Nonparametric Estimation of Single Factor Heath-Jarrow-Morton Term Structure Models and a Test for Path Independence*

Andrew Jeffrey[†], Oliver Linton[‡], and Thong Nguyen[§]

March, 2000

Abstract

A nonparametric estimator of the term structure's volatility is provided for a class of Heath-Jarrow-Morton term structure models. This class of models is important because it captures, as a special case, all term structure models where the short term interest rate follows a time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy; such a structure is frequently considered in the term structure literature. A test for this 'Markovian short-term interest rate based structure' is developed. We find that this Markovian short-term interest rate based structure is not rejected. The result suggests that a sufficiently flexible one factor Markov short-term interest rate based model is able to capture the dynamics of the empirical terms structure of interest rates in the US bond market. A caveat of course is that nonparametric procedures inherently require large sample sizes.

*VERY PRELIMINARY. Comments are welcome. We would like to thank Konstantin Tyurin, Anjun Zhou and in particular, Federico Bandi and Peter Phillips, and seminar participants at the Yale School of Management and the Economics Department at Yale University for very helpful comments and suggestions.

[†]Yale School of Management, New Haven, CT. 06520-8200. E-mail: Andrew.Jeffrey@yale.edu

[‡]LSE and Yale University. Address for correspondence: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: lintono@lse.ac.uk. Web page: http://econ.lse.ac.uk/staff/olinton/. Research was supported by the National Science Foundation, and the Danish Social Science Research Council.

[§]Economics Department, Yale University, New Haven, CT 06520-8268. E-mail: Thong.Nguyen@yale.edu

1 Introduction

The advent of the Heath-Jarrow-Morton (1992) (hereafter HJM) framework has introduced an interesting way of thinking about the term structure's evolution. Within this framework an arbitrary initial term structure is taken as given, presumably generated by market equilibrium, and then the intertemporal transitions of the whole term structure are determined by characterizing how each point on the term structure reacts to movements in the set of Wiener processes that form the underlying sources of uncertainty in the bond market. This is distinctly different from most previous frameworks where the focus has been to model the evolution of specific economics variables that act as state variables for the fundamental sources of uncertainty in the bond market. In these other frameworks the evolution of the state variables is modelled as a joint Markov diffusion process and the term structure at time t is constructed so that it is a function of these state variables at time t only. For convenience we refer to such models as those from a Markov based paradigm.

In the last few years the empirical term structure literature has turned towards studying the intertemporal behavior of the short term interest rate (which we denote by $r(t)$). This is because $r(t)$ is almost always used as a state variable in the above models, and further, because once the evolution of $r(t)$ is known in the equivalent risk-neutral economy, then the whole term structure can be recovered.[1] Aït-Sahalia (1996), Stanton (1997), Jiang-Knight (1997), Jiang (1998), and Bandi (1998) have all adopted a general approach to ascertain the behavior of the short term interest rate by starting with the assumption that it follows a time-homogeneous univariate Markov diffusion process and then nonparametrically estimating the drift and volatility coefficients of this process. Motivation for this work stems from the lack of a theoretical underpinning regarding the parametric form of the short term interest rate process and the mixed results obtained from several competing parametric term structure models. This need to reduce the arbitrary parametric restrictions imposed by previous term structure models is provided by the above nonparametric approaches that allow the "data to speak for itself". Once the nature of the short term interest rate process is understood, the market price of interest rate risk is required to determine the short term interest rate process in the equivalent risk-neutral economy. Restricting the market price of interest rate risk at time t to be a function only of $r(t)$ implies that the short term interest rate will also follow a time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy and the resulting term structure model will then fall within the above mentioned Markov based paradigm.

A natural question at this point is the following: can a reasonable depiction of the term

[1] That is, the evolution of the term structure under the equivalent risk-neutral probability measure which can be interpeted as an equivalent economy where agents behave in a risk-neutral fashion. If the bond market is complete then there is only one such probability measure.

structure's evolution be obtained by assuming that it is governed by a time-homogeneous univariate Markov diffusion model for the short term interest rate? To answer this question we deviate from the above focus of estimating the short term interest rate process by examining whole term structure transitions instead. This is achieved by adopting the more general HJM framework and nonparametrically estimating the volatility of every point along the term structure (hereafter referred to as the "volatility structure"). Once the volatility structure has been obtained we can determine if it is consistent with a term structure that has been generated by a time-homogeneous univariate Markov diffusion model for $r(t)$. The primary implication of the answer to the above question is whether it is necessary to look further afield for models of intertemporal term structure behavior.

Given the generality of the HJM framework it is impossible to provide a nonparametric procedure that allows for the volatility structure to depend on all possible information. Consequently, for tractability, we restrict attention to the HJM framework in which only one Wiener process introduces uncertainty into the bond market and the volatility structure of the forward rate curve can depend on the contemporaneous short term interest rate and time to maturity. It may appear that our restrictions are severe. This is a deceptive impression since the class of term structure models permitted under these restrictions captures, as a special case, all models where the short term interest rate follows a time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy.[2] In fact, generically the class of models that we consider is path-dependent. Consequently a rejection of the time-homogeneous univariate Markov diffusion assumption for $r(t)$ in this setting suggests incorporating path-dependence in the term structure's evolution.

We do not consider the possibility of multiple stochastic factors in either a Markov based setting nor a path-dependent setting. The incorporation of multiple factors is related to the following fundamental questions. How many sources of uncertainty are required to adequately depict the intertemporal behavior of the term structure? Given the answer to this, is the evolution of the term structure path-dependent?[3] These questions are beyond the scope of this paper. However the work herein will hopefully provide some insights towards a means of addressing them.

[2] And if the market price of interest rate risk at time t is a function of $r(t)$ only then then short-term interest rate also follows a time-homogeneous univariate Markov diffusion in the observed economy.

[3] Path-dependence is meant in the following sense. If there are N sources of uncertainty in the bond market with N state variables tracking these sources of uncertainty then to recover the term structure at time t we must observe not only the state variables at time t but also their realized path up to time t. It is possible for a path-dependent model to have a path-independent representation when the number of state variables used exceeds the number of sources of uncertainty, for example Cheyette (1992,1996), Ritchken-Sankarasubramanian (1995), Jeffrey (1995a), and Inui-Kijima (1998). These models are still considered path-dependent since the additional state variables are not additional sources of uncertainty, they simply act as sufficient statistics for the path-dependent information.

To the best of our knowledge the only work that has applied nonparametric techniques to estimate the volatility structure in the HJM setting is the work by Zhou and Pearson (1998), who have extended Stanton's methodology. However, their analysis does not include the asymptotics conducive to statistical testing. In relation to the issue of whether the short term interest rate follows a time-homogeneous univariate Markov diffusion process, Aït-Sahalia (1997) provides preliminary evidence rejecting such behavior. His analysis is based on computing nonparametric estimators for conditional densities of the short term interest rate and testing inequality constraints imposed upon them implied by the properties of continuous-time diffusions. The data that his conclusions are based upon is a constructed time series of monthly short term interest rates from 1857 to 1995 and, as Aït-Sahalia states, "as a caveat, these preliminary empirical results are sensitive to the choice of data set" (page 32). Our analysis can be seen as an alternative means of searching for the possibility of path-dependence in the term structure's evolution.

In this paper we provide a consistent nonparametric estimator for the volatility structure and provide its asymptotic distribution. The form of the estimator is motivated by standard kernel smoothing techniques making use of the properties of the variance of term structure transitions, similar to Stanton (1997)'s estimators. Observing that expected term structure transitions also depend on the volatility structure, as a consequence of the no-arbitrage condition, we consider the possibility of using this restriction to estimate the volatility structure. We show that such an estimator will have a larger asymptotic variance relative to our original estimator (under a common nonparametric specification), and hence we do not advocate its use. Instead, we construct a χ^2 test statistic to determine if our original volatility structure estimate is consistent with the no-arbitrage restriction imposed on expected term structure transitions. Using zero-coupon bond yield curves extracted daily from Treasury bills, notes and bonds obtained from the Center for Research in Security Prices (CRSP) over the period June 1960 to December 1998, the estimated volatility structure of the yield curve appears to be consistent with the no-arbitrage restriction.

To ascertain the appropriateness of using a Markovian based paradigm to model intertemporal term structure behavior where the short term interest rate follows a time-homogeneous univariate Markov diffusion, we develop a χ^2 test statistic to determine if the estimated volatility structure is consistent with such an assumption. Rather surprisingly, we do not find strong support for rejecting this assumption in favor of a single factor path-dependent alternative.

The remainder of the paper proceeds as follows. In Section 2 the term structure's dynamics are discussed and our model specification is provided. A consistent estimator for the volatility structure is provided in Section 3 along with its asymptotic distribution. Section 4 contains the construction of a test to determine if the estimated volatility structure is consistent with the no-arbitrage restriction embedded in expected term structure transitions. Section 5 contains the construction of a test to ascertain if the estimated volatility structure is consistent with a term structure that has been generated in a Markovian based paradigm where the short term interest rate follows a time-homogeneous univariate Markov diffusion. The empirical estimation of the volatility structure and testing for consistency with both the no-arbitrage condition and the time-homogeneous univariate Markov diffusion assumption for $r(t)$ is reported in Section 6. Conclusions and suggestions for future research are in the final section. Technical proofs are in appendices.

2 Term Structure Dynamics and Model Specification

To set the framework and introduce notation let $P(t,T)$ denote the price of a one dollar face value, default free, zero coupon bond at time t that will mature at time T. A continuum of such bonds trade continuously with differing maturities (one for each date T). The instantaneous forward rate at time t for date T, denoted $f(t,T)$, is defined by $f(t,T) = -\partial \ln(P(t,T))/\partial T$. For convenience we will further assume that $-\partial f(t,T)/\partial T$ exists for all T. The HJM (1992) framework represents the term structure in terms of forward rates and for a single source of uncertainty in the bond market, introduced by the Wiener process $W(t)$, the uncertain evolution of each forward rate with fixed maturity date T satisfies the stochastic differential equation

$$df(t,T) = \alpha(\omega,t,T)dt + \gamma(\omega,t,T)dW(t), \tag{1}$$

where ω indicates the possible dependence on the term structure's realization up to time t. More specifically $\omega \in \mathcal{F}_t$ where $\mathcal{F}_t$ indicates all available information at time t generated by the term structure's evolution. Given the complete markets structure of this framework the no-arbitrage restriction between bonds of different maturities manifests itself in equation (1) with the following restriction on the forward rate's drift term:

$$\alpha(\omega,t,T) = \gamma(\omega,t,T)\left(\int_t^T \gamma(\omega,t,v)dv + \lambda(\omega,t)\right), \tag{2}$$

where $\lambda(\omega,t)$ is the market price of risk at time t for the single source of uncertainty in the bond market. Note that $\lambda(\omega,t)$ is maturity independent; intuitively this is because all assets (bonds with different maturities) are subject to the same source of uncertainty $W(t)$ and hence command the same risk premium per unit of risk. In this framework the short term interest rate at time t, denoted $r(t)$, can be defined as the instantaneously maturing forward rate at time t; that is $r(t) = f(t,t)$. The evolution of $r(t)$ resulting from equations (1) and (2) is

$$dr(t) = \left(\left.\frac{\partial f(t,T)}{\partial T}\right|_{T=t} + \gamma(\omega,t,t)\lambda(\omega,t)\right)dt + \gamma(\omega,t,t)dW(t). \tag{3}$$

One of our goals in this paper is to estimate the term structure's volatility which in the above setting is represented by the forward rate volatility structure $\gamma(\omega, t, T)$. The estimation procedure considered in subsequent sections relies on observing whole term structures at discrete points in time. These term structures are obtained by resorting to the curve fitting literature where an estimate of the whole term structure is extracted from a sample of both zero-coupon and coupon bearing government securities. In particular we adopt the Jeffrey-Linton-Nguyen (1999) implementation of the nonparametric term structure estimator proposed by Linton-Mammen-Nielsen-Tanggaard (1998). Typically, estimation of the forward rate curve is more difficult than yield curve estimation since forward rates require knowledge of the derivative of the yield curve. The yield at time t with maturity date T, denoted $y(t, T)$, is defined by $y(t, T) = -\frac{1}{T-t} \ln (P(t, T))$ and consequently the relationship between the forward rate curve and the yield curve is $f(t, T) = y(t, T) + (T - t)\partial y(t, T)/\partial T$. The difficulty associated with estimating $\partial y(t, T)/\partial T$ is primarily evident with highly flexible procedures for extracting the term structure, since the estimate of this derivative is sensitive to the procedure adopted. Further, it is well known that if a nonparametric technique is used then the derivative of an estimate has a slower rate of convergence than the estimate itself.

To circumvent the issue of potentially poor forward rate curve estimates we deviate from typical implementations of HJM based models by portraying the term structure with yields instead of forward rates.[4] The dynamics of the yield curve resulting from (1) and (2) is[5]

$$dy(t, T) = m(\omega, t, T)dt + \eta(\omega, t, T)dW(t), \tag{4}$$

where
$$\eta(\omega, t, T) = \frac{1}{T - t} \int_t^T \gamma(\omega, t, v)dv$$

$$m(\omega, t, T) = \frac{y(t, T) - r(t)}{T - t} + \lambda(\omega, t)\eta(\omega, t, T) + \frac{1}{2}(T - t)\eta(\omega, t, T)^2.$$

The short term interest rate process can be obtained from (3) by observing that $r(t) = y(t, t)$, $\gamma(\omega, t, t) = \eta(\omega, t, t)$ and $\partial f(t, T)/\partial T|_{T=t} = 2\ \partial y(t, T)/\partial T|_{T=t}$. In this setting the volatility of the term structure is now represented via the yield volatility structure $\eta(\omega, t, T)$ for which a nonparametric estimator is developed in the following section.[6]

[4] For example Heath-Jarrow-Morton (1990), Amin-Morton (1994) and Bühler-Uhrig-Homburg-Walter-Weber (1999) all use estimated forward rate curves in their analyses.

[5] See Appendix A1.

[6] If it is necessary to obtain an estimate for the forward rate volatility structure $\gamma(\omega, t, T)$ then one can either i) adopt the estimation procedure advocated in the following section of this paper except apply them to forward rates instead of yields, or ii) derive an estimate for $\gamma(\omega, t, T)$ from the yield volatility structure estimate by observing that $\gamma(\omega, t, T) = (T - t)\frac{\partial \eta(\omega, t, T)}{\partial T} + \eta(\omega, t, T)$.

Equation (4) above depicts the evolution of each yield with fixed maturity date T. However, at times it will be convenient to consider the evolution of a yield with a fixed time to maturity instead. Denote the yield at time t with time to maturity τ as $\widetilde{y}(t,\tau)$, that is $\widetilde{y}(t,\tau) = y(t, t+\tau)$. The intertemporal dynamics of $\widetilde{y}(t,\tau)$ implied by (4) is[7]

$$d\widetilde{y}(t,\tau) = \widetilde{m}(\omega,t,\tau)dt + \widetilde{\eta}(\omega,t,\tau)dW(t), \tag{5}$$

where

$$\widetilde{\eta}(\omega,t,\tau) = \eta(\omega,t,t+\tau)$$

$$\widetilde{m}(\omega,t,\tau) = \frac{\partial\widetilde{y}(t,\tau)}{\partial\tau} + \frac{\widetilde{y}(t,\tau) - r(t)}{\tau} + \lambda(\omega,t)\widetilde{\eta}(\omega,t,\tau) + \frac{1}{2}\,\tau\,\widetilde{\eta}(\omega,t,\tau)^2.$$

Despite the appearance of $\partial\widetilde{y}(t,\tau)/\partial\tau$ in the drift term of $\widetilde{y}(t,\tau)$ this representation will prove useful for developing both an estimate for the volatility structure and a test of the time-homogeneous univariate Markov diffusion assumption without actually requiring the computation of the derivative of the yield curve.

Potential functional forms for the term structure's volatility are vast. In its most general form it can depend on time t, maturity T, and all realized interest rates up to time t (indicated by ω). For our estimation we restrict attention to the class

$$\eta(\omega,t,T) = \eta(r(t),\tau), \text{ where } \tau = T - t, \tag{6}$$

in which case $\widetilde{\eta}(\omega,t,\tau) = \eta(r(t),\tau)$ as well. This class of volatility structures is quite general in the sense that it encompasses a wide variety of term structure models proposed in the literature. Of particular interest is that it captures all models from a Markov based paradigm where the short term interest rate follows a time-homogeneous univariate Markov diffusion.[8] Other specific examples include versions of Hull-White (1990) extended-Vasicek and extended-CIR models where the short term interest rate follows a univariate Markov diffusion with time-inhomogeneity introduced into the drift term only[9], the constant and exponential decay forward rate volatility structure models given as examples in HJM (1992), and a large subset of volatility structure forms suggested in Cheyette (1992), Brace-Musiela (1994), Ritchken-Sankarasubramanian (1995), Bliss-Ritchken (1996) and Jeffrey (1997a,b). Generically the volatility structure (6) results in a path-dependent term structure evolution[10] with the short term interest becoming non-Markov. This non-Markov behavior in $r(t)$

[7] See Appendix A1.

[8] See Appendix A2.

[9] See Appendix A2.

[10] Jeffrey (1995b) shows that in addition to restricting $\eta(\omega,t,T)$ to be a function of $r(t), t$, and T, severe functional form restrictions are also required to assure that the volatility structure is admissable in a Markov based paradigm where $r(t)$ is the state variable.

manifests itself in the drift term of (3) with the slope of the forward rate curve at the origin $\partial f(t,T)/\partial T|_{T=t}$ carrying information about the path that the term structure has taken.[11]

3 Yield Curve Volatility Structure Estimation

We now proceed to develop a nonparametric estimator for the yield volatility structure $\eta(r(t),\tau)$ without imposing any structure on the market price of interest rate risk $\lambda(\omega,t)$. This section provides the motivation for the estimator using kernel smoothing arguments [see Härdle (1991)] and in Appendix A3 it is shown that this estimator is asymptotically consistent and normally distributed. The estimator is based on observing the short term interest rate $r(t_i)$ and yield transitions $\Delta\widetilde{y}(t_i,\tau)=\widetilde{y}(t_{i+1},\tau)-\widetilde{y}(t_i,\tau)$ at $n-1$ points in time indexed by t_i for $i=1,\ldots,n-1$. However it is also possible, with minor modifications, to construct the estimator based on observing $r(t_i)$ and yield transitions $\Delta y(t_i,T)=y(t_{i+1},T)-y(t_i,T)$ at times t_i for $i=1,\ldots,n-1$.[12] For simplicity it is assumed that all time intervals t_i to t_{i+1} are equally spaced, that is $\Delta t=t_{i+1}-t_i$ for all i.

Consider the expectation of the square of an instantaneous transition in the yield curve characterized by (5), that is $E\left[(d\widetilde{y}(t,\tau))^2\middle|\mathcal{F}_t\right]$. It is apparent that this is related to the square of the yield volatility structure via

$$\widetilde{\eta}(\omega,t,\tau)^2=\lim_{\Delta t\downarrow 0}\left(\frac{1}{\Delta t}E\left[(\widetilde{y}(t+\Delta t,\tau)-\widetilde{y}(t,\tau))^2\middle|\mathcal{F}_t\right]\right), \tag{7}$$

observing that the drift term of the yield curve dynamics becomes negligible in determining $\widetilde{\eta}(\omega,t,\tau)^2$ as $\Delta t\downarrow 0$; that is, the random shocks in the bond market dominate changes in yields from one instant to the next. An obvious interpretation of (7) is that the yield volatility structure is the standard deviation of yield curve transitions. Since we restrict attention to volatility structures of the form $\widetilde{\eta}(\omega,t,\tau)=\eta(r(t),\tau)$, it is enough to condition the expectation in (7) on $r(t)$ only; that is, the realization of $r(t)$ is enough to determine the volatility structure at time t.[13] Consequently we have the following approximation for the

[11] This can be seen by observing, from equations (1) and (2), that

$$\left.\frac{\partial f(t,T)}{\partial T}\right|_{T=t}=\frac{\partial f(t_0,t)}{\partial t}+\int_{t_0}^{t}\gamma(\omega,s,t)^2+\frac{\partial\gamma(\omega,s,t)}{\partial t}\left(\int_s^t\gamma(\omega,s,v)dv+\lambda(\omega,s)\right)ds+\int_{t_0}^{t}\frac{\partial\gamma(\omega,s,t)}{\partial t}dW(s)$$

which generically involves information from time t_0 to t. It is also possible, but not necessary, that the preference parameter $\lambda(\omega,t)$ can depend on information other than $r(t)$ forcing non-Markov behavior in the short-term interest rate.

[12] The former type of yield transitions is chosen purely because it simplifies the derivation of our estimator's asymptotic variance.

[13] This is obtained by applying $E[\cdot|\,r(t)]$ to both sides of (7) and invoking iterated expectations.

yield volatility structure:

$$\eta(r(t),\tau)^2 \simeq \frac{1}{\Delta t} E\left[\left(\widetilde{y}(t+\Delta t,\tau) - \widetilde{y}(t,\tau)\right)^2 \middle| \, r(t) \right]. \tag{8}$$

Replacing the expectation on the right hand side with the Nadaraya-Watson kernel smoothing estimator suggests the following estimate for the yield volatility structure:[14]

$$\widehat{\eta}(r,\tau)^2 = \frac{1}{\Delta t} \frac{\sum_{i=1}^{n-1} K_h\left(r - r(t_i)\right) \left(\Delta\widetilde{y}(t_i,\tau)\right)^2}{\sum_{i=1}^{n-1} K_h\left(r - r(t_i)\right)} \quad \text{for each fixed } \tau, \tag{9}$$

where $K_h(u) = K(u/h)/h$, and $K(\cdot)$ is a symmetric kernel function that integrates to one, while h is the bandwidth parameter (which decreases to zero as sample size $n \to \infty$) that controls the amount of smoothing.[15] The asymptotic distribution of this estimator is as follows.

Corollary 1 *Given Assumption 1-3 stated in Appendix A3, the asymptotic distribution of*

[14] See Härdle (1991) chapter 7 for a discussion of kernel smoothing in a time series context.

[15] If a continuum of yield transitions in the maturity dimension is not observed, that is for each i we only observe $\Delta\widetilde{y}(t_i,\tau_j)$ for $j = 1,\ldots,J_i$, then interpolation can be achieved by further kernel smoothing across the maturity dimension which results in the estimator

$$\widehat{\eta}(r,\tau)^2 = \frac{1}{\Delta t} \frac{\sum_{i=1}^{n-1}\sum_{j=1}^{J_i} K_{h_1}\left(r - r(t_i)\right) K_{h_2}\left(\tau - \tau_j\right) \left(\Delta\widetilde{y}(t_i,\tau_j)\right)^2}{\sum_{i=1}^{n-1}\sum_{j=1}^{J_i} K_{h_1}\left(r - r(t_i)\right) K_{h_2}\left(\tau - \tau_j\right)},$$

where $K_{h_1}(u) = \frac{1}{h_1} K_1\left(\frac{u}{h_1}\right)$ and $K_{h_2}(u) = \frac{1}{h_2} K_2\left(\frac{u}{h_2}\right)$ for two kernel functions $K_1(\cdot)$ and $K_2(\cdot)$, each integrating to one, with respective bandwidth parameters h_1 and h_2 depending on the sample size in their respective dimensions.

the yield volatility structure estimator is a mixed normal[16,17]

$$\sqrt{\frac{h\ \overline{L}\left(t_n,r\right)}{\Delta t}}\left(\widehat{\eta}(r,\tau)^2-\eta(r,\tau)^2\right)\Longrightarrow N\left(0,4\left(\int_{-\infty}^{+\infty}K\left(u\right)^2du\right)\eta(r,\tau)^4\right), \tag{10}$$

where $\overline{L}\left(t_n,r\right)$ *is the chronological local time of point* r *for the process* $r(t)$*, which is a measure of the absolute amount of time that the short term interest rate process spends in the vicinity of point* r *over the time interval* 0 *to* t_n*. An estimate for the asymptotic variance of* $\widehat{\eta}(r,\tau)^2$ *can be obtained by replacing* $\eta(r,\tau)^4$ *with* $\widehat{\eta}(r,\tau)^4$*, and from equation (48) in Appendix 3* $\overline{L}\left(t_n,r\right)$ *can be estimated by* $\Delta t\sum_{i=1}^{n}K_h\left(r-r(t_i)\right)$*. The implied asymptotic variance of* $\widehat{\eta}(r,\tau)$ *is*

$$\operatorname{var}\left(\widehat{\eta}(r,\tau)\right)=\frac{\Delta t\ \left(\int_{-\infty}^{+\infty}K\left(u\right)^2du\right)\eta(r,\tau)^2}{h\ \overline{L}\left(t_n,r\right)}. \tag{11}$$

PROOF. See Appendix A4.1. ■

It is important to note that stationarity is not assumed for interest rates as the concept of chronological local time generalizes the density concept usually associated with stationarity. The interesting observation regarding the above estimator is that the drift term of the yield curve process is negligible. However, the complete markets construct of the HJM framework along with the no-arbitrage condition implies that the drift of the yield curve process depends on the yield volatility structure in a nontrivial manner. This observation is investigated below.

[16]This result does not require the *long span* assumption, that is $t_n\to\infty$ is not needed.

[17]The above approach can be also be applied to the case where the volatility of the yield with time to maturity τ is a function of that yield and τ, that is $\eta(\omega,t,T)=\eta(\widetilde{y}(t,\tau),\tau)$. In this case it can be established that an estimator is

$$\widehat{\eta}(\widetilde{y},\tau)^2=\frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1}K_h\left(\widetilde{y}-\widetilde{y}(t_i,\tau)\right)\left(\Delta\widetilde{y}(t_i,\tau)\right)^2}{\sum_{i=1}^{n-1}K_h\left(\widetilde{y}-\widetilde{y}(t_i,\tau)\right)}$$

which is asymptotically distributed as a mixed normal

$$\sqrt{\frac{h\ \overline{L}_\tau\left(t_n,\widetilde{y}\right)}{\Delta t}}\left(\widehat{\eta}(\widetilde{y},\tau)^2-\eta(\widetilde{y},\tau)^2\right)\Longrightarrow N\left(0,4\left(\int_{-\infty}^{+\infty}K\left(u\right)^2du\right)\eta(\widetilde{y},\tau)^4\right).$$

where $\overline{L}_\tau\left(t_n,\widetilde{y}\right)$ is the chronological local time of point $\widetilde{y}$ for the process $\widetilde{y}(t,\tau)$. Another class of volatility structures sometimes considered in the Heath-Jarrow-Morton literature is $\gamma(\omega,t,T)=\gamma(f(t,T),T-t)$. Such a volatility structure is often referred to as a "proportional volatility structure" (see Amin-Morton (1994)). One can estimate such models by applying the the above estimation procedure to forward rates instead of yields.

3.1 Use of the No-Arbitrage Restriction in Estimation

The no-arbitrage condition constrains the first moment of yield curve transitions (the drift term of yield curve dynamics) to be a function of the term structure's volatility and the risk preference parameter $\lambda(\omega, t)$. Consequently, if a reliable nonparametric estimate of both $\lambda(\omega, t)$ and the drift term of the yield curve dynamics exists then the volatility structure can be recovered. Below we investigate this scheme and show that it is not advisable.

For ease of exposition we abstract from the difficulty associated with requiring the market price of risk and assume that it is a known constant; that is, set $\lambda(\omega, t) = \lambda$. Now consider the no-arbitrage constrained yield curve transitions depicted by equation (5), which implies that

$$E\left[\left(\lambda\,\widetilde{\eta}(\omega,t,\tau)+\frac{1}{2}\tau\,\widetilde{\eta}(\omega,t,\tau)^2\right)dt\middle|\,\mathcal{F}_t\right]=E\left[d\widetilde{y}(t,\tau)-\left(\frac{\partial\widetilde{y}(t,\tau)}{\partial\tau}+\frac{\widetilde{y}(t,\tau)-r(t)}{\tau}\right)dt\middle|\,\mathcal{F}_t\right].$$

Observing that we consider volatility structures of the form $\widetilde{\eta}(\omega, t, \tau) = \eta(r(t), \tau)$, take the conditional expectation of both sides of the above equation with respect to $r(t)$ and invoke iterated expectations to give

$$\lambda\,\eta(r(t),\tau)+\frac{1}{2}\tau\,\eta(r(t),\tau)^2=\frac{1}{dt}E\left[d\widetilde{y}(t,\tau)-\left(\frac{\partial\widetilde{y}(t,\tau)}{\partial\tau}+\frac{\widetilde{y}(t,\tau)-r(t)}{\tau}\right)dt\middle|\,r(t)\right].$$

An estimate of $\eta(r(t), \tau)$, denoted $\widehat{\eta}_{NA}(r(t), \tau)$, can be constructed as the solution of the following quadratic equation:

$$\lambda\,\widehat{\eta}_{NA}(r,\tau)+\frac{1}{2}\tau\,\widehat{\eta}_{NA}(r,\tau)^2=\frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1}K_h\left(r-r(t_i)\right)\left[\Delta\widetilde{y}(t_i,\tau)-\left(\frac{\partial\widetilde{y}(t_i,\tau)}{\partial\tau}+\frac{\widetilde{y}(t_i,\tau)-r(t_i)}{\tau}\right)\Delta t\right]}{\sum_{i=1}^{n-1}K_h\left(r-r(t_i)\right)} \tag{12}$$

that is,

$$\widehat{\eta}_{NA}(r(t),\tau)=\begin{cases}\frac{-\lambda\pm\sqrt{\lambda^2+2\tau C}}{\tau} & \text{if } \tau\neq 0\\ \frac{C}{\lambda} & \text{if } \tau=0,\end{cases} \tag{13}$$

where C is the right-hand side of the equation (12) above (we assume that there exists a real root to the quadratic equation in (12)). The asymptotic distribution of this estimator is as follows.

Corollary 2 *Given Assumption 1-3 stated in Appendix A3, the asymptotic distribution of the no-arbitrage based estimator $\widehat{\eta}_{NA}(r, \tau)$ is a mixed normal*[18]

$$\sqrt{h\,\overline{L}\left(t_n,r\right)}\left(\widehat{\eta}_{NA}(r,\tau)-\eta(r,\tau)\right)\Longrightarrow N\left(0,\frac{\left(\int_{-\infty}^{+\infty}K\left(u\right)^2du\right)\eta(r,\tau)^2}{\left(\lambda+\tau\eta(r,\tau)\right)^2}\right). \tag{14}$$

[18] For the case when $\lambda + \tau\eta(r, \tau) \simeq 0$, and/or when $\lambda^2 + 2\tau C < 0$, a new estimate can be alternatively considered as an updating scheme based on the original estimator:

$$\lambda\,\widehat{\eta}_{NA}(r,\tau)=C-\frac{1}{2}\tau\,\widehat{\eta}(r,\tau)^2$$

Proof. See Appendix A4.2. ■

Simply by comparing the rates of convergence of the no-arbitrage based estimator $\widehat{\eta}_{NA}(r,\tau)$ in (14) and the original estimator $\widehat{\eta}(r,\tau)$ in (10) it is apparent that $\widehat{\eta}_{NA}(r,\tau)$ has a much larger asymptotic variance than $\widehat{\eta}(r,\tau)$; asymptotically it is $\frac{1}{\Delta t}$ times larger. For example, if daily data is to be used, as it is often the case, the variance of the no-arbitrage based estimator will be approximately 250 times larger than that of the original estimator. It is also worth noting that for asymptotic consistency the estimator $\widehat{\eta}_{NA}(r,\tau)$ requires both $\Delta t \downarrow 0$ (the *infill* assumption) and $t_n \to \infty$ (the *long span* assumption), whereas $\widehat{\eta}(r,\tau)$ only requires the former.[19] This is consistent with the well known observation that the drift term of a stochastic process is more difficult to estimate than its volatility. Finally, the above argument is based on knowing the risk preference parameter $\lambda(\omega,t)$ and as a consequence we are likely to be even worse off when $\lambda(\omega,t)$ must be estimated. In short, we advocate the estimation of the yield volatility structure using the estimator $\widehat{\eta}(r,\tau)$ which is based on the second moment of term structure transitions as opposed to estimation from the first moment of term structure transitions in conjunction with the no-arbitrage condition. A possible means by which both first and second moment information can be combined to obtain a new estimate of the volatility structure which can potentially improve the bias and/or the variance is to consider a weighted combination such as $w\widehat{\eta}(r,\tau)+(1-w)\,\widehat{\eta}_{NA}(r,\tau)$. Such analysis is beyond the scope of the present paper and is left for possible future research.

4 A Test of the No-Arbitrage Restriction

Since the original volatility structure estimate $\widehat{\eta}(r,\tau)$ provided by (9) does not rely on the no-arbitrage restriction we now turn to the construction of a test to determine whether it is consistent with this restriction. In this section it is convenient to consider yields with fixed maturity dates, that is $y(t,T)$ for fixed T's, as this will eliminate the need to compute derivatives of yield curves in the formation of the test statistic. In particular we consider a data structure that contains observed short term interest rates $r(t_i)$, yields $y(t_i,T)$, and

which is still a consistent estimate of $\widehat{\eta}(r,\tau)$ with asymptotic variance of

$$\frac{\left(\int_{-\infty}^{+\infty} K\left(u\right)^2 du\right) \eta(r,\tau)^2}{\lambda^2}$$

The conclusion remains valid for this special case.

[19] $\Delta t \downarrow 0$ is required to ensure the local normality of term structure transitions and it is also enough to provide asymptotic consistency of the $\widehat{\eta}(r,\tau)$ estimator even for a finite time span $[0,t_n]$. However for the $\widehat{\eta}_{NA}(r,\tau)$ estimator to be asymptotically consistent it requires the chronological local time $\overline{L}\left(t_n,r\right)$ to approach infinity which generically is only achieved by letting $t_n \to \infty$. For further details refer to Appendix A3.

yield transitions $\Delta y(t_i, T) = y(t_{i+1}, T) - y(t_i, T)$ at $n-1$ points in time indexed by t_i for $i = 1, \ldots, n-1$, where for simplicity $t_{i+1} - t_i = \Delta t$ for all i. To further facilitate analysis the market price of risk $\lambda(\omega, t)$ is hereafter restricted to be of the form $\lambda(r(t))$; this allows the short term interest rate at time t to be the sole conditioning variable.

Consider instantaneous yield curve transitions characterized by (4) where the yield volatility structure is of the form $\eta(r(t), T-t)$ and the no-arbitrage restriction has constrained the drift term. This can be expressed as

$$dy(t,T) - \left(\tfrac{y(t,T)-r(t)}{T-t}\right) dt = q(r(t), T-t)dt + \eta(r(t), T-t)dW(t), \tag{15}$$

where the no-arbitrage condition imposes the following testable restriction

$$\psi(r(t), T-t) = q(r(t), T-t) - \lambda(r(t))\eta(r(t), T-t) - \frac{1}{2}(T-t)\eta(r(t), T-t)^2 = 0. \tag{16}$$

To test (16), estimates of the terms involved, namely q, λ and η, are required. Estimate of η can be obtained as in (9). For q, an expression for $q(r(t), T-t)$ resulting directly from (15) is

$$q(r(t), T-t) = \lim_{\Delta t \downarrow 0} \left(\frac{1}{\Delta t} E\left[(y(t+\Delta t, T) - y(t,T)) - \left(\tfrac{y(t,T)-r(t)}{T-t}\right) dt \,\middle|\, r(t) \right] \right),$$

noting that in the above expectation it is sufficient to condition on $r(t)$ since the only random variable that $q(r(t), T-t)$ depends upon is the short term interest rate at time t. Given that term structures are observed at discrete points in time t_i, and replacing the expectation operator with the Nadaraya-Watson kernel smoothing estimator, suggests the following estimate for $q(r(t), T-t)$:

$$\widehat{q}(r, \tau) = \frac{\frac{1}{\Delta t}\sum_{i=1}^{n-1} K_h\left(r - r(t_i)\right)\left[\Delta y(t_i, \tau) - \left(\frac{y(t_i,\tau)-r(t_i)}{T-t_i}\right)\Delta t\right]}{\sum_{i=1}^{n-1} K_h\left(r - r(t_i)\right)}, \tag{17}$$

where $K_h(u) = K(u/h)/h$ for a given symmetric kernel function $K(\cdot)$ and h is the bandwidth parameter that decreases to zero as the sample size n grows to infinity. Note, at this point the estimate for $q(r(t), \tau)$ has been obtained without the imposition of the no-arbitrage restriction.

As for the unobserved market price of risk, $\lambda\left(r\left(t\right)\right)$, first note that it is maturity independent. For each maturity level, we can obtain an estimate of $\lambda\left(r\left(t\right)\right)$ by observing that, provided the no-arbitrage restriction holds

$$dy(t,T) - \left(\tfrac{y(t,T)-r(t)}{T-t}\right) dt - \frac{1}{2}(T-t)\eta(r(t), T-t)^2 = \lambda(r(t))\eta(r(t), T-t)dt + \eta(r(t), T-t)dW(t).$$

This results in

$$\lambda(r(t)) = \frac{1}{\eta(r(t), T-t)} E\left[dy(t,T) - \left(\tfrac{y(t,T)-r(t)}{T-t}\right) dt - \frac{1}{2}(T-t)\eta(r(t), T-t)^2 \left| r\left(t\right) \right. \right],$$

which we can use to approximate $\lambda(r(t))$ by kernel smoothing

$$\widehat{\lambda}(r,\tau_j) = \frac{1}{\Delta t}\sum_{i=1}^{n-1}\frac{K_h(r-r(t_i))}{\sum_{i=1}^{n-1}K_h(r-r(t_i))}\times\frac{\left[\Delta y(t_i,\tau_j)-\left(\frac{y(t_i,\tau_j)-r(t_i)}{T-t}+\frac{1}{2}\tau\,\eta(r(t_i),\tau_j)^2\right)\Delta t\right]}{\eta(r(t_i),\tau_j)}, \tag{18}$$

where we denote $\widehat{\lambda}(r,\tau_j)$ as the market price of risk that we obtain based on the time series $\Delta y(t_i,\tau_j)$ with some fixed τ_j.

However, since the market price of risk is maturity independent, we will construct its estimate based on information aggregated across J different maturity levels as follows. Observing that

$$\begin{aligned}&\sum_{j=1}^{J}\left[dy(t,T_j)-\left(\tfrac{y(t,T_j)-r(t)}{T_j-t}\right)dt-\frac{1}{2}(T_j-t)\eta(r(t),T_j-t)^2dt\right]\\ =\;& \lambda(r(t))\sum_{j=1}^{J}\eta(r(t),T_j-t)dt+\sum_{j=1}^{J}\eta(r(t),T_j-t)dW(t),\end{aligned}$$

and following the previous derivation, we obtain

$$\lambda(r(t)) = \frac{1}{\sum_{j=1}^{J}\eta(r(t),T_j-t)dt}E\left[\sum_{j=1}^{J}\left[dy(t,T_j)-\left(\tfrac{y(t,T_j)-r(t)}{T_j-t}\right)dt-\frac{1}{2}(T_j-t)\eta(r(t),T_j-t)^2dt\right]|r(t)\right],$$

which yields an estimate of $\lambda(r)$ as

$$\widehat{\lambda}(r) = \frac{1}{\Delta t}\sum_{i=1}^{n-1}\frac{K_h(r-r(t_i))}{\sum_{i=1}^{n-1}K_h(r-r(t_i))}\times\frac{\left[\sum_{j=1}^{J}\left[\Delta y(t_i,\tau_j)-\left(\frac{y(t_i,\tau_j)-r(t_i)}{T-t}+\frac{1}{2}\tau\,\eta(r(t_i),\tau_j)^2\right)\Delta t\right]\right]}{\sum_{j=1}^{J}\eta(r(t_i),\tau_j)}. \tag{19}$$

(17) and (19), together with the estimate of the volatility structure established in section 3 will allow us to test the moment restriction (16). Note that the rate of convergence of $\widehat{\eta}(r,\tau)$ defined in (9) is $\sqrt{\frac{1}{\Delta t}h\,\overline{L}(t_n,r)}$, while the rate of convergence of $\widehat{q}(r,\tau)$ and $\widehat{\lambda}(r)$ is $\sqrt{h\,\overline{L}(t_n,r)}$(see Appendix A5.1), so the variance of $\widehat{q}(r,\tau)$ and $\widehat{\lambda}(r)$ will dominate that of $\widehat{\eta}(r,\tau)$ and consequently determine the variance of our test statistic

$$\widehat{\psi}(r,\tau) = \widehat{q}(r,\tau)-\widehat{\lambda}(r)\,\widehat{\eta}(r,\tau)-\frac{1}{2}\tau\widehat{\eta}(r,\tau)^2,$$

$$\mathrm{var}\left[\widehat{\psi}(r,\tau)\right] = \mathrm{var}\left[\widehat{q}(r,\tau)\right]+\eta(r,\tau)^2\mathrm{var}\left[\widehat{\lambda}(r)\right]-2\eta(r,\tau)\mathrm{cov}\left[\widehat{q}(r,\tau),\widehat{\lambda}(r)\right].$$

Unfortunately, it turns out that $\widehat{q}(r,\tau)$ and $\widehat{\lambda}(r)$ are asymptotically perfectly correlated since every point along the yield curve is driven by the same Brownian motion, and the large sample test of thus becomes very problematic, to say the least[20]. To tackle this difficulty, we develop some estimates of $q(r)$ and $\lambda(r)$, which are asymptotically independent, by sampling alternatively from the sample to break the dependency. Namely, we first use the first half sample consisting of all even (odd) observations to estimate $q(r,\tau)$, then use the other half sample consisting of odd (even) observations to estimate $\lambda(r(t))$. The independence property of $dW(t)$ will render two estimates $\widehat{q}(r,\tau)$ and $\widehat{\lambda}(r)$ independent from each other[21]. Admittedly, this is not the most efficient way to use the data for estimation purpose, but for the need of inference, it is our choice of strategy here.

Concretely, the sub-sample consisting of even (odd) observations yields the following estimate of $q(r,\tau)$,which we denote as $\widehat{\widehat{q}}(r,\tau)$, similar to (17)

$$\widehat{\widehat{q}}(r,\tau) = \frac{1}{\Delta t}\sum_{i=1}^{[(n-1)/2]} \frac{K_h(r-r(t_{2i}))}{\sum_{i=1}^{[(n-1)/2]} K_h(r-r(t_{2i}))} \times \left[\Delta y(t_{2i},\tau) - \left(\begin{array}{c} \frac{y(t_{2i},\tau)-r(t_{2i})}{T-t} \\ +\frac{1}{2}\tau\eta(r(t_{2i}),\tau)^2 \end{array} \right) \Delta t\right].$$

As shown in Appendix A5.1, $\widehat{\widehat{q}}(r,\tau)$ is a consistent estimator of $q(r,\tau)$ with the asymptotic variance

$$\operatorname{var}\left[\widehat{\widehat{q}}(r,\tau)\right] = 2\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)\eta(r,\tau)^2}{h\,\overline{L}(t_n,r)}, \tag{20}$$

which is twice as large as the variance that we would obtain if we use the whole sample to estimate $q(r,\tau)$ as in (17).

Next, to estimate $\lambda(r)$, similar to what we have done earlier in (19), we aggregate information across the maturity spectrum by constructing

$$\begin{aligned} \widehat{\widehat{\lambda}}(r) &= \frac{1}{\Delta t}\sum_{i=1}^{[(n-1)/2]} \frac{K_h(r-r(t_{2i-1}))}{\sum_{i=1}^{[(n-1)/2]} K_h(r-r(t_{2i-1}))} \\ &\times \frac{\left[\sum_{j=1}^{J}\left[\Delta y(t_{2i-1},\tau_j) - \left(\frac{y(t_{2i-1},\tau_j)-r(t_{2i-1})}{\tau_j} + \frac{1}{2}\tau\,\eta(r(t_{2i-1}),\tau_j)^2\right)\Delta t\right]\right]}{\sum_{j=1}^{J}\eta(r(t_{2i-1}),\tau_j)}, \end{aligned} \tag{21}$$

which is based only on the subsample consists of odd (even) observations.

As shown in Appendix A5.1, $\widehat{\widehat{\lambda}}(r)$ is a consistent estimator of $\lambda(r)$ with the asymptotic variance

$$\operatorname{var}\left[\widehat{\widehat{\lambda}}(r)\right] = 2\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)}{h\,\overline{L}(t_n,r)},$$

and is asymptotically uncorrelated with $\widehat{\widehat{q}}(r,\tau)$ for all τ.

[20] See Appendix A3.4 for details.

[21] See Appendix 3.4.

Test statistic can be obtained via the following moment restriction:

$$E\left[\widehat{\widehat{\psi}}(r,\tau)\right] = E\left[\widehat{\widehat{q}}(r,\tau) - \widehat{\widehat{\lambda}}(r)\,\widehat{\eta}(r,\tau) - \frac{1}{2}\tau\widehat{\widehat{\eta}}(r,\tau)^2\right] = 0, \tag{22}$$

if no-arbitrage holds, where

$$\operatorname{var}\left[\widehat{\widehat{\psi}}(r,\tau)\right] = \operatorname{var}\left[\widehat{\widehat{q}}(r,\tau)\right] + \eta(r,\tau)^2 \operatorname{var}\left[\widehat{\widehat{\lambda}}(r)\right] = 4\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)\eta(r,\tau)^2}{h\,\overline{L}(t_n,r)}. \tag{23}$$

5 Test of the Time-Homogeneous Univariate Markov Diffusion Assumption

All the recent nonparametric methods that have been proposed to estimate the drift and volatility of the short term interest rate have constructed their estimators under the presumption that $r(t)$ follows a time-homogeneous univariate Markov diffusion. Imposing the additional restriction that the market price of risk $\lambda(\omega,t)$ is a function of $r(t)$ only implies that the short term interest rate will also follow a time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy. This structure is advantageous since the yield curve at time t must be a function of the short term interest rate at time t and time to maturity only, that is $y(t,T)$ must be of functional form $y(r(t),T-t)$, which greatly simplifies the modelling of fixed-income securities.[22] The nonparametric estimator of the volatility structure constructed in section 3, $\widehat{\eta}(r,\tau)$, encapsulates all such models as a special case. Consequently the objective of this section is to construct a test statistic to ascertain if $\widehat{\eta}(r,\tau)$ is consistent with $r(t)$ following a time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy. In this section it will be convenient to consider yields with fixed time to maturities, that is $\widetilde{y}(t,\tau)$ for fixed τ's, as this will eliminate the need to compute derivatives of the yield curve in the formation of our test statistic.

Suppose the short term interest rate does follow time-homogeneous univariate Markov diffusion in the equivalent risk-neutral economy. If this is the case then the evolution of $r(t)$ in the observable economy is

$$\begin{aligned} dr(t) &= \mu(r(t))\,dt + \sigma(r(t))dW(t) \\ &= [\theta(r(t)) + \lambda(r(t))\sigma(r(t))]\,dt + \sigma(r(t))dW(t), \end{aligned}$$

where we again assume the market price of risk $\lambda(r(t))$ is a function of $r(t)$ only, and $\theta(r(t))$

[22] For example, partial differential equations can often be constructed to represent the price of interest rate contingent claims.

represents the drift of $r(t)$ in the equivalent risk-neutral economy. The yield $\widetilde{y}(t,\tau)$ must then be of the functional form $\widetilde{y}(r(t),\tau)$ with the following dynamics:[23]

$$\begin{aligned} d\widetilde{y}(r(t),\tau) &= \left(\frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)}\mu\left(r(t)\right) + \frac{1}{2}\sigma(r(t))^2 \frac{\partial^2 \widetilde{y}(r(t),\tau)}{\partial r(t)^2}\right) dt \\ &\quad + \frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)}\sigma(r(t))\ dW(t) \\ &= m\left(r\left(t\right),\tau\right) dt + \eta(r(t),\tau)\ dW(t). \end{aligned} \tag{24}$$

The Markov assumption implies the following restrictions on the diffusion function $\eta(r(t),\tau)$ and on the drift function $m\left(r\left(t\right),\tau\right)$ of $d\widetilde{y}(r(t),\tau)$

$$\eta(r(t),\tau) = \frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)}\sigma(r(t)), \text{ or}$$

$$\phi_1(r(t),\tau) = \frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)} - \frac{\eta(r(t),\tau)}{\sigma(r(t))} = 0 \tag{25}$$

and

$$\phi_2(r(t),\tau) = m\left(r\left(t\right),\tau\right) - \xi\left(t,T\right)\mu\left(r(t)\right) - \frac{1}{2}\sigma(r(t))^2 \frac{\partial \xi\left(t,T\right)}{\partial r(t)} = 0, \tag{26}$$

where $\xi\left(t,T\right) = \partial \widetilde{y}(r(t),\tau)/\partial r(t)$.

In the rest of this section, we will develop large sample tests for these two moment restrictions implied by the Markov hypothesis.

5.1 Test for the first restriction

To test the first restriction (25), an estimate for $\partial \widetilde{y}(r(t),\tau)/\partial r(t)$ is required (note that estimates of $\eta(r(t),\tau)$ and $\sigma(r(t))$ are readily available from section 3). As

$$\widetilde{y}\left(r(t),\tau\right) = y\left(r(t),T-t\right),$$

we have

$$\frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)} = \frac{\partial y(r(t),\tau)}{\partial r(t)},$$

and we like to take advantage of this equality by using the dynamics of y in (4) rather than that of $\widetilde{y}$ in (5) for convenience.

Provided the no-arbitrage restriction on the drift function is not rejected, we will exploit it to get an estimate for $\partial y(r(t),\tau)/\partial r(t)$. Observing (4), we have

$$\frac{y(t,T)-r(t)}{T-t} = m(\omega,t,T) - \lambda(\omega,t)\eta(\omega,t,T) - \frac{1}{2}(T-t)\eta(\omega,t,T)^2,$$

[23]This is obtained using Itô's lemma and noting that τ is fixed when considering the evolution of $\widetilde{y}(t,\tau)$.

which, under Markov assumption, can be written as

$$\frac{y(r(t),T-t)-r(t)}{T-t}=m(r(t),T-t)-\lambda(r(t))\eta(r(t),T-t)-\frac{1}{2}(T-t)\eta(r(t),T-t)^2.$$

Differentiate both sides with respect to $r(t)$

$$\frac{\partial y(r(t),T-t)}{\partial r(t)}=1+(T-t)\left[\begin{array}{c}\frac{\partial m(r(t),T-t)}{\partial r(t)}-\frac{\partial\lambda(r(t))}{\partial r(t)}\eta(r(t),T-t)\\ -\frac{\partial\eta(r(t),T-t)}{\partial r(t)}\lambda(r(t))-\tau\frac{\partial\eta(r(t),T-t)}{\partial r(t)}\eta(r(t),T-t)\end{array}\right]. \quad (27)$$

So an estimate of $\frac{\partial y(r,\tau)}{\partial r}$ can be formed via (27)

$$\frac{\partial\widehat{y}(r,\tau)}{\partial r}=1+\tau\left[\begin{array}{c}\frac{\partial\widehat{m}(r,\tau)}{\partial r}-\frac{\partial\widehat{\lambda}(r)}{\partial r}\widehat{\eta}(r,\tau)\\ -\frac{\partial\widehat{\eta}(r,\tau)}{\partial r}\widehat{\lambda}(r)-\tau\frac{\partial\widehat{\eta}(r,T-t)}{\partial r}\widehat{\eta}(r,\tau)\end{array}\right], \quad (28)$$

if estimates of all terms in the RHS of (27) can be constructed. These estimates can be obtained in a similar fashion to what we have done earlier in section 4. First of all, an estimate of $\eta(r(t),\tau)$ is readily available from section 3, while an estimate of $\lambda(r)$ is also available from (19). As shown in Appendix A5.2.1, their derived derivative estimators are consistent estimators of the respective derivatives, but converging with the rate $1/h$ time slower than the function estimators themselves

$$\operatorname{var}\left[\frac{\partial\widehat{\eta}(r,\tau)}{\partial r}\right]=\frac{\Delta t\left(\int_{-\infty}^{+\infty}K'(u)^2\,du\right)\eta(r,\tau)^2}{h^3\,\overline{L}(t_n,r)},$$

and

$$\operatorname{var}\left[\frac{\partial\widehat{\lambda}(r,\tau)}{\partial r}\right]=\frac{\left(\int_{-\infty}^{+\infty}K'(u)^2\,du\right)}{h^3\,\overline{L}(t_n,r)}.$$

For an estimate of $m(r,\tau)$, note that under Markov assumption, (4) can be re-written as

$$dy(t,T)=m(r(t),T)dt+\eta(r(t),T)dW(t), \quad (29)$$

where the drift function is just a function of $r(t)$ now.

This yields

$$m(r(t),T)=\frac{1}{dt}E\left[dy(t,T)\,|r(t)\right],$$

which can be used to estimate $m(r,\tau)$ by the now familiar kernel smoothing scheme

$$\widehat{m}(r,\tau)=\frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1}K_h\left(r-r(t_i)\right)\Delta y(t_i,\tau)}{\sum_{i=1}^{n-1}K_h\left(r-r(t_i)\right)}. \quad (30)$$

Under certain assumptions, this estimator is consistent, with the asymptotic variance

$$\operatorname{var}\left[\widehat{m}(r,\tau)\right]=\frac{\left(\int_{-\infty}^{+\infty}K(u)^2\,du\right)\eta(r,\tau)^2}{h\,\overline{L}(t_n,r)}$$

As shown in A5.2.1, this estimator can be utilized to provide a consistent estimate for $\frac{\partial m(r,\tau)}{\partial r}$, whose asymptotic variance is of order $h^3\ \overline{L}\left(t_n, r\right)$, or converging with the rate $1/h$ time slower than $\widehat{m}(r,\tau)$ itself

$$\operatorname{var}\left[\frac{\partial \widehat{m}(r,\tau)}{\partial r}\right] = \frac{\left(\int_{-\infty}^{+\infty} K'\left(u\right)^2 du\right) \eta(r,\tau)^2}{h^3\ \overline{L}\left(t_n, r\right)}$$

Thus, for the asymptotic distribution of our test statistics $\widehat{\phi}_1(r,\tau)$, as a variance of a derivative estimator dominates the variance of its corresponding function estimator, and as reasoned elsewhere, variances of second moment based estimators can be ignored when first moment based estimators are present (the former converge $\sqrt{\Delta t}$ faster the the latter), so the variance of $\widehat{\phi}_1(r,\tau)$ will be driven only by variations of $\partial\widehat{m}(r(t),\tau)/\partial r(t)$ and $\partial\widehat{\lambda}(r(t),\tau)/\partial r(t)$

$$\operatorname{var}\left[\widehat{\phi}_1(r,\tau)\right] = \tau^2\left[\operatorname{var}\left(\tfrac{\partial\widehat{m}(r,\tau)}{\partial r}\right) + \eta^2(r,\tau)\operatorname{var}\left(\tfrac{\partial\widehat{\lambda}(r,\tau)}{\partial r}\right) - 2\eta(r\left(t\right),\tau)\operatorname{cov}\left(\tfrac{\partial\widehat{m}(r,\tau)}{\partial r}, \tfrac{\partial\widehat{\lambda}(r,\tau)}{\partial r}\right)\right].$$

Similar to the situation encountered in section 4, if we use the whole sample to estimate $\partial m(r,\tau)/\partial r$ and $\partial\lambda(r,\tau)/\partial r$ in order to construct $\widehat{\phi}_1(r,\tau)$, then $\operatorname{var}\left[\widehat{\phi}_1(r,\tau)\right]$ turns out to be asymptotically zero, which hinders our ability to construct a test. We consequently utilize the approach employed in the previous section to deal with this problem: use the half sample consisting of every other points to estimate $m(r,\tau)$ (and its derived derivative), and use the other half to estimate $\lambda(r,\tau)$ (and its derived derivative) as

$$\widehat{\widehat{m}}(r,\tau) = \frac{1}{\Delta t}\frac{\sum_{i=1}^{[(n-1)/2]} K_h\left(r - r(t_{2i})\right)\Delta y(t_{2i},\tau)}{\sum_{i=1}^{[(n-1)/2]} K_h\left(r - r(t_{2i})\right)}, \tag{31}$$

and (21).

By construction the two estimates are independent, and we have the new estimate for $\frac{\partial y(r,\tau)}{\partial r}$

$$\frac{\partial\widehat{\widehat{y}}(r,\tau)}{\partial r} = 1 + \tau\left[\begin{array}{c} \frac{\partial\widehat{\widehat{m}}(r,\tau)}{\partial r} - \frac{\partial\widehat{\widehat{\lambda}}(r)}{\partial r}\widehat{\eta}(r,\tau) \\ -\frac{\partial\widehat{\eta}(r,\tau)}{\partial r}\widehat{\widehat{\lambda}}(r) - \tau\frac{\partial\widehat{\eta}(r,\tau)}{\partial r}\widehat{\eta}(r,\tau) \end{array}\right].$$

The testable moment restriction now is

$$E\left[\widehat{\widehat{\phi}}_1(r,\tau)\right] = E\left[\frac{\partial\widehat{\widehat{y}}(r,\tau)}{\partial r} - \frac{\widehat{\eta}(r,\tau)}{\widehat{\sigma}(r,\tau)}\right] = 0, \tag{32}$$

if no-arbitrage holds, where

$$\begin{aligned} \operatorname{var}\left[\widehat{\widehat{\phi}}_1(r,\tau)\right] &= \tau^2\left[\operatorname{var}\left(\frac{\partial\widehat{\widehat{m}}(r,\tau)}{\partial r}\right) + \eta^2(r,\tau)\operatorname{var}\left(\frac{\partial\widehat{\widehat{\lambda}}(r,\tau)}{\partial r}\right)\right] \\ &= \tau^2\frac{4\left(\int_{-\infty}^{+\infty} K'\left(u\right)^2 du\right)\eta(r,\tau)^2}{h^3\ \overline{L}\left(t_n, r\right)}. \end{aligned}$$

5.2 Test for the second restriction

To test the second restriction, there are some considerations in place. If we are to estimate $\xi(t,T) = \frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)}$ and $\frac{\partial \xi(t,T)}{\partial r(t)} = \frac{\partial^2 \widetilde{y}(r(t),\tau)}{\partial r^2(t)}$ directly via (27), as we have done for $\frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)}$ above, the variance of $\frac{\partial^2 \widetilde{y}(r(t),\tau)}{\partial r^2(t)}$ will be very large, of order $(1/h^5)$, potentially rendering a very weak test. So we opt to test this restriction jointly with the first restriction, i.e. provided that the first restriction holds up with empirical data, then we can use

$$\xi(t,T) = \frac{\partial \widetilde{y}(r(t),\tau)}{\partial r(t)} = \frac{\eta(r(t),\tau)}{\sigma(r(t))},$$

thus

$$\widehat{\xi}(t,T) = \frac{\widehat{\eta}(r(t),\tau)}{\widehat{\sigma}(r(t))},$$

to substitute in for the related terms in (26). Note that the variance of $\widehat{\phi}_2(r,\tau)$ then is of order $\left(1/h\,\overline{L}(t_n,r)\right)$, and induced only by variations of $\widehat{m}(r(t),\tau)$ and $\widehat{\mu}(r(t))$, since the estimate of $\sigma(r(t))$ constructed in section 3 has much faster rate of convergence of $\sqrt{h\,\overline{L}(t_n,r)/\Delta t}$, while $\partial\xi(t,T)/\partial r(t)$ has the asymptotic variance of zero if it is to be obtained from (9); see appendix A5.2.2.

We construct the test statistic $\phi_2(r(t),\tau)$ through (26)

$$\widehat{\phi}_2(r,\tau) = \widehat{m}(r,\tau) - \widehat{\xi}(r,\tau)\,\widehat{\mu}(r) - \frac{1}{2}\widehat{\sigma}(r)^2 \frac{\partial \widehat{\xi}(r,\tau)}{\partial r},$$

where $\widehat{\sigma}$ and $\widehat{\xi}$ are from section 3, $\widehat{m}$ is from (30), and

$$\widehat{\mu}(r) = \frac{1}{\Delta t} \frac{\sum_{i=1}^{n-1} K_h(r - r(t_i))\,\Delta r(t_i)}{\sum_{i=1}^{n-1} K_h(r - r(t_i))}.$$

The variance of $\widehat{\phi}_2(r,\tau)$ is then

$$\mathrm{var}\left[\widehat{\phi}_2(r,\tau)\right] = \mathrm{var}\left[\widehat{m}(r,\tau)\right] + \xi(r,\tau)^2\,\mathrm{var}\left[\widehat{\mu}(r)\right] - 2\xi(r,\tau)\,\mathrm{cov}\left[\widehat{m}(r,\tau),\widehat{\mu}(r)\right]$$

and we run into the similar problem with var $\left[\widehat{\phi}_2(r,\tau)\right]$ approaches 0.

We will resort to the same tactic used earlier to get away with this asymptotic perfect correlation, namely sampling every other observations to form two independent estimates of $m(r,\tau)$ and $\mu(r)$. We first use all even observation to construct an estimate of $m(r,\tau)$ as in (30), and then use the odd (even) observations to estimate the drift function of the short rates by

$$\widehat{\widehat{\mu}}(r) = \frac{1}{\Delta t} \frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i-1}))\,\Delta r(t_{2i-1})}{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i-1}))}, \tag{33}$$

which has the variance

$$\operatorname{var}\left(\widehat{\widehat{\mu}}(r)\right) = 2\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\, du\right) \sigma(r)^2}{h\, \overline{L}(t_n, r)}.$$

Again, we now have $\operatorname{cov}\left(\widehat{\widehat{m}}(r,\tau), \widehat{\widehat{\mu}}(r)\right) = 0$. Test statistic can be obtained via the following moment restriction:

$$E\left[\widehat{\widehat{\phi}}_2(r,\tau)\right] = E\left[\widehat{\widehat{m}}(r,\tau) - \widehat{\xi}(r,\tau)\,\widehat{\widehat{\mu}}(r) - \frac{1}{2}\widehat{\sigma}(r)^2 \frac{\partial \widehat{\xi}(r,\tau)}{\partial r}\right] = 0, \tag{34}$$

if Markov hypothesis hold, where

$$\operatorname{var}\left[\widehat{\widehat{\phi}}_2(r,\tau)\right] = \operatorname{var}\left(\widehat{\widehat{m}}(r,\tau)\right) + \xi^2(r,\tau) \operatorname{var}\left(\widehat{\widehat{\mu}}(r)\right) = 4\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\, du\right) \eta(r,\tau)^2}{h\, \overline{L}(t_n, r)}.$$

6 Empirical Implementation

In this section, CRSP daily bond data is used to estimate the volatility structure of the yield curve via the estimator derived in section 3. The first step in estimating a dynamic model of interest rates is to back out the yield curve at a point in time. McCulloch (1971,1975) has been used to implement this step (Buler et al.(1999), Zhou and Pearson (1998), among others). Jeffrey-Linton-Nguyen (1999) however advocates using Linton et al. (1998)'s methodology, a recently-introduced nonparametric procedure to fit the yield curves. In their paper, nonparametric method is shown to work best. We will use the same method to obtain the "observed" yield curves. Only the segment out to 10–year along the maturity dimension is used in our test as it is well documented that extracting the yield curves at the long end is subject to much more noise due to the lack of data. When we go back to the 1960s and the 1970s, the farthermost maturity we would consider is even much shorter, just around 4 to 5 years.

Estimates of the volatility is obtained via (9). The kernel of our choice is the Gaussian kernel.[24] Figure 1 depicts the empirical volatility functions that we obtain using the bandwidth of 1% for CRSP bond data from January 1960 to December 1998. We also estimate the volatility structure using CRSP bond data for the period from January 1970 to December 1998, and the period from January 1983 to December 1998 respectively, which have similar structure and thus not presented here. For the short rate $r(t)$, as well as yields at other maturity level $y(t,\tau)$, the volatility function displays the features commonly observed in other studies: interest rates become more volatile when the level increases (the drastic

[24] Note that when the Gaussian kernel is used $\int_{-\infty}^{+\infty} K(q)^2 dq = \frac{1}{2\sqrt{\pi}}$, $\int_{-\infty}^{+\infty} K'(q)^2 dq = \frac{1}{4\sqrt{\pi}}$.

increase at extremely high level interest rate level, around 15% should be interpreted with caution however, since there are very few observations in that area). As shown in Figure 2, where chronological local time of the short rates are estimated by (48), most data are observed in the 3% to 8% region. To minimize this boundary effect in our empirical study, we will make inference based only on this data range.

We specify a 6-point grid of the short rate r ranging from 3% to 8%, where most of short rates are clustered to alleviate potential bias at the boundary, which is nonparametric technique is well-known for. Along the maturity dimension, we set a grid of 4 $\tau's = 1, 2, \ldots, 4$ years if the sampling period is 1961-1998, a grid of 5 $\tau's = 1, 2, \ldots, 5$ for the 1970-1998 period, and a grid of 9 $\tau's = 1, 2, \ldots, 9$ for the 1983-1998 period. The reason is that in the early years, i.e. the 1960s and 1970s, long bonds over 5 year-maturity were scarce, while in the latter decades, they have become more available. Again, it is aimed at minimizing the impact of the bias at the boundary.

6.1 Test of No-Arbitrage Restriction

A χ^2-test is constructed for each maturity level, based on (22)[25]

$$\chi^2(\tau_j) = \sum_{i=1}^{K} \frac{\widehat{\widehat{\psi}}(r_i, \tau_j)^2}{\operatorname{var}\left[\widehat{\widehat{\psi}}(r_i, \tau_j)\right]} \sim \chi_K^2,$$

a χ^2 with K degree of freedom. As proved in Appendix A.5.1, $\widehat{\widehat{\psi}}(r_i, \tau_j)$ is not asymptotically correlated with any $\widehat{\widehat{\psi}}(r_{i'}, \tau_j)$ when $r_i \neq r_{i'}$, so construction of the above χ^2 is valid. 6 interest rate levels from 3% to 8% results in a χ^2 with 5 degree of freedom, which has the critical value at 95% of 12.59.

As shown in Table 1, we fail to reject the hypothesis that $E\left[\widehat{\widehat{\psi}}(r_i, \tau_j)\right] = 0$ at 95% confidence level or the no-arbitrage restriction does hold up with empirical data for the sample period. The results are consistent over the 3 periods that we test (except for one single case for 1-year yield if data from the period from 1983 to 1998 is used, which is however very likely to be just a spurious result.)

The results reported here are robust to a wide bandwidth range from 0.5% to 2%. We also switch the sub-samples, for example use odd and even observations to estimate $q(r, \tau)$ and $\lambda(r)$ used in (22) respectively, then repeat the experiment but using even and odd

[25] An alternative test can be constructed as $\frac{1}{K}\sum_{i=1}^{K}\left[\frac{\widehat{\widehat{\psi}}(r_i,\tau_j)}{\operatorname{std}\left(\widehat{\widehat{\psi}}(r_i,\tau_j)\right)}\right]^2 \sim \chi_1^2$. The results of our no-abitrage test and Markov tests do not change with this alternative.

observations to estimate $q(r,\tau)$ and $\lambda(r)$ respectively. The test results are not sensitive to this switching.

6.2 Test of Markovian property

Similar to the above section, a χ^2-test is constructed for each maturity level, based on (32) and (34)

$$\chi_1^2(\tau_j) = \sum_{i=1}^{K} \frac{\widehat{\widehat{\phi}}_1(r,\tau)^2}{\mathrm{var}\left[\widehat{\widehat{\phi}}_1(r,\tau)\right]} \sim \chi_K^2,$$

and

$$\chi_2^2(\tau_j) = \sum_{i=1}^{K} \frac{\widehat{\widehat{\phi}}_2(r,\tau)^2}{\mathrm{var}\left[\widehat{\widehat{\phi}}_2(r,\tau)\right]} \sim \chi_K^2$$

with the same construction along the maturity dimension and the short rate grid used in the no-arbitrage restriction test.

The χ^2 test here is also valid due to the same reason as above: each term $\widehat{\widehat{\phi}}_1$ ($\widehat{\widehat{\phi}}_2$) at an interest rate level is asymptotically independent from $\widehat{\widehat{\phi}}_1$ ($\widehat{\widehat{\phi}}_2$) at another interest rate level. We also have a χ^2 test with 6 degree of freedom here for each restriction. As explained in section 5.2, the second test is in effect a joint test of the first restriction and the second one.

The test results are reported in Table 2 and 3, for the first restriction and second restriction respectively. In both cases, perhaps rather surprisingly, we can not reject the Markovian hypothesis at 95% confidence level. Again, the results reported here are robust to a wide range of bandwidth choice from 0.5% to 2%, and to subsample switching as well.

7 Conclusion

In this paper we have provided a non-parametric estimator for the term structure's volatility over the period June 1961 to December 1998. The estimate does not incorporate the restriction imposed by the no-arbitrage restriction because, as we show, incorporation of this restriction drastically reduces the efficiency of the estimate. We do however provide a test to determine whether the volatility estimate is consistent with the no-arbitrage constraint. The results show that our estimate is consistent with this constraint.

A test for path independence is also constructed, and interestingly, it turns out that we can not reject the path independence. There are several possible interpretations for this result. From the robustness of our results, it seems to suggest that by allowing a sufficient flexibility in dynamics of the yield curves, as we have done here via nonparametric specification, one factor Markovian model can capture the empirical yield curves just as well as a more complex one factor path dependent HJM model. One can also reasonably suspect the power of the large sample test with finite sample observed, which renders the test results doubtful. However, recent simulation evidence on nonparametric methods on diffusion process estimation seems to suggest that the sample size of around 28 years as we have here is rather reliable for inference (see Bandi and Nguyen (1999) for an example, where estimates of diffusion process based on 20 years of daily data are shown to capture the asymptotic behavior very well).

Our final caveat, which is a direction for future research, is that our estimation methodology presumes only a single factor framework. Perhaps a multi-factor model is more appropriate and will provide more power to test whether a single factor model is sufficient. This issue is not further pursued in this present paper, but the econometric framework in this paper can possibly be extended to multi-factor cases.

8 Appendix

A1-YIELD CURVE EVOLUTION

From equation (1) the forward rate dynamics can be written as

$$f(t,T) = f(0,T) + \int_0^t \alpha(\omega,s,T)ds + \int_0^t \gamma(\omega,s,T)dW(s).$$

Noting that the yield curve $y(t,T) = \frac{1}{T-t}\int_t^T f(t,v)dv$ implies

$$(T-t)\cdot y(t,T) = \int_t^T f(0,v)dv + \int_0^t \int_t^T \alpha(\omega,s,v)\,dv\,ds + \int_0^t \int_t^T \gamma(\omega,s,v)\,dv\,dW(s).$$

Differentiating both sides with respect to t provides

$$\begin{aligned}(T-t)\cdot dy(t,T) - y(t,T)dt \;&=\; -f(0,t)\,dt - \left(\int_0^t \alpha(\omega,s,t)ds\right)dt - \left(\int_0^t \gamma(\omega,s,t)dW(s)\right)dt \\ &\quad + \left(\int_t^T \alpha(\omega,t,v)dv\right)dt + \left(\int_t^T \gamma(\omega,t,v)dv\right)dW(t).\end{aligned}$$

Observing that $r(t) = f(t,t)$ and rearranging the above yields

$$dy(t,T) = \frac{1}{T-t}\left(y(t,T) - r(t) + \int_t^T \alpha(\omega,t,v)dv\right)dt + \left(\frac{1}{T-t}\int_t^T \gamma(\omega,t,v)dv\right)dW(t).$$

Letting $\eta(\omega,t,T) = \frac{1}{T-t}\int_t^T \gamma(\omega,t,v)dv$, which may be interpreted as the yield volatility structure, and imposing the no-arbitrage constraint on $\alpha(\omega,t,T)$ provided by equation (2) results in the following yield curve dynamics implied by the HJM framework:

$$\begin{aligned} dy(t,T) &= m(\omega,t,T)dt + \eta(\omega,t,T)dW(t) \\ \text{where} \quad m(\omega,t,T) &= \frac{y(t,T)-r(t)}{T-t} + \lambda(\omega,t)\eta(\omega,t,T) + \frac{1}{2}(T-t)\eta(\omega,t,T)^2. \end{aligned}$$

The above dynamics is for a yield with a fixed maturity date T. To determine the dynamics of a yield with a fixed time-to-maturity τ, denoted $\widetilde{y}(t,\tau)$ where $\widetilde{y}(t,\tau) = y(t,t+\tau)$, consider the following:

$$\widetilde{y}(t,\tau) = y(0,t+\tau) + \int_0^t m(\omega,s,t+\tau)ds + \int_0^t \gamma(\omega,s,t+\tau)dW(s).$$

Differentiating both sides with respect to t provides

$$\begin{aligned} d\widetilde{y}(t,\tau) &= \left(\frac{\partial y(0,t+\tau)}{\partial\,(t+\tau)} + \int_0^t \frac{\partial m(\omega,s,t+\tau)}{\partial\,(t+\tau)}ds + \int_0^t \frac{\partial\gamma(\omega,s,t+\tau)}{\partial\,(t+\tau)}dW(s)\right)dt \\ &\quad + m(\omega,t,t+\tau)dt + \gamma(\omega,t,t+\tau)dW(t) \\ &= \frac{\partial\widetilde{y}(t,\tau)}{\partial\tau}dt + m(\omega,t,t+\tau)dt + \gamma(\omega,t,t+\tau)dW(t). \end{aligned}$$

Consequently the dynamics of $\widetilde{y}(t,\tau)$ can be expressed as

$$\begin{aligned} d\widetilde{y}(t,\tau) &= \widetilde{m}(\omega,t,\tau)dt + \widetilde{\eta}(\omega,t,\tau)dW(t) \\ \text{where} \quad \widetilde{\eta}(\omega,t,\tau) &= \eta(\omega,t,t+\tau), \text{ and} \\ \widetilde{m}(\omega,t,\tau) &= \frac{\partial\widetilde{y}(t,\tau)}{\partial\tau} + \frac{\widetilde{y}(t,\tau)-r(t)}{\tau} + \lambda(\omega,t)\widetilde{\eta}(\omega,t,\tau) + \frac{1}{2}\,\tau\,\widetilde{\eta}(\omega,t,\tau)^2. \end{aligned}$$

A2-CAPTURED MARKOV MODELS

Consider the case where the short term interest rate follows a time-homogeneous univariate Markov diffusion

$$dr(t) = \theta(r(t))dt + \sigma(r(t))dW(t)$$

under the risk-neutral probability measure. Given complete bond markets the price (denoted P) of a zero-coupon bond at time t with a face value of one dollar and maturity date T can be computed using the risk-neutral valuation formula $P = E_{P^*}\left[\exp\left(-\int_t^T r(s)ds\right)\middle|\,\mathcal{F}_t\right]$ where $E_{P^*}\left[\cdot|\,\mathcal{F}_t\right]$ represents the expectation operator under the risk-neutral probability measure given information at time t. Since $r(t)$ follows a Markov diffusion process the price P

must satisfy the following partial differential equation (see Vasicek (1977) and/or Musiela-Rutkowski (1997) page 296):

$$\frac{\partial P}{\partial r(t)}\theta(r(t)) + \frac{\partial P}{\partial t} + \frac{1}{2}\sigma(r(t))^2\frac{\partial^2 P}{\partial r(t)^2} - r(t)P = 0.$$

Letting $\tau = T - t$ the term $\frac{\partial P}{\partial t}$ in the above partial differential equation can be replaced with $-\frac{\partial P}{\partial \tau}$ and hence the price of the zero-coupon bond at time t with maturity date T can be expressed as a function of $r(t)$ and τ only. Consequently the yield at time t with maturity date T, defined by $y(t,T) = -\frac{1}{T-t}\ln P$, can also be expressed as a function of $r(t)$ and τ only; that is $y(t,T)$ is of the functional form $Y(r(t),\tau)$. Applying Itô's lemma to $Y(r(t),\tau)$ indicates that for this class of term structure models the yield volatility structure is of the form $\frac{\partial Y(r(t),\tau)}{\partial r(t)}\sigma(r(t))$ which is a function of $r(t)$ and τ only. Consequently the yield volatility structure $\eta(\omega,t,T) = \eta(r(t),\tau)$ contains the class of term structure models where the short-tern interest rate follows a time-homogeneous univariate Markov diffusion under the equivalent risk-neutral probability measure.

Now consider the yield volatility structure $\eta(\omega,t,T) = \eta(r(t),\tau)$ which, via (4) under suitable regularity conditions, is equivalent to stating that the forward rate volatility structure is of the form $\gamma(\omega,t,T) = \gamma(r(t),\tau)$. Further assume that $\gamma(r(t),\tau)$ is admissible in a framework where the short term interest rate follows a Markov diffusion process under the risk-neutral probability measure; that is, it satisfies Condition 1 in Jeffrey (1995)[26] which, for this class of volatility structures, implies that the following must be satisfied

$$\begin{aligned}\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}\theta(r(t),t) + h(t,T) &= \gamma(r(t),\tau)\int_0^\tau \gamma(r(t),v)dv + \int_0^{r(t)} \frac{\partial}{\partial\tau}\left(\frac{\gamma(m,\tau)}{\gamma(m,0)}\right)dm \\ &\quad + \frac{1}{2}\gamma(r(t),0)^2\frac{\partial}{\partial r(t)}\left(\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}\right), \end{aligned} \tag{35}$$

where $\tau = T - t$, $\theta(r(t),t)$ is the drift of the short term interest rate under the risk neutral probability measure, and $h(t,T)$ is a particular function of t and T. Evaluating (35) at $T = t$ further shows that $\theta(r(t),t)$ must be of the form

$$\theta(r(t),t) = \left(\int_0^{r(t)} \frac{\partial}{\partial\tau}\left(\frac{\gamma(m,\tau)}{\gamma(m,0)}\right)dm\bigg|_{\tau=0} - h(t,t)\right). \tag{36}$$

Substituting (36) into (35) implies

$$\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}h(t,t) - h(t,T) = X(r(t),\tau) \tag{37}$$

[26] Condition 1 in Jeffrey (1995) is a necessary and sufficient condition for $\gamma(\omega,t,T)$ so that it is allowable in a framework where the short term interest rate follows a Markov diffusion process under the risk-neutral probability measure.

for the appropriate function $X(r(t),\tau)$. For the case where $\frac{\partial}{\partial r(t)}\left(\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}\right)\neq 0$ differentiating both sides of the above with respect to $r(t)$ indicates that $h(t,t)$ must be a constant indicating that the drift of the short term interest rate $\theta(r(t),t)$ is time-homogeneous. However the case where $\frac{\partial}{\partial r(t)}\left(\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}\right)=0$ is a special case of the example studies in Jeffrey (1995) page 628 where a volatility structure of the form $\gamma(r(t),\tau)$ with the property $\frac{\partial}{\partial r(t)}\left(\frac{\gamma(r(t),\tau)}{\gamma(r(t),0)}\right)=0$ must have a generically time-inhomogeneous short term interest rate process of the form

$$dr(t)=(\kappa r(t)+c(t))\,dt+\sqrt{a+br(t)}\,dW(t), \tag{38}$$

where $c(t)$ is an arbitrary function which can be chosen to calibrate the resulting term structure model to a given initial term structure. In summary the yield volatility structure $\eta(\omega,t,T)=\eta(r(t),\tau)$ admits the class of affine time-inhomogeneous short term interest rate processes (38) which is a subset of the extended Vasicek and extended CIR models considered in Hull-White (1990).

A3- ASYMPTOTIC DISTRIBUTIONS OF DRIFT AND DIFFUSION RELATED ESTIMATORS

To cover various processes in the body of this paper for which drift and diffusion related estimators are required, it is sufficient to consider a semi-martingale of the form

$$\begin{aligned} dZ(t) &= \alpha(\omega,t)\,dt+\eta(r(t))\,dW(t) \\ &= [\zeta(r(t))+\varphi(\omega,t)]\,dt+\eta(r(t))\,dW(t) \end{aligned} \tag{39}$$

where $r(t)$ also follows a semi-martingale of the form

$$dr(t)=\mu(\omega,t)\,dt+\sigma(r(t))\,dW(t), \tag{40}$$

$W(t)$, $t\geq 0$, is a one-dimensional standard Brownian Motion defined on the probability space $(\Omega,\{\mathcal{F}_t;t\geq 0\},P)$, and all drift and diffusion coefficients are $\mathcal{F}_t$-measurable with $\omega\in\mathcal{F}_t$ and sufficient regularity conditions assumed to ensure a unique non-explosive solution for $Z(t)$; for example see HJM (1992) for such regularity conditions in the context of interest rate models. The objects of interest are i) an estimate of $\eta(r(t))$ without any functional form restrictions placed on $\alpha(\omega,t)$, and ii) an estimate of $a(r(t))$ when $b(\omega,t)$ is known. In particular we consider the following two respective *kernel* based estimators which are adaptations of those proposed by Stanton (1997);

$$\widehat{\eta}(r)^2 = \frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1}K_h(r-r(t_i))\left[\Delta Z(t_i)\right]^2}{\sum_{i=1}^{n-1}K_h(r-r(t_i))} \tag{41}$$

$$\widehat{\zeta}(r) = \frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1}K_h(r-r(t_i))\left[\Delta Z(t_i)-n(\omega,t_i)\,\Delta t\right]}{\sum_{i=1}^{n-1}K_h(r-r(t_i))} \tag{42}$$

where $K_h(u) = K(u/h)/h$ for any *kernel function* $K(\cdot)$ which is a continuous, bounded, symmetric real function that integrates to one, and h is a positive scaling factor which is interpreted at the *bandwidth* of the kernel estimators. In the above n is the number of observations, $t_1 = 0$ without loss of generality, t_{n+1} is the time span of observations, $\Delta Z(t_i) = Z(t_{i+1}) - Z(t_i)$, and for simplicity $t_{i+1} - t_i = \Delta t$ for all i.

Before proceeding with the development of the asymptotic distributions of the estimators (41) and (42) we first introduce some preliminary results and required assumptions that are used extensively in the following proofs. The structure of the proofs follows Bandi and Phillips (1998) and Bandi (1999), however their analysis is restricted to Markov processes whereas ours are not.

Definition *(Chronological Local Time. See Phillips and Park (1998))*

The chronological local time of the semi-martingale r defined by (40) at point a over the time interval $[0,t]$ is defined as

$$\begin{aligned}\overline{L}(t,a) &= \frac{1}{\sigma^2(a)}\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_o^t \mathbf{1}_{|r(s)-a|<\epsilon}\sigma(r(s))^2 ds \\ &= \frac{1}{\sigma^2(a)}L(t,a)\end{aligned}$$

where $L(t,a)$ is the local time of r at point a over the time interval $[0,t]$.

Lemma *(The Occupation Time Formula. See Revuz and Yor (1999))*

For the semi-martingale r defined by (40) with quadratic variation process $\langle r,r\rangle_s$, and for every Borel function f of r

$$\int_o^t f(r(s))\, d\langle r,r\rangle_s = \int_{-\infty}^{+\infty} f(a)L(t,a)da$$

where $L(t,a)$ is the local time of r at point a over the time interval $[0,t]$.

Direct applications of the Occupation Time Formula along with the definition of Chronological Local Time provide the following two results:

$$\begin{aligned}\int_o^t f(r(s))\, ds &= \int_o^t \frac{f(r(s))}{\sigma^2(r(s))^2} d\langle r,r\rangle_s \\ &= \int_{-\infty}^{+\infty} f(a)\overline{L}(t,a)da\end{aligned} \tag{43}$$

and further, for any kernel function $K(\cdot)$ and continuous bounded function $f(\cdot)$,

$$\begin{aligned}\lim_{h\downarrow 0}\frac{1}{h}\int_o^t K\left(\frac{x-r(s)}{h}\right) f(r(s))\, ds &= \lim_{h\downarrow 0}\frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{x-a}{h}\right) f(a)\overline{L}(t,a)da \\ &= \lim_{h\downarrow 0}\int_{-\infty}^{+\infty} K(q)\, f(x+hq)\overline{L}(t,x+hq)dq \\ &\to \overline{L}(t,x)f(x) = O_{a.s.}\left(\overline{L}_r(t,r)\right) \ .\end{aligned} \tag{44}$$

Assumptions that must be imposed to study the asymptotic distributions of the estimators (41) and (42) are the following:

Assumption 1. (Recurrence)

The process $\{r(t); t \geqslant 0\}$ *defined in (40) is recurrent; that is for every point* a *on the support of this process the chronological local time* $\overline{L}(t,a) \to \infty$ *as* $t \to \infty$.

Assumption 2. (Boundedness)

The drift and diffusion functions are (locally) bounded:

$$|\alpha(\omega, t_i) - \alpha(\omega, t_j)| + |\eta(r(t_i)) - \eta(r(t_j))| \leq C\,|Z(t_i) - Z(t_j)| \tag{45}$$

for constants C, *and there exists some* $0 < \nu < \frac{1}{2}$ *such that*

$$\frac{(\Delta_n)^{\nu}}{\overline{L}(t,a)} \int_0^{t_n} |K_h(r - r(s))\alpha(\omega, s)ds| = O(1) \tag{46}$$

.

Assumption 3. (Sampling Conditions)

Let both the discretization width Δt *and bandwidth for the kernel estimators* h *depend on the sample size* n *which grows to infinity. To indicate this we will hereafter denote these quantities as* Δ_n *and* h_n *respectively. The sample frequency* $\frac{1}{\Delta_n} \to \infty$ *as* $n \to \infty$*; referred to as 'infill assumption'. The time span of observations* $t_{n+1} = n\Delta_n \to \infty$ *as* $n \to \infty$*; referred to as 'long span assumption'.*[27] *The bandwidth parameter* $h \downarrow 0$ *in such a way that* $\frac{h_n}{\Delta_n} \to \infty$ *and* $(\Delta_n)^{\beta} \frac{1}{h_n}\overline{L}(t,a) = O(1)$ *for all* $0 < \beta < \frac{1}{2}$ *and every point* a *on the support of the process* $\{r(t); t \geqslant 0\}$ *defined in (40);* $\overline{L}(t,a)$ *is the chronological local time of the process* $\{r(t); t \geqslant 0\}$.

Given the above construct we now proceed to prove the following two theorems which provide the basis for the analysis of all estimators and test statistics in this paper.

Theorem 1 *Given Assumption 1-3, the asymptotic distribution of* $\widehat{\eta}(r)^2$ *is a mixed normal:*

$$\sqrt{\frac{h_n\,\overline{L}(t_n, r)}{\Delta_n}} \left(\widehat{\eta}(r)^2 - \eta(r)^2\right) \Longrightarrow N\left(0, 4\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)\eta(r)^4\right)$$

for every point r *on the support of the process* $\{r(t); t \geqslant 0\}$ *in (40).*

Theorem 2 *Given Assumption 1-3, the asymptotic distribution of* $\widehat{m}(r)$ *is a mixed normal:*

$$\sqrt{h_n\,\overline{L}(t_n, r)}\left(\widehat{\zeta}(r) - \zeta(r)\right) \Longrightarrow N\left(0, 4\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)\eta(r)^2\right)$$

for every point r *on the support of the process* $\{r(t); t \geqslant 0\}$ *in (40).*

[27] This assumption can be dropped for the diffusion coefficient estimator provided Theorem 1 below.

We first introduce some preliminaries recently established in the literature, which will be utilized intensively in our proofs. The structure of the proofs for the estimators follows Bandi and Phillips (1998) and Bandi (1999), however their analysis is restricted to Markov processes while ours are not.

A3.1. Volatility Function Estimator

Let $\kappa_{n,t_n} = \max_{i\leq n} \max_{i\Delta_n \leq s\leq (i+1)\Delta_n} \left|Z(s) - Z(t_i)\right|$. By Holder's inequality

$$\Pr\left[t \geq 0 : \lim_{\Delta_n\to 0} \sup \frac{|Z(t+\Delta_n) - Z(t)|}{(\Delta_n)^{\beta}} > 0\right] = 0 \; a.s. \; for \; every \; \beta < \frac{1}{2},$$

thus

$$\frac{\kappa_{n,t_n}}{(\Delta_n)^{\beta}} = o_{a.s.}(1) \text{ for every } \beta < \frac{1}{2}. \tag{47}$$

Another useful consequence, under our condition that $\frac{L_r(t_n,r)}{h}(\Delta t)^{\beta} = O(1)$, is

$$\frac{\kappa_{n,t_n}}{h} L_r(t_n, r) = o_{a.s.}(1).$$

We first prove point-wise consistency for $\widehat{\eta}$. As shown in Phillips and Park (1998) and Bandi and Phillips (1999)

$$\Delta_n \sum_{i=1}^{n-1} K_h(r - r_i) \longrightarrow^p \overline{L}_r(t_n, r), \tag{48}$$

provided the above sampling conditions are observed.

Consider the quantity

$$\begin{aligned} \Delta Z^2(t_i) &= \left[Z(t_i + \Delta_n) - Z(t_i)\right]^2 \\ &= \left[Z(t_i + \Delta_n)^2 - Z(t_i)^2\right] - 2Z(t_i)\left[Z(t_i + \Delta_n) - Z(t_i)\right] \end{aligned} \tag{49}$$

Observing that

$$Z(t_i + \Delta_n) - Z(t_i) = \int_{t_i}^{t_i+\Delta_n} \alpha(\omega, s)ds + \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s),$$

and from Ito's lemma

$$Z(t_i + \Delta_n)^2 - Z(t_i)^2 = 2\int_{t_i}^{t_i+\Delta_n} Z(s)\alpha(\omega, s)ds + 2\int_{t_i}^{t_i+\Delta_n} Z(s)\eta(r(s))dW(s) + \int_{t_i}^{t_i+\Delta_n} \eta(r(s))^2 ds,$$

we consequently have

$$\begin{aligned} \Delta Z(t_i)^2 &= 2\int_{t_i}^{t_i+\Delta_n} \left[Z(s) - Z(t_i)\right]\alpha(\omega, s)ds \\ &\quad + 2\int_{t_i}^{t_i+\Delta_n} \left[Z(s) - Z(t_i)\right]\eta(r(s))dW(s) + \int_{t_i}^{t_i+\Delta_n} \eta(r(s))^2 ds. \end{aligned} \tag{50}$$

Therefore,

$$
\begin{aligned}
& \frac{1}{\Delta_n}\frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\Delta Z\left(t_i\right)^2}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}-\eta^2\left(r\right) \\
= & \frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\left[\frac{1}{\Delta_n}\Delta Z(t_i)^2-\eta\left(r\left(t_i\right)\right)^2\right]}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}+\frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\eta\left(r\left(t_i\right)\right)^2}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}-\eta^2\left(r\right) \\
= & \frac{\sum_{i=1}^{n-1}\frac{1}{\Delta_n}K_h(r-r(t_i))}{\sum_{i=1}^{n-1}K_h(r-r(t_i))}\left\{2\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right)-Z\left(t_i\right)\right]\alpha(\omega,s)ds+2\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right)-Z\left(t_i\right)\right]\eta(r\left(s\right))dW_s\right. \\
& \left.+\int_{t_i}^{t_i+\Delta_n}\left[\eta(r\left(s\right))^2-\eta(r\left(t_i\right))^2\right]ds\right\} \\
& +\frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\eta\left(r\left(t_i\right)\right)^2}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}-\eta^2\left(r\right). \qquad (51)
\end{aligned}
$$

Consider the term

$$
\begin{aligned}
& \frac{2\sum_{i=1}^{n-1}\frac{1}{\Delta_n}K_h(r-r\left(t_i\right))\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right)-Z\left(t_i\right)\right]\alpha(\omega,s)ds}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))} \\
\leq\; & \frac{\frac{2\kappa_{n,t_n}}{\Delta_n}\left|\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\int_{t_i}^{t_i+\Delta_n}\alpha(\omega,s)ds\right|}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))} \\
=\; & \frac{\frac{2\kappa_{n,t_n}}{\Delta_n}\left|\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\left[\int_{t_i}^{t_i+\Delta_n}\left(\alpha(\omega,s)-\alpha(\omega,t_i)\right)ds+\int_{t_i}^{t_i+\Delta_n}\left(\alpha(\omega,t_i)\right)ds\right]\right|}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))} \\
\leq\; & \mathrm{C}\kappa_{n,t_n}^2+2\kappa_{n,t_n}\left|\frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\alpha(\omega,t_i)}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}\right|,
\end{aligned}
$$

where the two inequalities follow definition of κ_{n,t_n} and the local boundedness assumption (45) that we impose on the drift term, respectively. Under (46)

$$
\begin{aligned}
\kappa_{n,t_n}\left|\frac{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))\alpha(\omega,t_i)}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}\right| &\leq \frac{\kappa_{n,t_n}}{\overline{L}\left(t_n,r\right)}\int_0^{t_n}\left|K_h(r-r\left(s\right))\alpha(\omega,s)\right|ds \\
&= \frac{\kappa_{n,t_n}}{\Delta_n^{\nu}}\to 0,
\end{aligned}
$$

when $\Delta_n\to 0$.

We next analyze the term

$$
\frac{\sum_{i=1}^{n-1}\frac{2}{\Delta_n}K_h(r-r\left(t_i\right))\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right)-Z\left(t_i\right)\right]\eta(r\left(s\right))dW\left(s\right)}{\sum_{i=1}^{n-1}K_h(r-r\left(t_i\right))}
$$

$$\leq \frac{2\kappa_{n,t_n}}{\Delta_n} \frac{\left|\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)\right|}{\sum_{i=1}^{n-1} K_h(r - r(t_i))}$$

$$= \frac{2\kappa_{n,t_n}}{\Delta_n} \Delta_n^{1/2} \frac{\left|\sum_{i=1}^{n-1} K_h(r - r(t_i)) \frac{1}{\Delta_n^{1/2}} \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)\right|}{\sum_{i=1}^{n-1} K_h(r - r(t_i))}.$$

Each element of the numerator

$$\frac{1}{\Delta_n^{1/2}} \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)$$

is a martingale difference sequence with zero expectation and finite variance

$$E\left[\frac{1}{\Delta_n^{1/2}} \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)\right]^2 = E\left[\frac{1}{\Delta_n} \int_{t_i}^{t_i+\Delta_n} \eta(r(s))^2 ds\right] < \infty,$$

where the equality follows Ito isometry. Invoking the law of large numbers for martingale differences we obtain

$$\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i)) \frac{1}{\Delta_n^{1/2}} \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)}{\sum_{i=1}^{n-1} K_h(r - r(t_i))} \to 0.$$

when $n \to \infty$, while

$$\frac{2\kappa_{n,t_n}}{\Delta_n} \Delta_n^{1/2} = \frac{2\kappa_{n,t_n}}{\Delta_n^{1/2}} = O(1),$$

so the whole term will approach zero.

For the next term in (51)

$$\frac{1}{\Delta_n} \frac{\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_{i+}\Delta_n} \left[\eta(r(s))^2 - \eta(r(t_i))^2\right] ds}{\sum_{i=1}^{n-1} K_h(r - r(t_i))}$$

Using the mean-value theorem and the occupation formula, we have

$$\begin{aligned}
& \sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_{i+}\Delta_n} \left[\eta(r(s))^2 - \eta(r(t_i))^2\right] ds \\
= & \sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_{i+}\Delta_n} 2\eta(r^*(t_i))\eta'(r^*(t_i))(r(s) - r(t_i)) ds \\
\leq & \; 2\kappa_{n,t_n} \left|\int_0^{t_n} K_h(r - r(s))\eta(f(r(s), r(t_i)))\eta'(f(r(s), r(t_i)))ds\right| \\
= & \; \frac{2\kappa_{n,t_n}}{h_n} \left|\int_{-\infty}^{+\infty} K\left(\frac{r-a}{h_n}\right)\eta(f(a))\eta'(f(a))\frac{L_r(t_n, a)}{\sigma(a)^2} da\right|
\end{aligned} \qquad (52)$$

$$= \quad 2\kappa_{n,t_n}\left|\int_{-\infty}^{+\infty} K(q)\eta(f(r+h_nq,r))\eta'(f(r+h_nq,r))\frac{L_r(t_n,r+h_nq)}{\sigma(r+h_nq)^2}dq\right|$$

$$\to \quad 2\kappa_{n,t_n}\left|\eta(r)\eta'(r)\frac{L_r(t_n,r)}{\sigma(r)^2}\int_{-\infty}^{+\infty} K(q)dq\right| = \mathrm{C}\kappa_{n,t_n}\overline{L}_r(t_n,r)$$

as $h_n \to 0$ ($r^*(t_i) = f(r(s), r(t_i)) \in [r(t_i), r(t_{i+1})]$ when $s \in [r(t_i), r(t_{i+1})]$, and will approach $r(t_i)$ when $\Delta_n \to 0$). Again, this bound approaches $0 : \kappa_{n,t_n} = o_{a.s.}\left(\Delta_{n,t_n}^{1/2-\delta}\right) \to 0$, when $\Delta_{n,t_n} \to 0$.

For the last term in (51)

$$\frac{\sum_{i=1}^{n-1} K_h(r-r(t_i))\eta(r(t_i))^2}{\sum_{i=1}^{n-1} K_h(r-r(t_i))} - \eta(r)^2,$$

we first consider the quantity

$$\frac{\Delta_n}{h_n}\sum_{i=1}^{n-1} K\left(\frac{r-r(t_i)}{h_n}\right)\eta(r(t_i))^2 - \frac{1}{h}\int_0^{t_n} K\left(\frac{r-r(s)}{h_n}\right)\eta(r(s))^2 ds$$

$$= \frac{1}{h_n}\sum_{i=1}^{n-1}\int_{t_i}^{t_i+\Delta_n}\left[K\left(\frac{r-r(t_i)}{h_n}\right)\eta(r(t_i))^2 - K\left(\frac{r-r(s)}{h_n}\right)\eta(r(s))^2\right]ds.$$

Adding and subtracting

$$\frac{1}{h_n}\sum_{i=1}^{n-1}\int_{t_i}^{t_i+\Delta_n}\left[K\left(\frac{r-r(s)}{h_n}\right)\eta(r(t_i))^2\right]ds$$

from the above quantity, we obtain

$$\begin{aligned}
= \quad & \frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n}\left[K\left(\frac{r-r(t_i)}{h_n}\right)\eta(r(t_i))^2 - K\left(\frac{r-r(s)}{h_n}\right)\eta(r(t_i))^2\right]ds \\
& +\frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n}\left[K\left(\frac{r-r(s)}{h_n}\right)\eta(r(t_i))^2 - K\left(\frac{r-r(s)}{h_n}\right)\eta^2(r(s))\right]ds \\
= \quad & \frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n} K'\left(\frac{r-r(t_i)}{h_n}\right)\left(\frac{r(s)-r(t_i)}{h_n}\right)\eta(r(t_i))^2 ds \\
& +\frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n} K''\left(\frac{r-r(t_i)}{h_n}\right)\left(\frac{r(s)-r(t_i)}{h_n}\right)^2\eta(r(t_i))^2 ds \\
& +\frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n} K\left(\frac{r-r(s)}{h_n}\right)\left[\eta(r(t_i))^2 - \eta(r(s))^2\right]ds \\
\leq \quad & \frac{\kappa_{n,t_n}}{h_n}\left|\frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n} K'\left(\frac{r-r(t_i)}{h_n}\right)\eta(r(t_i))^2 ds\right|
\end{aligned}$$

$$+\left(\frac{\kappa_{n,t_n}}{h_n}\right)^2\left|\frac{1}{h_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_i+\Delta_n}K''\left(\frac{r-r(t_i)}{h_n}\right)\eta(r(t_i))^2ds\right|+\mathrm{C}\kappa_{n,t_n}$$

(the third term, using (52), is of order κ_{n,t_n})

$$\leq \frac{\kappa_{n,t_n}}{h_n}\left|\frac{1}{h_n}\int_0^{t_n}K'\left(\frac{r-r(s)}{h_n}\right)\eta(r(s))^2ds\right|+\left(\frac{\kappa_{n,t_n}}{h_n}\right)^2\left|\frac{1}{h}\int_0^{t_n}K''\left(\frac{r-r(s)}{h_n}\right)\eta(r(s))^2ds\right|$$
$$+\mathrm{C}\kappa_{n,t_n}$$
$$= \frac{\kappa_{n,t_n}}{h_n}O_{a.s.}\left(\overline{L}_r(t_n,r)\right)+\left(\frac{\kappa_{n,t_n}}{h_n}\right)^2O_{a.s.}\left(\overline{L}_r(t_n,r)\right)+\mathrm{C}\kappa_{n,t_n}.$$

Under our sampling conditions, all of the terms in the above bound approaches 0.

In conclusion, under the stated assumptions,

$$\frac{\sum_{i=1}^{n-1}K_h(r-r(t_i))\eta(r_i)^2}{\sum_{i=1}^{n-1}K_h(r-r(t_i))}-\eta(r)^2\to 0,$$

which concludes our consistency proof. ■

The variance of the estimator will be driven only by the term

$$\sum_{i=1}^{n-1}\frac{\frac{1}{\Delta_n}K_h(r-r(t_i))}{\sum_{i=1}^{n-1}K_h(r-r(t_i))}2\int_{t_i}^{t_i+\Delta_n}\left[Z(s)-Z(t_i)\right]\eta(s)dW(s);$$

which has variation as

$$4\frac{\sum_{i=1}^{n-1}\frac{1}{\Delta_n^2}K_h(r-r(t_i))^2}{\left[\sum_{i=1}^{n-1}K_h(r-r(t_i))\right]^2}\int_{t_i}^{t_i+\Delta_n}\left[Z(s)-Z(r(t_i))\right]^2\eta(r(s))^2ds$$
$$\simeq \frac{4\sum_{i=1}^{n-1}K_h(r-r(t_i))^2\eta(r(t_i))^4}{\left[\sum_{i=1}^{n-1}K_h(r-r(t_i))\right]^2}.$$
$$= \frac{4\Delta_n}{h_n}\frac{\frac{1}{h_n}\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})^2\eta(r(t_i))^4\Delta_n}{\left[\Delta_n\sum_{i=1}^{n-1}K_{h_r}(r-r(t_i))\right]^2}$$

As $n\to\infty$ and $\Delta_n\to 0$, the numerator approaches

$$\frac{1}{h_n}\int_0^{t_n}K\left(\frac{r-r(s)}{h_n}\right)^2\eta(r(s))^4ds$$
$$= \frac{1}{h_n}\int_0^{t_n}K\left(\frac{r-r(s)}{h_n}\right)^2\eta(r(s))^4\frac{d[r(s)]}{\sigma(r(s))^2}=\frac{1}{h_n}\int_{-\infty}^{+\infty}K\left(\frac{r-a}{h_n}\right)^2\eta(a)^4\frac{L_r(t_n,a)}{\sigma(a)^2}da$$
$$= \int_{-\infty}^{+\infty}K(q)^2\eta(qh_n+r)^4\frac{L_r(T,qh_n+r)}{\sigma(qh_n+r)^2}dq\underset{h_n\to 0}{\to}\left(\int_{-\infty}^{+\infty}K(q)^2dq\right)\eta(r)^4\frac{L_r(t_n,r)}{\sigma(r)^2}$$
$$= \left(\int_{-\infty}^{+\infty}K(q)^2dq\right)\eta(r)^4\overline{L}_r(t_n,r),$$

while the denominator approaches, from (48), $\overline{L}_r^2(t_n, r)$.

In conclusion, the asymptotic variance of $\widehat{\eta}^2$ is

$$\frac{4\Delta_n}{h_n \overline{L}_r(t_n, r)} \left(\int_{-\infty}^{+\infty} K(q)^2 \, dq \right) \eta(r)^4.$$

The variance of $\widehat{\eta}$ is then obtained by a straightforward application of delta method

$$\frac{\Delta_{n,t_n}}{h_n \overline{L}_r(t_n, r)} \left(\int_{-\infty}^{+\infty} K(q)^2 \, dq \right) \eta(r)^2. \tag{53}$$

■

A3.2. Asymptotic covariance

Suppose we have two semi-martingale with similar structure to the Z_t that we introduced earlier, Z_{1t} and Z_{2t}, where

$$\begin{aligned} dZ_1(t) &= \alpha_1(\omega, t)\, dt + \eta_1(\omega, t)\, dW(t) \\ &= [\zeta_1(r(t)) + \varphi_1(\omega, t)]\, dt + \eta_1(r(t))\, dW(t) \end{aligned}$$

and

$$\begin{aligned} dZ_2(t) &= \alpha_2(\omega, t)\, dt + \eta_2(\omega, t)\, dW(t) \\ &= [\zeta_2(r(t)) + \varphi_2(\omega, t)]\, dt + \eta_2(r(t))\, dW(t). \end{aligned}$$

Estimates of $\widehat{\eta}_{1,2}$ are obtained by kernel smoothing as in (41). We will prove the claim that

$$\operatorname{cov}(\widehat{\eta}_1(r), \widehat{\eta}_2(r')) \to \begin{cases} 0 \text{ when } r \neq r' \\ \operatorname{std}(\widehat{\eta}_1(r)) \operatorname{std}(\widehat{\eta}_2(r')) \text{ when } r = r' \end{cases}$$

under the sampling conditions. From (51) and (50)

$$\begin{aligned} &\operatorname{cov}\left(\widehat{\eta}_1(r)^2, \widehat{\eta}_2(r')^2\right) \\ = \; & \frac{4}{\Delta_n^2} \sum_{i=1}^{n-1} K_h(r - r(t_i)) K_h(r' - r(t_i)) \\ & \times \frac{\int_{t_i}^{t_i + \Delta_n} [Z_1(s) - Z_1(t_i)] [Z_2(s) - Z_2(t_i)] \eta_1(r(s)) \eta_2(r(s)) ds}{\left[\sum_{i=1}^{n-1} K_h(r - r(t_i))\right] \left[\sum_{i=1}^{n-1} K_h(r' - r(t_i))\right]} \\ \simeq \; & \frac{4\Delta_n}{h_n} \frac{\frac{1}{h_n} \sum_{i=1}^{n-1} K(\frac{r - r(t_i)}{h_n}) K(\frac{r' - r(t_i)}{h_n}) \eta_1(r(t_i))^2 \eta_2(r(t_i))^2 \Delta_n}{\left[\Delta_n \sum_{i=1}^{n-1} K_h(r - r(t_i))\right] \left[\Delta_n \sum_{i=1}^{n-1} K_h(r' - r(t_i))\right]}. \end{aligned} \tag{54}$$

For the numerator, observe that, as $n \to \infty$ and $\Delta_n \to 0$

$$\frac{1}{h_n} \sum_{i=1}^{n-1} K(\frac{r - r(t_i)}{h_n}) K(\frac{r' - r(t_i)}{h_n}) \eta_1(r(t_i))^2 \eta_2(r(t_i))^2 \Delta_{n,t_n}$$

$$
\begin{aligned}
&\to \frac{1}{h_n}\int_0^{t_n} K(\frac{r-r(s)}{h_n})K(\frac{r'-r(s)}{h_n})\eta_1(r(s))^2\eta_2(r(s))^2 \\
&= \frac{1}{h_n}\int_0^{t_n} K(\frac{r-r(s)}{h_n})K(\frac{r'-r(s)}{h_n})\eta_1(r(s))^2\eta_2(r(s))^2\frac{d[r(s)]}{\sigma(r(s))^2} \\
&= \frac{1}{h_n}\int_{-\infty}^{+\infty} K(\frac{r-a}{h_n})K(\frac{r'-a}{h_n})\eta_1(a)^2\eta_2(a)^2\frac{L_r(t_n,a)}{\sigma^2(a)}da \\
&= \int_{-\infty}^{+\infty} K(q)K(q+\frac{r-r'}{h_n})\eta(qh_n+r)^2\eta(qh_n+r')^2\frac{L_r(t_n,qh_n+r)}{\sigma^2(qh_n+r)}dq \\
&\underset{h_n\to 0}{\to} \overline{L}_r(t_n,r)\eta(r)^2\eta(r')^2 \lim_{h_n\to 0}\int_{-\infty}^{+\infty} K(q)K(q+\frac{r-r'}{h_n})dq.
\end{aligned}
$$

Note that the integral can be written as

$$
\begin{aligned}
\int_{-\infty}^{+\infty} K(q)K(q+\frac{r-r'}{h_n})dq &= \int_{-\infty}^{+\infty} K(q)K(\frac{r'-r}{h_n}-q)dq \\
&= (K*K)\left(\frac{r'-r}{h_n}\right) \to \begin{cases} 0 & \text{if } r'\neq r \\ \int_{-\infty}^{+\infty} K(q)^2\, dq & \text{if } r'=r, \end{cases}
\end{aligned}
$$

where $(K*K)$ is the convolution of the kernel K. For concreteness, in case of the normal kernel, since $(K*K)=N(0,2)$, the above quantity approaches, as $h_n\to 0$,

$$
\frac{1}{2\sqrt{\pi}}\exp\left[-\left(\frac{r'-r}{h_n}\right)^2\right] \to \begin{cases} 0 & \text{if } r'\neq r \\ \frac{1}{2\sqrt{\pi}} = \int_{-\infty}^{+\infty} K(q)^2\, dq & \text{if } r'=r. \end{cases}
$$

The results we want to prove is obtained by substituting the above limit and $\Delta_n \sum_{i=1}^n K_{h_r}(r-r(t_i)) \to \overline{L}(t_n,r)$ and $\Delta_{n,t_n}\sum_{i=1}^n K_{h_r}(r'-r(t_i)) \to \overline{L}(t_n,r')$ into (54). Intuitively, our pointwise estimators, as local averages, are asymptotically serially independent for different point estimates when $h_n\to 0$. When $r'=r$, then since in our model the same Brownian motion drives every points on the whole yield curve, our estimators by construction will be asymptotically perfectly correlated. ■

A3.3. Estimator of part of the drift function

We motivate our estimator of by observing that

$$
\begin{aligned}
\widehat{\zeta}(r(t)) &= \lim_{\Delta t\downarrow 0}\left(\frac{1}{\Delta t}E\left[(Z(t+\Delta t)-Z(t))-\varphi(\omega,t)\Delta t\,|\,\mathcal{F}_t\right]\right) \\
&= \lim_{\Delta t\downarrow 0}\left(\frac{1}{\Delta t}E\left[(Z(t+\Delta t)-Z(t))-\varphi(\omega,t)\Delta t\,|\,r(t)\right]\right),
\end{aligned}
$$

and thus, can be estimated by kernel smoothing

$$
\widehat{\zeta}(r) = \frac{1}{\Delta t}\frac{\sum_{i=1}^{n-1} K_h(r-r(t_i))\left[\Delta Z(t_i)-\varphi(\omega,t_i)\Delta t\right]}{\sum_{i=1}^{n-1} K_h(r-r_i)},
$$

The asymptotic properties of $\widehat{\zeta}$ can be derived in a fashion similar to the above section. Since

$$\Delta Z(t_i) = Z(t_i + \Delta_n) - Z(t_i) = \int_{t_i}^{t_i+\Delta_n} [\zeta(r(s)) + \varphi(\omega, s)]\, ds + \int_{t_i}^{t_i+\Delta_{n,t_n}} \eta(r(s))dW(s),$$

the numerator of $\widehat{\zeta}$ can be written as

$$\begin{aligned} &\frac{1}{\Delta_n}\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_i+\Delta_n} [\alpha(\omega, s) - \alpha(\omega, t_i)]\, ds \\ &+\frac{1}{\Delta_n}\sum_{i=1}^{n-1} K_h(r - r(t_i))\, \zeta(r(t_i))\Delta_n + \frac{1}{\Delta_n}\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s). \end{aligned}$$

Assume that the drift function of the process $Z(.)$ satisfies the local boundedness assumption in (45), then consistency of $\widehat{\zeta}$ is obtained, similar to the consistency proof for the volatility function estimator in A3.1. Explicitly, the first term is:

$$\begin{aligned} &\frac{1}{\Delta_n}\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_i+\Delta_n} [\alpha(\omega, s) - \alpha(\omega, t_i)]\, ds}{\sum_{i=1}^{n-1} K_h(r - r(t_i))} \\ \leq\ & \frac{1}{\Delta_n} C\kappa_{n,t_n}\Delta_n \frac{\sum_{i=1}^{n-1} K_h(r - r(t_i))}{\sum_{i=1}^{n-1} K_h(r - r(t_i))} \\ =\ & C\kappa_{n,t_n} = o_{a.s.}\left(\Delta_n^{1/2-\delta}\right) \to 0, \end{aligned}$$

when $\Delta_n \to 0$, where the last equality is due to (47). The term

$$\frac{1}{\Delta_n}\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i)) \int_{t_i}^{t_i+\Delta_n} \eta(r(s))dW(s)}{\sum_{i=1}^{n-1} K_h(r - r(t_i))}$$

approaches zero by an application of the law of large number for martingale differences, similar to what we have done earlier. For the last term, under the assumptions and follows the derivation in the above section, it can be proved that

$$\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i))\zeta(r_i)}{\sum_{i=1}^{n} K_h(r - r(t_i))} \to \zeta(r),$$

which completes the proof of consistency for $\widehat{\zeta}$. ■

The variance of $\widehat{\zeta}$ is driven by the last term, whose variation is

$$\begin{aligned} \frac{1}{\Delta_n^2}\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i))^2 \int_{t_i}^{t_i+\Delta_n} \eta(r(s))^2 ds}{\left[\sum_{i=1}^{n-1} K_h(r - r(t_i))\right]^2} &\simeq \frac{1}{\Delta_n}\frac{\sum_{i=1}^{n-1} K_h(r - r(t_i))^2\, \eta(r(t_i))^2}{\left[\sum_{i=1}^{n-1} K_h(r - r(t_i))\right]^2} \\ &= \frac{1}{h_n}\frac{\frac{1}{h_n}\sum_{i=1}^{n-1} K^2\left(\frac{r-r(t_i)}{h}\right) \eta(r(t_i))^2\Delta_n}{\left[\Delta_n \sum_{i=1}^{n-1} K_h(r - r(t_i))\right]^2}. \end{aligned}$$

As $n \to \infty$ and $\Delta_n \to 0$, the numerator approaches

$$\begin{aligned}
& \frac{1}{h_n} \int_0^{t_n} K\left(\frac{r - r(s)}{h_n}\right)^2 \eta(r(s))^2 ds \\
= & \frac{1}{h_n} \int_0^{t_n} K\left(\frac{r - r(s)}{h_n}\right)^2 \eta(r(s))^2 \frac{d[r(s)]}{\sigma(r(s))^2} = \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{r-a}{h_n}\right)^2 \eta(a)^2 \frac{L_r(t_n, a)}{\sigma(a)^2} da \\
= & \int_{-\infty}^{+\infty} K(q)^2 \eta(qh_n + r)^2 \frac{L_r(T, qh_n + r)}{\sigma(qh_n + r)^2} dq \underset{h_n \to \infty}{\to} \left(\int_{-\infty}^{+\infty} K(q)^2 dq\right) \eta(r)^2 \frac{L_r(t_n, r)}{\sigma(r)^2} \\
= & \left(\int_{-\infty}^{+\infty} K(q)^2 dq\right) \eta(r)^2 \overline{L}_r(t_n, r).
\end{aligned}$$

As $\Delta_n \sum_{i=1}^{n-1} K_h(r - r(t_i)) \to \overline{L}_r(t_n, r)$ under the stated sampling assumptions, the variance is

$$\frac{1}{h_n \overline{L}(t_n, r)} \left(\int_{-\infty}^{+\infty} K(u)^2 du\right) \eta(r)^2. \tag{55}$$

Two observations are in order. First, comparing (53) and (55), it is obvious that the rate of convergence (or divergence) of $\widehat{\zeta}(r)$ is $\sqrt{\Delta_n}$ slower than that of $\widehat{\eta}$. Secondly, for $\widehat{\zeta}(r)$ to converge, it is absolutely necessary that $\overline{L}(t_n, r) \to \infty$ when $t_n \to \infty$. Sufficient condition for this is that the process $r(t)$ is recurrent, as we assume in assumption 3.

A3.4. Asymptotic Covariance

Observe that

$$\begin{aligned}
& \operatorname{cov}\left(\widehat{\zeta}_1(r), \widehat{\zeta}_2(r')\right) \\
= & \frac{4}{\Delta_n^2} \frac{\sum_{i=1}^{n-1} K_h(r - r(t_i)) K_h(r' - r(t_i)) \int_{t_i}^{t_i + \Delta_n} \eta_1(r(s)) \eta_2(r(s)) ds}{\left[\sum_{i=1}^{n-1} K_h(r - r(t_i))\right] \left[\sum_{i=1}^{n-1} K_h(r' - r(t_i))\right]},
\end{aligned}$$

a proof for asymptotic perfect correlation between the cross-sectional estimates and asymptotic serial uncorrelatedness, i.e.

$$\operatorname{cov}\left(\widehat{\zeta}_1(r), \widehat{\zeta}_2(r')\right) \to \begin{cases} 0 & \text{if } r' \neq r \\ \operatorname{std}\left(\widehat{\zeta}_1(r)\right) \operatorname{std}\left(\widehat{\zeta}_2(r')\right) & \text{if } r' = r. \end{cases}$$

where $\zeta_1(r)$ and $\zeta_2(r)$ are defined in A3.3, is very similar to A3.2 and for space consideration is not shown here. This important result will underlie our statistical tests in a significant way. All of our tests can be interpreted as based on some form variations of $\widehat{\zeta}$. Consequently, when we examine a sequence of $\widehat{\zeta}_1(r_j)'s$, where $j = 1, ...K$, then they are asymptotically independent, and a χ^2-test with K degree of freedom can be constructed easily, as being done in our empirical investigation. However, as any pair of $\left(\widehat{\zeta}_1(r), \widehat{\zeta}_2(r)\right)$ will be asymptotically perfectly correlated as our claim above, a large sample test based on them for example to examine some linear relationship between the two, is very problematic, to say the least. To break this curse of dependency is the subject of the next sub-section.

■

A3.5. Sampling from sub-samples

We examine the asymptotic behavior of the estimator $\widehat{\zeta}$ when estimation is conducted based on just half of the sample consists of every second observations. Without loss of generality, suppose we sample from even observations:

$$\widehat{\zeta}(r) = \frac{1}{\Delta_n} \frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i})) \left[\Delta Z(2t_i) - \varphi(\omega, 2t_i)\Delta t\right]}{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))}.$$

A distinctive characteristics of this sampling scheme, as mentioned else where in the paper, is that although we sample at every other time intervals, or every $2\Delta t$, we are still only interested in the increment of Z over Δt. In other words, our $\Delta Z(2t_i)$ is $[Z(2t_i + \Delta t) - Z(2t_i)]$, not $[Z(2t_i + 2\Delta t) - Z(2t_i)]$.

It is apparent that consistency will remain, with all the boundedness arguments go through, while

$$\frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))\zeta(r(t_{2i}))}{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))} \to \zeta(r)$$

still holds under our sampling conditions. For the variance, analyze the term

$$\begin{aligned}
\frac{1}{\Delta_n^2} \frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))^2 \int_{2t_i}^{2t_i+\Delta_n} \eta(r(s))^2 ds}{\left[\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))\right]^2} &\simeq \frac{1}{\Delta_n} \frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))^2 \eta(r(t_{2i}))^2}{\left[\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))\right]^2} \\
&= \frac{1}{h_n} \frac{\frac{1}{h_n} \sum_{i=1}^{[(n-1)/2]} K\left(\frac{r(t_{2i})-r}{h_n}\right)^2 \eta(r(t_{2i}))^2 \Delta_n}{\left[\Delta_n \sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))\right]^2}.
\end{aligned}$$

As $n \to \infty$ and $\Delta_n \to 0$, the numerator approaches

$$\begin{aligned}
&\frac{1}{2}\frac{1}{h_n} \int_0^{t_n} K\left(\frac{r - r(s)}{h_n}\right)^2 \eta(r(s))^2 ds \\
&\underset{h_n \to \infty}{\to} \frac{1}{2}\left(\int_{-\infty}^{+\infty} K(q)^2 dq\right) \eta(r)^2 \overline{L}_r(t_n, r),
\end{aligned}$$

which is half of the limit we obtain before when we sample from the whole sample. For the denominator, using (48)

$$\Delta_n \sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i})) \longrightarrow^p \frac{1}{2}\overline{L}_r(t_n, r).$$

Thus the estimator obtained from this sampling scheme will still be consistent but having the variance twice as large as the original one.

The desired effect, i.e. to have

$$\mathrm{cov}\left(\widehat{\zeta}_1(r), \widehat{\zeta}_2(r)\right) = 0$$

is achieved now if we estimate $\widehat{\zeta}_1(r)$ from the even observations sample and estimate $\widehat{\zeta}_2(r)$ from the odd observations sample as follows

$$\begin{aligned}\mathrm{cov}\left(\widehat{\zeta}_1(r), \widehat{\zeta}_2(r)\right) &= \mathrm{cov}\left[\begin{array}{c} \frac{1}{\Delta_n}\frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i}))\int_{2t_i}^{2t_i+\Delta_n}\eta(r(s))dW(s)}{\left[\sum_{i=1}^{[(n-1)/2]} K_h(r-(t_{2i}))\right]^2}, \\ \frac{1}{\Delta_n}\frac{\sum_{i=1}^{[(n-1)/2]} K_h(r - r(t_{2i-1}))\int_{2t_i-\Delta_n}^{2t_i}\eta(r(s))dW(s)}{\left[\sum_{i=1}^{[(n-1)/2]} K_h(r-(t_{2i-1}))\right]^2} \end{array}\right] \\ &= 0,\end{aligned}$$

by the independence property of the Brownian motion.

■

A4-PROOFS FOR ASYMPTOTICS OF OUR HJM ESTIMATORS

A4.1. Volatility estimator

This is just a straightforward application of A3.1, where the yield process $y(t, \tau)$ plays the role of $Z(t)$, while the spot rate process $r(t)$ is $r(t)$. ■

A4.2. No-arbitrage based estimator

We will apply the results established in A3.3 here. For the process $\widetilde{y}$ in (5), the drift can be broken up into two parts as:

$$\zeta(r(t)) = \lambda\,\eta(r(t), \tau) + \frac{1}{2}\tau\,\eta(r(t), \tau)^2$$

and

$$\varphi(\omega, t) = \left(\frac{\partial\widetilde{y}(t, \tau)}{\partial\tau} + \frac{\widetilde{y}(t, \tau) - r(t)}{\tau}\right)$$

Then the estimate of $\zeta(r)$ defined in (12), in a straightforward application of the results established in A3.3, is a consistent estimate of $\zeta(r(t))$, with the asymptotic variance as in (55). Then, by the delta method, using (13), the variance of $\widehat{\eta}_{NA}(r, \tau)$ is

$$\frac{1}{(\lambda + \tau\,\eta(r, \tau))^2}\frac{\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)\eta(r, \tau)^2}{h\,\overline{L}(t_n, r)}.$$

■

A5-PROOFS FOR ASYMPTOTICS OF THE TEST STATISTICS

A5.1.No-arbitrage restriction test

Observing (15), we can apply the results in A3.3 with

$$\zeta(r(t)) = q(r(t), T-t) = \lambda\,\eta(r(t), T-t) + \frac{1}{2}(T-t)\eta(r(t), T-t)^2$$

and

$$\varphi(\omega, t) = \tfrac{y(t,T)-r(t)}{T-t}.$$

Thus, $\widehat{q}(r, T-t)$, obtained by kernel smoothing as in (17), is a consistent estimator of $q(r, T-t)$ with the variance shown in (55), or, if we are to sample from half of the sample, twice as large as that variance. If an estimate of the market price of risk, $\lambda(r(t))$, is to be extracted from this $\widehat{q}(r, T-t)$ ((18) can be interpreted this way), then the variance of this $\widehat{\lambda}(r(t))$ is

$$\operatorname{var}\left[\widehat{\lambda}(r(t))\right] = \frac{1}{\eta(r(t), T-t)^2}\operatorname{var}\left[\widehat{q}(r, T-t)\right] = \frac{\left(\int_{-\infty}^{+\infty} K(u)^2\,du\right)}{h\,\overline{L}(t_n, r)}.$$

This result does not change even when we aggregate information across maturities; we can simply apply the above result for the following process

$$dY(t) = \sum_{j=1}^{J} dy(t, T_j),$$

which has the dynamics

$$\begin{aligned} dY(t) &= \left[\sum_{j=1}^{J}\frac{y(t,T_j)-r(t)}{T_j-t} + \frac{1}{2}\sum_{j=1}^{J}(T_j-t)\eta(r(t), T_j-t)^2 + \lambda(r(t))\sum_{j=1}^{J}\eta(r(t), T_j-t)\right]dt \\ &\quad + \left[\sum_{j=1}^{J}\eta(r(t), T_j-t)\right]dW(t) \end{aligned}$$

If $\widehat{\lambda}(r(t))$ is estimated from just half of the sample, as we have done, then its variance will increase twice, as shown in A3.4.

If we are to estimate $\widehat{\widehat{q}}$ and $\widehat{\widehat{\lambda}}$ from separate subsamples as described in A3.5, then this sampling scheme, as proved in A3.5, render $\widehat{\widehat{q}}$ and $\widehat{\widehat{\lambda}}$ independent and validate the claim on the variance of the test statistic in (22). For the empirical implementation as done in section 6.1, as proved in A3.4, for different levels of r, $\widehat{\widehat{q}}(r)$ and $\widehat{\widehat{\lambda}}(r)$ are uncorrelated with $\widehat{\widehat{q}}(r')$ and $\widehat{\widehat{\lambda}}(r')$, and similar for $\widehat{\widehat{\lambda}}(r)$ and $\widehat{\widehat{\lambda}}(r')$, thus the series $\left(\widehat{\widehat{q}}(r) - \widehat{\lambda}(r)\right)$ is asymptotically serially uncorrelated. Our χ^2-test is thus validated. ■

A5.2.Markov test statistics

A5.2.1. Asymptotics for the derivative estimators

We will establish the asymptotic distribution of the derivative estimators obtained by differentiate the estimators (41) and (42)

Theorem 3 *Given Assumption 1-3, the asymptotic distributions of* $\partial\widehat{\eta}(r)^2/\partial r$ and $\partial\widehat{\zeta}(r)/\partial r$ *are mixed normal:*

$$\sqrt{\frac{h^3\ \overline{L}\left(t_n, r\right)}{\Delta t}}\left(\partial\widehat{\eta}(r)^2/\partial r - \partial\eta(r)^2/\partial r\right) \Longrightarrow N\left(0, 4\left(\int_{-\infty}^{+\infty} K'\left(u\right)^2 du\right)\eta(r)^4\right),$$

$$\sqrt{h^3\ \overline{L}\left(t_n, r\right)}\left(\partial\widehat{\zeta}(r)/\partial r - \partial\zeta(r)/\partial r\right) \Longrightarrow N\left(0, \left(\int_{-\infty}^{+\infty} K'\left(u\right)^2 du\right)\eta(r)^2\right).$$

This implies that:

$$\operatorname{var}\left(\partial\widehat{\eta}(r)/\partial r\right) = \frac{\Delta t\ \left(\int_{-\infty}^{+\infty} K'\left(u\right)^2 du\right)\eta(r)^2}{h^3\ \overline{L}\left(t_n, r\right)}. \tag{56}$$

We also have

$$\operatorname{cov}\left(\frac{\partial\widehat{\zeta}_1(r)}{\partial r}, \frac{\partial\widehat{\zeta}_2(r')}{\partial r}\right) \to \begin{cases} 0 & \text{if } r' \neq r \\ \operatorname{std}\left(\frac{\partial\widehat{\zeta}_1(r)}{\partial r}\right)\operatorname{std}\left(\frac{\partial\widehat{\zeta}_2(r')}{\partial r}\right) & \text{if } r' = r. \end{cases} \tag{57}$$

where $\zeta_1(r)$ *and* $\zeta_2(r)$ *are parts of the drift functions of any two* $Z\left(t\right)$ *processes defined as in A3.2.*

The proof is similar to the proof for asymptotics of derivative estimator in regression context in Pagan and Ullah (1999), theorem 4.2, p. 179.We first establish the asymptotic distribution of $\partial\widehat{\eta}(r)/\partial r$. From (9), we have

$$\begin{aligned} \frac{\partial\left(\widehat{\eta}^2(r)\right)}{\partial r} &= \frac{1}{\Delta_n}\frac{\left[\sum_{i=1}^{n-1} K_h'(r - r\left(t_i\right))\Delta Z\left(t_i\right)^2\right]}{\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))} \\ &\quad -\frac{1}{\Delta_n}\frac{\left[\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))\Delta Z\left(t_i\right)^2\right]\left[\sum_{i=1}^{n-1} K_h'(r - r\left(t_i\right))\right]}{\left[\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))\right]^2}. \end{aligned} \tag{58}$$

Using (50), and steps similar to the derivation of the volatility structure estimator' asymptotics done in A3.1, the first term is

$$\begin{aligned} &\left[\frac{\sum_{i=1}^{n-1} K_h'(r - r\left(t_i\right))\left[\frac{1}{\Delta_n}\Delta Z(t_i)^2 - \eta\left(r\left(t_i\right)\right)^2\right]}{\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))}\right] + \left[\frac{\sum_{i=1}^{n-1} K_h'(r - r\left(t_i\right))\eta\left(r\left(t_i\right)\right)^2}{\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))}\right] \\ = &\ \frac{\sum_{i=1}^{n-1}\frac{1}{\Delta_n}K_h'(r - r\left(t_i\right))}{\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))} \\ &\times\left\{\begin{array}{c} 2\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right) - Z\left(t_i\right)\right]\alpha(\omega, s)ds + 2\int_{t_i}^{t_i+\Delta_n}\left[Z\left(s\right) - Z\left(t_i\right)\right]\eta(r\left(s\right))dW\left(s\right) \\ +\int_{t_i}^{t_i+\Delta_n}\left[\eta(r\left(s\right))^2 - \eta(r\left(t_i\right))^2\right]ds \end{array}\right\} \\ &+\left[\frac{\sum_{i=1}^{n-1} K_h'(r - r\left(t_i\right))\eta\left(r\left(t_i\right)\right)^2}{\sum_{i=1}^{n-1} K_h(r - r\left(t_i\right))}\right]. \end{aligned}$$

Denote the first term and second term of the above expression as ψ_1 and ψ_2, respectively. Boundedness of ψ_1, or $\psi_1 \to 0$, follows exactly the same arguments for the volatility function estimator. For ψ_2, its numerator can be manipulated as below, with integration by parts

$$\begin{aligned}
& \sum_{i=1}^{n-1} K_h'(r - r(t_i))\eta(r(t_i))^2 \Delta_n, \text{ as } n \to \infty \textit{ and } \Delta_n \to 0, \\
\to\ & \frac{1}{h_n}\int_0^{t_n} K'\left(\frac{r - r(s)}{h_n}\right)\eta(r(s))^2 ds = \frac{1}{h_n}\int_0^{t_n} K'\left(\frac{r - r(s)}{h_n}\right)\eta(r(s))^2 \frac{d[r(s)]}{\sigma(r(s))^2} \\
=\ & \frac{1}{h_n}\int_{-\infty}^{+\infty} K'\left(\frac{r-a}{h}\right)\eta(a)^2\frac{L_r(t_n,a)}{\sigma(a)^2}da = \frac{1}{h_n}\int_{-\infty}^{+\infty} K\left(\frac{r-a}{h_n}\right)\left[\eta(a)^2\frac{L_r(t_n,a)}{\sigma(a)^2}\right]' da \\
=\ & \frac{1}{h_n}\int_{-\infty}^{+\infty} K\left(\frac{r-a}{h_n}\right)\left[\eta(a)^2\right]' \frac{L_r(t_n,a)}{\sigma(a)^2}da + \frac{1}{h_n}\int_{-\infty}^{+\infty} K\left(\frac{r-a}{h_n}\right)\eta(a)^2\left[\frac{L_r(t_n,a)}{\sigma(a)^2}\right]' da, \\
\to\ & \left[\frac{\partial\left(\eta(r)^2\right)}{\partial r}\overline{L}_r(t_n,r) + \eta(r)^2\left[\frac{L_r(t_n,r)}{\sigma(r)^2}\right]'\right] \\
=\ & \left[\frac{\partial\left(\eta(r)^2\right)}{\partial r}\overline{L}_r(t_n,r) + \eta(r)^2\overline{L}_r'(t_n,r)\right]. \qquad (59)
\end{aligned}$$

as $h_n \to \infty$

For the second term in (58), utilizing the previous result $\left[\frac{1}{\Delta_n}\sum_{i=1}^{n-1} K_{h_r}(r - r_i)\Delta Z(t_i)^2\right] \to \eta(r)^2$ while

$$\sum_{i=1}^{n-1} K_h'(r - r_i)\Delta_{n,t_n}, \text{ as } n \to \infty \text{ and } \Delta_{n,t_n} \to 0, \qquad (60)$$

$$\begin{aligned}
\to\ & \frac{1}{h_n}\int_0^{t_n} K'\left(\frac{r - r(s)}{h_n}\right)ds = \int_0^{t_n} K'\left(\frac{r - r(s)}{h_n}\right)\frac{d[r_s]}{\sigma(s)^2} \\
=\ & \frac{1}{h_n}\int_{-\infty}^{+\infty} K'\left(\frac{a-r}{h}\right)\frac{L_r(t_n,a)}{\sigma(a)^2}da = \frac{1}{h_n}\int_{-\infty}^{+\infty} K\left(\frac{a-r}{h}\right)\left[\frac{L_r(t_n,a)}{\sigma(a)^2}\right]' da,
\end{aligned}$$

which approaches

$$\left[\frac{L_r(t_n,r)}{\sigma(r)^2}\right]' = \overline{L}_r'(t_n,r)$$

as $h_n \to 0$.

Substituting (59) and (60) into (58), point-wise consistency of $\partial\widehat{\eta}(r)^2/\partial r$ is thus achieved. ■

For the asymptotic variance of $\partial\widehat{\eta}^2(r)/\partial r$, consider

$$\begin{aligned}
& \sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left[\frac{\partial\left(\widehat{\eta}(r)^2\right)}{\partial r} - \frac{\partial\left(\eta(r)^2\right)}{\partial r}\right] \\
=\ & \sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left\{\frac{1}{\Delta_n}\frac{\frac{1}{h_n}\left[\sum_{i=1}^{n} K'(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]}{\sum_{i=1}^{n} K(\frac{r-r(t_i)}{h_n})} - \frac{\partial\left(\eta(r)^2\right)}{\partial r}\right\}
\end{aligned}$$

$$
\begin{aligned}
&-\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\frac{\frac{1}{h_n}\left[\sum_{i=1}^n K(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]\left[\sum_{i=1}^n K'(\frac{r-r(t_i)}{h_n})\right]}{\left[\sum_{i=1}^n K(\frac{r-r(t_i)}{h_n})\right]^2} \\
= \quad & \sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left\{\frac{1}{\Delta_n}\frac{\frac{1}{h_n}\left[\sum_{i=1}^{n-1} K'(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]}{\sum_{i=1}^{n-1} K(\frac{r-r(t_i)}{h_n})}-\frac{\partial(\eta(r)^2)}{\partial r}\right\} \\
&-\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\sum_{i=1}^{n-1}\frac{\frac{1}{h_n}K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(r-r(t_i))}\left\{\frac{\frac{1}{\Delta_n}\left[\sum_{i=1}^{n-1} K(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]}{\sum_{i=1}^{n-1} K(\frac{r-r(t_i)}{h_n})}-\eta(r)^2\right\} \\
&+\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\sum_{i=1}^{n-1}\frac{\frac{1}{h_n}K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}\eta(r)^2 \\
= \quad & \sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left\{\frac{1}{\Delta_n}\frac{\frac{1}{h_n}\left[\sum_{i=1}^{n-1} K'(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]}{\sum_{i=1}^{n-1} K(\frac{r-r(t_i)}{h_n})}-\frac{\partial(\eta(r)^2)}{\partial r}\right\} \\
&-h_n\sqrt{\frac{h_n\ \overline{L}(t_n,r)}{\Delta_n}}\sum_{i=1}^{n-1}\frac{\frac{1}{h_n}K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}\left\{\widehat{\eta}(r)^2-\eta(r)^2\right\} \\
&+\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\sum_{i=1}^{n-1}\frac{K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}\eta(r)^2 \\
= \quad & \sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left\{\frac{1}{\Delta_{n,t_n}}\frac{\frac{1}{h_n}\left[\sum_{i=1}^{n-1} K'(\frac{r-r(t_i)}{h_n})\Delta Z(t_i)^2\right]}{\sum_{i=1}^{n-1} K(\frac{r-r(t_i)}{h_n})}-\frac{\partial(\eta(r)^2)}{\partial r}\right\} \\
&-h_n\sum_{i=1}^{n-1}\frac{\frac{1}{h_n}K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}\left\{\sqrt{\frac{h_n\ \overline{L}(t_n,r)}{\Delta_{n,t_n}}}\left\{\widehat{\eta}(r)^2-\eta(r)^2\right\}\right\} \\
&+\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\sum_{i=1}^{n-1}\frac{K'(\frac{r-r(t_i)}{h_n})}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}\eta(r)^2.
\end{aligned}
$$

Proceed similarly to the proof of the asymptotics of the volatility function estimator, then the first term will have the variance as

$$
\frac{\Delta_n\ 4\left(\int_{-\infty}^{+\infty} K'(u)^2\,du\right)\eta(r)^4}{h_n^3\ \overline{L}(t_n,r)}. \tag{61}
$$

Since $\sqrt{\frac{h_n\ \overline{L}(t_n,r)}{\Delta_n}}\left\{\widehat{\eta}(r)^2-\eta(r)^2\right\}$ tends to a random variable in distribution while $h_n\to 0$ as proved in A3.1, the asymptotic distribution of

$$
\sqrt{\frac{h_n^3\ \overline{L}(t_n,r)}{\Delta_n}}\left[\frac{\partial(\widehat{\eta}(r)^2)}{\partial r}-\frac{\partial(\eta(r)^2)}{\partial r}\right]
$$

is the same as that of

$$\sqrt{\frac{h_n^3\,\overline{L}\left(t_n,r\right)}{\Delta_n}}\left\{\frac{1}{\Delta_n}\frac{\frac{1}{h_n}\left[\sum_{i=1}^{n-1}K'(\frac{r-r(t_i)}{h_n})\Delta Z\left(t_i\right)^2\right]}{\sum_{i=1}^{n-1}K(\frac{r-r(t_i)}{h_n})}-\frac{\partial\left(\eta(r)^2\right)}{\partial r}\right\},$$

i.e., the asymptotic variance of $\partial\widehat{\eta}(r)^2/\partial r-\partial\eta(r)^2/\partial r$ is as in (61). The results here are consistent with a well-known feature of nonparametric technique that the derivative functional estimator has slower rate of convergence than the functional estimator itself.

The proof for $\partial\widehat{m}(r)/\partial r$ is almost identical, and is omitted here. ■

A5.2.2 Markov test

For the first Markov restriction, first note that for ϕ_1 defined in (25), the variance of its estimate constructed in (28) will be dominated by the variations of the derivative estimators, which are as follows, from Theorem 3

$$\begin{aligned}\operatorname{var}\left[\frac{\partial\widehat{m}(r,\tau)}{\partial r}\right] &= \frac{1}{h^3\,\overline{L}\left(t_n,r\right)}\left(\int_{-\infty}^{+\infty}K'\left(u\right)^2du\right)\eta(r,\tau)^2\\ \operatorname{var}\left[\frac{\partial\widehat{\lambda}(r)}{\partial r}\right] &= \frac{1}{h^3\,\overline{L}\left(t_n,r\right)}\left(\int_{-\infty}^{+\infty}K'\left(u\right)^2du\right)\\ \operatorname{var}\left[\frac{\partial\widehat{\eta}(r,\tau)}{\partial r}\right] &= \frac{\Delta_n}{h^3\,\overline{L}\left(t_n,r\right)}\left(\int_{-\infty}^{+\infty}K'\left(u\right)^2du\right)\eta(r,\tau)^2.\end{aligned}$$

Obviously that the first two variances dominate the last one by a factor of $(1/\Delta_n)$, so the variance of $\widehat{\phi}_1$ will be driven by the first 2 terms mainly. To break the dependency between $\widehat{m}$ and $\widehat{\lambda}$, the sampling scheme utilized earlier is employed again, and the rest of our proofs, namely (32) and the consequent constructed χ^2-test used in the empirical application follows through as in A5.1 ((57) plays a similar role as the result established in A3.2, i.e. shows each element of the χ^2-test is asymptotically serially uncorrelated).

For the second restriction, namely (26), first note that the asymptotics of the estimate drift function of the short rate process defined in (33) can be conveniently established by an application of A3.2.1 and A.3.2.2 on the short rate process $r(t)$, where

$$\zeta\left(r\left(t\right)\right)=\mu\left(r\left(t\right)\right)$$

and

$$\varphi\left(\omega,t\right)=0$$

So the asymptotic variance of $\widehat{\widehat{\mu}}\left(r\left(t\right)\right)$ resulted from sampling from half of the sample, is

$$2\frac{\left(\int_{-\infty}^{+\infty}K\left(u\right)^2du\right)\sigma(r)^2}{h\,\overline{L}\left(t_n,r\right)}.$$

while that of $\widehat{\widehat{m}}(r(t))$ is similarly

$$2\frac{\left(\int_{-\infty}^{+\infty} K(u)^2 \, du\right) \eta(r,\tau)^2}{h\,\overline{L}(t_n, r)}.$$

For other terms of $\widehat{\phi}_1$, first prove that the asymptotic variance of both $\widehat{\xi}(r(t), \tau)$ and $\frac{\partial \widehat{\xi}(t,\tau)}{\partial r}$ are zero (note that all analysis here are only first order asymptotics. The claim does not hold when we consider second order) For $\widehat{\xi}(r(t), \tau)$,

$$\begin{aligned}
& \operatorname{var}\left[\frac{\widehat{\eta}(r,\tau)}{\widehat{\sigma}(r)}\right] \\
= & \left[\frac{1}{\widehat{\sigma}(r)}\right]^2 \operatorname{var}\left[\widehat{\eta}(r,\tau)\right] + \left[\frac{\widehat{\eta}(r,\tau)}{\widehat{\sigma}^2(r)}\right]^2 \operatorname{var}\left[\widehat{\sigma}(r,\tau)\right] \\
& -2\left[\frac{1}{\widehat{\sigma}(r)}\right]\left[\frac{\widehat{\eta}(r,\tau)}{\widehat{\sigma}^2(r)}\right] \operatorname{cov}\left[\widehat{\eta}(r,\tau), \widehat{\sigma}(r,\tau)\right].
\end{aligned}$$

Substituting, using the results from A3.1 and A3.2

$$\begin{aligned}
\operatorname{var}\left[\widehat{\eta}(r,\tau)\right] &= \frac{\Delta t \, \left(\int_{-\infty}^{+\infty} K(u)^2 \, du\right) \eta(r,\tau)^2}{h\,\overline{L}(t_n, r)} \\
\operatorname{var}\left[\widehat{\eta}(r,\tau)\right] &= \frac{\Delta t \, \left(\int_{-\infty}^{+\infty} K(u)^2 \, du\right) \sigma(r)^2}{h\,\overline{L}(t_n, r)} \\
\operatorname{cov}\left[\widehat{\eta}(r,\tau), \widehat{\sigma}(r,\tau)\right] &= \operatorname{std}\left(\widehat{\eta}(r,\tau)\right) \operatorname{std}\left(\widehat{\sigma}(r)\right),
\end{aligned}$$

then

$$\operatorname{var}\left[\widehat{\xi}(r(t), \tau)\right] = 0.$$

As for $\partial\widehat{\xi}(r)/\partial r$ first note that

$$\frac{\partial\left(\widehat{\xi}(r)\right)}{\partial r} = \frac{\partial\left(\widehat{\eta}(r)\right)}{\partial r}\frac{1}{\widehat{\sigma}(r)} - \frac{\partial\left(\widehat{\sigma}(r)\right)}{\partial r}\frac{\widehat{\eta}(r)}{\widehat{\sigma}(r)},$$

but its variance will be dominated by variations of $\partial\widehat{\eta}(r)/\partial r$ and $\partial\widehat{\sigma}(r)/\partial r$, whose asymptotic variance is $1/h^2$ larger than the asymptotic variance of other terms in the right hand side, $\widehat{\eta}(r)$ and $\widehat{\sigma}(r)$ (compare (56) and (53)). By the delta method, then the variance of $\partial\widehat{\xi}(r)/\partial r$ is

$$\begin{aligned}
& \left[\frac{1}{\widehat{\sigma}(r)}\right]^2 \operatorname{var}\left[\frac{\partial\left(\widehat{\eta}(r)\right)}{\partial r}\right] + \left[\frac{\widehat{\eta}(r)}{\widehat{\sigma}(r)}\right]^2 \operatorname{var}\left[\frac{\partial\left(\widehat{\sigma}(r)\right)}{\partial r}\right] \\
& -2\left[\frac{1}{\widehat{\sigma}(r)}\right]\left[\frac{\widehat{\eta}(r)}{\widehat{\sigma}(r)}\right] \operatorname{cov}\left[\frac{\partial\left(\widehat{\eta}(r)\right)}{\partial r}, \frac{\partial\left(\widehat{\sigma}(r)\right)}{\partial r}\right] \\
= & \; 0,
\end{aligned}$$

as a consequence of Theorem 3.

Thus the variance of $\widehat{\widehat{\phi}}_1$ from (26) is driven by variations of $\widehat{\widehat{\mu}}(r)$ and $\widehat{\widehat{m}}(r)$, i.e.:

$$\begin{aligned}
\operatorname{var}\left[\widehat{\widehat{\phi}}_1(r)\right] &= \operatorname{var}\left[\widehat{\widehat{m}}(r)\right] + \widehat{\xi}^2(r)\operatorname{var}\left[\widehat{\widehat{\mu}}(r)\right] \\
&= 2\frac{\Delta t\ \left(\int_{-\infty}^{+\infty} K(u)^2\, du\right)\eta(r,\tau)^2}{h\,\overline{L}(t_n,r)} + \frac{\eta(r,\tau)^2}{\sigma(r)^2}2\frac{\Delta t\ \left(\int_{-\infty}^{+\infty} K(u)^2\, du\right)\sigma(r)^2}{h\,\overline{L}(t_n,r)} \\
&= 4\frac{\Delta t\ \left(\int_{-\infty}^{+\infty} K(u)^2\, du\right)\eta(r,\tau)^2}{h\,\overline{L}(t_n,r)},
\end{aligned}$$

which is what we claimed earlier in section 5.

Proofs for (34) and the χ^2-test now simply follow arguments used in A5.1 and the first Markov restriction.

■

References

[1] Aït-Sahalia, Y. (1996). Nonparametric Pricing of Interest rate Derivative Securities. Econometrica 64, 3, 527-560.

[2] Aït-Sahalia, Y. (1997). Do Interest Rates Really Follow Continuous-Time Markov Diffusions? Working Paper.

[3] Amin, K. I. and A. Morton (1994). Implied Volatility Function in Arbitrage-Free Term Structure Models. Journal of Financial Economics 35, 141-180.

[4] Bandi, F. (1998). Short term Interest Rate Dynamics: A Spatial Approach. Unpublished paper, Yale University.

[5] Bandi, F. (1999). Stanton's (JF, 1997) Approximations to the Infinitesimal First and Second Moments of a Diffusion: Statistical Properties and Implications for Pricing. Unpublished paper, Yale University.

[6] Bandi, F. and T. Nguyen (1999). Fully Nonparametrics Estimations for Diffusions: a Small Sample Analysis. Unpublished paper, University of Chicago and Yale University.

[7] Bandi, F. and P. Phillips (1998). Econometric Estimation of Diffusion models. Unpublished paper, Yale University.

[8] Bliss, R. and P. Ritchken (1996). Empirical Test of Two State-Variable Heath-Jarrow-Morton Model. Journal of Money, Credit and Banking 28, No.3, 452-481.

[9] Brace A., and M. Musiela. A Multifactor Gauss Markov Implementation of Heath, Jarrow, and Morton. Mathematical Finance, 4 (1994), 259-283.

[10] Buhler, W., M. Uhrig-Homburg, U. Walter, and T. Weber (1999). An Empirical Comparison of Forward and Spot-Rate Models for Valuing Interest Rate Option. Journal of Finance.

[11] Cox, J., J. Ingersoll and S. Ross (1985). A Theory of the Term Structure of Interest Rates. Econometrica, 53, 385-406.

[12] Florens-Zmirou (1993). On Estimating the Diffusion Coefficient from Discrete Observations. Applied Probability,790-804.

[13] Hardle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

[14] Heath,D., R. Jarrow, A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. Econometrica 60, 77-105.

[15] Hull, J. and A. White (1990). Pricing Interest Rate Derivative Securities. Review of Financial Studies 3 (4), 573-592.

[16] Jeffrey, A. (1995). Single Factor Heath-Jarrow-Morton Term Structure Models based on Markov Spot Interest Rate Dynamics. Journal of Financial and Quantitative Analysis, 30 (4), 619-642.

[17] Jeffrey, A. (1997). An Empirical Examination of a Path-Dependent Term Structure Model. Working paper, Yale School of Managment.

[18] Jeffrey, A., O. Linton and T. Nguyen (1999). Term Structure Estimation and a Proxy for the Instantaneous Spot Interest Rate". Unpublished paper, Yale University, 1999

[19] Jiang, G. and J. Knight (1997). A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short Term Interest Rate Model. Econometric Theory, 13, 616-645.

[20] Jiang, G. (1998). Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Prices of Derivative Securities, Journal of Financial and Quantitative Analysis, 33(4), pages 465-97..

[21] Linton, O., E. Mammen, J. Nielsen and C. Tanggard (1998). Estimating Yield Curves by Kernel Smoothing Methods. Working paper.

[22] McCulloch, J. H. (1975). The Tax-Adjusted Yield Curve. The Journal of Finance, 30, 811-830.

[23] Musiela, M. and M. Rutkowski (1997). Martingale Methods in Financial Modelling. Springer.

[24] Pagan, A and A. Ullah (1999). Nonparametric Econonometrics. Cambridge University Press.

[25] Phillips, P. and J. Park (1998). Nonstationary Density Estimation and Kernel Autorgression. Working paper. Cowles Foundation for Research in Economics, Yale University.

[26] Revuz and Yor (1999). Continuous Martingales and Brownian Motion. Spinger-Verlarg: New York.

[27] Ritchken, P. and L. Sankarasubramanian (1995). Volatility Structures of Forward Rates and the Dynamics of the Term Structure Mathematical-Finance 5(1), 55-72.

[28] Stanton, R. (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. Journal of Finance 52, No. 5,1973-2002.

[29] Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. Journal of Financial Economics, 5(2), 177-88.

[30] Zhou, A. and N. Pearson (1998) A Nonparametric Analysis of the Forward Rate Volatilities. Working paper, University of Illinois at Urbana-Champaign.

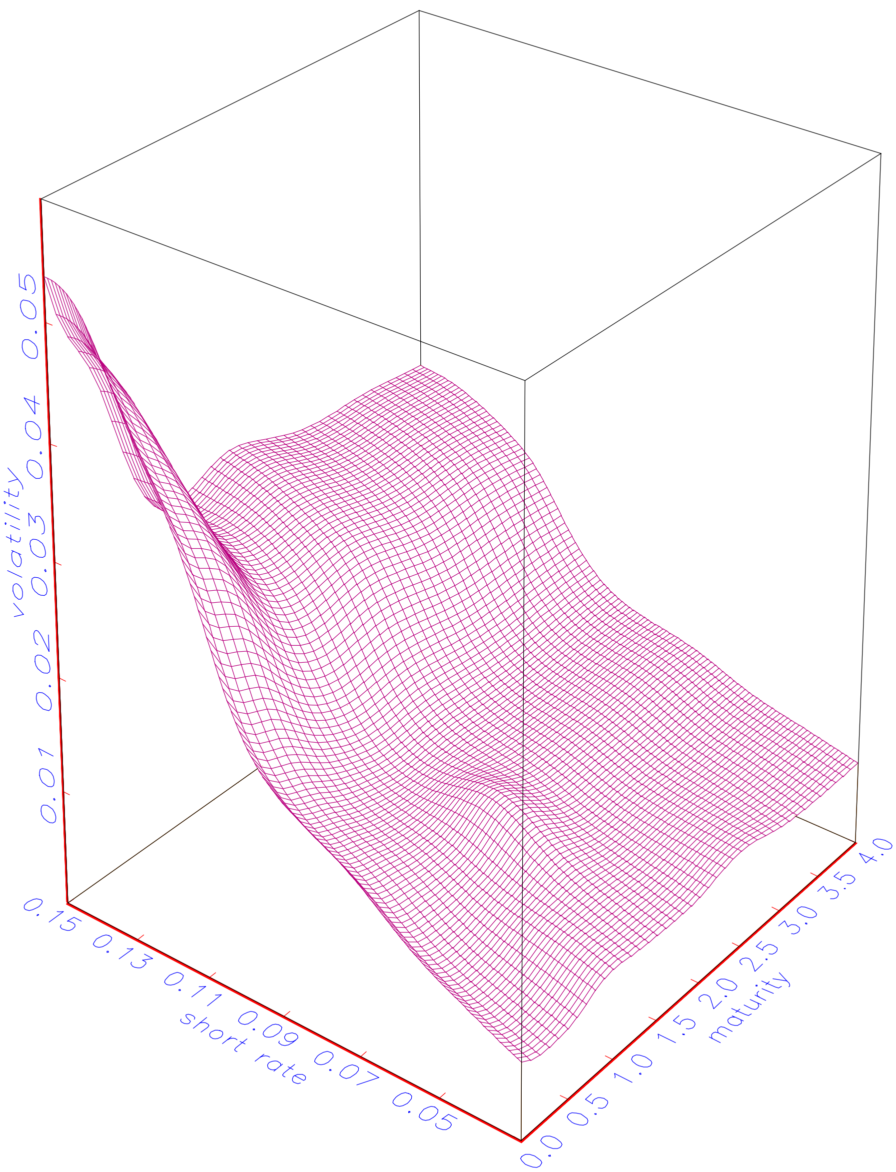

Figure 1: Volatility structure of the yield curves estimated nonparametrically from CRSP bond daily data from January 1961 to December 1998 where the yield curve is extracted by Linton et al. (1998) kernel smoothing based method. Bandwith for the volatility structure estimation is 1%. Maturity is from 0 to 5 years, and the short rate is from 0% to 15%.

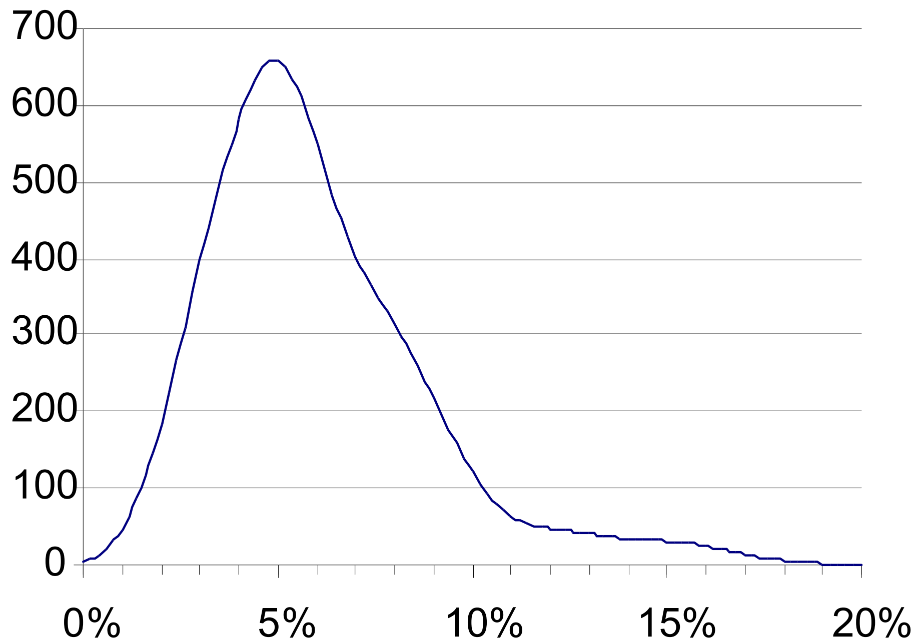

Figure 2: Chronological local time of the short rate estimated from CRSP bond daily data from January 1961 to Dec 1998. Short rates extracted by Linton et al. (1998) kernel smoothing based method. Local time is estimated by formula (47) with bandwith of 1%.

Maturity\Years	**83-98**($6df$)	**70-98**($6df$)	**61-98**($6df$)
1	17.78	1.30	4.23
2	4.27	4.90	3.37
3	2.12	5.50	5.89
4	2.29	8.19	3.72
5	2.31	7.93	
6	2.40		
7	2.38		
8	2.24		
9	2.36		

Table 1: No-arbitrage restriction test results. For construction of the test, 6 interest rate levels are chosen, from 3% to 8%, where most of data are observed (see Figure 2). The longest maturity is period dependent, which results in the table being asymmetric. Critical value for a $\chi^2(6)$ test at 95% confidence level is 12.59.

Maturity\Years	**83-98**$(6df)$	**70-98**$(6df)$	**61-98**$(6df)$
1	1.34	2.99	2.49
2	3.94	5.73	8.07
3	2.40	6.15	7.61
4	2.77	5.10	5.24
5	2.50	4.91	
6	1.85		
7	1.77		
8	1.80		
9	1.62		

Table 2: Univariate Markovian first restriction test. For construction of the test, 6 interest rate levels are chosen, from 3% to 8%, where most of data are observed (see Figure 2). The longest maturity is period dependent, which results in the table being asymmetric. Critical value for a $\chi^2(6)$ test at 95% confidence level is 12.59.

Maturity\Years	**83-98**$(6df)$	**70-98**$(6df)$	**61-98**$(6df)$
1	1.27	2.69	1.02
2	1.57	5.21	4.66
3	1.51	4.25	6.98
4	2.27	3.48	4.20
5	2.39	3.06	
6	2.56		
7	2.35		
8	2.11		
9	1.98		

Table 3: Univariate Markovian second restriction test. For construction of the test, 6 interest rate levels are chosen, from 3% to 8%, where most of data are observed (see Figure 2). The longest maturity is period dependent, which results in the table being asymmetric. Critical value for a $\chi^2(6)$ test at 95% confidence level is 12.59.