Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models.

Matt Pritsker*

November 21, 1997

Abstract

A number of recent papers have used nonparametric density estimation or nonparametric regression to study the instantaneous spot interest rate, and to test term structure models. However, little is known about the performance of these methods when applied to persistent time-series, such as U.S. interest rates. This paper uses the Vasicek [1977] model to study the performance of kernel density estimates of the ergodic distribution of the instantaneous spot rate. The model's tractability allows me to analyze the MISE of the kernel estimate as a function of persistence, variance of the ergodic distribution, span of the data, sampling frequency, and kernel bandwidth. Our principle result is that persistence has an important impact on optimal bandwidth selection and on finite sample performance. We also find that sampling the data more frequently has little effect on estimator quality. We also examine one of Ait-Sahalia's [1996a] new nonparametric tests of parametric continuous-time Markov models of the instantaneous spot interest rate. The test is based on the distance between parametric and nonparametric (kernel) estimates of the ergodic distribution of the interest rate process. Our principal result is that the test rejects too often when using asymptotic critical values and 22 years of data. The reason for the high rejection rate is probably because the asymptotic distribution of the test does not depend on persistence, but the finite sample performance of the estimator does. After critical values are adjusted for size, the test has low power in distinguishing between the Vasicek and Cox-Ingersoll-Ross models when compared with a conditional moment based specification test.

Keywords: Interest Rate, Nonparametric, Bandwidth, Specification Test.
JEL Classification Numbers: C12, C14, G12.

*Board of Governors of the Federal Reserve System. The author thanks Yacine Ait-Sahalia, Torben Andersen, and Mark Fisher for useful comments and thanks Ruth Wu for valuable research assistance. All errors are the author's fault. The views expressed in this paper reflect those of the author and not those of the Board of Governors of the Federal Reserve System or other members of its staff. Address correspondence to Matt Pritsker, The Federal Reserve Board, Mail Stop 91, Washington DC 20551. The author may also be reached by telephone: (202) 452-3534, FAX: (202) 452-5296, and email: mpritsker@frb.gov.

I Introduction.

A number of recent papers (Stanton [1995], Ait-Sahalia [1996a,1996b,1996c], Siddique[1994]) have used nonparametric density estimation or nonparametric regression to study the behavior of the instantaneous spot interest rate, and to test theories of the term structure. These nonparametric methods have potential to enhance our understanding of the term structure because they may capture features of term structure dynamics which are missed or misspecified in parametric approaches. Although nonparametric methods have advantages over more parametric approaches, they also have important drawbacks including: (1) They require very large amounts of data; (2) most results are based on asymptotics; relatively little is known about finite sample properties; (3) little is known about their performance (except asymptotically) when applied to persistent time series of the type encountered in interest rate studies; and (4) the methods require a choice of bandwidth (smoothing parameter), but most results on optimal bandwidth selection have only been worked out for i.i.d. data, or asymptotically for time series data. Choosing the bandwidth optimally with finite sample time series data is far more difficult.

As noted in Ball and Torous [1996] drawbacks one through three also apply to parametric estimation methods when time series processes (such as interest rates) contain a root that is close to unity.[1] The purpose of this paper is to begin an examination of the importance of near unit root behavior for nonparametric kernel density estimation. More specifically, in this paper we examine the finite sample performance of kernel density estimation when the spot rate follows a continuous time AR(1) (Ornstein-Uhlenbeck process) as in Vasicek. We chose to focus on Vasicek's model not because it is fully realistic, but rather because it is highly tractable. This tractability allows us to study the role of non-i.i.d. data and persistence in kernel density estimation of the process' ergodic distribution in a finite sample setting. In particular, the tractability allows us to analytically solve for the finite sample Bias, Variance, Covariance, and Mean Integrated Squared Error (MISE) of the kernel density estimate as a function of the persistence of the data generating process, variance of the ergodic distribution, span of the data, the frequency with which the data is sampled, and the kernel bandwidth. We also solve (not analytically) for the kernel bandwidth which minimizes the Mean Integrated Squared Error.

The second contribution of this paper is that we use the insights gained from studying kernel density estimation of the Vasicek process in order to examine the properties of one

[1] It is well known that interest rates are highly persistent. Using monthly data from McCulloch and Kwon [1993], Pagan, Hall, and Martin [1995], estimate that one-month zero coupon rates have a first order autocorrelation coefficient of 0.98. They fail to reject the null hypothesis that one month zero coupon rates are integrated, and conclude that nominal interest rates are integrated or nearly integrated.

of Yacine Ait-Sahalia's [1996a] tests of parametric continuous time Markov models of the spot interest rate. This test is based on the distance between parametric and nonparametric estimates of the ergodic distribution of the interest rate process. When this distance is too large, the parametric model is rejected. Ait-Sahalia derived the distribution of his test statistic asymptotically. In this paper we examine the small sample properties of the test.

The remainder of the paper contains three sections. Section II discusses kernel estimation and bandwidth selection in general. This section also presents our results on kernel estimation and bandwidth selection for the Vasicek model. Section III describes Ait-Sahalia's test, and describes our results on the size and power of the test in finite samples. Section IV concludes.

II Kernel Density Estimation.

A. The kernel estimation problem.

Suppose the instantaneous spot rate of interest r follows a stationary Markov diffusion process of form:

$$dr = \mu(r)dt + \sigma(r)dw.$$

For times t and s, $t < s$, denote the probability that r at time s is equal to $r_{(s)}$ given r at time t is equal to $r_{(t)}$ by the transition density $\pi(r_{(s)}, s|r_{(t)}, t)$, and assume $\mu(r)$ and $\sigma(r)$ satisfy regularity conditions sufficient to guarantee the existence of joint distribution function $\pi(r_{(s)}, s, r_{(t)}, t)$, and long run ergodic distribution $\pi(u)$ where $\pi(u)$ represents the unconditional probability that r is equal to u.

A kernel density estimate of $\pi(u)$ is a smoothed estimate of the ergodic density of r at point u. With N discrete observations of the interest rate process, the kernel density estimate has form:

$$\hat{\pi}(u) = \frac{1}{Nh}\sum_{i=1}^{N} K(\frac{u - r_{(i)}}{h})$$

where K is the kernel function and h ,the kernel bandwidth, is a parameter set to choose the degree of smoothing in the kernel density estimate.

For time series estimation, the quality of the kernel density estimate will depend on the span of the data(T) and the sampling frequency (ΔT). To explicitly incorporate these factors in the kernel density estimate, suppose there are T years of data which are sampled every ΔT years. Then there are $N = \frac{T}{\Delta T}$ observations, and the kernel density estimate can be expressed as:

$$\hat{\pi}(u) = \frac{1}{Th}\sum_{s=\Delta T}^{N\Delta T} K(\frac{u - r_{(s)}}{h})\Delta T. \tag{1}$$

For a given span of data, it may be possible to improve the kernel density estimate by sampling more frequently. In the limit, the best one could do is sample continuously. This corresponds to allowing $\Delta T \to 0$ which causes the density estimate in (1) to converge to the continuous sampling kernel density estimate:

$$\hat{\pi}(u) = \frac{1}{Th}\int_0^T K(\frac{u - r_{(s)}}{h})ds \tag{2}$$

I will present theoretical results below for both the discrete and continuous sampling kernel estimates.[2]

The most important element of kernel density estimation is choice of bandwidth. Asymptotic results on kernel density estimation place restrictions on the rate at which the bandwidth shrinks with sample size. In a sample with fixed size, all bandwidth choices produce bonified density estimates. This leaves the choice of bandwidth to the discretion of the econometrician. For data description purposes, an extra element of discretion is fine since choosing different bandwidths emphasize different aspects of the data.[3], but for some purposes, such as hypothesis testing, a single density estimate is desirable.[4] Moreover, in order to scientifically replicate the testing method, it is important to choose the bandwidth in a non-subjective way that can be implemented by other researchers. This requires choosing the bandwidth by some criterion, such as minimizing a statistical loss function. The most common loss function considered is the Mean Integrated Squared Error (MISE) of the density estimate:

$$\begin{aligned} MISE &= E\int_u [\hat{\pi}(u) - \pi(u)]^2\, du \\ &= E\int_u [\hat{\pi}(u) - E\hat{\pi}(u)]^2\, du + E\int_u [E\hat{\pi}(u) - \pi(u)]^2\, du \end{aligned}$$

[2] The continuous sampling kernel estimate is a theoretical abstraction, but it is a very useful one because it provides an upper bound on the performance of kernel density estimates with discrete sampling. In addition, it is very tractable, and sometimes provides a good approximation for the behavior of kernel density estimates that involve frequent discrete sampling.

[3] When estimating $\pi(u)$, larger bandwidths involve more averaging of data across observations far from u. The averaging reduces the variance of $\hat{\pi}(u)$ and emphasizes more "global" features of the density. However, by smoothing across realizations of r that are relatively far from u, local features of the data are lost; this introduces bias. Smaller bandwidths have the opposite effect. Silverman [1986] and Scott [1992] suggest examining density estimates using several different bandwidths since each reveals different aspects of the data.

[4] Multiple density estimates may generate conflicting tests which need to be reconciled based on some criterion.

$$= MIVAR + MISB.$$

The MISE is the expected squared difference between the estimated and true densities integrated over the range of r, and is one measure of the distance between the true and estimated densities. As shown in the second and third lines, the Mean Integrated Squared Error is the sum of the Mean Integrated Variance (MIVAR) and the Mean Integrated Squared Bias (MISB).

Many bandwidth selection methods are designed to choose the bandwidth to minimize a constructed proxy for MISE.[5] One common method assumes the data is drawn from a particular parametric distribution (such as all draws are i.i.d. Gaussian) and chooses an optimal bandwidth based on parametric estimates of the parameters of the distribution.[6] Other methods, such as least squares cross-validation, construct a proxy for MISE, and then choose the bandwidth to minimize that proxy.[7] In finite samples, the first of these methods is only appropriate for i.i.d. data; the second method is valid asymptotically, but is known to generate downward biased bandwidth estimates in finite samples with dependent observations. We will investigate the performance of these bandwidth selection methods in more detail for the specific case of the Vasicek model.

B. Kernel estimation and the Vasicek model.

Vasicek [1977] used the absence of arbitrage to solve for bond prices when interest rates evolve according to the continuous time AR(1):

$$dr = \kappa(\theta - r)dt + \sigma dw. \tag{3}$$

Under this process, r has ergodic distribution:

$$\pi(r) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2\kappa}}} \exp\left[-.5\left(\frac{(r-\theta)}{\sqrt{\frac{\sigma^2}{2\kappa}}}\right)^2\right], \tag{4}$$

[5] MISE is not observable.

[6] For example, the MISE minimizing bandwidth for i.i.d. draws using Gaussian data is $1.06\sigma N^{-1/5}$. The bandwidth that is used in practice would be $1.06\hat{\sigma}N^{-1/5}$, where $\hat{\sigma}$ is a parametric estimate of σ.

[7] This procedure is appealing because Stone[1984] showed for i.i.d. data that the bandwidth chosen by least squares cross validation converges to one that minimizes the mean integrated squared error of the density estimate as sample size grows. However, Park and Marron [1990] show that it produces very noisy estimates of optimal bandwidth in small samples.

and r has conditional distribution:

$$
\begin{aligned}
\pi(r_{(s)}, s|r_{(t)}, t) &= \frac{1}{\sqrt{2\pi V(r_{(s)}|r_{(t)})}} \exp\left[-.5\left(\frac{(r-\mu(r_{(s)}|r_{(t)}))}{\sqrt{V(r_{(s)}|r_{(t)})}}\right)^2\right] \qquad (5)\\
&\text{where:}\\
\mu(r_{(s)}|r_{(t)}) &= \theta + (r_{(t)} - \theta)e^{-\kappa(s-t)}\\
V(r_{(s)}|r_{(t)}) &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(s-t)}).
\end{aligned}
$$

The Vasicek model has three convenient properties. The first is that the ergodic and conditional distributions of r given in equations (4)and (5) are gaussian. When the kernel is also gaussian, this greatly simplifies computation of the expected kernel density estimate and MISE. The second and third properties are that the ergodic distribution is homogenous of degree 0 in σ^2 and κ, and the conditional distribution is homogenous of degree 0 in κ, σ^2, and $\frac{1}{t-s}$. As I will show below, these features make it possible to study the interaction of persistence, sampling frequency, and span of the data on kernel estimation while holding the ergodic distribution fixed.

Our results on the theoretical properties of kernel estimation in the Vasicek model are broken into two subsections. The first summarizes our finite sample results on kernel density estimation; the second contrasts the finite sample properties with the asymptotic properties.

Finite sample results.

This section provides analytical or near analytical finite sample expressions for the Bias, Variance, Covariance, and Mean Integrated Squared Error of the kernel density estimate as a function of the data generating process, span of the data, frequency with which it is sampled, and the kernel bandwidth. These analytical expressions are then used to study how MISE and the optimal bandwidth change with sampling frequency, span of the data, and persistence of the data generating process.

Proposition I presents analytical expressions for the bias, and MISE for kernel estimates of the ergodic distribution of the Vasicek process when the process is sampled at N discrete intervals each of length Δt. The expression for MISE was derived earlier by Wand[92] using a technique that is different than the one used here. Expressions for the finite sample variance and covariance of the kernel density estimate are contained in appendices C and D respectively.

Proposition I. If $r_{(0)}$ is drawn from the ergodic distribution of the Vasicek process given in equation (4), and r then evolves according to the diffusion (3), and $\hat{\pi}(u) = \frac{1}{Nh}\sum_{s=\Delta t}^{N\Delta t} K(\frac{u-r_{(s)}}{h})$, *where K(.) is the standard Gaussian kernel then:*

$$
\begin{aligned}
E\hat{\pi}(u) &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h^2+V_E}}\exp\left[-.5\left(\frac{u-\theta}{\sqrt{h^2+V_E}}\right)^2\right], && (6)\\
MISB(\hat{\pi}(u)) &= \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sqrt{2h^2+2V_E}}+\frac{1}{\sqrt{2V_E}}-\frac{2}{\sqrt{(h^2+2V_E)}}\right), && (7)\\
MIVAR(\hat{\pi}(u)) &= \frac{1}{N^2\sqrt{2\pi}}\left[\frac{N}{\sqrt{2h^2}}+2\sum_{l=2}^{N}\sum_{m=1}^{l-1}\frac{1}{\sqrt{2h^2+2V_E(1-e^{-\kappa(l-m)\Delta t})}}\right.\\
&\quad \left. -\frac{N^2}{\sqrt{2h^2+2V_E}}\right] && (8)\\
MISE(\hat{\pi}(u)) &= MISB(\hat{\pi}(u))+MIVAR(\hat{\pi}(u)), && (9)
\end{aligned}
$$

where, $V_E = \frac{\sigma^2}{2\kappa}$.
Proof: See the appendix.

More information about the Vasicek process is learned if the data is sampled at more frequent intervals for a given span of data. An upper bound of the information in the sample comes from a kernel density estimate with continuous sampling. Expressions for the bias and MISE for kernel estimation with continuous sampling are below; results on variance and covariance are in appendices C and D respectively.

Proposition II. If $r_{(0)}$ is drawn from the ergodic distribution of the Vasicek process given in equation (4), and r then evolves according to the diffusion (3), and $\hat{\pi}(u) = \frac{1}{Th}\int_0^T K(\frac{u-r_{(s)}}{h})ds$, where K(.) is a Gaussian kernel, then $E\hat{\pi}(u)$ and $MISB$ are as in equations (6) and (7), and:

$$
\begin{aligned}
MIVAR(\hat{\pi}(u)) &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2h^2+2V_E}}\frac{4}{\kappa T^2}\int_0^T \ln\left[\frac{1+\sqrt{1-\psi(t)}}{1+\sqrt{1-\psi(0)}}\right]dt, && (10)\\
MISE(\hat{\pi}(u)) &= MISB(\hat{\pi}(u))+MIVAR(\hat{\pi}(u)), && (11)
\end{aligned}
$$

where, $V_E = \frac{\sigma^2}{2\kappa}$, and $\psi(t) = \left(\frac{V_E}{h^2+V_E}\right)\exp(-\kappa t)$.
Proof: See the appendix.

The expression for MISE in propositions I is analytical and very easy to compute. The expression for MISE in proposition II is not closed form, but can be computed easily.[8] This tractability shows that in this case the form of the bias of the kernel density estimate[9] is

[8]The expected kernel density estimate and its mean integrated squared bias (MISB) are typically not available in closed form, but they are for the Vasicek model. Similarly, the general expression for mean integrated variance (MIVAR) in proposition II is a five dimensional integral, but it reduces to a one-dimensional integral for the Vasicek model.

[9]Kernel density estimates produce biased estimates of the true density in finite samples.

that the expected kernel density estimate (see equation (6)) has a larger variance than the true distribution.[10]

This tractability is especially useful for analyzing the MISE and the choice of optimal bandwidth as a function of mean reversion (κ), data span (T), sampling frequency (ΔT), and variance of the ergodic distribution (V_E). Let $MISE(\kappa, T, \Delta T, V_E)$ represent the MISE as functions of these parameters, and let h^* be the MISE minimizing bandwidth. Some of the properties of MISE(.) and h^* are summarized in proposition III:

Proposition III. Under the conditions of proposition I, for fixed N and V_E,:

1. $MISE(\kappa, T, \Delta T, V_E)$ *and* h^* *are homogenous of degree 0 in* κ, $\frac{1}{\Delta T}$ *and* $\frac{1}{T}$.
2. $\lim_{\kappa \to \infty} MIVAR(\hat{\pi(u)}) = N^{-1}(2\pi)^{-.5}[(2h^2)^{-.5} - (2h^2 + 2V_E)^{-.5}]$.
3. $\lim_{\kappa \to 0} MIVAR(\hat{\pi(u)}) = (2\pi)^{-.5}[(2h^2)^{-.5} - (2h^2 + 2V_E)^{-.5}]$.

Proof. Result 1 follows from inspection of equation (8) and results 2 and 3 follow from taking limits.

Proposition III examines the role of mean reversion, as measured by κ on the quality of the kernel density estimate while holding the ergodic distribution fixed. The first result is useful for thinking about the span of data necessary to achieve reasonable estimates. My natural intuition says that if the span of the data is very long and N is large, then the estimates will probably be reasonable. It is well understood that this intuition is wrong. This result provides an additional illustration of the problem. Specifically, suppose for N observations (N could be very large), a given dependence between the data, and a given span of data, a kernel density estimate contains an unacceptably high MISE. By repeatedly halving the mean reversion, doubling the time between observations, and doubling the span of the data, it is possible to generate arbitrarily long spans of data which also generate identically poor density estimates.

Results 2 and 3 examine the quality of the kernel density estimate in extreme cases. Both of these results are intuitive. In result 2, as κ goes to infinity, the rate of mean reversion goes to infinity and the correlation between the interest rate observations goes to zero. In this limiting case, because the interest rate observations are gaussian and their correlations go to 0, the observations become i.i.d. gaussian random variables. Moreover, my expressions for MIVAR and MISE converge to known expressions (see Silverman 1986) for MIVAR and MISE when there are N i.i.d. gaussian observations and a gaussian kernel is used. In result 3, as κ goes to zero, mean reversion goes to 0 and the correlation between discretely spaced observations of the process approaches 1. This means, in the limit, additional observations beyond the first don't contribute any new information. Therefore, mean integrated variance in this case approaches the same limit for every sample size N.

[10]This result is contained in DeHeuvels [1977] for i.i.d. data.

For the continuous sampling kernel density estimate, the first result in proposition III does not apply since it is not possible to meaningfully halve or double the sampling frequency in this case. The second and third results are sort of replicated for the continuous sampling case. When κ goes to infinity while V_E is held fixed, the MIVAR goes to 0. Intuition for this result is that as κ goes to infinity, the continuous sampling kernel estimate uses a continuum of i.i.d. observations. This should generate an MIVAR of 0. I have not been able to generate limiting results as κ goes to 0 in the continuous sampling case. However, in the analysis below I computed the effects of reducing κ while holding V_E fixed. I found that this increases MISE.

To further examine the role of persistence in the quality of kernel estimates, and in bandwidth selection, I used the formulas in proposition II to compute the optimal (MISE minimizing) bandwidth, and resulting Mean Integrated Squared Error, Mean Integrated Variance, and Mean Integrated Squared Bias of kernel density estimates for 5 different parameterizations of the Vasicek model using 22 years of continuously sampled data.[11] Table 1 reports the results for all five sets of Vasicek parameter values. The primary, or baseline case is contained in the row labeled model 0, and used parameters $\kappa = .85837, \theta = .089102$, and $\sigma^2 = .0021854$.[12] These parameters are baseline because they are actual GMM estimates of the Vasicek parameters reported in Ait-Sahalia [1996b] and hence provide a set of parameters that are relevant to real world analysis of interest rate processes. The rows labeled models (-2), (-1), (1), and (2) provide results in which the baseline κ and σ^2 were doubled, quadrupled, halved, and quartered respectively. Models towards the top of the table involve less persistence, but all models have the same ergodic distribution.

Table 1 contains two basic insights. First, the slower is the mean reversion, the higher is the mean integrated squared error, mean integrated squared variance, and mean integrated squared bias. This shows we should put less faith in our density estimates when persistence is high. From proposition I we know that the mean integrated squared bias does not depend directly on persistence. Therefore, if mean integrated squared bias goes up with persistence it must happen because the optimal response to more persistence is a higher bandwidth. This is the second insight from table 1 because it shows explicitly that the optimal bandwidth is increasing in persistence. This is an important result because it shows that the optimal bandwidth does not just depend on the sample size, and on the ergodic distribution of the data, but it also depends on a time series property of the data generating process. This shows that kernel density estimation should be guided by the time series properties of the data generating process.

[11] 22 years of data were used in this table to replicate the sample size used in Ait-Sahalia [1996a].

[12] All parameters are at an annual rate.

Tables 2 and 3 expand these insights by examining kernel density estimation for both continuous and discrete sampling for six different spans of data and six different sampling frequencies. Table 2 presents results on optimal bandwidth choice when the bandwidth is chosen to minimize MISE. Table 3 presents results on the resulting MISE when the bandwidth is chosen optimally. The Vasicek parameters and models used in tables 2 and 3 are identical to those used in table 1.[13]

Two important stylized facts are immediately apparent from table 2: For the amounts of mean reversion considered in the table, the choice of optimal bandwidth is highly insensitive to the frequency with which the data is sampled, and is very sensitive to the persistence of the data generating process. If either of these facts is ignored in bandwidth selection, the selected choice of bandwidth could be quite poor.

More specifically, if the bandwidth selection rule for i.i.d. gaussian observations was followed, then someone who sampled the data every 30 days would choose the bandwidth as $1.06\sqrt{V_E}N^{(-1/5)}$ where N is the number of time series observations that are spaced 30 days apart and V_E is an estimate of the variance of the ergodic distribution. Someone who followed the same rule but sampled the data every day would have chosen a bandwidth that is roughly half as large although the results here indicate the appropriate bandwidth for daily sampling should be roughly the same as for thirty day sampling in our baseline case.

Similarly, the table shows that the optimal bandwidth increases by 50% or more when moving from Model (-2), the least persistent model, to Model (2), the most persistent model. If the bandwidth is chosen without accounting for persistence, this shows there is the possibility the bandwidth could be chosen quite poorly.

The last stylized fact from table 2 is that the span of the data matters for bandwidth selection. This is not a surprise. In addition, the table shows that the MISE minimizing bandwidth holding sampling frequency fixed shrinks at an average rate of about $N^{-.3}$.[14]

Table 3 shows that MISE is not very sensitive to sampling frequency, but is very sensitive to the amount of persistence in the data. This result is very important for analyzing the behavior of interest rates nonparametrically, especially when the series is short. For example, for the baseline case and 20 years of data, sampling the data daily instead of monthly reduces the MISE by 0.25%; i.e. sampling daily instead of monthly produces almost no reduction in MISE. For 50 years of data in the baseline case, sampling the data daily instead of monthly

[13] Tables 2 and 3 assume there are 240 business days per year. This is a bit short of 250, but in order to obtain an even number of evenly spaced discrete observations, I needed a number of days that is divisible by all of the sampling frequencies. 240 does this and 250 does not.

[14] Computation shows this rate of bandwidth shrinkage is consistent with the asymptotic rate of shrinkage necessary to implement Ait-Sahalia's test that is based on the ergodic distribution. If this shrinkage rule is used to choose the bandwidth, it implies that the unnormalized test statistic converges in probability to zero at a rate of about $N^{.15}$ under the null.

reduces the MISE by 0.53%, which is also very small.

While changing the sampling frequency has minimal effects on MISE, changing the span of the data has large effects which appear to be nonlinear in the span of the data. For the baseline model (model 0), doubling the span of the data from 5 years to 10 years decreases MISE by an average of 31%. Doubling the span from 10 to 20 years decreases MISE by 37%; and doubling the span from 50 to 100 years decreases MISE by 43%.

A final insight that can be gleaned from these results concerns choosing the bandwidth using automatic methods such as least squares cross-validation. The basic principle behind cross validation is to estimate the kernel density while omitting some observations, and then use the quality of the fit for the omitted observations in choosing the kernel bandwidth. In i.i.d. data this type of procedure is intuitive: if the bandwidth is too small, the data will be over fitted, and the omitted data will be fitted poorly; conversely, if the bandwidth is too large, the data will be over smoothed and the fit for the omitted data will also be poor.

The key to the i.i.d. reasoning is that the cross-validated data adds significant new information to the estimation problem. Table 3 shows that for our baseline case, data which is sampled daily adds almost no new information above and beyond that for data which is sampled monthly. In effect, this means the observations from persistent data sampled at a high frequency will be clustered very close together, but will not necessarily provide significant information about the long run ergodic distribution. The cross validation algorithm will interpret the clustering as fine detail in small samples, and will fit this spurious fine detail by choosing a bandwidth which is too small. This downward bias phenomenon is also documented in Hart and Vieu [1994]; they propose a fix which involves cross-validating on observations which are spread less closely in time. Silverman [1986] also notes the possibility of severe downward bias in bandwidths chosen by cross-validation when the data has been rounded or when there are large numbers of observations that are nearly identical to other observations in the data. I will also discuss least squares cross-validation below in the context of Ait-Sahalia's ergodic distribution based test.

Finite Sample vs. Asymptotic Results.

In this section I will contrast the finite sample and asymptotic properties of kernel density estimates of the ergodic distribution of interest rates in the Vasicek model.

Robinson[1983] provides conditions on the kernel function, time series dependence, true ergodic density, and the rate of bandwidth shrinkage, such that the distribution of the kernel density estimates at n discretely separated points $u_1, ...u_n$ are asymptotically unbiased, normally distributed, and uncorrelated. Moreover, this asymptotic distribution is the same as that from a kernel density estimate that uses i.i.d. draws from the ergodic distribution of the interest rate process. Robinson[1983] emphasizes that these asymptotic results should not be taken too seriously because in finite samples, the quality of the asymptotic approximation will depend on both serial dependence and the size of the bandwidth.

Our results here show that the asymptotic approximation of the finite sample properties understate both the bias and variance of the kernel density estimate. Figure 1 plots the true and expected kernel density estimates for the five models from table 1. The difference between the true and expected densities is the bias. Although this bias is zero asymptotically, the figure shows the finite sample bias appears to be important. One measure of this importance is the extent to which the bias affects statistical inference in asymptotic testing procedures that ignore it.

To illustrate the effect of the bias, figure 1 contains asymptotic plus or minus 1 standard deviation confidence bands around the true distribution function.[15] If the points of an estimated distribution function lie mostly outside of this confidence band, a test based on the asymptotic distribution will almost always reject. Because the finite sample bias is much larger than this confidence band, most density estimates will lie mostly outside of the asymptotic confidence bands when the null is true. This will cause the null to be rejected too often.[16]

Figure 2 illustrates that the asymptotic distribution not only understates the finite sample bias, it also severely understates the finite sample variance of the kernel density estimate. The dashed lines in the figure are the true finite sample plus or minus 1 standard deviation confidence bands based on daily sampling. The dotted lines are the asymptotic plus or minus 1 standard deviation confidence bands. Based on the figure the true density estimates are far more variable than implied by the asymptotic distribution. This will cause the null to be

[15]The confidence band in the figure uses the unnormalized asymptotic standard deviation, which is equal to $(Nh)^{-.5}\sqrt{\frac{1}{2\sqrt{\pi}}\pi(u)}$. It is important to emphasize the asymptotic confidence bands are not true confidence bands.

[16]Formally, let $r_1, ...r_M$ be a set of evenly spaced r values on [-.02,.20]. The asymptotic approximation to the finite sample distribution of kernel estimate $\hat{\pi}$ implies:

$$\hat{\pi}(r_i) - \pi(r_i) \sim \text{independently}\mathcal{N}(0, \sigma^2(r_i)),$$

where $\sigma^2(r_i)$ is the variance of the kernel density estimate at r_i. This implies for large M under the null the test statistic

$$\frac{1}{\sqrt{2M}}\sum_{i=1}^{M}\left[\left(\frac{\hat{\pi}(r_i) - \pi(r_i)}{\sigma(r_i)}\right)^2 - 1\right] \approx \mathcal{N}(0,1)$$

But this ignores the finite sample bias. If it is taken into account, the mean of the above distribution is $\frac{1}{\sqrt{2M}}\sum_{i=1}^{M}[\frac{b(r_i)}{\sigma(r_i)}]^2$ where $b(r_i)$ is the bias of the kernel density estimate at point i. From figure 1 it is clear for models 0, 1, and 2, that $[\frac{b(r_i)}{\sigma(r_i)}]^2 > 1$ for nearly all i. Therefore the mean of the bias corrected distribution is approximately $\sqrt{M/2}$. If this bias adjustment is not made, the null will be erroneously rejected most of the time using the asymptotic distribution theory.

rejected far too often using the asymptotic variance. Figure 3 shows this remains a problem even with 100 years of data. Both figures show the understatement of variance worsens with the persistence of the data generating process.

An additional feature of the asymptotic distribution is that kernel estimates at different points of the support of the distribution are asymptotically uncorrelated. To contrast this result with finite sample experience, figures 4 and 5 present contour plots of the correlation function of the kernel density estimate for Models -2 through Model 2 when the data is sampled daily. In both figures, the contours represent gradations of 0.2 in correlation. For example, the white band down the center of each graph represents the points that have correlations of at least 0.8. The band just below this one has correlations from 0.6 to 0.8, and so on. The lowest correlation shown in figures 4 and 5 is -0.8 and the highest is 0.8. Both figures show that the density estimates at different points of the support of the distribution are highly correlated, even in samples with 100 years of data.

Figures 4 and 5 have two additional interesting features. First, density estimates above the mean of the distribution of the short rate are usually negatively correlated with density estimates below the mean, and positively correlated with density estimates above the mean. An intuitive explanation for this feature of the correlation function is that starting points matter in small samples. For example, in the Vasicek model if the starting value of the process is above the mean of the distribution, future values are expected to stay above the mean as well. In small samples this will cause too much probability mass to be assigned to observations above the mean and too little to observations below it. This is consistent with the observed correlation pattern. The second interesting feature of the figures is that the size of the region with relatively high correlations grows in persistence and shrinks in the span of the data. This feature is probably also related to small sample biases due to dependence on starting values because it is well known that starting values tend to be more important in processes that have more persistence.

Finally, figures 6 and 7 plot the variance-covariance function for the kernel density estimate. Each contour in the figures corresponds to gradations of 0.4 in covariance. The neutral gray in the corners of each contour plot correspond to a covariance of zero. Covariances are negative in areas with darker shading and positive in areas with lighter shading. The dominant feature in the figures is that the absolute magnitude of the variance-covariance function appears to increase with persistence and decrease with the span of the data; this is as expected. The other interesting feature of figures 6 and 7 is that the variance function (i.e. the covariance function along the 45 degree line) is bimodal. Finite sample variance has a local minimum near θ, the mean of the interest rate process; variance then rises on either side of θ before falling again. By contrast, the asymptotic variance is proportional to the ergodic density. Since the ergodic density is unimodal so is the asymptotic variance. Therefore, the bimodality is an interesting finite sample feature of the variance function. It is not clear why the distribution is bimodal in finite samples. The bimodality of the variance does not occur with i.i.d. draws from the ergodic distribution. Therefore, the bimodality is somehow related to the time series properties of the interest rate process in finite samples. I conjecture the reason for the higher variance away from the mean is related to the small sample problems caused by the choice of starting value. However, since the process reverts to the mean from above and below, the small sample problem at the mean may be less severe than at other points in the support of the distribution. This is just conjecture however. It definitely remains a topic for further research.

This concludes my discussion of the finite sample properties of kernel density estimates in the Vasicek model. Because these results are analytical, or nearly analytical, they should prove to be very useful for gauging the performance of other nonparametric methods. The next section examines the finite sample behavior of Yacine Ait-Sahalia's [1996a] nonparametric test of parametric models of the short term interest rate.

III Ait-Sahalia's test.

Ait-Sahalia [1996a] introduced a test of the null hypothesis that the instantaneous spot interest rate is generated by a parametric class of markov models. The test is based on the distance between two estimates of the ergodic distribution of the spot interest rate. The first is a nonparametric kernel estimate which is consistent over a wide class of continuous time models. The second is the closest density that could have been generated under the parametric null. The null is rejected whenever the distance between the density estimates is too large. Using this test, Ait-Sahalia rejected virtually every major single factor model of the instantaneous short term interest rate, but failed to reject a more general alternative single factor model.

Ait-Sahalia reached his conclusions using 22 years of daily data on short term interest rates, however the critical values for his tests come from asymptotic theory. The purpose of this and the next section is to examine the finite sample size of the test when data is generated under the Vasicek model, and to examine the finite sample power of the test when the data is generated under a plausible parameterization of the Cox-Ingersoll-Ross model. Again I want to emphasize that I am focusing on these models because they are relatively tractable, and this tractability aids in interpreting the results from the test.

It is possible that the finite sample size of the test will differ from its asymptotic size. As we will see below, this is because the asymptotic distribution of the test statistic depends on the ergodic distribution of the interest rate process, but not on its persistence. By contrast, in a 22 year span of data, Tables 1 and 3 show that, holding the ergodic distribution fixed, the distance (as measured by MISE) between the true Vasicek density and the MISE minimizing kernel estimate is increasing in the persistence of the process. Since the test is based on a measure of this distance[17], it is reasonable to conjecture that the persistence of the interest rate series that we observe may bias the asymptotic test towards rejection in a sample that only contains 22 years of data. This latter point is especially poignant since the results in table 3 for the baseline case show that sampling every day or sampling once every 30 days generates kernel results of nearly the same quality as measured by MISE. However, 22 years of data sampled once every 30 business days is only about 185 observations. This would be considered a small N for the purposes of nonparametric estimation. The remainder of this section proceeds in three parts. Part A provides a description of Ait-Sahalia's test, part B presents our results on the size of the test, and part C presents our results on the power of the test.

A. Description of Ait-Sahalia's test.

Ait-Sahalia tests the null hypothesis that the instantaneous short rate follows the single factor parametric diffusion:

$$dr = \mu(r, \theta)dt + \sigma(r, \theta)dw$$

where θ is an element of a parameter set Θ, against the alternative that dr is generated by a more general single factor diffusion of form:

$$dr = \mu(r)dt + \sigma(r)dw.$$

The null restricts r's ergodic distribution to have functional form $\pi(u, \theta)$, while the ergodic distribution under the alternative has the more general functional form $\pi(u)$. Ait-Sahalia's test is based on the weighted squared distance between the two distributions:

$$M = \int_u (\pi(u, \theta) - \pi(u))^2 \pi(u) du,$$

where the weighting function is $\pi(u)$. A normalized estimate of this distance is:

$$\hat{M} = N h_N \int_u (\pi(u, \hat{\theta}) - \hat{\pi}(u))^2 \hat{\pi}(u) du.$$

[17]The test is actually based on the distance between the kernel density estimate and the closest density consistent with a particular parametric family. As we will see later, in a special case it is possible for the distance between the true and nonparametric densities to grow, but for the distance between the nonparametric density and the closest member of the parametric family to shrink.

where N is the number of interest rate observations in the sample, h_N is the kernel bandwidth, $\hat{\pi}(u)$ is a second order nonparametric kernel density estimate of $\pi(u)$ which is consistent under both the null and alternative, and $\pi(u,\hat{\theta})$ is that estimate of $\pi(u)$ which is closest to $\hat{\pi}(u)$, but is restricted to satisfy the null.[18] Under the null hypothesis, $\hat{\pi}(u)$ and $\pi(u,\hat{\theta})$ are both consistent and thus their distance will shrink to zero as the sample size grows. Under the alternative hypothesis, the parametric restrictions will bind, preventing the distance from shrinking.

Ait-Sahalia uses this feature of the distance metric to create a test statistic. In particular, he shows that under the null hypothesis and certain regularity conditions, if the bandwidth h_N is chosen so that $\lim_{N\to\infty} Nh_N = \infty$ and $lim_{N\to\infty} Nh_N^{4.5} = 0$, then

$$h_N^{-1/2}(\hat{M} - E_M) \xrightarrow{d} \mathcal{N}(0, V_M),$$

where,

$$E_M = \left(\int_{-\infty}^{+\infty} K^2(x)dx\right)\left(\int_{\underline{x}}^{\overline{x}} \pi^2(u)du\right),$$

and

$$V_M = 2\left(\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty} K(u)K(u+x)du\right\}^2 dx\right)\left(\int_{\underline{x}}^{\overline{x}} \hat{\pi}^4(u))du\right),$$

and $\underline{x}$ and $\overline{x}$ represent the upper and lower bounds of the observed interest rates.

As indicated earlier, the test is based on a measure of the distance between two estimates of the ergodic distribution of the interest rate process. The most important point for our analysis is that the asymptotic distribution depends on the choice of kernel, the true ergodic distribution, and the rate at which the bandwidth shrinks, but it does not depend on the persistence of the process. However, as indicated in Tables 1 and 3, persistence has an important influence on the distance (as measured by MISE) between the nonparametric estimate and the true distribution. Because this finite sample property is not accounted for in the asymptotics, the mean asymptotic distance under the null probably understates the mean distance in finite samples. Intuitively, this makes it likely that the test will reject too often in finite samples when there is large amounts of persistence in the data.

To implement the test and examine this it is necessary to construct consistent estimators of M, V_M, and E_M. There are many ways to do this. To do so we computed the kernel density

[18] Formally, $\hat{\theta}$ is the solution to:

$$\min_{\theta\in\Theta}\int_u (\pi(u,\hat{\theta}) - \hat{\pi}(u))^2\hat{\pi}(u)du.$$

estimate $\hat{\pi}(u)$ and then used it to numerically approximate the following expressions:

$$\hat{M} = \min_{\hat{\theta}\in\Theta} N h_N \int_{\underline{x}}^{\overline{x}} (\pi(\hat{\theta}, u) - \hat{\pi}(u))^2 \hat{\pi}(u) du,$$

$$\hat{E}_M = \left(\int_{-\infty}^{+\infty} K^2(x)dx\right)\left(\int_{\underline{x}}^{\overline{x}} \hat{\pi}^2(u)du\right), \text{ and}$$

$$\hat{V}_M = 2\left(\int_{-\infty}^{+\infty} \left\{\int_{-\infty}^{+\infty} K(u)K(u+x)du\right\}^2 dx\right)\left(\int_{\underline{x}}^{\overline{x}} \hat{\pi}(u)^4 du\right)$$

Ait-Sahalia suggested using the estimators above or an alternative set of estimators.[19] Table 4 contains the results using the estimators above, while Table 4A contains results using the alternative estimator. Because the results are so similar, we focus only on the results from table 4 in our discussion below.

B. Test size.

To analyze the size of the test I conducted 500 Monte Carlo simulations for each of the five sets of Vasicek parameters contained in Table 1. Each Monte Carlo used 22 years of data sampled daily.[20] The bandwidth used for each kernel density estimate is the bandwidth which minimizes the MISE for the given data generating process when the data is sampled continuously. Based on table 2, I know this is virtually the same as the optimal bandwidth when the data is sampled daily. Ait-Sahalia [1996a] also chooses the bandwidth in finite samples to minimize some proxy for MISE. For purposes of comparison, we also conducted a Monte Carlo analysis for our baseline Vasicek model in which the bandwidth was chosen using least squares cross validation.[21]

[19] The alternate method estimates $\hat{\pi}(u)$ using nonparametric kernel density estimation, but then approximates the other quantities as:

$$\hat{M} = \min_{\hat{\theta}\in\Theta} N h_N (1/N) \sum_{i=1}^{N} (\pi(\hat{\theta}, x_i) - \hat{\pi}(x_i))^2,$$

$$\hat{E}_M = \left(\int_{-\infty}^{+\infty} K^2(x)dx\right)(1/N)\sum_{i=1}^{N} \hat{\pi}(x_i)$$

$$\hat{V}_M = 2\left(\int_{-\infty}^{+\infty} \left\{\int_{-\infty}^{+\infty} K(u)K(u+x)du\right\}^2 dx\right)(1/N)\sum_{i=1}^{N} \hat{\pi}^3(x_i)$$

[20] I assumed 250 business days per year, for a total of 5500 daily observations.

[21] Least squares cross validation was used to choose among 10 bandwidths that ranged from .5 to 1.5 times the optimal bandwidth for the continuous sampling case.

Under the null hypothesis, as the sample size N goes to infinity, the test statistic:

$$\hat{V}_M^{-.5} h_N^{-.5}(\hat{M} - \hat{E}_M) \xrightarrow{d} \mathcal{N}(0,1).$$

This asymptotic distribution is the basis of the test statistic. Thus, the finite sample critical values for a one sided 5% and 1% test are 1.645 and 2.33, respectively. Table 4, Panels A and B, present Monte Carlo estimates of the appropriate finite sample critical values at the 5% and 1% confidence levels, respectively. Next to each Monte Carlo critical value estimate are upper and lower bounds of a 95% confidence interval for the true critical values.[22] Panel A shows that the finite sample critical values for each data generating process are on the order of 10x the size of the asymptotic critical values for tests at the 5% level. Panel B shows that the asymptotic critical values are more than 10x too small for tests at the 1% level. Finally, panel C shows estimated empirical rejection frequencies with standard errors when asymptotic critical values are used in the tests. The panel shows the finite sample size of the test is very different from the asymptotic size; the test rejects the null far too often, on the order of 50% of the time for our baseline model at the 5% confidence level.

Although the results in Panel C show the test rejects too often in small samples, consistent with our intuition on finite sample properties and on persistence, the panel also shows that rejection rates increase with persistence at first (in moving from model (-2) to model (-1)) but then perversely tend to move down as the interest rate process becomes even more persistent. This doesn't mean our basic intuition on size is wrong, but another phenomenon is occuring at the same time. Basically, as the persistence of the series is increased, the optimal bandwidth that is used in the test is increased as well. This change in persistence has two effects. The first effect is that more persistence increases the MISE, MIVAR, and MISB of the density estimate. This is shown in Table 1 and favors a rejection frequency that increases with persistence. However, the second implication of persistence in this case is that it makes the kernel density estimate appear more gaussian, i.e. closer in a distance sense to the set of gaussian distribution functions. The reasons for this are twofold. First, all else equal, larger bandwidths make the density estimates more closely resemble the shape of the kernel, which in this case is gaussian. Second, when persistence is increased while holding the ergodic distribution fixed, this has the effect of concentrating the data more tightly. This further strengthens the tendency of the density estimate to appear like the kernel.[23] Since Ait-Sahalia's test in this case is actually a test of the distance between the kernel estimate

[22]The confidence intervals are based on the order statistics of the Monte Carlo estimates and are fully nonparametric. Details on how to create these confidence intervals are contained in David [1981].

[23]As an extreme example, if all of the data in the sample is concentrated at or very close to the number 5, and the kernel function is gaussian with positive bandwidth h, then the kernel density estimate will be approximately equal to the gaussian distribution with a mean of 5 and a standard deviation of h.

and the closest gaussian density, this distance goes down with persistence after a point. This somewhat strange phenomenon is an artifact of using a gaussian kernel and the Vasicek model. It in no way invalidates our main point that the test rejects too frequently and that higher persistence significantly degrades the quality of the density estimate, and biases the test towards rejection.

It is important to emphasize at this point that all of the results we have presented are for infeasible bandwidth selectors, i.e. bandwidths that are selected to minimize the mean integrated squared error when the true data generating process is known. More realistically, it is useful to examine the size of the test when the true data generating process is not known. We examined the size of the test when the data is generated under model 0 and the bandwidth is selected using least squares cross validation.[24] The bandwidth choices using this method were almost always half the size of the optimal bandwidth and probably would have been lower if we did not constrain the cross validation algorithm. Moreover, the size of the test was unacceptable. Specifically, using least squares cross validation with asymptotic critical values, the null hypothesis for Model 0 was rejected 91.8% of the time at the 5% level and 88.2% of the time at the 1% level.[25]

C. Power.

To investigate the power of the test, we conducted 500 Monte Carlo experiments. In each experiment, we generated 5500 draws of data under the CIR model:

$$dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dw.$$

with parameters $\kappa = .89218, \theta = .090495$, and $\sigma^2 = 3.2742$. These parameters were taken out of Ait-Sahalia [1996b] and represent GMM estimates of the CIR parameters based on real data. Therefore, we treat this as the baseline CIR model. We tested the null hypothesis that the data was generated by our baseline Vasicek model (model 0 in table I) using two different methods. The first method used Ait-Sahalia's test with bandwidth choice .0217661 (see Table 1) and the lower bound finite sample critical value estimates for model 0 from panels A and B of Table 2.[26] The second test is a conditional moment test that the likelihood

[24] We've cheated a little bit here by restricting our least squares validation search to bandwidths between .5 times and 1.5 times the optimal bandwidth for continuous sampling. So, this is not a fully feasible bandwidth selector either, but should give some flavor for results using least squares cross validation.

[25] The estimated finite sample 5% critical value using least squares cross validation is 61.33, with a 95% confidence interval of [51.51, 71.83]. The estimated finite sample 1% critical value is 114.86 with a 95% confidence interval of [92.85, 239.18]. The 5% and 1% asymptotic critical values are 1.645 and 2.33 respectively.

[26] This bandwidth choice is appropriate under the null hypothesis. However, the critical value estimates are likely to be less than the true finite sample critical values since they are the lower bounds of 95% confidence

function implied by the Vasicek model is appropriate.[27] To implement the test, I estimated the parameters of the Vasicek model via maximum likelihood. If the likelihood function is properly specified, then for each observation t the following moment restriction should be satisfied:

$$E\left(\frac{\partial\mathcal{L}(r_t|r_{t-\Delta t},\sigma,\kappa,\theta)}{\partial\sigma}\middle| r_{t-\Delta t}\right)=0,$$

where $\mathcal{L}(r_t|r_{t-\Delta t},\kappa,\theta,\sigma)$ is the conditional log likelihood of r_t given $r_{t-\Delta t}$ and the parameters that generated the data. To test this moment restriction, I estimated the regression:

$$\frac{\partial\mathcal{L}(r_t|r_{t-\Delta t})}{\partial\sigma}=\beta_0+\beta_1 r_{t-\Delta t}+u_t,$$

and then performed a two-sided Z-test of whether $\beta_1=0$.[28] In all fairness, this specification test makes a lot of sense when testing the Vasicek model vs the CIR model because the expected value of the score function is approximately linear in r if the CIR model is correct and the likelihood function is appropriate for the Vasicek model. Monte Carlo analysis using 500 simulations under the baseline Vasicek null shows this test has the appropriate size under the null hypothesis.[29]

When the data was generated by the baseline CIR model, Ait-Sahalia's size corrected test rejects the Vasicek null 39.8% of the time at the 5% level and 13.8% of the time at the 1% level. Thus it displays modest power against the CIR model after critical values are adjusted for size. By contrast, the GMM test rejected the null for all 500 Monte Carlo runs at the 1% level or better. This suggests the likelihood based GMM test is far more powerful for distinguishing between the Vasicek null and the CIR alternative. This differences in test power makes intuitive sense because Ait-Sahalia's test only exploits information about the unconditional ergodic distribution, but ignores useful information on the conditional distribution of each observation.

These results show that for power against parametric alternatives, tests that are tailored to have power against that alternative will provide better performance than using a non-parametric test. This suggests that an overall model testing procedure should first test a model against parametric alternatives using tests specifically tailored for those alternatives. If the model passes those tests, the nonparametric test should be applied because it may have power in directions not captured by the parametric test. Even in this case, the non-

intervals for these critical values.

[27] See Newey [1985a] and Newey [1985b] for additional details on these tests.

[28] Standard errors were computed using Newey-West's method with a lag length of 15.

[29] Using data generated under the Vasicek null, this specification test rejected the null hypothesis 4.8% of the time at the 5% level in 500 Monte Carlo runs and it rejected the null hypothesis 1.6% of the time at the 1% level.

parametric tests need to be used cautiously because the results in section II of the paper show that the finite sample and asymptotic properties of these estimators are very different.

IV Summary and Conclusion.

The application of nonparametric methods to analyze interest rate data provides new opportunities for estimating interest rate dynamics, and for testing interest rate theories. In addition to the opportunities, however, interest rate data present difficult challenges because the nonparametric techniques were developed for i.i.d. observations, while interest rates are dependent and highly persistent.

In this paper I have provided an in-depth examination of kernel density estimation of the ergodic distribution of the short term interest rate when interest rates follow a continuous time AR(1) process as in Vasicek [1977]. This functional form makes it possible to compute many useful finite sample time-series results, and to contrast finite sample properties of the kernel density estimate with its asymptotic properties.

Our principal results for the Vasicek model in finite samples are that holding the ergodic distribution of the interest rate process fixed, the Mean Integrated Squared Error (MISE) of the kernel density estimate is highly sensitive to the persistence of the data generating process, and to the span of the data; however, it is very insensitive to the frequency with which the process is sampled. Similarly, holding the ergodic distribution fixed, the optimal, i.e. MISE minimizing, kernel bandwidth is sensitive to the persistence of the interest rate process and to the span of the data, but is very insensitive to the data sampling frequency. This suggests bandwidth selection rules which are sensitive to the data sampling frequency, such as rules which are appropriate for i.i.d. draws, are likely to generate very poor density estimates when the data is generated by a persistent time-series process.

When contrasting asymptotic and finite sample results we found that the asymptotic distribution seriously understates the finite sample bias and seriously understates the finite sample variance of the kernel density estimates. The asymptotic approach also understates the absolute magnitude of the finite sample correlation between density estimates at non-adjacent points in the support of the interest rate distribution. Also, asymptotically, the variance of the kernel density estimate is unimodal. By contrast the finite sample variance is bimodal, even with 100 years of data.

The paper also provided an in-depth examination of the size and power of one of Ait-Sahalia's test of parametric interest rate models. The test we examined is based on the distance between a nonparametric estimate of the ergodic distribution of the interest rate process and the closest estimate that is consistent with a particular parametric family. Examining this test when the data was generated by the Vasicek model, we found the size of the test when using asymptotic critical values is far too large. Also, we conjecture that the high rate of rejection occurs because the persistence of the data generating process does not affect the asymptotic critical values, but has an important effect on the finite sample performance of the kernel density estimates. After the test is corrected for size, it appears to have low power in distinguishing between the CIR and Vasicek models when compared with an alternative conditional moment test of the restrictions imposed by the Vasicek model.

The most important implication of this paper is that statistical inference using nonparametric density estimation is difficult in a time series context, especially for highly persistent time series such as interest rates. Our results on estimation of the ergodic distribution of the short rate in the Vasicek process, show that the problem is not that the nonparametric kernel estimates are necessarily bad, but rather that hypothesis tests and other inferences based on asymptotics may be very poor because the asymptotic approximations of the kernel density estimate's bias, variance, and correlation are so far off from their true finite sample values, even when the sample size is as large as 100 years. This suggests that the results from nonparametric techniques should be interpreted with caution and probably should be used as part of a larger research effort that includes parametric estimation to learn more about the persistence of a process, as well as significant bootstrapping, and other Monte Carlo analysis to simulate the behavior of nonparametric estimators in finite samples. Despite the difficulties highlighted here, I do believe that the principle advantage of nonparametric density estimation, its ability to describe the data, remains intact. I also conjecture that the best prospects for future work in this area will be for approaches that find good ways to combine the information from nonparametric and parametric estimation methods.

BIBLIOGRAPHY

Ait-Sahalia, Yacine, 1996a, "Testing Continuous Time Models of the Spot Interest Rate," *Review of Financial Studies* 2, No. 9, (1996): 385-426.

Ait-Sahalia, Yacine, 1996b, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica* 64, No.3, (May 1996): 527-60.

Ait-Sahalia, Yacine, 1996c, "Do Interest Rates Really Follow Continuous-Time Markov Diffusions?," Working Paper, Graduate School of Business, University of Chicago, 1996.

Ball, Clifford A., and W.N. Torous, 1996, "Unit Roots and the Estimation of Interest Rate Dynamics," Journal of Empirical Finance 3 (1996): 215-238.

Broze, Laurence, O. Scaillet, and J.-M. Zakoian, 1995, "Testing for Continuous Time Models of the Short Term Interest Rate," *Journal of Empirical Finance* 2 (1995): 199-223.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985 "A Theory of the Term Structure of Interest Rates", *Econometrica* 53, no. 2 (March 1985): 385-407.

Hart, J.D. and P. Vieu, "Data-Driven Bandwidth Choice for Density Estimation Based on Dependent Data," *The Annals of Statistics* 18, no. 2 (1990): 873-890.

David, Herbert A., 1981, *Order Statistics,* John Wiley and Sons, Inc., New York, 1981.

Hardle, W., 1990, *Applied nonparametric regression,* Cambridge University Press, Cambridge, 1990.

Karlin, Samuel, and Howard M. Taylor, 1981, *A Second Course in Stochastic Processes,* Academic Press, Inc., New York, 1981.

McCulloch, J.H., and H.-L. Kwon, 1993 "U.S. Term Structure Data, 1947-1991," Working Paper 93-6, Ohio State University.

Newey, Whitney K., 1985a, "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica* 53, no. 5 (September 1985): 1047-1070.

Newey, Whitney K., 1985b, "Generalized Method of Moments Specification Testing," *Journal of Econometrics* 29 (1985): 229-256.

Newey, Whitney K., and K.D. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, no. 3 (May 1987): 703-708.

Pagan, A.R., A.D. Hall, and V. Martin, 1995, "Modeling the Term Structure," Working Paper, 1995.

Park, Byeong U., and J.S. Marron, 1990, "Comparison of Data-Driven Bandwidth Selectors," *Journal of the American Statistical Association,* 85 no. 409 (March 1990): 66 - 72.

Robinson, Peter M., 1986, "On the Consistency and Finite-Sample Properties of Nonparametric Kernel Time Series Regression, Autoregression, and Density Estimators," *Annals of the Institute of Statistical Mathematics,* 38, No. 3, A, (1986): 539-549.

Robinson, Peter M., 1983, "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis,* 4. no. 3 (1983): 185-207.

Scott, David W., 1992 *Multivariate Density Estimation, Theory, Practice, and Visualization,* John Wiley and Sons, Inc., New York, 1992.

Siddique, Akhtar R., 1994, "Nonparametric Estimation of Mean and Variance and Pricing of Securities," Working Paper, Georgetown School of Business, 1994.

Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, London, 1986.

Stanton, Richard, 1995, "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk", Working Paper, University of California at Berkeley, September 1995.

Stone, Charles J., 1984 "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *The Annals of Statistics,* 12 (1984): 1285-1297.

Vasicek, Oldrich, 1977 "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5 (1977): 177-188.

Wand, M.P., 1992, "Finite sample performance of density estimators under moving average dependence," *Statistics and Probability Letters* 13 (1992): 109-115.

Appendices

This appendix provides detailed results on kernel density estimation when the time series of interest rates is generated under the Vasicek process. The appendix is divided into four main parts. Part A lays out notation and provides some basic results that simplify the derivation of the results in the next three sections. Part B presents our results on Mean Integrated Squared Error, Mean Integrated Squared Variance, and Mean Integrated Squared Error when the interest rate process is sampled continuously. Part C presents similar results when the interest rate process is sampled at evenly spaced discrete time intervals. In part C we also derive the Variance of the Kernel Density estimate for both discrete and continuous sampling. Finally, part D works out formulae that are useful for computing the covariance of the density estimates.

A Notation and Four Basic Results.

As a preliminary, this section fixes notation and states four results that will be used in the sections that follow.

Throughout our exposition,$\pi(r_s, s, |r_t, t)$ will denote the probability that r takes the realization r_s at time s given that r was equal to r_t at time t. Similarly, $\pi(\tilde{r})$ will denote the unconditional probability that r takes on the realization $\tilde{r}$.

In certain sections we will derive results on kernel estimation for the Vasicek process. We will use the term Vasicek process to represent the following stochastic process for r:

$$dr = \kappa(\theta - r)dt + \sigma dw.$$

Under the Vasicek process r's ergodic distribution is gaussian of form:

$$r \sim \mathcal{N}(\theta, \frac{\sigma^2}{2\kappa}).$$

The results we will use below are as follows:

Result 1:

$$\pi(r_s) = \int \pi(r_s, s|r_t, t)\pi(r_t)dr_t$$

This result is well known and shows that the ergodic distribution of the process is the unconditional distribution of the process.

Result 2:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-.5(\frac{y-u}{\sigma_1})^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-.5(\frac{u-z}{\sigma_2})^2} du = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}} e^{-.5\left(\frac{y-z}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2}$$

Result 3:

$$\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-.5(\frac{y-u}{\sigma_1})^2}\right)^2 dy = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_1}.$$

Result 4:

$$\frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-.5(\frac{u-y}{\sigma_1})^2} e^{-.5(\frac{y-\theta}{\sigma_2})^2} = \frac{1}{\sqrt{2\pi}S_1} \frac{1}{\sqrt{2\pi}S_2} e^{-.5(\frac{y-(\lambda u+(1-\lambda)\theta)}{S_1})^2} e^{-.5(\frac{u-\theta}{S_2})^2}$$

where $S_1 = \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}$, $S_2 = \sqrt{\sigma_1^2+\sigma_2^2}$, and $\lambda = \frac{\sigma_2^2}{\sigma_2^2+\sigma_1^2}$.

Results 2,3, and 4 are especially useful when the kernel function is gaussian and the ergodic and conditional distribution of r are Gaussian, as they are with the Vasicek process. Our basic results for the continuous sampling kernel estimate are below.

B Kernel Estimation with Continuous Sampling.

This section motivates kernel estimation of the ergodic distribution of a stochastic process r when the realizations of the process are continuously sampled over a span of T years. I will then derive expressions for the expected kernel density estimate, the Mean Integrated Squared Error (MISE), Mean Integrated Squared Bias (MISB), and Mean Integrated Variance (MIVAR). I assume throughout that the process for r begins with a draw from its unconditional (ergodic) distribution at time 0, and then follows a strictly stationary Markov diffusion process of form:

$$dr = \mu(r)dt + \sigma(r)dw.$$

To motivate continuous sampling, suppose that the process is sampled discretely at N evenly spaced times so that the time between observations is ΔT. This implies that $1/N = \frac{\Delta T}{T}$. With this notation, the traditional kernel density estimate has the form:

$$\hat{\pi}(u) = \frac{1}{Th} \sum_{s=\Delta t}^{N\Delta t} K\left(\frac{u-r_s}{h}\right) \Delta t.$$

As the sampling is conducted more and more frequently, Δt goes to zero and the kernel density estimate contverges to:

$$\hat{\pi}(u) = \frac{1}{Th}\int_0^T K\left(\frac{u - r_s}{h}\right) ds$$

This is the expression for the kernel density estimate with continuous sampling.

Given the continuous sampling kernel density estimate, below we derive the Expected Kernel Density Estimate, Mean Integrated Squared Bias, and Mean Integrated Variance.

The Expected Kernel Density Estimate.

To compute the expected kernel density estimate, we will exploit the linearity of the integral operator and the fact that r follows a Markov process. Therefore, if r starts from r_0 at time 0, then

$$E(\hat{\pi}(u)|r_0) = \frac{1}{Th}\int_0^T [EK(\frac{u - r_s}{h})|r_0]ds = \frac{1}{Th}\int_0^T \left[\int_{r_s} K(\frac{u - r_s}{h})\pi(r_s, s|r_0, 0)dr_s\right] ds$$

The unconditional expected value of the kernel density estimate is given by the expression:

$$E\hat{\pi}(u) = \int_{r_0}\left(\frac{1}{Th}\int_0^T \left[\int_{r_s} K(\frac{u - r_s}{h})\pi(r_s, s|r_0, 0)dr_s\right] ds\right)\pi(r_0)dr_0$$

By Result 1 and Fubini's theorem this simplifies considerably to become:

$$E\hat{\pi}(u) = \int_{r_s} K(\frac{u - r_s}{h})\pi(r_s)dr_s.$$

If the kernel is gaussian, and r evolves according to the Vasicek process, then by result 2 we get an analytical expression for the expected kernel density estimate:

$$E\hat{\pi}(u) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h^2 + \frac{\sigma^2}{2\kappa}}}e^{-.5\left(\frac{u-\theta}{\sqrt{h^2+\frac{\sigma^2}{2\kappa}}}\right)^2}$$

This shows the unconditional expected kernel density estimate for the Vasicek process has the form of a gaussian distribution with mean θ and variance $h^2 + \frac{\sigma^2}{2\kappa}$. This is nearly identical to the expression for the ergodic distribution for r. The difference between the two expressions is the inclusion of h in the expression for the expected density. In this case we can see how choosing a nonzero bandwidth introduces bias in the kernel density estimate.

The Mean Integrated Squared Error of the density estimate is the sum of the Mean Integrated Squared Bias and the Mean Integrated Variance. The methods that are used to compute each of these are presented below.

Mean Integrated Squared Bias.

The expression for the Mean Integrated Squared Bias is:

$$MISB = \int_u (E\hat{\pi}(u) - \pi(u))^2 du.$$

If the kernel is gaussian, and r evolves according to the Vasicek process, then using results 2 and 3, the formula for the Mean Integrated Squared Bias simplifies to have the following analytical form:

$$MISB(Vasicek) = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{2h^2 + \frac{\sigma^2}{\kappa}}} + \frac{1}{\sqrt{2(\frac{\sigma^2}{2\kappa})}} - \frac{2}{\sqrt{(h^2 + \frac{\sigma^2}{\kappa})}} \right)$$

Mean Integrated Variance.

The expression for the Mean Integrated Variance of the kernel estimate is:

$$MIVAR = E \int_u (\hat{\pi}(u) - E\hat{\pi}(u))^2 \, du.$$

By Fubini's theorem, and simplification, this can be written as:

$$MIVAR = \int_u \left(E\left[\hat{\pi}^2(u)\right] - (E\hat{\pi}(u))^2 \right) du.$$

If the kernel is gaussian, and r follows the Vasicek process then by result 3 this simplifies further to become:

$$MIVAR(Vasicek) = \int_u E\left[\hat{\pi}^2(u)\right] du - \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{2h^2 + \frac{\sigma^2}{\kappa})}} \right).$$

Computing the first term in the expression for Mean Integrated Variance is very difficult, and must be done numerically, even with a process as simple as the Vasicek model. However, it is relatively simple to compute for the Vasicek process. In what follows we will present the general formula for the first expression, and then the formula for this expression under the Vasicek model.

It is useful to define some auxilary expressions to compute the first term of the expression for MIVAR. Define:

$$F(t) = \int_0^t K\left(\frac{u - r_s}{h}\right) ds.$$

Although the function F(t) depends on the random realizations of the process r, over each increment of time dt, the function is locally deterministic with derivative[30]:

$$dF(t) = K\left(\frac{u - r_t}{h}\right) dt.$$

Using elementary rules of calculus this implies:

$$d(F(t)^2) = 2F(t)dF(t),$$

and an alternative expression for the square of the kernel density estimate is:

$$\hat{\pi}(u)^2 = \frac{1}{T^2h^2}\int_0^T 2F(t)dF(t)$$

Substituting in for the expressions for F(t) and dF(t) and rearranging yields:

$$\hat{\pi}(u)^2 = \frac{2}{T^2h^2}\int_0^T \left[\int_0^t K\left(\frac{u-r_s}{h}\right) K\left(\frac{u-r_t}{h}\right) ds\right] dt$$

The above expression is very convenient because it allows us to express $E\int_u \hat{\pi}(u)^2 du$ using the law of iterated conditional expectations, and result 1. This yields:

$$E\int_u \hat{\pi}(u)^2 du = \int_u \left\{\frac{2}{T^2h^2}\int_0^T \left(\int_0^t E\left[K\left(\frac{u-r_s}{h}\right) K\left(\frac{u-r_t}{h}\right)\right] ds\right) dt\right\} du$$

By the law of iterated conditional expectations, and by Result 1, the expectation of the term inside the inner braces can be written as:

$$E\left[K\left(\frac{u-r_s}{h}\right) K\left(\frac{u-r_t}{h}\right)\right] = \int_{r_s} K\left(\frac{u-r_s}{h}\right)\left\{\int_{r_t} K\left(\frac{u-r_t}{h}\right)\pi(r_t, t|r_s, s)dr_t\right\}\pi(r_s)dr_s$$

Substituting, the above expression into the one before it gives us our final, and very large expression for the first term of the MIVAR:

$$E\int_u \hat{\pi}(u)^2 du =$$

[30]In the next increment of time, F will grow by the amount

$$K\left(\frac{u - r_{t+dt}}{h}\right) dt.$$

Applying Ito's lemma to r_{t+dt}, the above expression can be written in Taylor series form as:

$$K\left(\frac{u - r_{t+dt}}{h}\right) dt = K\left(\frac{u - r_t}{h}\right) dt + K_r(.)dr * dt + .5K_{rr}(.)(dt^2).$$

All of the stochastic terms in this expansion are of smaller order than dt, and thus can be ignored. This implies that F is locally deterministic with the derivative given in the text.

$$\frac{2}{T^2h^2}\int_u \left\{ \int_0^T \left(\int_0^t \left[\int_{r_s} K\left(\frac{u-r_s}{h}\right) \left\{ \int_{r_t} K\left(\frac{u-r_t}{h}\right) \pi(r_t,t|r_s,s)dr_t \right\} \pi(r_s)dr_s \right] ds \right) dt \right\} du$$

The above expression is not pretty, but it is correct. Moreover the expression inside the outer { } very closely resembles the expression for the expected squared occupation time of a Markov diffusion in Karlin and Taylor (1981). More importantly for my purposes, with a gaussian kernel, if the data is generated by the Vasicek model, then this expression can be computed relatively easily numerically. The reason is that the transition probabilities in the Vasicek model are conditionally and unconditionally gaussian. This combined with Fubini's theorem allows me to apply Result 2 three times, allowing me to reduce the above expression to the following simplified expression:

$$E\int_u \hat{\pi}(u)^2 du = \frac{2}{T^2}\frac{1}{\sqrt{2\pi}}\int_0^T \left[\int_0^t \frac{1}{\sqrt{2h^2+\frac{\sigma^2}{\kappa}(1-e^{-\kappa(t-s)})}} ds \right] dt.$$

This expression can be rewritten as:

$$E\int_u \hat{\pi}(u)^2 du = \frac{2}{T^2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2h^2+\frac{\sigma^2}{\kappa}}}\int_0^T \left[\int_0^t \frac{ds}{\sqrt{1-\left(\frac{\frac{\sigma^2}{\kappa}}{2h^2+\frac{\sigma^2}{\kappa}}\right)e^{-\kappa(t-s)}}} \right] dT$$

Making the substitution:

$$z = \sqrt{\left(\frac{\frac{\sigma^2}{\kappa}}{2h^2+\frac{\sigma^2}{\kappa}}\right)e^{-\kappa(t-s)}},$$

The above integral simplifies to become:

$$E\int_u \hat{\pi}(u)^2 du = \frac{2}{\kappa}\frac{2}{T^2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2h^2+\frac{\sigma^2}{\kappa}}}\int_0^T \left[\int_{\underline{z}}^{\overline{z}} \frac{dz}{z\sqrt{1-z^2}} \right] dT.$$

where

$$\underline{z} = \sqrt{\left(\frac{\frac{\sigma^2}{\kappa}}{2h^2+\frac{\sigma^2}{\kappa}}\right)e^{-\kappa t}} \text{ and } \overline{z} = \sqrt{\left(\frac{\frac{\sigma^2}{\kappa}}{2h^2+\frac{\sigma^2}{\kappa}}\right)}.$$

To compute the integral in terms of z, we make the additional substitution $z = cos(\rho)$. This substitution and algebra produce the result:

$$E\int_u \hat{\pi}(u)^2 du = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2h^2+\frac{\sigma^2}{\kappa}}}\left\{ 1+\frac{2}{\kappa}\frac{2}{T^2}\int_0^T ln\left[\frac{1+\sqrt{1-\psi(t)}}{1+\sqrt{1-\psi(0)}}\right] dt \right\},$$

where

$$\psi(t) = \left(\frac{\frac{\sigma^2}{\kappa}}{2h^2 + \frac{\sigma^2}{\kappa}} \right) e^{-\kappa t}$$

Substituting the above expression as the first piece of the Mean Integrated Variance for the Vasicek process with a gaussian kernel produces our final result:

$$MIVAR(Vasicek) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2h^2 + \frac{\sigma^2}{\kappa}}} \frac{2}{\kappa} \frac{2}{T^2} \int_0^T ln \left[\frac{1 + \sqrt{1 - \psi(t)}}{1 + \sqrt{1 - \psi(0)}} \right] dt,$$

where $\psi(t)$ is as above.

The final integral in the above expression can be numerically computed very rapidly in Mathematica and produces the Mean Integrated Variance of the kernel density estimate.

C Discrete Sampling Kernel Density Estimates.

In this section, we will derive expressions for the first and second moments of

$$\hat{\pi}(u) = \frac{1}{Nh} \sum_{s=\Delta s}^{N\Delta s} K(\frac{u - r_s}{h})$$

when the interest rate at time 0 is drawn from density $\pi(u)$ and then r evolves as in equation (3) in the text. These first and second moments can then be used to compute the mean and variance of the kernel density estimate, and can also be used to compute Mean Integrated Squared Bias and Mean Integrated Squared Error.

The Expected Kernel Density Estimate

The first moment is $E\hat{\pi}(u)$; solving for this moment is easy:

$$\begin{aligned}
E\hat{\pi}(u) &= E\frac{1}{Nh} \sum_{s=\Delta s}^{N\Delta s} K(\frac{u - r_s}{h}) \\
&= \frac{1}{N} \sum_{s=\Delta s}^{N\Delta s} \int_{r_0} \left(\int_{r_s} \frac{1}{h} K\left(\frac{u - r_s}{h}\right) \pi(r_s, s|r_0, 0) dr_s \right) \pi(r_0) dr_0 \\
&= \frac{1}{N} \sum_{s=\Delta s}^{N\Delta s} \int_{r_s} \frac{1}{h} K\left(\frac{u - r_s}{h}\right) \pi(r_s) dr_s \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{h^2 + \frac{\sigma^2}{2\kappa}}} e^{-.5\left(\frac{u-\theta}{\sqrt{h^2 + \frac{\sigma^2}{2\kappa}}}\right)^2}
\end{aligned}$$

The second equality is by the law of iterated conditional expectations. The third equality is by Fubini's theorem and result 1. The final equality follows by result 2.

The Expected Squared Kernel Density Estimate

Second moments are much more difficult than first moments. The expression for the expected square of the kernel density estimate is:

$$\begin{aligned} E\hat{\pi}(u)^2 &= E\left(\frac{1}{Nh}\sum_{s=\Delta s}^{N\Delta s} K(\frac{u-r_s}{h})\right)^2 \\ &= \left(\frac{1}{N}\right)^2 \left(\sum_{s=\Delta s}^{N\Delta s} E[\frac{1}{h}K(\frac{u-r_s}{h})]^2 + 2\sum_{t=\Delta t}^{N\Delta t}\sum_{s=\Delta s}^{t-\Delta t} E\frac{1}{h}K(\frac{u-r_s}{h})\frac{1}{h}K(\frac{u-r_t}{h})\right) \end{aligned}$$

The first piece of this expectation is relatively easy to find. We know for a Gaussian kernel that:

$$\begin{aligned} \left[\frac{1}{h}K(\frac{u-r_s}{h})\right]^2 &= \frac{1}{\sqrt{2\pi}h}\frac{1}{\sqrt{2\pi}h}e^{-.5\left(\frac{u-r_s}{\frac{h}{\sqrt{2}}}\right)^2} \\ &= \frac{1}{h\sqrt{2}\sqrt{2\pi}}\frac{1}{\frac{h}{\sqrt{2}}}K\left(\frac{u-r_s}{\frac{h}{\sqrt{2}}}\right) \end{aligned}$$

Therefore, using the method used to derive the formula for $E\hat{\pi}(u)$ we have:

$$\begin{aligned} E\left[\frac{1}{h}K(\frac{u-r_s}{h})\right]^2 &= \frac{1}{h\sqrt{2}\sqrt{2\pi}}E\frac{1}{\frac{h}{\sqrt{2}}}K\left(\frac{u-r_s}{\frac{h}{\sqrt{2}}}\right) \\ &= \frac{1}{h\sqrt{2}\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}}e^{-.5\left(\frac{u-\theta}{\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}}\right)^2} \end{aligned}$$

The second piece of the expectation is more difficult because it involves a double sum of terms with general form:

$$\begin{aligned} &E\left[\frac{1}{h}K\left(\frac{u-r_s}{h}\right)\frac{1}{h}K\left(\frac{u-r_t}{h}\right)\right] = \\ &= \int_{r_0}\left(\int_{r_s}\frac{1}{h}K\left(\frac{u-r_s}{h}\right)\left\{\int_{r_t}\frac{1}{h}K\left(\frac{u-r_t}{h}\right)\pi(r_t,t|r_s,s)dr_t\right)\pi(r_s,s|r_0,0)dr_s\right)\pi(r_0)dr_0. \\ &= \int_{r_s}\frac{1}{h}K\left(\frac{u-r_s}{h}\right)\left\{\int_{r_t}\frac{1}{h}K\left(\frac{u-r_t}{h}\right)\pi(r_t,t|r_s,s)dr_t\right\}\pi(r_s)dr_s. \\ &= \int_{r_s}\frac{1}{h}K\left(\frac{u-r_s}{h}\right)\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h^2+V(r_t,t|r_s,s)}}\exp^{-.5\left(\frac{u-\mu(r_t,t|r_s,s)}{\sqrt{h^2+V(r_t,t|r_s,s)}}\right)^2}\pi(r_s)dr_s \end{aligned}$$

where $V(r_t,t|r_s,s)$ is the variance of r_t given r_s, and $\mu(r_t,t|r_s,s)$ is the mean of r_t given r_s.

The first equality above follows by the law of iterated expectations; the second equality follows from result 1; and the third equality follows by result 2. This generates an answer that is the product of three gaussian density functions multiplied by some additional factors. This makes it possible to apply additional results to further simplify the integral. Substitution of the expressions for $V(r_t,t|r_s,s)$ and $\mu(r_t,t|r_s,s)$, and some algebra makes it possible to represent the above integral as the product of three gaussian density functions each of which involve r_s minus some mean divided by some variance. This makes it possible to apply result 4 to any two of these density functions. The resulting expression contains the product of only two gaussian density functions that involve r_s. Result 2 can be applied to these functions to integrate out r_s. This produces our final complicated expression which, for the sake of notational compactness, we will write as $G(\kappa,t-s,V_E,h)$:

$$G(\kappa,t-s,V_E,h)=E\left[\frac{1}{h}K\left(\frac{u-r_s}{h}\right)\frac{1}{h}K\left(\frac{u-r_t}{h}\right)\right]=$$
$$C(\kappa,t-s,V_E,h)\frac{1}{\sqrt{2\pi}V_1(\kappa,t-s,V_E,h)}e^{-.5\left(\frac{u-\theta}{V_1(\kappa,t-s,V_E,h)}\right)^2}\frac{1}{\sqrt{2\pi}}\frac{1}{V_2(h,V_E)}e^{-.5\left(\frac{u-\theta}{V_2(h,V_E)}\right)^2}$$

where,

$$\begin{aligned}
V_E &= \frac{\sigma^2}{2\kappa}\\
C(\kappa,t-s,V_E,h) &= \frac{e^{\kappa(t-s)}}{e^{\kappa(t-s)}-\frac{V_E}{V_E+h^2}}\\
V_1(\kappa,t-s,V_E,h) &= \sqrt{\frac{(h^2+V_E)e^{2\kappa(t-s)}-V_E+\frac{h^2V_E}{h^2+V_E}}{\left(e^{\kappa(t-s)}-\frac{V_E}{V_E+h^2}\right)^2}}\\
V_2(h,V_E) &= \sqrt{h^2+V_E}
\end{aligned}$$

Although the final form of this expression appears onerous, it only involves a leading term multiplied by the product of two gaussian density functions. More importantly, it is fully analytical which means it is possible to generate (for a given bandwidth) an analytical expression for the variance of the kernel density estimate at each point u. This is done below.

Variance of the Discrete Sampling Kernel Density Estimate.

The variance of $\hat{\pi}(u)$ is given by $E[\hat{\pi}(u)]^2 - [E\hat{\pi}(u)]^2$. This variance has the following form:

$$
\begin{aligned}
Var[\hat{\pi}(u)] &= \left(\frac{1}{N^2}\right)\left(\frac{N}{h\sqrt{2}\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}}\exp^{-.5\left(\frac{u-\theta}{\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}}\right)^2} + 2\sum_{t=1}^{N}\sum_{s=1}^{t-1}G(\kappa, t-s, V_E, h)\right) \\
&- \frac{1}{\sqrt{2}\sqrt{2\pi}\sqrt{V_E+h^2}}\frac{1}{\sqrt{2\pi}\sqrt{\frac{V_E+h^2}{2}}}\exp^{-.5\left(\frac{u-\theta}{\sqrt{\frac{V_E+h^2}{2}}}\right)^2}
\end{aligned}
$$

where $G(\kappa, t-s, V_E, h)$ is as given above.

Variance of the Continuous Sampling Kernel Density Estimate.

To compute the Variance of the Continuous Sampling Kernel Density Estimate, it suffices to take limits of the variance for the discrete sampling kernel density estimate as the sampling interval goes to 0. Define sampling interval Δt and Δs so that

$$\frac{1}{N} = \frac{\Delta t}{T} = \frac{\Delta s}{T},$$

and redefine $G(.)$ to remove the discreteness captured by the term ΔT, i.e. define $G^*(\kappa, t-s, V_E, h)$ as:

$$
\begin{aligned}
&G^*(\kappa, t-s, V_E, h) = \\
&\quad C^*(\kappa, t-s, V_E, h)\frac{1}{\sqrt{2\pi}V_1^*(\kappa, t-s, V_E, h)}e^{-.5\left(\frac{u-\theta}{V_1^*(\kappa,t-s,V_E,h)}\right)^2}\frac{1}{\sqrt{2\pi}}\frac{1}{V_2(h, V_E)}e^{-.5\left(\frac{u-\theta}{V_2(h,V_E)}\right)^2}
\end{aligned}
$$

where,

$$
\begin{aligned}
V_E &= \frac{\sigma^2}{2\kappa} \\
C^*(\kappa, t-s, V_E, h) &= \frac{e^{\kappa(t-s)}}{e^{\kappa(t-s)} - \frac{V_E}{V_E+h^2}} \\
V_1^*(\kappa, t-s, V_E, h) &= \sqrt{\frac{(h^2+V_E)e^{2\kappa(t-s)} - V_E + \frac{h^2 V_E}{h^2+V_E}}{\left(e^{\kappa(t-s)} - \frac{V_E}{V_E+h^2}\right)^2}} \\
V_2(h, V_E) &= \sqrt{h^2+V_E}
\end{aligned}
$$

With this change in notation, the variance of the continous sampling kernel density estimate has form:

$$\lim_{\Delta t \to 0} \lim_{\Delta s \to 0} \left(\frac{1}{T^2}\right) \left(\frac{T\Delta s \Delta t}{h\sqrt{2}\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}} \exp^{-.5\left(\frac{u-\theta}{\sqrt{\frac{h^2}{2}+\frac{\sigma^2}{2\kappa}}}\right)^2} + 2\sum_{t=\Delta t}^{T} \left\{ \sum_{s=\Delta s}^{t-\Delta s} G^*(\kappa, t-s, V_E, h)\Delta s \right\} \Delta t \right)$$

$$- \frac{1}{\sqrt{2}\sqrt{2\pi}\sqrt{V_E+h^2}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{V_E+h^2}{2}}} \exp^{-.5\left(\frac{u-\theta}{\sqrt{\frac{V_E+h^2}{2}}}\right)^2}$$

The first term inside the parenthesis goes to 0 while the second term converges to a double integral. The term outside the parenthesis remains unchanged. This yields our final result for the variance of the continuous sampling kernel density estimate:

$Var[\hat{\pi}(u)] =$

$$\frac{2}{T^2}\int_{t=0}^{T}\left\{\int_{s=0}^{t} G^*(\kappa, t-s, V_E, h)ds\right\}dt - \frac{1}{\sqrt{2}\sqrt{2\pi}\sqrt{V_E+h^2}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{V_E+h^2}{2}}} \exp^{-.5\left(\frac{u-\theta}{\sqrt{\frac{V_E+h^2}{2}}}\right)^2}$$

Unfortunately, this integral is too difficult to reduce further.

MIVAR of Discrete Sampling Kernel Density Estimate.

The mean integrated variance of the kernel density estimate is simply the integral of the variance with respect to u. The first and third terms of the expressions for variance are trivial to integrate. The middle term involves the product of two gaussian densities and hence can be integrated using result 2. Simplification produces the result that is contained in proposition II.

Mean Integrated Squared Bias.

Since the expected kernel density estimate is the same in both the continuous and discrete sampling cases, the mean integrated squared bias is also the same in both cases and is presented in the results for the kernel density estimate with continuous sampling.

D Covariance of the kernel density estimates.

For kernel density estimates $\hat{\pi}(u)$ and $\hat{\pi}(v)$ of the ergodic distribution of r at points u and v, this section derives expressions for $Cov(\hat{\pi}(u), \hat{\pi}(v))$. Robinson [1984] provides a central limit theorem in which this covariance goes to zero asymptotically. However, he cautions that in applications, the covariance will be nonzero due to positive bandwidth and due to the dependence in the data. Because we derive analytical (yet complicated) expressions for this covariance in the case of the Vasicek model, the magnitude of this covariance as function of bandwidth size and data dependence can be studied for finite sample kernel density estimates.

Covariance with discrete sampling kernel estimate.

The formula for the covariance is:

$$Cov(\hat{\pi}(u), \hat{\pi}(v)) = E[\hat{\pi}(u)\hat{\pi}(v)] - [E\hat{\pi}(u)][E\hat{\pi}(v)].$$

The only piece of this formula that has not been solved for is $E[\hat{\pi}(u)\hat{\pi}(v)]$. This can be expressed as:

$$E[\hat{\pi}(u)\hat{\pi}(v)] = E\left(\frac{1}{N}\sum_{t=1}^{N}\frac{1}{h}K(\frac{u-r_t}{h})\frac{1}{N}\sum_{s=1}^{N}\frac{1}{h}K(\frac{v-r_s}{h})\right)$$

$$= \frac{1}{N^2h^2}E\left(\sum_{t=1}^{N}K(\frac{u-r_t}{h})K(\frac{v-r_t}{h}) + \sum_{t=1}^{N}K(\frac{u-r_t}{h})\sum_{s=1}^{t-1}K(\frac{v-r_s}{h}) + \sum_{t=1}^{N}K(\frac{v-r_t}{h})\sum_{s=1}^{t-1}K(\frac{u-r_s}{h})\right)$$

The right hand side of this expression involves three sets of terms. However, the second and third set are virtually identical since one can be obtained from the other by switching u and v. Therefore, it is only necessary to compute

$$E\left(K(\frac{u-r_t}{h})K(\frac{v-r_t}{h})\right) \text{ and } E\left(K(\frac{u-r_t}{h})K(\frac{v-r_s}{h})\right).$$

Using results 1 and 4, we find:

$$E\left(K(\frac{u-r_t}{h})K(\frac{v-r_t}{h})\right) = J_1(u,v,h,\theta,V_E) =$$

$$\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{h^2}{2}+V_E}}\exp^{-.5\left(\frac{.5(u+v)-\theta}{\sqrt{\frac{h^2}{2}+V_E}}\right)^2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2h^2}}\exp^{-.5\left(\frac{u-v}{\sqrt{2h^2}}\right)^2}$$

Similarly, using results 1 and 4, we find that:

$$E\left(K(\frac{u-r_t}{h})K(\frac{v-r_s}{h})\right) = J_2(u,v,h,\theta,V_E,\kappa,(t-s)) =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{h^2 + V_E}} \exp^{-.5\left(\frac{v-\theta}{\sqrt{h^2+V_E}}\right)^2} \quad \text{x}$$

$$\frac{1}{\sqrt{2\pi}} \frac{e^{\kappa(t-s)}}{\sqrt{\left(\frac{h^2 V_E}{h^2+V_E}\right) + (h^2+V_E)e^{2\kappa(t-s)} - V_E}} \exp^{-.5\left(\frac{(u-\theta)e^{\kappa(t-s)} + (\theta-v)\frac{V_E}{h^2+V_E}}{\sqrt{\left(\frac{h^2 V_E}{h^2+V_E}\right) + (h^2+V_E)e^{2\kappa(t-s)} - V_E}}\right)^2}$$

Therefore, $E[\hat{\pi}(u)\hat{\pi}(v)]$

$$= \frac{1}{N^2}\left(\sum_{t=1}^{N} J_1(u,v,h,\theta,V_E) + \sum_{t=1}^{N}\sum_{s=1}^{t-1} J_2(u,v,h,\theta,V_E,\kappa,(t-s)) + \sum_{t=1}^{N}\sum_{s=1}^{t-1} J_2(v,u,h,\theta,V_E,\kappa,(t-s))\right)$$

This is enough information to compute the covariance in the case of discrete sampling.

Covariance with Continuous Sampling

In the case of continuous sampling, we can take limits as above. This yields $E[\hat{\pi}(u)\hat{\pi}(v)]$:

$$= \frac{1}{T^2}\left(\int_{t=0}^{T}\int_{s=0}^{t} J_2^*(u,v,h,\theta,V_E,\kappa,(t-s))dsdt + \int_{t=0}^{T}\int_{s=0}^{t} J_2^*(v,u,h,\theta,V_E,\kappa,(t-s))dsdt\right)$$

Table 1

Optimal Bandwidth and MISE for Vasicek Process

Model	Vasicek Parameters			Optimal Bandwidth	MISE	MISVAR	MISB
	κ	θ	σ^2				
(-2)	3.43348	0.089102	0.0087416	0.0140979	0.176687	0.146629	0.0300585
(-1)	1.71674	0.089102	0.0043708	0.0175509	0.306968	0.241167	0.0658013
(0)	0.85837	0.089102	0.0021854	0.0217661	0.511568	0.375094	0.136474
(1)	0.42917	0.089102	0.0010927	0.0268048	0.806781	0.543768	0.263014
(2)	0.21459	0.089102	0.00054635	0.0325055	1.17959	0.721825	0.457764

Notes: The table presents theoretical results that would be expected when computing kernel density estimates of the long run distribution of the spot interest rate using 22 years of continuously sampled data when the observation of the spot interest rates begin with a draw from their ergodic distribution at time 0 and then evolve as in the Vasicek model:

$$dr = \kappa(\theta - r)dt + \sigma dW.$$

Parameters for five models are considered. Model (0) is a baseline model since the parameters for this model are empirical estimates reported in Ait-Sahalia [1996a]. The long run distribution of the spot interest rate is the same for all five models, but the rate at which interest rates revert to their long run distribution differs. Columns (2)-(4) list the parameters of the Vasicek model for each specification. A gaussian kernel is used for all estimates. For each specification, column (5) reports the bandwidth choice that minimizes the Mean Integrated Squared Error of the kernel density estimate. Columns (6)-(8) report the Mean Integrated Squared Error (MISE), Mean Integrated Squared Variance (MISVAR), and Mean Integrated Squared Bias (MISB) of the kernel density estimates that use the bandwidth in column (5). Details on bandwidth selection are presented in the appendix.

Table 2: MISE Minimizing Bandwidth as a function of Data Span (T), and Sampling Frequency for Five Parameterizations of Vasicek Model

T = 5 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0224	0.0276	0.0333	0.0389	0.0434
1 day	0.0224	0.0276	0.0333	0.0389	0.0434
5 day	0.0224	0.0276	0.0333	0.0389	0.0434
10 day	0.0225	0.0276	0.0333	0.0389	0.0434
20 day	0.0229	0.0277	0.0333	0.0389	0.0434
30 day	0.0234	0.0278	0.0334	0.0389	0.0434
T = 10 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0181	0.0224	0.0276	0.0333	0.0389
1 day	0.0181	0.0224	0.0276	0.0333	0.0389
5 day	0.0181	0.0224	0.0276	0.0333	0.0389
10 day	0.0183	0.0224	0.0276	0.0333	0.0389
20 day	0.0187	0.0225	0.0276	0.0333	0.0389
30 day	0.0192	0.0227	0.0276	0.0333	0.0389
T = 20 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0145	0.0181	0.0224	0.0276	0.0333
1 day	0.0145	0.0181	0.0224	0.0276	0.0333
5 day	0.0146	0.0181	0.0224	0.0276	0.0333
10 day	0.0148	0.0181	0.0224	0.0276	0.0333
20 day	0.0154	0.0183	0.0224	0.0276	0.0333
30 day	0.0160	0.0185	0.0225	0.0276	0.0333
T = 50 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0108	0.0135	0.0169	0.0209	0.0258
1 day	0.0108	0.0135	0.0169	0.0209	0.0258
5 day	0.0110	0.0136	0.0169	0.0209	0.0258
10 day	0.0113	0.0136	0.0169	0.0209	0.0258
20 day	0.0120	0.0139	0.0169	0.0209	0.0258
30 day	0.0126	0.0141	0.0170	0.0209	0.0258
T = 100 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0087	0.0108	0.0135	0.0169	0.0209
1 day	0.0087	0.0108	0.0135	0.0169	0.0209
5 day	0.0089	0.0109	0.0135	0.0169	0.0209
10 day	0.0093	0.0110	0.0136	0.0169	0.0209
20 day	0.0101	0.0113	0.0136	0.0169	0.0209
30 day	0.0107	0.0117	0.0137	0.0169	0.0209

Table 3: MISE as a function of Data Span (T), and Sampling Frequency for Five Parameterizations of the Vasicek Model

T = 5 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.5466	0.8544	1.2337	1.6090	1.9007
1 day	0.5467	0.8544	1.2337	1.6090	1.9007
5 day	0.5474	0.8546	1.2338	1.6090	1.9007
10 day	0.5498	0.8553	1.2340	1.6091	1.9007
20 day	0.5587	0.8579	1.2348	1.6094	1.9009
30 day	0.5719	0.8622	1.2361	1.6099	1.9011
T = 10 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.3302	0.5466	0.8544	1.2337	1.6090
1 day	0.3302	0.5466	0.8544	1.2337	1.6090
5 day	0.3310	0.5468	0.8544	1.2337	1.6090
10 day	0.3331	0.5474	0.8546	1.2338	1.6091
20 day	0.3409	0.5498	0.8553	1.2340	1.6091
30 day	0.3520	0.5536	0.8564	1.2343	1.6092
T = 20 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.1910	0.3302	0.5466	0.8544	1.2337
1 day	0.1911	0.3302	0.5466	0.8544	1.2337
5 day	0.1917	0.3304	0.5467	0.8544	1.2337
10 day	0.1937	0.3310	0.5468	0.8544	1.2337
20 day	0.2005	0.3331	0.5474	0.8546	1.2338
30 day	0.2094	0.3365	0.5484	0.8549	1.2338
T = 50 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0882	0.1590	0.2780	0.4673	0.7455
1 day	0.0882	0.1590	0.2780	0.4673	0.7455
5 day	0.0889	0.1592	0.2781	0.4673	0.7455
10 day	0.0906	0.1597	0.2782	0.4674	0.7455
20 day	0.0957	0.1616	0.2788	0.4675	0.7455
30 day	0.1019	0.1645	0.2797	0.4677	0.7456
T = 100 years	Model				
Sampling Frequency	(-2)	(-1)	(0)	(1)	(2)
Continuous	0.0478	0.0882	0.1590	0.2780	0.4673
1 day	0.0478	0.0882	0.1590	0.2780	0.4673
5 day	0.0478	0.0882	0.1590	0.2780	0.4673
10 day	0.0478	0.0882	0.1590	0.2780	0.4673
20 day	0.0478	0.0882	0.1590	0.2780	0.4673
30 day	0.0478	0.0882	0.1590	0.2780	0.4673

Table 4

Finite Sample Properties of Marginal Density Test Under Vasicek Model

A: Asymptotic and Finite Sample 5% Critical Values

Model	Vasicek Parameters			5% Critical Values			
	κ	θ	σ^2	Asymptotic	22 Year	Lower Bound	Upper Bound
(-2)	3.433480	0.089102	0.008742	1.645	11.52	10.21	13.36
(-1)	1.716740	0.089102	0.004371	1.645	17.00	13.43	20.25
(0)	0.858370	0.089102	0.002185	1.645	19.47	16.60	23.90
(1)	0.429185	0.089102	0.001093	1.645	19.17	15.41	25.43
(2)	0.214592	0.089102	0.000546	1.645	11.08	8.51	18.23

B: Asymptotic and Finite Sample 1% Critical Values

Model	Vasicek Parameters			1% Critical Values			
	κ	θ	σ^2	Asymptotic	22 Year	Lower Bound	Upper Bound
(-2)	3.433480	0.089102	0.008742	2.33	17.70	15.45	49.87
(-1)	1.716740	0.089102	0.004371	2.33	29.73	23.90	52.79
(0)	0.858370	0.089102	0.002185	2.33	44.09	37.30	127.25
(1)	0.429185	0.089102	0.001093	2.33	36.44	29.55	166.56
(2)	0.214592	0.089102	0.000546	2.33	39.94	28.67	130.53

C: Empirical Rejection Frequencies Using Asymptotic Critical Values

Model	Vasicek Parameters			5% level		1% level	
	κ	θ	σ^2	Rej. Freq.	Std. Err.	Rej. Freq.	Std. Err.
(-2)	3.433480	0.089102	0.008742	45.60%	2.23%	37.80%	2.17%
(-1)	1.716740	0.089102	0.004371	57.40%	2.21%	49.40%	2.24%
(0)	0.858370	0.089102	0.002185	51.60%	2.23 %	43.60%	2.22%
(1)	0.429185	0.089102	0.001093	40.80%	2.20%	34.20%	2.12%
(2)	0.214592	0.089102	0.000546	21.00%	1.82%	18.80%	1.75%

Notes: Panels A and B present asymptotic and finite sample critical values for the $\hat{M}$ test statistic, described in section III. of the text, when the short term interest rate is sampled once a day for 22 years. Lower and upper bounds of a 95% confidence interval for these critical values are also provided. Panel C presents estimates of the probability of rejecting the null when it is true using asymptotic critical values. All finite sample results are based on 500 monte-carlo similuations.For all tests, the short rate was generated by the Vasicek model:

$$dr = \kappa(\theta - r)dt + \sigma dW,$$

with the parameters shown. The kernel density bandwidths used to compute each test are listed in Table 1. Details on bandwidth selection and on the $\hat{M}$ statistic are provided in the text.

Table 4A

Finite Sample Properties of Marginal Density Test Under Vasicek Model

A: Asymptotic and Finite Sample 5% Critical Values

Model	Vasicek Parameters			5% Critical Values			
	κ	θ	σ^2	Asymptotic	22 Year	Lower Bound	Upper Bound
(-2)	3.433480	0.089102	0.008742	1.645	10.78	9.71	13.48
(-1)	1.716740	0.089102	0.004371	1.645	15.07	12.64	19.10
(0)	0.858370	0.089102	0.002185	1.645	18.04	15.89	22.69
(1)	0.429185	0.089102	0.001093	1.645	19.66	16.53	26.58
(2)	0.214592	0.089102	0.000546	1.645	12.45	9.71	23.35

B: Asymptotic and Finite Sample 1% Critical Values

Model	Vasicek Parameters			1% Critical Values			
	κ	θ	σ^2	Asymptotic	22 Year	Lower Bound	Upper Bound
(-2)	3.433480	0.089102	0.008742	2.33	17.41	14.66	49.09
(-1)	1.716740	0.089102	0.004371	2.33	29.25	23.86	56.34
(0)	0.858370	0.089102	0.002185	2.33	46.36	30.56	133.72
(1)	0.429185	0.089102	0.001093	2.33	40.91	33.32	160.55
(2)	0.214592	0.089102	0.000546	2.33	50.79	36.41	153.01

C: Empirical Rejection Frequencies Using Asymptotic Critical Values

Model	Vasicek Parameters			5% level		1% level	
	κ	θ	σ^2	Rej. Freq.	Std. Err.	Rej. Freq.	Std. Err.
(-2)	3.433480	0.089102	0.008742	43.40%	2.22%	36.20%	2.15%
(-1)	1.716740	0.089102	0.004371	55.00%	2.22%	45.20%	2.23%
(0)	0.858370	0.089102	0.002185	47.20%	2.23%	39.60%	2.19%
(1)	0.429185	0.089102	0.001093	37.00%	2.16%	32.00%	2.09%
(2)	0.214592	0.089102	0.000546	20.60%	1.81%	18.20%	1.73%

Notes: Panels A and B present asymptotic and finite sample critical values for the $\hat{M}$ test statistic, described in footnote 19 of section III. A, when the short term interest rate is sampled once a day for 22 years. Lower and upper bounds of a 95% confidence interval for these critical values are also provided. Panel C presents estimates of the probability of rejecting the null when it is true using asymptotic critical values. All finite sample results are based on 500 monte-carlo similuations.For all tests, the short rate was generated by the Vasicek model:

$$dr = \kappa(\theta - r)dt + \sigma dW,$$

with the parameters shown. The kernel density bandwidths used to compute each test are listed in Table 1. Details on bandwidth selection and on the $\hat{M}$ statistic are provided in the text.

FIGURE 1: TRUE AND EXPECTED DENSITY ESTIMATES
TRUE DENSITY: SOLID, EXPECTED DENSITY: DASHED
+/− 1 STD ASYM CONF BANDS: DOTTED, T = 22 YEARS, N = 5500

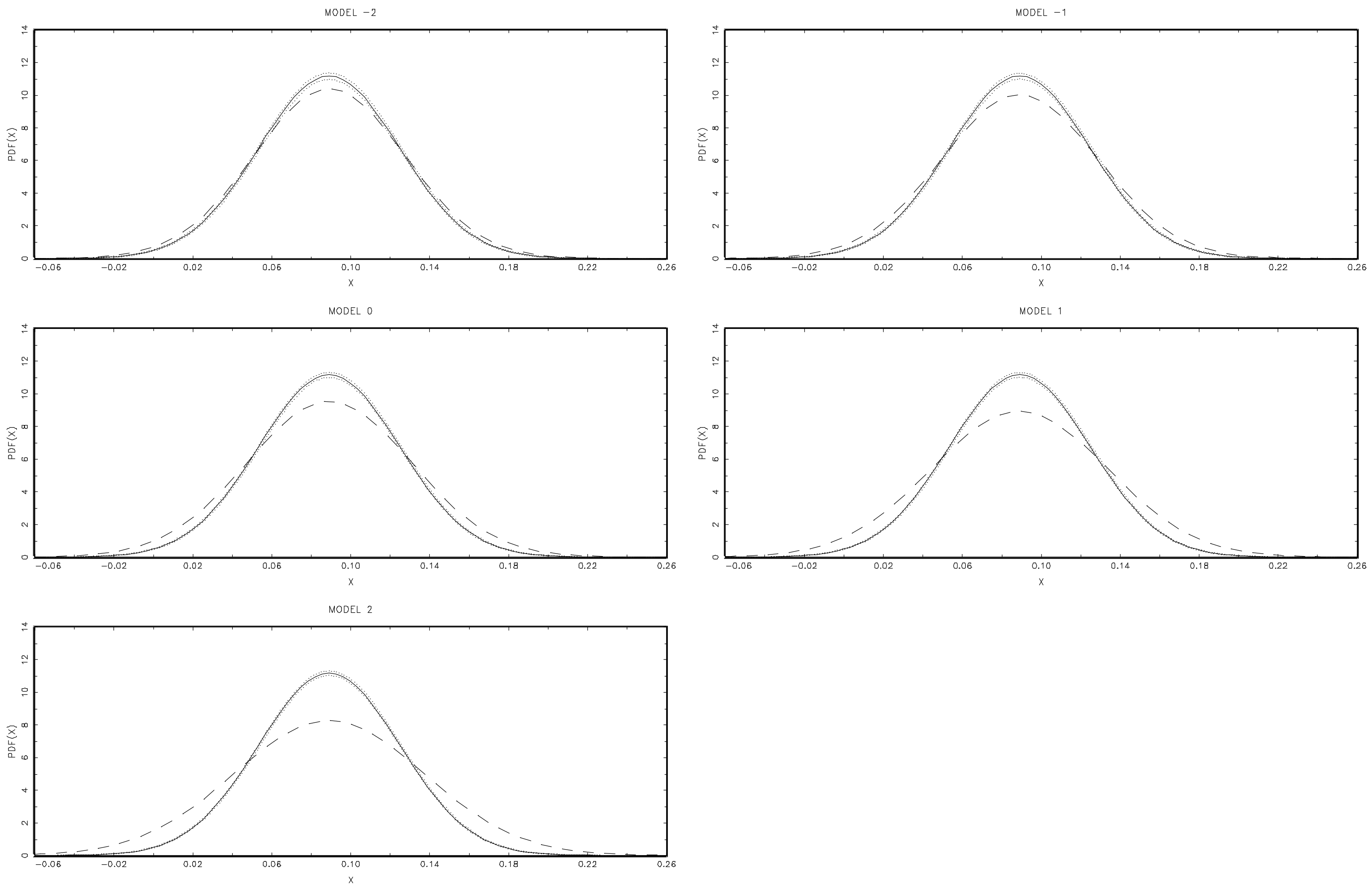

FIGURE 2: ASYMPTOTIC AND FINITE SAMPLE +/− 1 STD CONF BANDS
TRUE DENS: SOLID, ASYM BANDS: DOTTED, FINITE SAMPLE: DASHED
T = 22 YEARS, N=5500

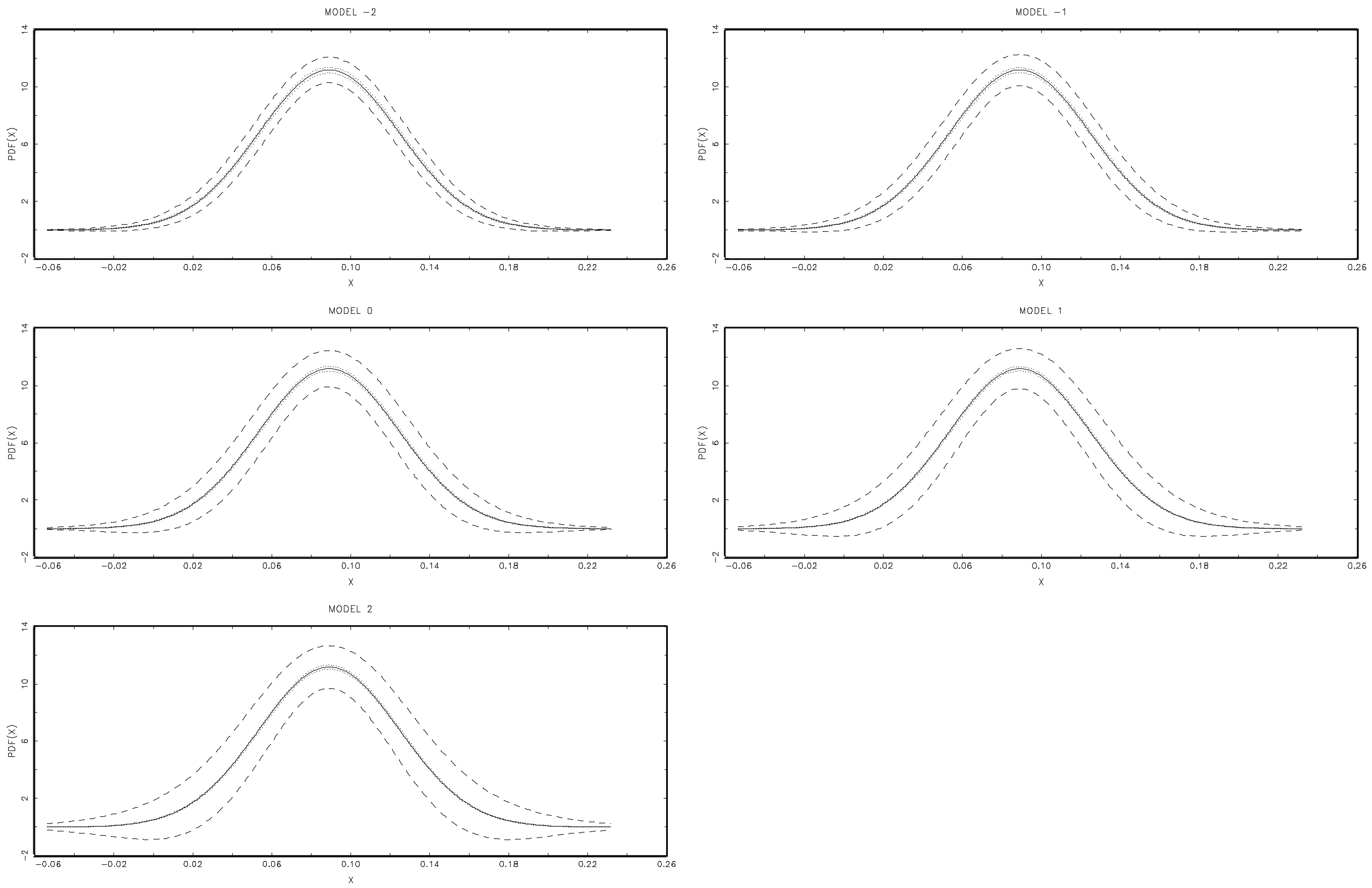

FIGURE 3: ASYMPTOTIC AND FINITE SAMPLE +/− 1 STD CONF BANDS
TRUE DENS: SOLID, ASYM BANDS: DOTTED, FINITE SAMPLE: DASHED
T = 100 YEARS, N=24000

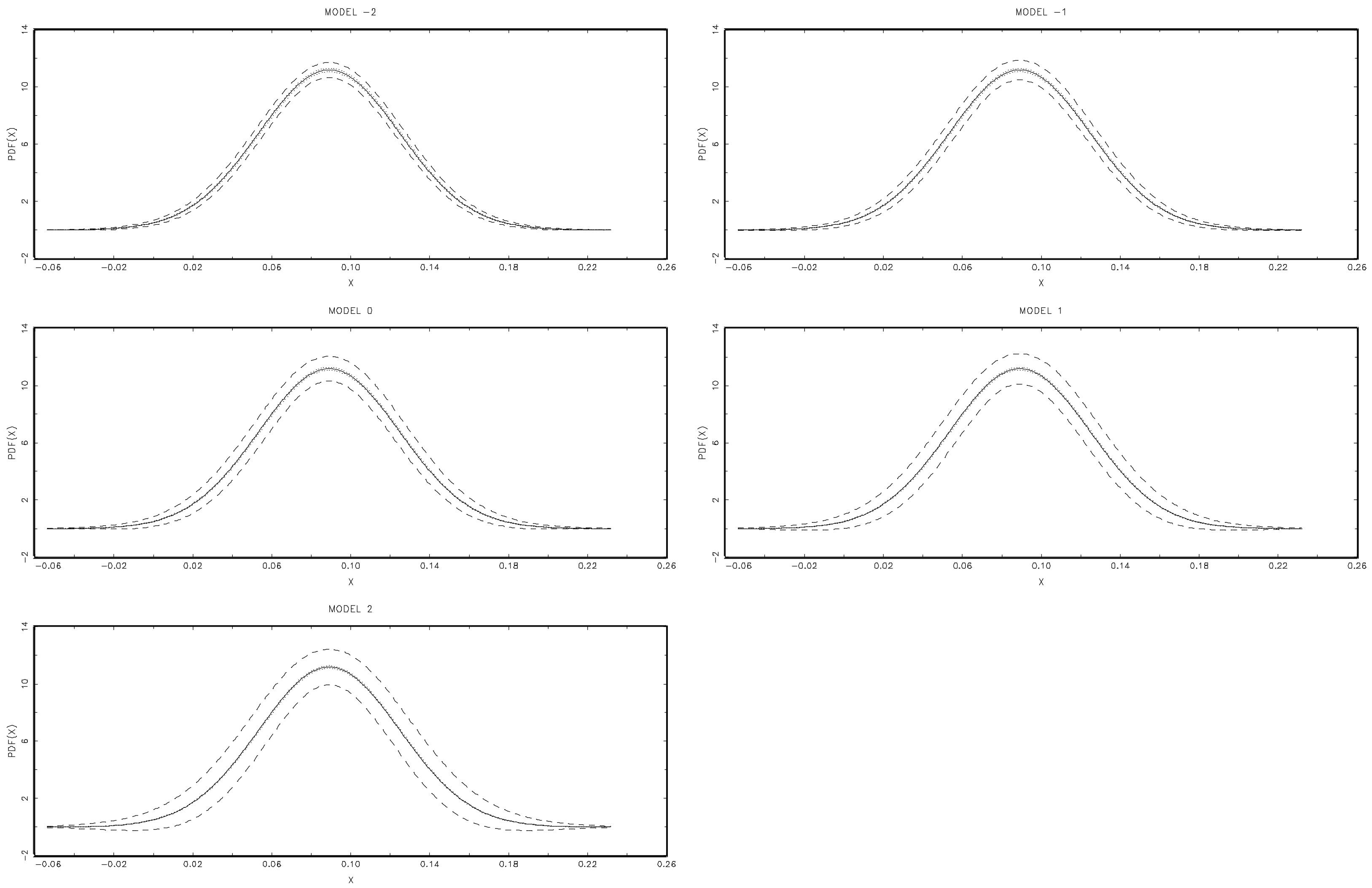

Figure 4: Correlation Function for Kernel Density Estimates
T = 22 Years, N = 5500

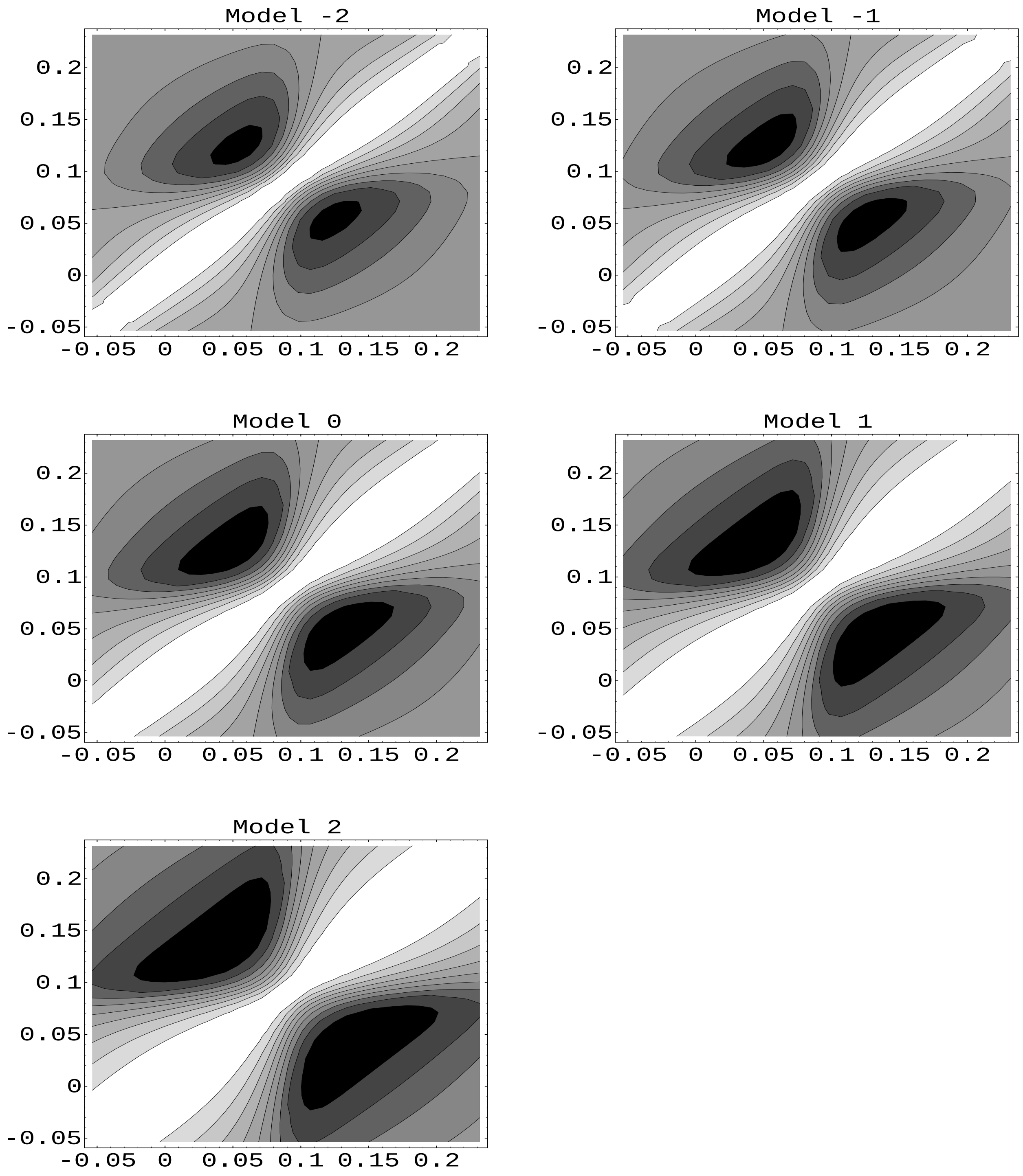

Figure 5: Correlation Function for Kernel Density Estimates
T = 100 Years, N = 24000

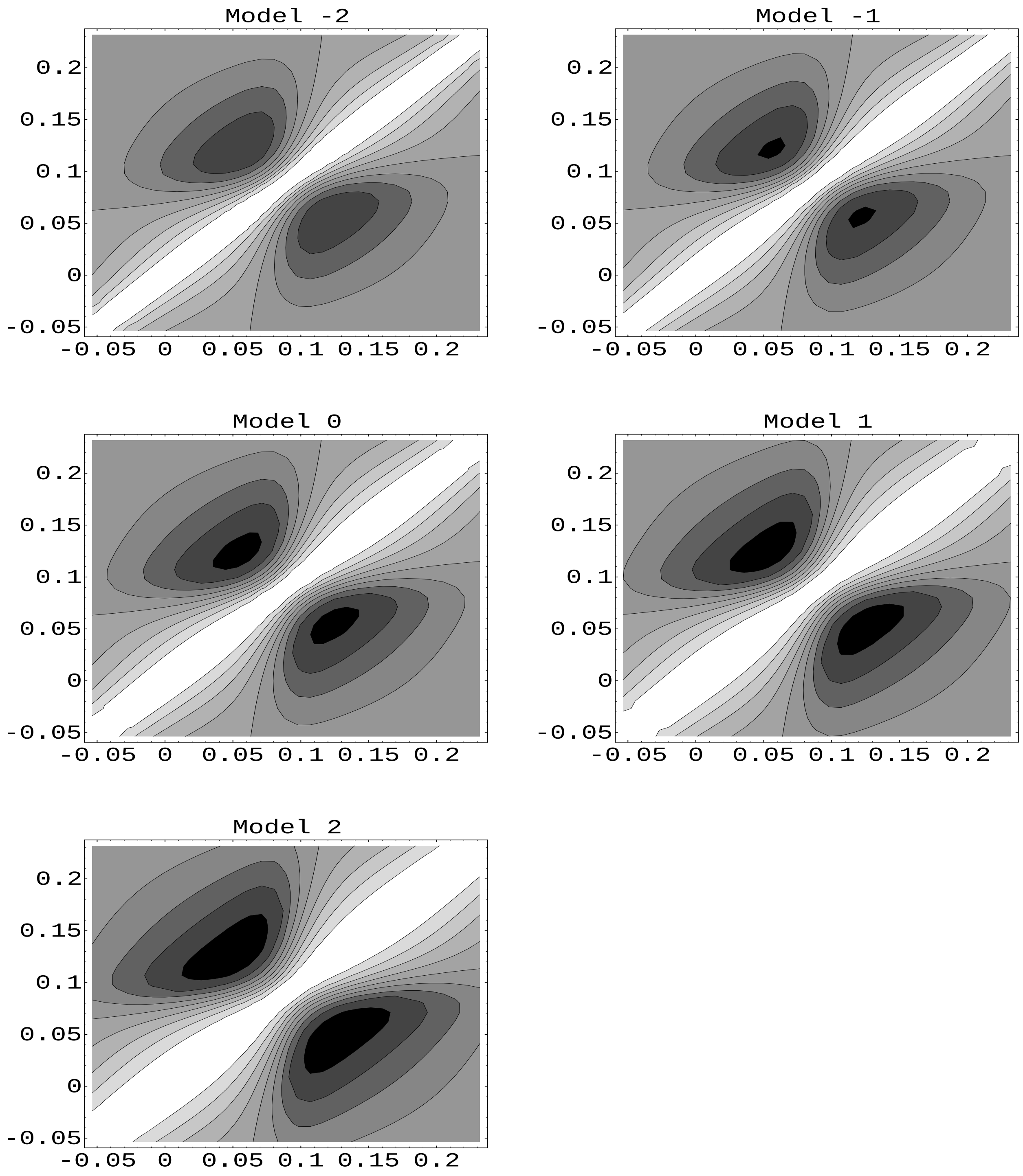

Figure 6: Covariance Function for Kernel Density Estimates
T = 22 Years, N = 5500

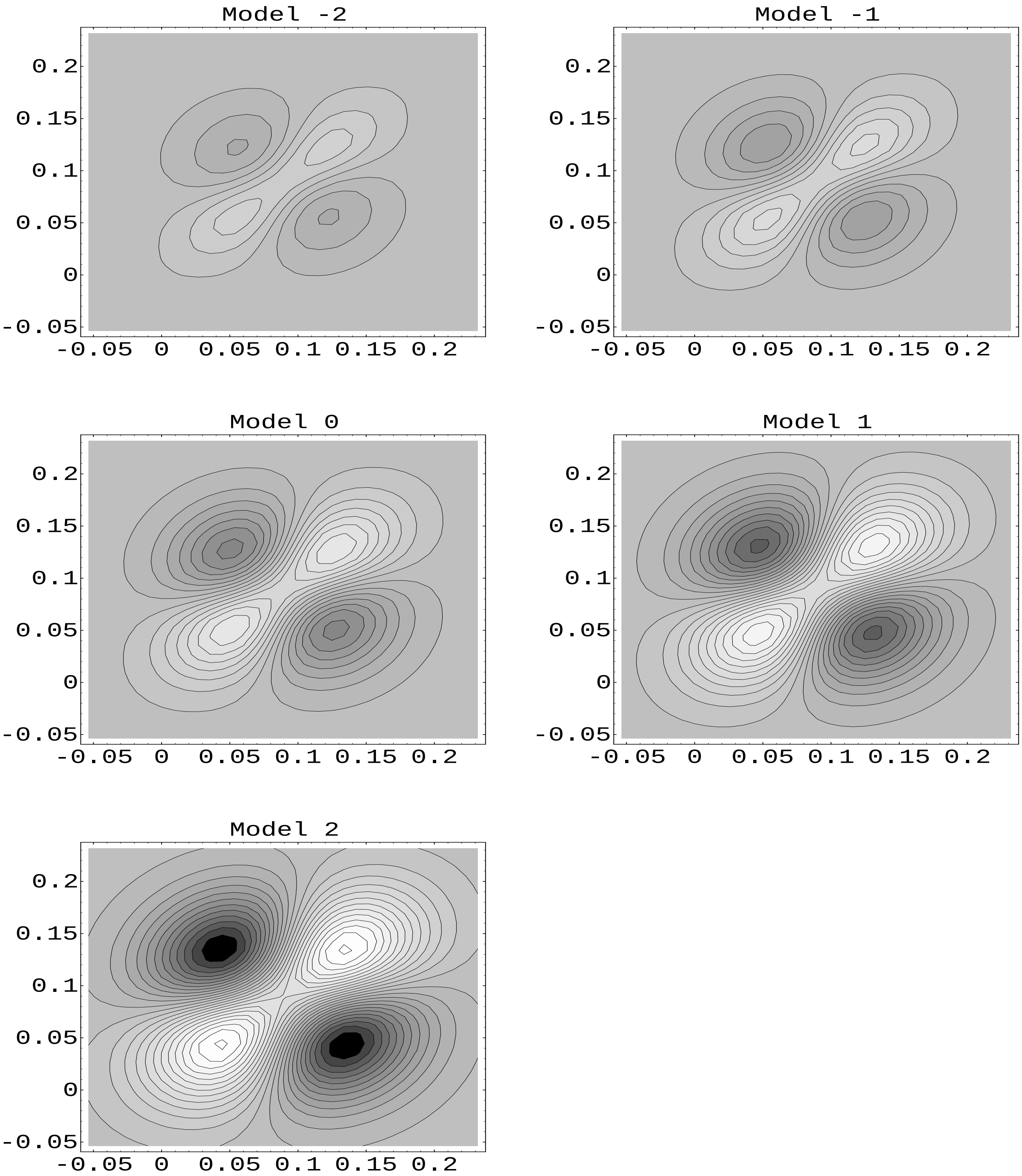

Figure 7: Covariance Function for Kernel Density Estimates
T = 100 Years, N = 24000