for negative returns rapidly changed. To improve comparisons over time, Figure 8b) shows the normalized net payoffs

$$\tilde{\omega}_t^c(R) = \frac{\omega_t^c(R)}{\int |\omega_t^c(y)| q_t(y) dy},$$

where the net profit curves $\omega_t^c(R) = \omega_t(R) - \pi_t(\omega_t)$ are centered by the market value of all positions:

$$\pi_t(\omega_t) = \int_0^\infty \omega_t(y) q_t(y)) dy.$$

By construction, the market value of the option portfolio with payoff function $\omega_t^c(R)$ is zero. The denominator in $\tilde{\omega}_t^c$ reflects the market value of the absolute payoffs, ensuring that $-1 \leq \tilde{\omega}_t^c(R) q_t(R) \leq 1$. The shape of the normalized payoff functions varies rapidly from month to month, with the signs of net positions for both small and large returns alternating frequently. Net positions for negative returns were most positive during the 2008-2009 financial crisis, yet turned negative afterwards.



(a) Total net payoffs.

(b) Normalized net payoffs.

Figure 8: Monthly non-market maker payoffs for the event $R_{t+1} = \kappa$ as traded in the preceding two months. Sample period shown is 1996-2017 and contains contracts with 28 days to maturity.

To construct the covariates in density ratio model (13), we estimate the coeffients in (14) for

[6] Using a cut-off prevents 'stale' positions from dominating the monthly demand curves. Using alternative cut-offs of one or three months yields similar findings.

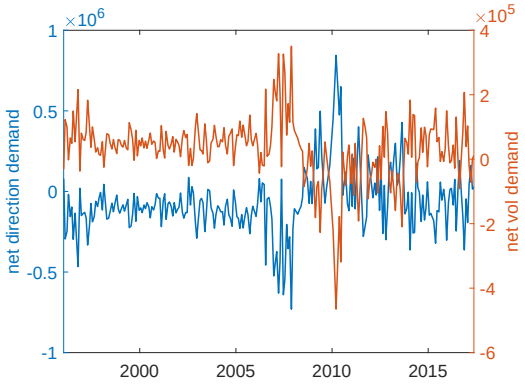


Figure 9: Monthly end-user net demand for synthetic return (left) and squared return (right), accumulating positions over the preceding two months.

some basis function $(g_k)_{k=1}^K$ by minimizing the weighted 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 $g(r) = (1, r, r^2)$, we can proxy net demand for 'directional' and 'variance risk' by their fitted coefficients $\hat{b}_t$. Figure 9 plots the resulting net demand proxies over time, which are strongly negatively correlated. Focusing on net variance demand, we see that end-users switched from buying protection against high values of the squared return, to selling it in the aftermath of the financial crisis, when the price of such protection rose dramatically.

4.3.3 Bandwidth choice

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

$$\text{MSE}_T(y, x) = b_m(y, x)^2 h^4 + \frac{\sigma_m^2(y, x)}{Th^2}.$$

Given a bandwidth function $h(y)$, the MSE-optimal bandwidth ratio h_x solves

$$\frac{\partial \text{MSE}_T(y, x)}{\partial c_x} = 2\mu_2(K)\left(m_{xx}(y, x) - \beta_{01}^2 m(y, x)\right) c_x b_m(y, x) h^4 - \frac{1}{c_x}\frac{\sigma_m^2(y, x)}{Th^2} = 0.$$

Bashtannyk and Hyndman (2001, Thm. 1) shows that for given h_y,

$$\arg\min_{h_x} \text{AMSE}(m(y,x), \hat{m}(y,x)) = \arg\min_{h_x} \text{AMSE}(\tilde{m}(y,x), \hat{m}(y,x)).$$

where $\tilde{m}(y,x) \equiv E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} \mid x_t = x\right)$. The latter equals

$$\text{AMSE}(\tilde{m}(y,x), \hat{m}(y,x)) = \tilde{b}(y,x)^2 h_x^4 + \frac{\tilde{\sigma}^2(y,x)}{T h_x},$$

where $\tilde{\sigma}^2(y,x) = \frac{R(K)}{f(x)}\text{Var}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} \mid x_t = x\right)$. Integrating over y, $\tilde{IMSE}(x) = \int \tilde{\text{MSE}}(y,x) f(y) dy$ is optimized by

$$h_x^{\tilde{\text{IMSE}}}(x) = \left(\frac{\int \tilde{\sigma}^2(y,x) f(y) dy}{T \int \tilde{b}(y,x)^2 f(y) dy}\right)^{\frac{1}{5}}, \tag{16}$$

where $\tilde{b}(y,x) = \frac{\mu_2(K)}{2}\left(\tilde{m}_{xx}(y,x) - \tilde{\beta}_{01}^2 \tilde{m}(y,x)\right)$. We can estimate the derivatives in the bias term using an initial exponential quadratic fit for every y: $\tilde{m}(y,x,\theta) = \exp\left(\theta_0(y) + \theta_1(y)x + \theta_2(y)x^2\right)$. Note $\tilde{m}_{xx}(y,x;\theta) - \tilde{\beta}_{01}(\theta)^2 \tilde{m}(y,x;\theta) = \theta_2(y)\tilde{m}(y,x;\theta)$. In particular, define

$$\hat{\tilde{\theta}}(y) = \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - \exp\left(\theta_0(y) + \theta_1(y)x_t + \theta_2(y)x_t^2\right)\right)^2,$$

and estimate $\tilde{\sigma}^2(y,x)$ by

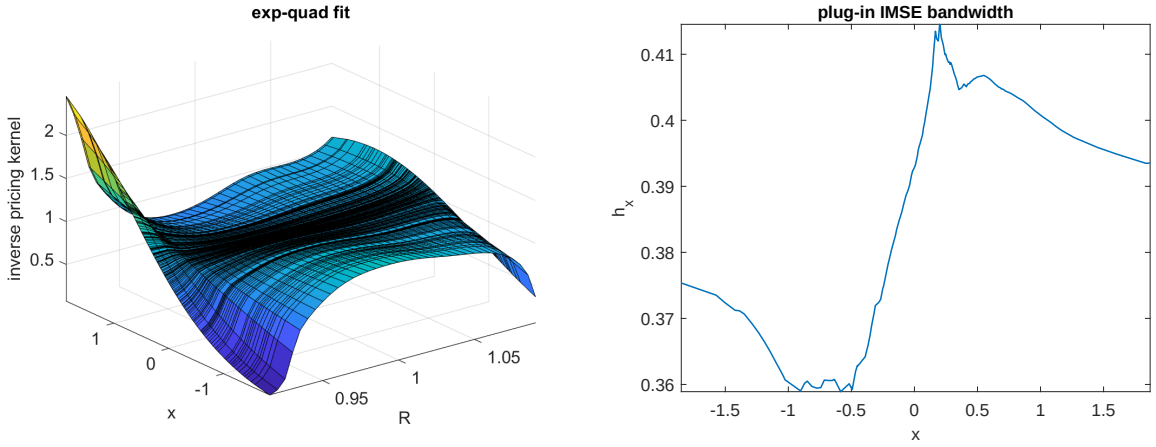
$$\hat{\tilde{\sigma}}^2(y,x) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - \exp\left(\hat{\tilde{\theta}}_0(y) + \hat{\tilde{\theta}}_1(y)x_t + \hat{\tilde{\theta}}_2(y)x_t^2\right)\right)^2 w_T(x_t, x),$$

and its weighted integral over y by $\tilde{\sigma}^2(x) = \int \hat{\tilde{\sigma}}^2(y,x)\hat{f}_h(y)dy$. The weights $w_T(x_t,x)$ require a pilot bandwidth, for example based on nearest-neighbors.

4.3.4 Results

We now fit the local exponential linear model (9) with covariate $x_t = \hat{b}_{t2}$, standardized to mean zero and variance one. To compute the plug-in bandwidth (16), we initially fit the exponential-quadratic regression of $\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}$ on x_t for any y. Figure 10a) shows the resulting fit, whose curvature explains most of the variation in x of the plug-in bandwidth in Figure 10b).

Figure 11 shows the estimated (inverse) conditional pricing kernels given the net variance demand proxies. When net variance demand is low, defined as its 25% percentile, the pricing kernel shows a similar asymmetric U-shape as when estimated without covariates, with a pronounced downward sloping left tail. Low demand is thus associated with overpriced left tail insurance.

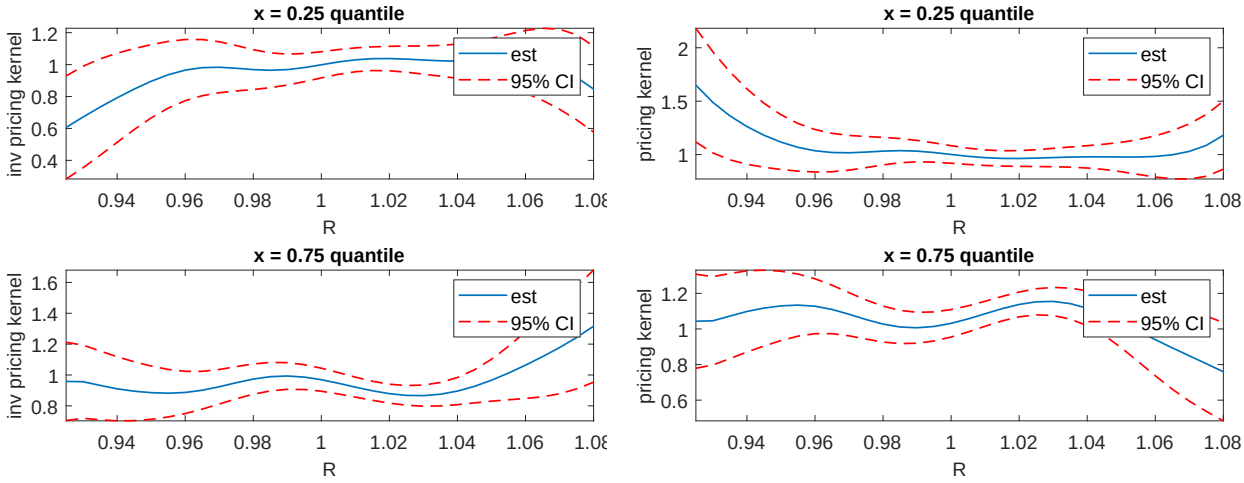


(a) Initial exp-quadratic fit.

(b) Plug-in optimal bandwidth $h_x^{\tilde{\text{IMSE}}}$.

Figure 10: Initial exponential-quadratic fit in x of the inverse pricing kernel given y using bandwidth $\hat{h}_y^{IMSE}$, and resulting $h_x^{\tilde{\text{IMSE}}}$, using data from 1996-2017.

Meanwhile, when the demand variance proxy is at its 75% percentile, the pricing kernel's left tail no longer clearly slopes downward. Thus high net variance demand, such as observed in the years before the financial crisis, is not associated with overpriced tail risks.



(a) Inverse conditional pricing kernels.

(b) Conditional pricing kernels.

Figure 11: Local exponential-linearly estimated pricing kernels (in right panels, inverse on left) conditional on net variance demand proxies at 25% and 75% percentiles, using data from 1996-2017. CIs based on plug-in standard errors and delta-method for the pricing kernel.

5 Conclusion

This paper propose a simple kernel-smoothing estimator of the multiplicative bias component in a general class of models for the ratio between observed and true conditional densities. The estimator is based on inverse density weighting, and we establish its asymptotic bias, variance, and Normality, in settings with and without conditioning variables. We also show how to perform inference on the multiplicatively corrected density forecasts. Our simulation study reports good performance of the estimator, even relative to the correctly specified parametric maximum likelihood estimator. We apply the estimator to extract conditional pricing kernels from option-implied risk-neutral densities for index returns. We find that low net variance demand is associated with a particularly high pricing kernel for left tail returns, suggesting substantial overpricing during times of financial market distress.

A Appendix

A.1 Proofs of results

Proof of Theorem 1. Given stationarity, the variance of the first term equals

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

The variance term equals

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

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 $C > 0$,

$$\sum_{j=1}^{T-1} |\text{Cov}\left(\frac{K_h(R_{j+1}-y)}{q_j(R_{j+1})}, \frac{K_h(R_1-y)}{q_0(R_1)}\right)| \leq \sum_{j=1}^{T-1} C\alpha(j)^{\frac{\delta}{4+2\delta}} \frac{1}{\sqrt{h}} = O\left(\frac{1}{\sqrt{h}}\right). \tag{17}$$

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

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

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

where the large and small blocks (ξ_i) and (η_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 ζ_T sums the remaining periods. The block sizes should be set such that

$$s_T/l_T \to 0, \quad l_T/T \to 0, \quad \frac{T}{l_T}\alpha(s_T) \to 0, \quad l_T/\sqrt{Th} \to 0 \tag{18}$$

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 (17) implies that

$$\sum_{j=1}^{T-1} |\text{Cov}(Z_{0,h}, Z_{j,h})| \to 0.$$

In combination with stationarity, this implies that $\text{Var}\,(m_T^s) = \frac{k_T s_T}{T}\Omega(y)(1+o(1)) \to 0$ and $\text{Var}\,(\xi_T) = \frac{l_T+s_T}{T}\Omega(y)(1+o(1)) \to 0$, while

$$\text{Var}\left(m_T^l\right) = \frac{k_T l_T}{T}\Omega(y)(1+o(1)) \to \Omega(y) \tag{19}$$

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 superscript τ indicate that quantities sum over $Z_{t,h}^\tau = \sqrt{h}Z_t^\tau$ rather than $Z_{t,h}$. Similar to (19), it can be shown that

$$\begin{aligned}
\text{Var}(m_T^{l\tau}) &\to R(K)m(y)E\left(\frac{c_t}{q_t(y)}1(q_t(y) \geq \tau)\right) \equiv \Omega_\tau(y) \\
\text{Var}(m_T^l - m_T^{l\tau}) &\to R(K)m(y)E\left(\frac{c_t}{q_t(y)}1(q_t(y) < \tau)\right).
\end{aligned} \tag{20}$$

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 $E\left(\left|\exp(iu(m_T^l - m_T^{l\tau})) - 1\right|\right) = O\left(\text{Var}(m_T^l - m_T^{l\tau})\right)$, which by (20) can be set arbitrary small by setting small enough τ. 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 (ξ_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

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

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 Ω_ω 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 Theorem 2. First, we show $\hat{\beta} \xrightarrow{p} \beta_0$. Write

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

where

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

Then $D(x, y, h_x, \beta_0) = 0$, while $D(x, y, h_x, \beta) \neq 0$ for any $\beta \neq \beta_0$. Therefore we establish that $\hat{\beta}$ is consistent by proving the uniform convergence

$$\sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} Q_T(y, x, \beta) - D(x, y, h_x, \beta) \right\| \xrightarrow{p} 0. \tag{21}$$

By the mean-value theorem and Cauchy-Schwartz inequality, for $j = 0, 1$, and any β and $\tilde{\beta}$

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

The Hessian is given by

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

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

$$\begin{aligned}
E\left(\sup_{\beta\in\Theta}\left|\frac{\partial^2 Q_T(x,y,\beta)}{\partial^2\beta_0}\right|\right) &\leq 2E\left(\sup_{\beta\in\Theta}\left|\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})} - 2\exp\left(\beta^T\tilde{x}_t\right)\right|\exp\left(\beta^T\tilde{x}_t\right)K_{h_x}(x_t-x)\right) \\
&\leq 2E\left(\sup_{\beta\in\Theta}\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\exp\left(\beta^T\tilde{x}_t\right)K_{h_x}(x_t-x)\right) \\
&\quad + 4E\left(\sup_{\beta\in\Theta}\exp\left(2\beta^T\tilde{x}_t\right)K_{h_x}(x_t-x)\right) \\
&\leq 2E\left(E\left(\frac{K_h(R_{t+1}-y)}{q_t(R_{t+1})}\mid x_t\right)\exp\left(\beta_{\cdot 0}^u + (|\beta_{\cdot 1}^u| + |\beta_{\cdot 1}^l|)|x_t-x|\right)K_{h_x}(x_t-x)\right) \\
&\quad + 4E\left(\exp\left(2\beta_{\cdot 0}^u + 2(|\beta_{\cdot 1}^u| + |\beta_{\cdot 1}^l|)|x_t-x|\right)K_{h_x}(x_t-x)\right) \\
&= O(1),
\end{aligned}$$

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 $\sup_{\beta\in\Theta}\left\|\frac{\partial^2 Q_T(x,y,\beta)}{\partial^2\beta_0}\right\| = O_p(1)$. Therefore $\frac{\partial}{\partial\beta}Q_T(y,x,\beta)$ is stochastically equicontinuous, which confirms (21) using Newey (1991, Corollary 2.2).

Consistency and β_0 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,

$$0 = \frac{\partial}{\partial\beta}Q_T(y,x,\hat{\beta}) = \frac{\partial}{\partial\beta}Q_T(y,x,\beta_0) + \frac{\partial^2 Q_T(x,y,\beta^*)}{\partial\beta\partial\beta^T}(\hat{\beta}-\beta_0), \tag{22}$$

for some β^* such that β_j^* lies in between $\hat{\beta}_j$ and $\beta_{0,j}$ almost surely. Re-arranging (22) and scaling by $\sqrt{Th^2}H^{-1}$ yields

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

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) $\sqrt{Th^2}S_T\left(x,y,\beta_0\right)$ satisfies a multivariate central limit theorem.

Step (1): Write

$$A_T\left(x,y,\beta^*\right) = R_{T1}\left(x,\beta_0\right) + \left\{R_{T1}\left(x,\beta^*\right) - R_{T1}\left(x,\beta_0\right)\right\} + R_{T2}\left(x,y,\beta^*\right),$$

where

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

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

Let $\|\cdot\|$ denote the Euclidean norm of a vector or matrix. We will show below that

$$\|R_{T1}(x,\beta_0) - A_{T0}(x)\| = \mathrm{o_p}(h_x), \tag{23a}$$

$$\|R_{T1}(x,\beta^*) - R_{T1}(x,\beta_0)\| = \mathrm{o_p}(1), \tag{23b}$$

$$\|R_{T2}(x,y,\beta^*)\| = \mathrm{o_p}(1), \tag{23c}$$

where $A_{T0}(x)$ is a positive-definite limit matrix given below.

Using second-order Taylor expansions around $x_t = x$, the means of the terms in $R_{T1}(x,\beta_0)$ are

$$\begin{aligned} E\left(\exp\left(2\beta_0^T\tilde{x}_t\right) K_{h_x}(x_t - x)\right) &= \int \exp\left(2\beta_{00} + 2\beta_{01}(x_t - x)\right) K_{h_x}(x_t - x) f(x_t)dx_t \\ &= \int \exp\left(2\beta_{00} + 2\beta_{01}h_x z\right) K(z) f(x + zh_x)dz \\ &= m(y,x)^2 f(x) + O(h_x^2), \end{aligned}$$

$$\begin{aligned} E\left(\exp\left(2\beta_0^T\tilde{x}_t\right)\left(\frac{x_t - x}{h_x}\right) K_{h_x}(x_t - x)\right) &= \int \exp\left(2\beta_{00} + 2\beta_{01}h_x z\right) zK(z) f(x + zh_x)dz \\ &= m(y,x)^2\mu_2(K)\left(2\beta_{01}f(x) + f'(x)\right)h_x + O(h_x^3), \end{aligned}$$

and

$$\begin{aligned} E\left(\exp\left(2\beta_0^T\tilde{x}_t\right)\left(\frac{x_t - x}{h_x}\right)^2 K_{h_x}(x_t - x)\right) &= \int \exp\left(2\beta_{00} + 2\beta_{01}h_x z\right) z^2K(z) f(x + zh_x)dz \\ &= m(y,x)^2\mu_2(K)f(x) + O(h_x^2). \end{aligned}$$

Therefore,

$$E\left(R_{T1}(x,\beta_0)\right) = 2m(y,x)^2 \begin{pmatrix} f(x) & \mu_2(K)r_{12}(x)h_x \\ \mu_2(K)r_{12}(x)h_x & \mu_2(K)f(x) \end{pmatrix} + \begin{pmatrix} O\left(h_x^2\right) & O\left(h_x^3\right) \\ O\left(h_x^3\right) & O\left(h_x^2\right) \end{pmatrix} \equiv A_{T0}(x) + O(h_x^2),$$

where $r_{12}(x) = 2\beta_{01}f(x) + f'(x)$. Furthermore, the variance of each element in $R_{T1}(x, \beta_0)$ is $O\left(\frac{1}{Th_x}\right)$ under the mixing condition. Therefore (23a) follows under the bandwidth condition 2g).

By the mean-value theorem, there exists some $\tilde{\beta}$ in between β^* and β_0 such that

$$R_{T1}(x, \beta^*) - R_{T1}(x, \beta_0) = \frac{4}{T}\sum_{t=1}^{T} (\beta^* - \beta_0)^T \tilde{x}_t \exp\left(2\tilde{\beta}^T\tilde{x}_t\right) H^{-1}\tilde{x}_t \left(H^{-1}\tilde{x}_t\right)^T K_{h_x}(x_t - x).$$

Therefore (23b) follows from

$$\begin{aligned} & E\left(\|R_{T1}(x, \beta^*) - R_{T1}(x, \beta_0)\|\right) \\ & \leq 4E\left(\| (\beta^* - \beta_0)^T \tilde{x}_t \exp\left(2\tilde{\beta}^T\tilde{x}_t\right) H^{-1}\tilde{x}_t \left(H^{-1}\tilde{x}_t\right)^T K_{h_x}(x_t - x)\|\right) \\ & \leq 4\|\beta^* - \beta_0\| \int \|(1, zh_x)\| E\left(\exp\left(2\tilde{\beta}^T(1, zh_x)\right)\right) \|\tilde{z}\tilde{z}^T\| K(z) f(x + zh_x) dz \\ & \to 0, \end{aligned} \tag{24}$$

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

To establish (23c), write

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

Using second-order expansions, $E\left(T_{T1}(x, y, \beta)\right) = O(h^2)$ for any β, while $E\left(T_{T2}(x, y, \beta_0)\right) = O(h_x^2)$. Furthermore, $\mathrm{Var}\left(T_{T1}(x, y, \beta)\right) = O\left(\frac{1}{Th^2}\right)$ and $\sup_{\beta\in\Theta}\frac{\partial}{\partial\beta}T_{T1}(x, y, \beta) = O_p(1)$, so that $T_{T1}(x, y, \beta) \xrightarrow{p} 0$ uniformly over β. Moreover, $\mathrm{Var}\left(T_{T2}(x, y, \beta_0)\right) = O\left(\frac{1}{T}\right)$, and $\|T_{T2}(x, y, \beta^*) - T_{T2}(x, y, \beta_0)\| = o_p(1)$ using similar steps as (24). Therefore $\|T_{Tj}(x, y, \beta^*)\| = o_p(1)$ for $j = 1, 2$, implying (23c).

Step (2): Write $S_T(x, y, \beta_0)$ as

$$S_T(x, y, \beta_0) = 2m(y, x)\left(S_{T1}(x, y, \beta_0) + S_{T2}(x, y, \beta_0) + S_{T3}(x, y, \beta_0)\right),$$

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

$$S_{T1}(x, y, \beta_0) = \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h}\tilde{x}_{t,h}K_{h_x}(x_t - x),$$
$$S_{T2}(x, y, \beta_0) = \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h}\left(\exp\left(\beta_{01}(x_t - x)\right) - 1\right)\tilde{x}_{t,h}K_{h_x}(x_t - x),$$
$$S_{T3}(x, y, \beta_0) = \frac{1}{T}\sum_{t=1}^{T}\left\{E\left(\frac{K_h(R_{t+1} - y)}{q_t(R_{t+1})} \mid x_t\right) - \exp\left(\beta_0^T\tilde{x}_t\right)\right\}\exp\left(\beta_{01}(x_t - x)\right)\tilde{x}_{t,h}K_{h_x}(x_t - x),$$

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 $E\left(S_{T1}(x, y, \beta_0)\right) = E\left(S_{T2}(x, y, \beta_0)\right) = 0$.

We will show below that

$$\sqrt{Th^2}S_{T1}(x, y, \beta_0) \xrightarrow{d} \mathrm{N}\left(0, V(x, y)\right), \tag{25a}$$
$$\left\|\sqrt{Th^2}S_{T2}(x, y, \beta_0)\right\| = \mathrm{O_p}(h), \tag{25b}$$
$$\left\|S_{T3}(x, y, \beta_0) - h^2 Ha(x, y)\right\| = \mathrm{O_p}\left(h^4\right), \tag{25c}$$

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,

$$\begin{aligned}\mathrm{Var}\left(\sqrt{Th^2}S_{T1}(x, y, \beta_0)\right) &= h^2\mathrm{Var}\left(u_{t+1,h}\tilde{x}_{t,h}K_h(x_t - x)\right) \\ &\quad + 2h^2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\mathrm{Cov}\left(u_{j+1,h}\tilde{x}_{j,h}K_h(x_j - x), u_{1,h}\tilde{x}_{0,h}K_h(x_0 - x)\right).\end{aligned}$$

The variance term equals

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

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

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

Therefore, for some constant $C>0$,

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

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

Asymptotic normality of $S_{T1}(x,y,\beta_0)$ 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 (18). 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 $\exp(\beta_{01}(x_t-x))\tilde{x}_t$ yield some x_t^* between $x_t - x$ and 0 such that

$$S_{T2}(x,y,\beta_0) = \frac{1}{T}\sum_{t=1}^{T} u_{t+1,h}\exp(\beta_{01}x_t^*) \begin{pmatrix} \beta_{01} \\ 2\beta_{01} + \beta_{01}^2 x_t^* \end{pmatrix} (x_t - x)\,\tilde{x}_{t,h}K_{h_x}(x_t-x),$$

using $\frac{\partial}{\partial x}\exp(\beta x)x = \exp(\beta x)(1+\beta x)$ and $\frac{\partial^2}{\partial^2 x}\exp(\beta x)x = \beta\exp(\beta x)(2+x)$. Using similar steps

as for $S_{T1}(x,y,\beta_0)$, it follows that $\text{Var}\left(S_{T2}(x,y,\beta_0)\right) = O\left(\frac{h_x^2}{Th^2}\right) = O\left(\frac{1}{T}\right)$.

Performing second- and third-order Taylor expansions for the constant and slope term, respectively, in $S_{T3}(x,y,\beta_0)$ allows obtaining its leading terms as

$$\begin{aligned}
S_{T3}(x,y,\beta_0) &= \frac{1}{2T}\sum_{t=1}^{T}\Big\{\mu_2(K)m_{yy}(y,x_t)h^2 + \left(m_{xx}(y,x) - \beta_{01}^2 m(y,x)\right)(x_t-x)^2 + \mu_2(K)m_{yyx}(y,x)(x_t-x) \\
&\quad + \frac{1}{3}\left(m_{xxx}(y,x) - \beta_{01}^3 m(y,x)\right)(x_t-x)^3\Big\}\left(1+\beta_{01}(x_t-x)\right)\tilde{x}_{t,h}K_{h_x}(x_t-x) + O_p(h^4), \\
&= \frac{1}{2}h^2 H \begin{pmatrix} \mu_2(K)f(x)\left(m_{yy}(y,x) + \left(m_{xx}(y,x) - \beta_{01}^2 m(y,x)\right)c_x^2\right) + O_p(h^2) \\ \left(\mu_2^2(K)m_{yyx}(y,x)c_x + \frac{1}{3}\mu_4(K)\left(m_{xxx}(y,x) - \beta_{01}^3 m(y,x)\right)c_x^3\right)f(x) \\ + \left(\mu_2^2(K)m_{yy}(y,x)c_x + \mu_4(K)\left(m_{xx}(y,x) - \beta_{01}^2 m(y,x)\right)c_x^3\right)\left(\beta_{01}f(x) + f'(x)\right) + O_p(h \end{pmatrix} \\
&\equiv h^2 H a(y,x) + O_p(h^4).
\end{aligned}$$

where the term $\beta_{01}f(x) + f'(x)$ 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

$$\begin{aligned}
E\left(\hat{\beta} - \beta_0\right) &= H^{-1}A_{T0}(x)^{-1}2m(y,x)h^2 H a(x,y) + o(h^2) \\
&= h^2 m(y,x)^{-1}f(x)^{-1}\begin{pmatrix} 1 & -f(x)^{-1}r_{12}(x)h_x \\ -f(x)^{-1}r_{12}(x) & \mu_2(K)^{-1}h_x^{-1} \end{pmatrix}\begin{pmatrix} a_1(x,y) \\ h a_2(x,y) \end{pmatrix} + o(h^2) \\
&= h^2 m(y,x)^{-1}f(x)^{-1}\begin{pmatrix} a_1(x,y) \\ -f(x)^{-1}r_{12}(x)a_1(x,y) + c_x^{-1}\mu_2(K)^{-1}a_2(x,y) \end{pmatrix} + o(h^2) \\
&\equiv h^2 b(x,y) + o(h^2),
\end{aligned}$$

while its asymptotic variance equals

$$\begin{aligned}
\text{Var}\left(\frac{1}{\sqrt{Th^2}}(\hat{\beta} - \beta_0)\right) &= A_{T0}(x)^{-1}4m(y,x)^2 V(x,y)A_{T0}(x)^{-1} + o(1) \\
&= m(y,x)^{-1}f(x)^{-1}c_x^{-1}R_0(K)\text{diag}(R_0(K), R_2(K)/\mu_2^2(K))E\left(\frac{c_t}{q_t(y)} \mid x_t = x\right) \\
&\equiv \Omega(x,y) + o(1).
\end{aligned}$$

□

A.2 Further simulation results

Table 3: 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. Optimal bandwidths chosen based on exponential-quadratic fit. Columns describe integrated weighted squared bias, variance, and mean squared error, truncated at the (0.025,0.975)-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	0.16	4.09	4.25	0.08	1.82	1.90	0.03	0.78	0.81
h-par	0.33	12.38	12.71	0.29	6.07	6.35	0.15	11.73	11.87
h-par iter	0.25	4.03	4.29	0.15	2.07	2.22	0.08	1.11	1.19
h-par trim	0.39	3.18	3.56	0.38	1.51	1.89	0.30	0.75	1.05
h-pilot	0.17	3.84	4.01	0.09	2.11	2.20	0.08	1.04	1.12
h-pilot iter	0.16	4.06	4.22	0.08	2.25	2.33	0.06	1.11	1.17
h-pilot trim	0.22	3.65	3.87	0.20	1.95	2.14	0.20	0.92	1.13

(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	0.13	2.95	3.08	0.05	1.26	1.31	0.02	0.53	0.55
h-par	0.37	6.70	7.07	0.33	4.59	4.91	0.19	1.93	2.12
h-par iter	0.28	3.25	3.53	0.17	2.26	2.42	0.06	1.16	1.21
h-par trim	0.35	2.70	3.05	0.33	1.49	1.82	0.19	0.66	0.85
h-pilot	0.18	2.78	2.96	0.09	1.85	1.94	0.04	1.03	1.06
h-pilot iter	0.13	2.91	3.04	0.08	1.93	2.01	0.03	1.09	1.12
h-pilot trim	0.16	2.61	2.76	0.13	1.65	1.78	0.07	0.82	0.89

References

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*.

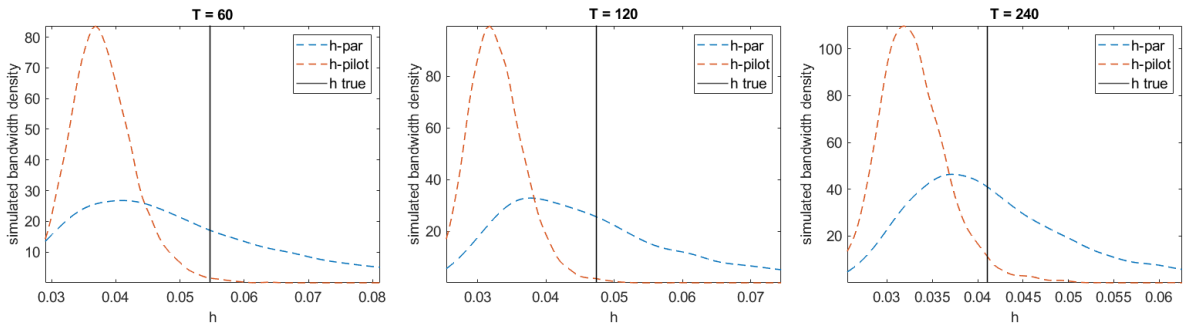


Figure 12: Simulated densities of plug-in asymptotic IMSE optimal fixed bandwidths under the AR(1)-GARCH(1,1)-t model stochastic volatility-model, using initial exponential-cubic (h-par) and local pilot (h-pilot) estimators, for $S = 10,000$ simulated time series and sample sizes $T = \{60, 120, 240\}$ months. IMSE truncated at (0.025,0.975)-unconditional quantiles.Vertical lines show corresponding optimal bandwidths if m_0 were known.

Bashtannyk, D. M. and Hyndman, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298.

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

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.

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.

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.

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.

Dalderop, J. (2021). Efficient estimation of pricing kernels and market-implied densities.

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.

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

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.

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.

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