Nonparametric Estimation of a Multifactor Heath-Jarrow-Morton Model: An Integrated Approach

By

Andrew Jeffrey, Oliver Linton,
Thong Nguyen, and Peter C.B. Phillips

July 2001

COWLES FOUNDATION DISCUSSION PAPER No. 1311

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Nonparametric Estimation of a Multifactor Heath-Jarrow-Morton model: An Integrated Approach

Andrew Jeffrey*, Oliver Linton†, Thong Nguyen‡, and Peter C.B. Phillips§

May 22, 2001

Abstract

We develop a nonparametric estimator for the volatility structure of the zero coupon yield curve in the Heath-Jarrow-Morton framework. The estimator incorporates cross-sectional restrictions along the maturity dimension, and also allows for measurement errors, which arise from the estimation of the yield curve from noisy data. The estimates are implemented with daily CRSP bond data.

1 Introduction

In this paper we propose a new estimator of the volatility structure in single factor and multifactor models of the term structure. A well known limitation of one factor models, regardless of whether they belong to the more traditional Markovian framework typified by Vasicek (1977) and Cox, Ingersoll, and Ross (1985) or the more recent approach pioneered by Ho and Lee (1983) and Heath, Jarrow and Morton (1992) (HJM hereafter) − which results in non-Markovian behavior of the interest rates − is that they poorly capture empirical dynamics either of short rates (the most commonly used state variable) or the whole term structure. This limitation has motivated the study of multi-factor models, especially in practical work. Two factor models with a rich variety of second factor specifications have been examined in the literature. The short term interest rate is usually the

*Yale School of Management, PO Box 208200, New Haven CT. 06520-8200. Phone: (203) 432 6029. Fax: (203) 432 6970. E-mail: Andrew.Jeffrey@yale.edu

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: lintono@lse.ac.uk.

‡Bank of America Securities, San Francisco, CA 94111, U.S.A. E-mail: thonghnguyen@yahoo.com

§Cowles Foundation, Yale University. E-mail: peter.phillips@yale.edu

first factor; some frequently chosen second factors include the long rate, the spread between long rates and short rates, inflation, central tendency (the long term mean of short term interest rates), and volatility of short term interest rates. Examples of two factor models can be found in Balduzzi, Das, Foresi and Sundaram (1998), Brennan and Schwartz (1979), Duffie and Singleton (1997), Knight, Li and Yuan (1999), Litterman and Scheinkman (1991), Longstaff and Schwartz (1992), Schaefer and Schwartz (1984), Pearson and Sun (1994), among others. More flexible three factor models and beyond are also studied by Boudoukh, Richardson and Stanton (1998), Balduzzi, Das, Foresi and Sundaram (1996), Chen and Scott (1993). A survey of multi-factor models can be found in Backus, Foresi and Telmer (1998).

Recent theoretical and empirical studies of non-Markovian models inside the Heath-Jarrow-Morton framework have also favored the multi-factor specification to enable richer dynamics, see for example Bliss and Ritchken (1996), Buhler, Uhrig-Homburg, Walter and Weber (1999), Inui and Kijima (1998), Jong and Santa-Clara (1999), Pearson and Zhou (1999).

Moving away from the simple one factor framework introduces extra analytical complexity. In consequence, out of mathematical convenience, most models so far studied are affine. That is, they have drift and diffusion functions defined as linear functions of the state variable(s), as proposed in Vasicek (1977) and typified in the work of Cox, Ingersoll, Ross (1985). The affine structure by itself imposes heavy restrictions on the specified dynamics, besides the fact that most affine models impose additional restrictions because they fall short of being “maximal” in the terminology of Dai and Singleton (1998).[1] Nonlinear models, such as Chan, Karoly, Longstaff and Saunder (1992) (CKLS hereafter), and more recently, nonparametric models as in Stanton (1997), are more flexible. At a cost of some loss in tractability, these models do capture the dynamics of the term structure better – see CKLS for an empirical demonstration to this effect. However, there are still restrictions that are imposed through seemingly innocuous specifications. For example, in the multi-factor Markovian model discussed in Knight, Li and Yuan (1999), only one Brownian motion drives the dynamics of each factor, or in the multi-factor non-Markovian model in Pearson and Zhou (1999), there is just one Brownian motion that drives the whole forward curve, although the volatility structure depends on more than one state variable.

In this study, we propose a flexible multi-factor generalized Heath-Jarrow-Morton model that encompasses most HJM models proposed so far in the literature. Our model is general in various ways. The nonparametric specification allows the functional forms to be minimally restricted. We also allow all volatility structures associated with each Brownian motion to depend on the whole set

[1] “Maximal” essentially means that the researcher does not impose any implicit and/or explicit restriction on the model beyond the affine structure. In practice, however, additional assumptions are often invoked. For instance, it is very common to assume that factors are driven by a set of orthogonal Brownian motions.

of specified state variables, departing from a common practice of associating each Brownian motion with a separate state variable. Most importantly, contrary to the usual approach of introducing measurement errors as an afterthought to fix the so-called *stochastic singularity* that is inherent in HJM models, our model directly incorporates measurement errors when the dynamics of the yield curves are initially specified. To our knowledge, this is the first study to analyze estimation issues of a multi-factor HJM model in a rigorous fashion.

The plan of the paper is as follows. In the next section, the prototype one factor model is specified and the estimator and its asymptotic properties is developed. We extend the methodology to cover the multi-factor case in section 3. Section 4 is devoted to the empirical estimation of a one factor and two factor model using US bond data. Section 5 concludes the paper. All technical derivations are given in the appendix.

2 One factor Nonparametric HJM model

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. The instantaneous forward rate at time t for date T, denoted $f(t,T)$, is defined by $f(t,T) = -\partial \ln(P(t,T))/\partial T$ and the yield at time t with maturity date T, denoted $y(t,T)$, is defined in terms of forward rates by $y(t,T) = -\frac{1}{T-t}\int_t^T f(t,v)dv$. The HJM framework represents the term structure in terms of forward rates. However, determining forward rates in practice via curve fitting procedures often proves to be sensitive to the method adopted. Estimation of yields is typically less sensitive to the method used, intuitively because yields are averages of forward rates. Consequently, we choose to portray the term structure using yields. Within the HJM framework where a single Wiener process $W(t)$ introduces uncertainty into the bond market, the uncertain evolution of each yield with fixed maturity date T is characterized by the stochastic differential equation

$$dy(t,T) = \alpha_y(\omega,t,T)dt + \gamma(\omega,t,T)dW(t),$$

where $W(t)$ is a one dimensional Brownian motion, while 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 just before time t generated by the term structure's evolution. We shall restrict our attention to volatility functions of the form

$$\gamma(\omega,t,T) = \gamma(r(t),T-t) = \gamma(r(t),\tau)$$

for some function $\gamma(.)$ that is smooth but otherwise of unknown functional form. This class of models is vast, covering most of the HJM specifications proposed so far in the literature. However, the technique introduced later is not restricted to this chosen state variable. Our specification is obviously nonparametric one with no restriction on the functional form, according to the usual convention.

As shown in the original HJM paper, the no arbitrage restriction dictates that the drift function of the forward curve evolution is just a function of the volatility structure if we are in risk neutral world. For yield curve evolution, this no arbitrage restriction also imposes that the drift function is a function of the yield volatility structure. specifically

$$\alpha_y(\omega, t, \tau) = \frac{\partial y(t,\tau)}{\partial \tau} + \frac{y(t,\tau) - r(t)}{\tau} + \frac{1}{2}\,\tau\,\gamma(\omega, t, \tau)^2;$$

see Jeffrey, Linton, and Nguyen (1999b) for a derivation of the above equation. Together with the knowledge of the market price of risk, the dynamics of the yield curve under the real world measure can be recovered. A common objective is to use the above dynamics to price fixed income instruments, however, and risk neutral martingale pricing only requires the dynamics in a risk neutral world, since the drift is simply a function of the volatility structure in a risk neutral world (it is not necessary to know the market price of risk in this world). So, for pricing based on an HJM model such as that considered here, we only need to estimate the volatility structure.

Using only observations from a time series of $y(.,\tau)$ with fixed τ, we can obtain an estimate of the volatility function $\gamma(r(t),\tau)$ as shown in a recent paper by Jeffrey, Linton, and Nguyen (1999b). Their estimator was

$$\widehat{\gamma}_1(r\,(t)\,,\tau)^2 = \frac{1}{\Delta_n}\frac{\sum_{i=1}^n K_{h_r}(r - r_i)\Delta y(t_i,\tau)^2}{\sum_{i=1}^n K_{h_r}(r - r_i)}, \qquad (1)$$

where $K_h(\cdot) = K(\cdot/h)/h$ and $K(\cdot)$ is a symmetric probability density, while h_r is a bandwidth sequence. Under various regularity conditions [including that $h_r \to 0$ and $nh_r \to \infty$] we have

$$\sqrt{\frac{h_r\,\overline{L}\,(t_n, r)}{\Delta t}}\left(\widehat{\gamma}_1(r,\tau)^2 - \gamma(r,\tau)^2\right) \xrightarrow{d} N\left(0, 4\,\|K\|^2\,\gamma(r,\tau)^4\right)$$

as $n \to \infty$, where $\|K\|^2 = \int_{-\infty}^{+\infty} K\,(u)^2\,du$, while $\overline{L}\,(t_n, r)$ is the chronological time of the process r, a concept that will be discussed in more detail later. The estimator can be considered as an extension to HJM models of the approach pioneered by Stanton (1977) and Jiang and Knight (1998) for Markovian interest rate models.

In the original HJM models, a finite set of Brownian motions serve as the source of randomness in the economy and drive the dynamics of the whole yield curves. This construction helps to make the market complete, but at the cost of inducing *stochastic singularity* into the model: we can easily find some linear combination of the dynamics of points along the yield curves that is deterministic, i.e., without the presence of the Brownian motion shocks.[2] Empirical data, of course, almost surely violate this relationship. To reconcile this internal consistency, other sources of randomness, such as measurement errors must be introduced. Generalized models of HJM are also motivated to overcome this undesirable inconsistency. In Kennedy (1994) and Kennedy (1997), the source of the shocks in the economy is a *Brownian sheet* and *random field,* respectively; or in more recent work by Santa Clara and Sornette (2000), *string shocks* are employed. Estimation issues with these new models are very challenging, unfortunately, and, perhaps in consequence, empirical studies with these new models have yet to be conducted.

In the following, we will introduce a new generalized model that is flexible enough to do away stochastic singularity, while retaining tractability and facilitating estimation. Our model is

$$dy(t_i, T_j) = \alpha_y(\omega, t_i, T_j)dt + \gamma(r(t_i), \tau_j)dW(t_i) + \sigma_\varepsilon dW_j(t_i), \tag{2}$$

where $(W_j)_{j=1}^J$ is a family of independent standard Brownian motions, independent of the standard Brownian motion W. The new family of $(W_j)_{j=1}^J$ will act as the extra source of shocks to the yield curves; they can be interpreted as measurement errors, which can arise from a variety of sources, a notable one being that the data used in the procedure were obtained from some preliminary *yield curve fitting* procedure such as splines in McCulloch (1971, 1975), bootstrapping as in Fama and Bliss (1987) or kernel smoothing as in Linton, Mammen, Nielsen and Tanggaard (1999).[3]

In contrast to more radical approaches such as that of Kennedy (1994, 1997), where the new source of shocks in the economy is completely different from the finite set of Brownian motions in the original HJM papers, the model proposed above can be considered as a stochastically extended version of the original HJM model. The model aims to incorporate measurement or observation errors while preserving as much as possible the HJM's shock structure. Put differently, the common Brownian motion(s) that drive the whole yield curve dynamics still provide the only source of economic uncertainty embedded in the underlying model, while the idiosyncratic shocks to each

[2] For example, in the one-factor model above, we can have the following deterministic relationship for any two yield dynamics:

$$dy(t, T_1)\gamma(\omega, t, T_2) - \alpha_y(\omega, t, T_1)\gamma(\omega, t, T_2) = dy(t, T_2)\gamma(\omega, t, T_1) - \alpha_y(\omega, t, T_2)\gamma(\omega, t, T_1)dt$$

[3] Zero-coupon yield curves are not observable with traded instruments in the fixed-income markets because bonds with time to maturity greater than 1 year are normally coupon-bearing bonds. So the (zero coupon) yield curves must be extracted by yield curve fitting procedures such as McCulloch (1971, 1975)'s splines. In any case, the data used to carry out these procedures do not correspond exactly to the theoretical price. Specifically, we usually observe quotes of bid and ask prices obtained from a telephone survey of the registered bond dealers. Also, there are tax difference and liquidity effects that can be interpreted as providing random errors in the observed bond prices.

point of the yield curves are statistical noise. The underlying economy is therefore still a complete market, where any instrument is hedgeable, while the observed economy is not necessarily complete any more. We thus take a half way approach between the original HJM models and that of Kennedy (1994, 1997) and Santa Clara and Sornette (2000).

The familiar homoskedastic orthogonal measurement error structure can be readily justified as an econometrically convenient device that is widely used, although relaxing this simple structure is quite feasible for our technique. This structure can also arise from the asymptotic properties established in LMNT (1999), that is, the error term is assumed to come from the kernel estimation procedure used in that paper. The error variance is small, and goes to zero as the number of observations used in extracting the single period term structure goes to infinity. The uncorrelatedness assumption is justified provided the grid of maturity points is either fixed or increases but at a slow enough rate so that the underlying kernel estimators are independent. We suppose that the grid $\{\tau_j\}_{j=1}^J$ becomes dense on the maturity interval as J is assumed to increase at a certain rate.

3 Estimation Method

In this section we introduce our estimation techniques based on the yield curve evolution. As argued in Jeffrey, Linton, Nguyen (1999b), we prefer to use yields instead of forwards since the former, which is an average of the latter, are easier to estimate. However, the estimator developed here can be adapted for forward curve evolution with minor modification.

Suppose we observe the yield curves at n points in time t_i, $i = 1, \ldots, n$. At any time t_i, we observe a random number, J_i, of points along the yield curve $\{y(t_i, T_1), \ldots, y(t_i, T_{J_i})\}$, where T_j is the time to maturity. For ease of notation, we will use the Musiela parameterization $\tau = T - t$; we will also drop the subscript i in J_i and assume for simplicity that we observe the yield curve at the same maturities for each time point. We shall further suppose that the observed maturities $\tau_1, \ldots, \tau_J$ can be thought of as being random draws from some continuous density function p. Let $\Delta y(t_i, \tau) = y(t_{i+1}, \tau) - y(t_i, \tau)$. For simplicity it is also assumed that all time intervals t_i to t_{i+1} are equally spaced, that is, $\Delta t_i = t_{i+1} - t_i \equiv \Delta t$ for all i and consequently $t_n = n\Delta t$.

It is straightforward to see from the characteristics of the driving Brownian motion that

$$E\left(dy(t, T_j)dy(t, T_k)\,|\mathcal{F}_{t-}\right) = \begin{cases} \gamma(r(t), \tau_j)\gamma(r(t), \tau_k)dt & \text{if } j \neq k \\ \\ [\gamma(r(t), \tau_j)^2 + \sigma_\varepsilon^2]dt & \text{if } j = k. \end{cases}$$

We will use these restrictions to generate estimating equations for the unknown volatility function

γ. Stacking all the cross-yield restrictions together we have at any time t

$$\frac{1}{\Delta_n}E\begin{pmatrix} \Delta y(t,\tau_1)^2 & \Delta y(t,\tau_1)\Delta y(t,\tau_2) & \cdots & \cdots \\ \Delta y(t,\tau_1)\Delta y(t,\tau_2) & \Delta y(t,\tau_2)^2 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \Delta y(t,\tau_1)\Delta y(t_1,\tau_J) & \Delta y(t,\tau_2)\Delta y(t_1,\tau_J) & \cdots & \Delta y(t,\tau_J)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \gamma^2(r,\tau_1)+\sigma_\varepsilon^2 & \gamma(r,\tau_1)\gamma(r,\tau_2) & \cdots & \cdots \\ \gamma(r,\tau_1)\gamma(r,\tau_2) & \gamma^2(r,\tau_2)+\sigma_\varepsilon^2 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \gamma(r,\tau_1)\gamma(r,\tau_J) & \gamma(r,\tau_2)\gamma(r,\tau_J) & \cdots & \gamma^2(r,\tau_J)+\sigma_\varepsilon^2 \end{pmatrix} \tag{3}$$

for a given set of maturity points $\tau_1, \ldots, \tau_J$. Letting $J \to \infty$ we obtain a population criterion function that the volatility structure $\gamma(r,\tau)$ and measurement error variance σ_ε^2 must satisfy, namely

$$\int E\left[\Delta y(t,\tau)\Delta y(t,\tau') - \gamma(r(t),\tau)\gamma(r(t),\tau)\Delta\right]^2 p(\tau)p(\tau')d\tau' d\tau, \tag{4}$$

where $p(.)$ is the marginal density distribution of maturity times. A modified sample version of this function, which ignores information along the diagonal, is proportional to

$$\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1 \\ k\neq j}}^{J}\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) - \gamma(r(t_i),\tau_j)\gamma(r(t_i),\tau_k)\Delta_n\right]^2, \tag{5}$$

which depends on the function γ but not on σ_ε^2. We deliberately do not use the diagonal restriction to avoid the need of estimating σ_ε^2 in this sample version. So, contrary to the approach in a related study, Jeffrey, Linton, Nguyen (1999b), where only information along the diagonal of the moment condition matrix is used, we here use only information from the off-diagonal elements. However, the resulting volatility structure estimate may be used together with the diagonal information afterwards to back out or make inference about σ_ε^2.

It is instructive to digress here for a discussion on the (probably implicit) difference between the Markovian and non-Markovian approach with regards to measurement errors. In the former approach, it is often the case that structural parameters controlling the dynamics of the short term interest rates (the often used state variables) are estimated, with the (implicit) assumptions that the short rates are observed without errors. Then, to reconcile the fact that other observed yields with maturity longer than zero are not perfectly correlated, which they should be, due to being driven by the same state variables, it is necessary to introduce observation errors for other yields. So in effect, we would "shift" all observation errors toward other yield levels, and make a (perhaps unrealistic or unreasonable) assumption that we could observe the short rate with utmost accuracy.

In the HJM framework, there is also an inevitable need to introduce observation errors if we are to reconcile for the less than perfect correlation observed in yield data across maturity dimension.[4] It is interesting to point out the fact that from our analysis, provided our specification is reasonably correct, one can see the potentially severe deficiency of not using the cross-sectional restriction along the maturity dimension in estimating an HJM model: the estimates thus may be severely biased due to the presence of the measurement errors.[5]

The localized version of the above criterion function is

$$\int \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \left[\Delta y(t_i, \tau_j)\Delta y(t_i, \widetilde{\tau}_k) - \gamma(\widetilde{r}_i, \widetilde{\tau}_j)\gamma(\widetilde{r}_i, \widetilde{\tau}_k)\Delta_n\right]^2 K_{h_r}(r_i - \widetilde{r}_i) K_{h_\tau}(\tau_j - \widetilde{\tau}_j) K_{h_\tau}(\tau_k - \widetilde{\tau}_k) d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k,$$

which is to be minimized with respect to functions γ. We will use the delta method to solve for the

[4] One possible approach, which is somewhat comparable to the common approach in the Markovian literature (see Duffie and Singleton (1997), among others), is to assume that we can observe a carefully-chosen number of points along the yield curves without measurement errors. For instance, for a two-factor HJM model, assume that we can observe 3 points with different maturities along the yield curves which we will conveniently assume to be observed without measument errors. Second-moment restriction on the dynamics of forward rates and cross-restrictions are:

$$\begin{aligned}
E\left[df\left(t, T_1\right)^2\right] &= \eta(r\left(t\right), l\left(t\right), T_1)^2 + \gamma(r\left(t\right), l\left(t\right), T_1)^2 = Y_{11} \\
E\left[df\left(t, T_2\right)^2\right] &= \eta(r\left(t\right), l\left(t\right), T_2)^2 + \gamma(r\left(t\right), l\left(t\right), T_2)^2 = Y_{22} \\
E\left[df\left(t, T_3\right)^2\right] &= \eta(r\left(t\right), l\left(t\right), T_3)^2 + \gamma(r\left(t\right), l\left(t\right), T_3)^2 = Y_{33} \\
E\left[df\left(t, T_1\right) df\left(t, T_2\right)\right] &= \eta(r\left(t\right), l\left(t\right), T_1)\eta(r\left(t\right), l\left(t\right), T_2) + \gamma(r\left(t\right), l\left(t\right), T_1)\gamma(r\left(t\right), l\left(t\right), T_2) = Y_{12} \\
E\left[df\left(t, T_1\right) df\left(t, T_3\right)\right] &= \eta(r\left(t\right), l\left(t\right), T_1)\eta(r\left(t\right), l\left(t\right), T_3) + \gamma(r\left(t\right), l\left(t\right), T_1)\gamma(r\left(t\right), l\left(t\right), T_3) = Y_{13} \\
E\left[df\left(t, T_2\right) df\left(t, T_3\right)\right] &= \eta(r\left(t\right), l\left(t\right), T_2)\eta(r\left(t\right), l\left(t\right), T_3) + \gamma(r\left(t\right), l\left(t\right), T_2)\gamma(r\left(t\right), l\left(t\right), T_3) = Y_{23}
\end{aligned}$$

The intuition here is that we have 6 functions of $\widehat{\eta}(r\left(t\right), l\left(t\right), T_j)$ and $\widehat{\gamma}(r\left(t\right), l\left(t\right), T_j)$, $j = 1, 2, 3$, so we need 6 moment conditions to exactly identify these functions. In general, if we have an $M-$factor model, we would need (at least) N points along the yield curves to generate $N + N * (N-1)/2 = N(N+1)/2 \geq M * N$ moment conditions so the model will be exactly (or over) identified ($M * N$ will be the number of unknown), i.e. $N = 2M - 1$. For a possible empricial implementation, since three factor models are well-documented to be able to capture the empirical behavior of the US bond market, the number of points we would look at will be $2 * 3 - 1 = 5$ for exact identification.

This approach is similar to that of the Markovian literature, in the sense that we assume that there are some "special" points which are observed without errors, and "shift" all observation errors to other points. The approach is however not very appealing for a number of reasons, one being that there is no natural candidate for these N points.

[5] For example, in our previous paper, if the new model with measurement errors is more appropriate, then the previous estimator, which effectively use only information along the diagonal, will have the asymptotics as

$$\left\{\left(\widehat{\gamma}(r\left(t\right), \tau)^2\right) - \left(\gamma(r\left(t\right), \tau)^2 + \sigma_\varepsilon^2\right)\right\} \xrightarrow{d} N\left(0, \frac{2\left(\gamma_1(r\left(t\right), \tau)^4\right)}{\frac{1}{\Delta_{n,t_n}}\sqrt{\pi} h_{n,T} \overline{L}\left(t_n, r(t)\right)} + O\left(\sigma_\varepsilon^4\right)\right).$$

It is apparent that the bias can be severe if the variance of the measurement errors is relatively large.

functional solution, $\widehat{\gamma}(r,\tau)$. The first order condition, obtained by differentiation with respect to the function $\widehat{\gamma}(.,.)$ as shown in the appendix, is

$$
\begin{aligned}
0 &= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r) \\
&\quad\times\left[K_{h_\tau}(\tau-\tau_j)\int\widehat{\gamma}(r,s)K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\widehat{\gamma}(r,s)K_{h_\tau}(s-\tau_j)ds\right] \\
&\quad-\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\widehat{\gamma}(r,\tau)K_{h_r}(r-r_i) \qquad (6)\\
&\quad\times\left[K_{h_\tau}(\tau-\tau_j)\int\widehat{\gamma}(r,s)^2K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\widehat{\gamma}(r,s)^2K_{h_\tau}(s-\tau_j)ds\right]\Delta_n.
\end{aligned}
$$

The function $\widehat{\gamma}\equiv 0$ is always a solution to (6). The solution $\widehat{\gamma}(r,\tau)$ can also be written (assuming $0/0=0$ in the pathological case) as

$$
\widehat{\gamma}(r,\tau)=\frac{\int\widehat{H}_1(r,\tau,s)\,\widehat{\gamma}(r,s)ds}{\int\widehat{H}_2(r,\tau,s)\,\widehat{\gamma}^2(r,s)ds}, \qquad (7)
$$

where

$$
\begin{aligned}
\widehat{H}_1(r,\tau,s) &= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r) \\
&\quad\times\left[K_{h_\tau}(\tau-\tau_j)K_{h_\tau}(s-\tau_k)+K_{h_\tau}(\tau-\tau_k)K_{h_\tau}(s-\tau_j)\right]
\end{aligned}
$$

and

$$
\widehat{H}_2(r,\tau,s)=\frac{1}{J^2}\Delta_n\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r_i-r)\left[K_{h_\tau}(\tau-\tau_j)K_{h_\tau}(s-\tau_k)+K_{h_\tau}(\tau-\tau_k)K_{h_\tau}(s-\tau_j)\right].
$$

The expression (7) is a nonlinear integral equation involving the linear operators $\widehat{H}_1$ and $\widehat{H}_2$. These quantities depend only on observed data and indeed are just kernel weighted sample averages. Relation (7) suggests the following iteration for the calculation of $\widehat{\gamma}(r,\tau)$:

$$
\begin{aligned}
\widehat{\gamma}^{[0]}(r,\tau) &= \pi(r,\tau) \\
\widehat{\gamma}^{[a+1]}(r,\tau) &= \frac{\int\widehat{H}_1(r,\tau,s)\,\widehat{\gamma}^{[a]}(r,s)ds}{\int\widehat{H}_2(r,\tau,s)\left(\widehat{\gamma}^{[a]}(r,s)\right)^2ds},\ a=0,1,\ldots,
\end{aligned}
$$

for some given starting value π. The integrations in $\widehat{\gamma}^{[a+1]}(r,\tau)$ are unidimensional and can be computed numerically. We discuss this further in the application section below.

Our iterative method is called successive approximation. For a detailed discussion we refer the reader to Kantorovich and Akilov (1964) and Luenberger (1969). See also Hastie and Tibshirani (1990), Mammen, Linton, and Nielsen (1999) and Linton, Mammen, Nielsen, and Tanggaard (2000) for related computations.

4 Asymptotic Properties in the Single Factor Case

Before proceeding with the development of the asymptotic distributions of the estimator defined above, we first introduce some notation and assumptions. Consider a semi-martingale defined as follows

$$dr(t) = \mu(r(t))dt + \sigma(r(t))\,dW(t), \tag{8}$$

where sufficient regularity conditions are assumed to ensure the above stochastic differential equation has a strong solution.

Definition *(Chronological Local Time. See Phillips and Park (1998)). The chronological local time of the semi-martingale r defined by (8) at point a over the time interval $[0,t]$ is defined as*

$$\begin{aligned}\overline{L}(t,a) &= \frac{1}{\sigma^2(a)}\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t \mathbf{1}_{|r(s)-a|<\epsilon}\sigma(r(s))^2 ds \\ &= \frac{1}{\sigma^2(a)}L(t,a),\end{aligned}$$

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 (8) with quadratic variation process $\langle r,r\rangle_s$, and for every Borel function f* of r

$$\int_0^t f(r(s))\,d\langle r,r\rangle_s = \int_{-\infty}^{+\infty} f(a)L(t,a)da,$$

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, which will be used repeatedly in our proofs

$$\begin{aligned}\int_0^t f(r(s))\,ds &= \int_0^t \frac{f(r(s))}{\sigma^2(r(s))^2}d\langle r,r\rangle_s \\ &= \int_{-\infty}^{+\infty} f(a)\overline{L}(t,a)da\end{aligned} \tag{9}$$

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

$$\lim_{h\downarrow 0}\frac{1}{h}\int_0^t K\left(\frac{x-r(s)}{h}\right)f(r(s))\,ds = \lim_{h\downarrow 0}\frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{x-a}{h}\right)p(a)\overline{L}(t,a)da \tag{10}$$

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

Assumptions that must be imposed to study the asymptotic distribution of the estimator are the following:

Assumption A1. (Recurrence). *The process $\{r(t);t\geqslant 0\}$ defined in (8) 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 A2. (Boundedness). *The drift and diffusion functions are (locally) bounded*

$$
|\alpha(\omega,t_i)-\alpha(\omega,t_j)| + |\eta(r(t_i))-\eta(r(t_j))| \leq C\,|y(t_i)-y(t_j)| \tag{11}
$$

for constants C, and there exists some $0<\nu<1/2$ such that

$$
\frac{(\Delta_n)^{\nu}}{\overline{L}(t,a)}\int_0^{t_n} |K_h(r-r(s))\alpha(\omega,s)ds| = O(1). \tag{12}
$$

Assumption A3. (Sampling Conditions along the time dimension). *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 $\Delta_n\to 0$ as $n\to\infty$, which is referred to as the 'infill assumption'. The time span of observations $t_{n+1}=n\Delta_n\to\infty$ as $n\to\infty$, referred to as the 'long span assumption'.*[6] *The bandwidth parameter $h_n\downarrow 0$ in such a way that $h_n/\Delta_n\to\infty$ 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 (8); $\overline{L}(t,a)$ is the chronological local time of the process $\{r(t);t\geqslant 0\}$.*

Assumption A4. *The kernel K has compact support($[-C_1,C_1]$, say), is symmetric about zero, and is continuously differentiable.*

As shown in the appendix, the estimator $\widehat{\gamma}(r,\tau)$ will have the asymptotic behavior stated formally in the following theorem:

Theorem 1. *Suppose that assumptions A.1-A4 hold. Then,*

$$
J\sqrt{\frac{h_r h_\tau \overline{L}(t_n,r)}{\Delta_n}}\left[\widehat{\gamma}(r,\tau)-\gamma(r,\tau)\right] \xrightarrow{d} N\left(0, 2\|K\|^4\,\frac{\gamma(r,\tau)^2\int\gamma(r,s)^4 p(s)\,ds}{p(\tau)\left(\int\gamma(r,s)^2 p(s)\,ds\right)^2}\right).
$$

Remark 1. Our approach is based on a combination of some recent advances in the nonparametric literature. Along the maturity dimension, we draw on the approach put forward by Linton, Mammen, Nielsen, and Tanggaard (1999), while information from the evolution of interest rates over time is

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

used in a fashion similar to the nonparametric estimations of stochastic diffusion processes proposed by Florens-Zmirou (1993), Stanton (1997), Jiang and Knight (1998), and Bandi and Phillips (1999).

REMARK 2. This model will cover all one factor HJM models that have been employed in the past literature, such as the one factor model in our previous paper, Jeffrey, Linton, and Nguyen (1999b). To cover the HJM model proposed in Pearson and Zhou (2000), we simply need to replace the kernel function $K_{h_r}(r_i - \widetilde{r}_i)$ that we use above by a two dimensional kernel $K_{h_r}((r_i, s_i) - (\widetilde{r}_i, \widetilde{s}_i))$. Note that, in contrast to the multi-factor model discussed in the next section, their model is very restrictive in the sense that the randomness is introduced into the model by just one Brownian motion, although the volatility is extended to a function of two variables. Their model can also be thought of as a special case of our two factor model, where the volatility structure on the second Brownian motion is simply zero.

5 Multi-Factor Models

Multi-factor models, both path independent and path dependent, have been widely studied in the literature, in an attempt to improve the fit of the model when measured against empirical data, for instance capturing dynamics of the observed short term interest rate.[7] Balancing between the flexibility provided by more factors and potential overfitting problems plus losing analytical tractability associated with more factors, researchers have commonly proposed two/three factor models, as cited earlier.[8]

The above approach could be extended to the multi-factor case in a straightforward fashion, at least for the step of setting up the criteria function. Derivation of the solution will quickly become very burdensome and tedious, as demonstrated for the two factor case later, although the basic idea is rather straightforward. However, complications that are introduced into the estimation procedures are not in any sense trivial and will have consequences on asymptotic behavior. The most difficult hurdle, in principle, arises from the non existence of local time in high dimensional cases for Brownian

[7] The practical implication of this particular limitation on pricing fixed-income assets is arguably debatable. Failure to capture the dynamics of the short rates accurately does not necessarily impair the ability to price fixed-income assets. For instance, Buser, Hendershott and Sanders (1990), Hull and White (1990) claim that one-factor interest rate models with flexible specification can generate interest-rate derivatives prices similar to those of 2-factor models. Nevertheless, even if their thesis is valid, the search for models that fit empirical dynamics of the underlying consensus is still of great interests, at least for internal consistency.

[8] More radical alternative aimed at accomodating more complex behaviors of the term structure's dynamics is to resort to stochastic processes different from Brownian motions to present shocks to the underlying economy. Among them, Levy process, fractional Brownian motion, and in particular, jump-diffusion process have become favorite candidates in many studies.

motion. In particular, for a d-dimensional Brownian motion with $d \geq 3$, local time does not exist and this makes a general solution using our approach challenging if not impossible. For the case of $d = 2$, although the local time does exist, its nonparametric estimate converges at a $\log(n)$ rate, which is too slow for reliable inference. In general, no parametric estimates have been proposed for a general diffusion process with $d \geq 2$, and most cases that are considered in the literature are severely restrictive – see Brugiere (1991) and Knight, Li and Yuan (1999) for related studies.

To proceed in light of these considerations, our strategy here is to make some additional assumptions. First we will be content that our volatility functions are of multiplicatively separable form, i.e.,

$$\gamma(\omega, t, T) = \gamma(\omega, t, T - t) = \prod_{m=1}^{M} \gamma(x_m(t), \tau)$$

where the set of $x_m, m = 1, \ldots, M$ are the state variables. We also assume that the local time can be decomposed into

$$\overline{L}(t_n, (x_m, m = 1, \ldots, M)) = \prod_{m=1}^{M} \overline{L}(t_n, x_m),$$

and since the one dimensional process $x_m(t)$ can be fairly assumed to be recurrent, $\overline{L}(t_n, x_m), m = 1, \ldots, M$ is well defined in this case, and thus the existence of $\overline{L}(t_n, (x_m, m = 1, \ldots, M))$ becomes viable. Convergence properties can be established under these extra assumptions, as shown below. We can now develop an estimate for a volatility structure of any m-factor HJM model, although the algebra will become excruciatingly involved, as evident from the calculations below, while the curse of dimensionality is likely to make convergence issue of any model with $m \geq 2$ problematic.

For concreteness with consideration of expositional ease, we will analyze a two factor HJM model in the following. The model, similar to the above setup, is

$$dy(t_i) = dy(t_i, T) = \alpha_y(\omega, t_i, T_j)dt + \gamma(\omega, t_i, T_j)dW_1(t) + \eta(\omega, t_i, T_j)dW_2(t_i) + \sigma_\varepsilon dW_j(t_i), \quad (13)$$

where again $(W_j)_{j=1}^J$ is a family of independent standard Brownian motions serving as measurement errors or statistical noise, independent of the standard Brownian motions W_1 and W_2, which serve as the stochastic shocks to the yield curves. We shall also restrict the volatility function to the form:

$$\begin{aligned} \gamma(\omega, t, T) &= \gamma(r(t), l(t), T - t) = \gamma(r(t), l(t), \tau) \\ \eta(\omega, t, T) &= \eta(r(t), l(t), T - t) = \eta(r(t), l(t), \tau), \end{aligned}$$

where the new extra state variables $l(t)$, can be the long term rate $l(t)$ along the line of Brennan and Schwarz; the forward rate itself $f(t, \tau)$; or the spread between long term rate and short term rate, $l(t) - r(t)$.

As mentioned earlier, the analysis in Dai and Singleton (1999) demonstrates that almost all (Markovian) affine models examined in the literature implicitly or explicitly impose some overidentifying restrictions on an underlying "maximal" model. This practice is also applicable to non affine models and non-Markovian models as well. For instance, Knight, Li and Yuan (1999)'s multi-factor model has severe restriction embedded in their specification. In our model, we have retained generality as far as possible, by allowing all volatility structures to be function of all state variables (albeit we still invoke independence of the Brownian motions), rather than simpler specifications such as

$$\begin{aligned}\gamma(\omega,t,T) &= \gamma(r(t),l\left(t\right),T-t)=\gamma(r(t),\tau),\\ \eta(\omega,t,T) &= \eta(r(t),l\left(t\right),T-t)=\eta(l(t),\tau),\end{aligned}$$

where each volatility depends on just a single state variable.

The moment conditions can be derived are

$$\begin{aligned}&E\left(dy(t,T_j)dy(t,T_k)\left|\mathcal{F}_{t-}\right.\right)\\ &= \begin{cases}\left[\gamma(r\left(t\right),l\left(t\right),\tau_j)\gamma(r\left(t\right),l\left(t\right),\tau_k)+\eta(r\left(t\right),l\left(t\right),\tau_j)\eta(r\left(t\right),l\left(t\right),\tau_k)\right]dt & \text{if } j\neq k\\ \\ \left[\gamma(r\left(t\right),l\left(t\right),\tau_j)^2+\eta(r\left(t\right),l\left(t\right),\tau_j)^2+\sigma_\varepsilon^2\right]dt & \text{if } j=k.\end{cases}\end{aligned}$$

We will incorporate the off-diagonal information based on these restrictions into the following localized criterion function

$$\begin{aligned}Q_n(\gamma,\eta) &= \sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\int\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)-\left(\gamma(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\gamma(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)+\eta(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\eta(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)\right)\Delta_n\right]^2\\ &\quad\times K_{h_r}(r_i-\widetilde{r}_i)K_{h_r}(l_i-\widetilde{l}_i)K_{h_\tau}(\tau_j-\widetilde{\tau}_j)K_{h_\tau}(\tau_k-\widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j d\widetilde{\tau}_k.\end{aligned}$$

Again, similar to the one factor case above, to avoid the need to estimate the measurement errors, we use only off-diagonal moment restrictions to estimate the object of interest, namely the volatility structures $\gamma(.)$ and $\eta(.)$.

Proceeding as in the one factor case, using the delta method, we obtain the following representation for the multifactor estimator

$$\begin{bmatrix}\widehat{\gamma}(r,l,\tau)\\ \widehat{\eta}(r,l,\tau)\end{bmatrix}=\begin{bmatrix}\int\widehat{H}_2\left(r,\tau,s\right)\widehat{\gamma}(r,s)^2ds & \int\widehat{H}_2\left(r,\tau,s\right)\widehat{\gamma}(r,s)\widehat{\eta}(r,s)ds\\ \int\widehat{H}_2\left(r,\tau,s\right)\widehat{\gamma}(r,s)\widehat{\eta}(r,s)ds & \int\widehat{H}_2\left(r,\tau,s\right)\widehat{\eta}(r,s)^2ds\end{bmatrix}^{-1}\begin{bmatrix}\int\widehat{H}_1\left(r,\tau,s\right)\widehat{\gamma}(r,s)ds\\ \int\widehat{H}_1\left(r,\tau,s\right)\widehat{\eta}(r,s)ds\end{bmatrix},$$

where

$$\begin{aligned}\widehat{H}_1 &= \sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)K_{h_r}(l_i-l)\\ &\quad\times\left[K_{h_\tau}(\tau-\tau_j)K_{h_\tau}(s-\tau_k)+K_{h_\tau}(\tau-\tau_k)K_{h_\tau}(s-\tau_j)\right],\end{aligned}$$

and

$$
\begin{aligned}
\widehat{H}_2 &= \Delta_n \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k\neq j}}^{J} K_{h_r}(r_i - r) K_{h_r}(l_i - l) \\
&\quad \times \left[K_{h_\tau}(\tau - \tau_j) K_{h_\tau}(s - \tau_k) + K_{h_\tau}(\tau - \tau_k) K_{h_\tau}(s - \tau_j) \right].
\end{aligned}
$$

This representation suggests an iterative solution as

$$
\begin{aligned}
\begin{bmatrix} \widehat{\gamma}^{[a+1]}(r,l,\tau) \\ \widehat{\eta}^{[a+1]}(r,l,\tau) \end{bmatrix} &= \begin{bmatrix} \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}^{[a]}(r,s)^2 ds & \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}^{[a]}(r,s) \widehat{\eta}^{[a]}(r,s) ds \\ \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}^{[a]}(r,s) \widehat{\eta}^{[a]}(r,s) ds & \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\eta}^{[a]}(r,s)^2 ds \end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix} \int \widehat{H}_1\left(r,\tau,s\right) \widehat{\gamma}^{[a]}(r,s) ds \\ \int \widehat{H}_1\left(r,\tau,s\right) \widehat{\eta}^{[a]}(r,s) ds \end{bmatrix}, \quad a = 0, 2, \ldots,
\end{aligned}
$$

starting from some given initial condition.

Similar to the one factor case, the asymptotics for this two factor estimator can be derived under the previously invoked assumptions of a mutiplicatively separable form of the volatility structure and the local time. As shown in the Appendix, under these assumptions, the estimates follow mixed normal asymptotics again.

Theorem 2. *Suppose that assumptions A.1-A4 hold. Then,*

$$
J \sqrt{\frac{h_r h_l h_\tau \overline{L}\left(t_n,(r,l)\right)}{\Delta_n}} \begin{bmatrix} \widehat{\gamma}(r,l,\tau) - \gamma(r,l,\tau) \\ \widehat{\eta}(r,l,\tau) - \eta(r,l,\tau) \end{bmatrix} \xrightarrow{d} N\left(0, \frac{\|K\|^6}{4\Upsilon_3 p(\tau)} \begin{bmatrix} \Upsilon_1 & 0 \\ 0 & \Upsilon_2 \end{bmatrix}\right),
$$

where Υ_1 , Υ_2 *and* Υ_3 *are defined as*

$$
\begin{aligned}
\Upsilon_1 &= 4\left[\gamma(r,l,\tau)^2 + \eta(r,l,\tau)^2\right] \int \left[\gamma(r,l,s)^2 + \eta(r,l,s)^2\right] \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\gamma(r,l,\tau)^2 \int \gamma(r,l,s)^2 \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\eta(r,l,\tau)^2 \int \eta(r,l,s)^2 \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\gamma(r,l,\tau)\eta(r,l,\tau) \int \gamma(r,l,s)\eta(r,l,\tau) \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds
\end{aligned}
$$

$$
\begin{aligned}
\Upsilon_2 &= 4\left[\gamma(r,l,\tau)^2 + \eta(r,l,\tau)^2\right] \int \left[\gamma(r,l,s)^2 + \eta(r,l,s)^2\right] \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\gamma(r,l,\tau)^2 \int \gamma(r,l,s)^2 \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\eta(r,l,\tau)^2 \int \eta(r,l,s)^2 \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds \\
&\quad + 2\gamma(r,l,\tau)\eta(r,l,\tau) \int \gamma(r,l,s)\eta(r,l,\tau) \left[\gamma(r,l,s)G_2 - \eta(r,l,s)G_3\right] p\left(s\right) ds
\end{aligned}
$$

$$\Upsilon_3 = \left[\int \gamma(r,s)^2 p(s)\, ds \int \eta(r,s)^2 p(s)\, ds - \left[\int \gamma(r,s)\eta(r,s) p(s)\, ds \right]^2 \right]^2,$$

and $G_1 = \int \gamma(r,s)^2 p(s)\, ds$, $G_2 = \int \eta(r,s)^2 p(s)\, ds$, $G_3 = \int \gamma(r,s)\eta(r,s)p(s)\, ds$.

Remark 3. The rate of convergence, as expected, is slowed down by a factor of h_l relatively to the case of one factor. With reasonable values of bandwidth, around 1% to 3%, as suggested by the historical interest rates, the rate of convergence decreases by a factor of hundreds, a significant change. This familiar curse of dimensionality would be rather detrimental to the estimation for multi-factors, or any useful inference based on this estimate such as testing for the number of factors necessary to depict the evolution of the empirical yield curves; parametric methods are deemed to be more suitable in this case, with the risk of model misspecification however.

6 Empirical Implementation

We apply the above techniques to estimate the volatility structure for one factor and two factor models from daily CRSP bond data from January 1961 to December 1998. As reasoned in Jeffrey, Linton, and Nguyen (1999b), and consistent with the derivation done previously, we deviate from the common practice of using the forwards for HJM model estimation and use the yield evolution instead, in light of the fact that the yield is easier to estimate. The first step in estimating a dynamic model of the yield curve is to extract the unobservable yield curves themselves from coupon bearing bonds observed in the market: LMNT (2000) kernel smoothing based yield curve fitting, whose implementation is shown in detail in Jeffrey, Linton and Nguyen (1999a), is our choice of yield curve extraction here.

Our choice of kernels is the commonly used Gaussian $K(t) = \exp(-0.5t^2)/\sqrt{2\pi}$, for which $\|K\|^2 = 1/2\sqrt{\pi}$. Choosing an optimal bandwidth for a nonparametric estimator is still an elusive question in the literature, there being no single scheme that is uniformly accepted although cross-validation is a frequently used procedure ; see Härdle (1990) and Pagan and Ullah (1999) for extensive discussion of the proliferation of proposed schemes. The existing methods are designed, however, for the standard regression context; for diffusion models, especially ones that allow for nonstationary processes, rules are yet to be developed. For our estimates, similar to that of LMNT (1999), since no closed form solutions are available, estimation of the models are computationally demanding, thus rendering cross-validation an unattractive option. So we opt instead to use a flat bandwidth (obtained by visual inspection) in our study and leave other bandwidth selection methods for future experiments. For the diagnostic nature of our empirical work, this appears not to be of great material import. Bandwidth for the time to maturity dimension is fixed at 1 year level, while that along the interest rate dimension is fixed at 1%.

The estimates are the solutions to the above derived first order conditions. The system of equations can be solved by minimization of the square of the first order condition. Note that we do not need to solve the system simultaneously, which would be computationally inhibitive. The trick is to solve for each "slice" associated with each fixed level of $\overline{r}$ (or in the case of two factors, with each fixed point of $(\overline{r}, \overline{l})$), i.e., solve for the discrete points along the curve $\widehat{\gamma}(r, s)$.

6.1 One factor

We find that to estimate our model, based on the first order condition (6), albeit iterative scheme suggested by (7) is feasible, it is much more convenient to solve for the first order condition (6) directly by minimization routine. This finding is consistent with the implementation procedure reported in Jeffrey, Linton, Nguyen (1999a), where a similar but somewhat simpler first order condition for LMNT (2000) is conducted.

The nonparametric estimates for the volatility structure of one factor HJM model using the method developed in this paper is shown in Figure 3.1, with interest rate data from January 1961 to December 1998.The well known feature of the volatility structure reported in the literature, that interest rates become more volatile when the level is high, is again observed here. For instance, volatility becomes as high as 4% when the short term hits 18% (this level only observed in the "Fed-experiment" period from 1979 to 1982). Along the maturity dimension, volatility tends to slowly increases with time to maturity when short term rates are low, but the pattern tends to reverse itself when interest rates drift to higher ranges. Overall, the shape of the volatility surface is rather consistent with what have been observed in the empirical literature; see Linton, Jeffrey, and Nguyen (1999b) for similar results.

We also experiment with different time periods, for instance, starting the data period from 1970, 1983 and from 1990 respectively. The divisions are motivated by the oil shock in the early 1970s, the so-called "Fed-experiment" from 1979 to 1983, where interest rates were floated by the central bank, and the relatively low and stable interest rate that prevailed in the 1990s. The volatility structures estimated for these 3 periods are reported in Figure 3.2, 3.3 and 3.4 respectively. Note that to avoid biasedness at the boundary and extrapolation, the volatility structures are estimated only in the ranges that data are observed. So in the first 2 figures, where the high interest rates period of 1979-1983 are included in the estimation procedures, the instantaneous interest rates (the state variable) can go from 0% to 18%, while in the last 2 figures, they only go high up to around 10%. Similar consideration is built into the time to maturity dimension. For instance, if the data period goes back as far as 1960, we only examine the yield curves with time to maturity up to 4 years, since bonds with longer maturities were rarely available then. This scale increases to 5 and 9 years if data from later periods are used instead. For ease of comparison, however, the scale along the volatility axis remains constant across the figures.

Comparing these figures, notwithstanding the scale differences (along the axis of short rates and time to maturity), the remarkable differences between the graphs seem to suggest some nonstationary behavior for interest rates. When data including the chaotic period of the Fed-experiment, as in Figure 3.1 and Figure 3.2, interest rates not only become more volatile when interest rates are higher, but even when interest rates are low, volatility is higher compared to later periods. After the Fed-experiment period, the volatility surface has become much more stable, although its typical shape is still observed; volatility increases with instantaneous interest rates and time to maturity.

In Figure 3.5, we compare the volatility surfaces obtained by different methodology, i.e., the one developed in this paper which accounts for measurement errors, and the simpler method employed in Pearson and Zhu (1998) and Jeffrey, Linton, and Nguyen (1999b). Yield curves from 1970 to 1980 are chosen to conduct this experiment. The first graph shows the estimate of the volatility surface using the new method, which uses only off diagonal information in the moment conditions developed above, while the second one shows that of the more "naive" method, which in fact uses only information along the diagonal of the moment conditions. Examining the two graphs, interestingly, using off diagonal restrictions as we have in this paper does not yield an estimate whose shape is dramatically different from that obtained using information from the diagonal alone. However, as expected, the latter does over estimate the volatility surface (by adding the variance of the measurement errors into its estimates; see the related footnote in section 2.1). When the short rate reaches 18% for instance, its estimate of the volatility is around 0.051, while the new method yields an estimate of 0.034. This magnitude of difference is definitely material when one uses the volatility surface to price fixed income instruments.

6.2 Two factor

As cautioned earlier, the variance of our estimates in this case (and in models with more factors) is relatively large, making statistical inference based on these estimates difficult, besides being computational burdensome in view of the complexity of the model. Consequently, we can not wholeheartedly endorse nonparametric estimation in multi-factor models for making inference and testing purposes. However, in a search for a reasonable parametric model, nonparametric methods even of high dimension can provide a good starting point.

For illustrative purposes, we implement a two factor model here, with the data period from January 1970 to December 1998. As mentioned earlier, which factors to be included is commonly rather ad hoc in the literature, with a whole array of existing specifications. In our following implementation, the chosen factors are the short term and the long term interest rate respectively, a rather common choice in the related literature. The same trick used earlier, i.e., solving for curves of $\eta(\overline{r}, \overline{l}, \tau)$ and $\gamma(\overline{r}, \overline{l}, \tau)$ at a fixed point $(\overline{r}, \overline{l})$ and different τ. Even with that, the first order condition can be difficult to solve, since we have to solve for the 2 curves $\eta(\overline{r}, \overline{l}, \tau)$ and $\gamma(\overline{r}, \overline{l}, \tau)$ simultaneously.

Graphical presentation in this case is quite problematic, since $\eta(.)$ and $\gamma(.)$ are functions of 3 variables – the short rate, the long rate and the maturity. So for illustration, we just report Tables 3.1 and 3.2, which contain values of $\eta(.)$ and $\gamma(.)$ respectively for one level of $r = 6.7\%$, where we allow long rates to vary from 2.5% to 10.8%, where maturities vary from 0 to 4 years. The ranges are chosen so empirical data can best accommodate the estimation procedure, pre-empting extrapolation and boundary issues. Volatility is reported in 1/1000.

Casual observations suggest that values of $\eta(.)$ and $\gamma(.)$ are rather close, and around $\eta^1(.)\sqrt{.5}$, where $\eta^1(.)$ is the value of the functional estimate of the volatility function when we specify one factor model. This expected results however still render a model capable of generating more complex behavior for the yield curve dynamics due to the independence of the two Brownian motions.[9] For each volatility, behaviors similar to the one factor case is displayed. Volatility increases with the stochastic state variable (the long rate), while along the maturity, volatility seems to positively correlate with time to maturity when the long rate is low, but turns to negatively correlate with time to maturity when the long rate reaches into the higher range.

7 Conclusion

The HJM approach has revolutionized dynamic models of the fixed income market, but specification and estimation issues in this framework remain a serious challenge.[10] In this study, we propose an extended version of the original HJM model to incorporate measurement errors directly into the model specification. A nonparametric estimate for the volatility structure, which is central to the dynamics of HJM, is provided and implemented with empirical data.

The technique can be readily adapted to multi-factor models, although limitations on the existence of multidimensional local time complicate the development of a general asymptotic theory. Under some simplifying assumptions, however, an asymptotic theory is possible. The theoretical results indicate rather slow convergence properties for nonparametric estimators, consistent with their counterparts in nonparametric regression. While we acknowledge these difficulties, nonparametric

[9] Intuitively, $\left[\sqrt{.5}\eta^1(.)\,dW_1 + \sqrt{.5}\eta^1(.)\,dW\right]$ will be able to generate more complex shocks than a single $\eta^1(.)\,dW$ even their total variations are the same.

[10] The technique has found itself aggressively migrated to the equity market too; see Bonchuner (1998) for a representative article from this rapidly expanding literature.

methods of the type presented here still seem to offer an attractive general tool to help researchers find appropriate parametric specifications.

8 Appendix

For two random variables X_n, Y_n, we say that $X_n \simeq Y_n$ whenever $X_n = Y_n(1 + o_p(1))$ as $n \to \infty$.

8.1 Proof of Theorem 1

We first establish the first order condition defining the estimator. Let $\delta_{r,\tau}(.,.)$ be the bivariate Dirac delta function at (r, τ). Defining

$$
\begin{aligned}
0 &= \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \int\int\int [\Delta y(t_i, \tau_j)\Delta y(t_i, \tau_k) - \widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k)\Delta_n] \\
&\quad \times \left(\delta\left(\widetilde{r}_i, \widetilde{\tau}_j\right)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k) + \delta\left(\widetilde{r}_i, \widetilde{\tau}_k\right)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)\right)\Delta_n K_{h_r}(r_i - \widetilde{r}_i)K_{h_\tau}(\tau_j - \widetilde{\tau}_j)K_\tau(\tau_k - \widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k.
\end{aligned}
$$

For the first terms, we have

$$
\begin{aligned}
&\int\int\int \delta\left(\widetilde{r}_i, \widetilde{\tau}_j\right)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k)K_{h_r}(r_i - \widetilde{r}_i)K_{h_\tau}(\tau_j - \widetilde{\tau}_j)K_{h_\tau}(\tau_k - \widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k \\
&= K_{h_r}(r - r_i)K_\tau(\tau - \tau_j)\int \widehat{\gamma}(r, s)K_{h_\tau}(s - \tau_k)ds
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\int\int\int \delta(\widetilde{r}_i, \widetilde{\tau}_k)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)K_{h_r}(r_i - \widetilde{r}_i)K_{h_\tau}(\tau_j - \widetilde{\tau}_j)K_{h_\tau}(\tau_k - \widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k \\
&= K_{h_r}(r - r_i)K_{h_\tau}(\tau - \tau_k)\int \widehat{\gamma}(r, s)K_{h_\tau}(s - \tau_j)ds.
\end{aligned}
$$

As for the second terms

$$
\begin{aligned}
&\int\int\int \widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k)\delta\left(\widetilde{r}_i, \widetilde{\tau}_j\right)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k)K_{h_r}(r_i - \widetilde{r}_i)K_{h_\tau}(\tau_j - \widetilde{\tau}_j)K_{h_\tau}(\tau_k - \widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k \\
&= K_{h_r}(r - r_i)K_{h_\tau}(\tau - \tau_j)\widehat{\gamma}(r, \tau)\int \widehat{\gamma}(r, s)^2 K_{h_\tau}(s - \tau_k)ds
\end{aligned}
$$

$$
\begin{aligned}
&\int\int\int \widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_k)\delta\left(\widetilde{r}_i, \widetilde{\tau}_k\right)\widehat{\gamma}(\widetilde{r}_i, \widetilde{\tau}_j)K_{h_r}(r_i - \widetilde{r}_i)K_{h_\tau}(\tau_j - \widetilde{\tau}_j)K_{h_\tau}(\tau_k - \widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{\tau}_j d\widetilde{\tau}_k \\
&= K_{h_r}(r - r_i)K_{h_\tau}(\tau - \tau_k)\widehat{\gamma}(r, \tau)\int \widehat{\gamma}(r, s)^2 K_{h_\tau}(s - \tau_j)ds.
\end{aligned}
$$

We now turn to the asymptotics of the procedure. In the sequel we will extensively use the following well established results:

$$\int g(.,s)K_{h_\tau}(s-\tau_k)ds \rightarrow g(.,\tau_k)$$

$$h_\tau \int g(.,s)K^2_{h_\tau}(s-\tau_k)ds \rightarrow g(.,\tau_k)\left\|K\right\|^2,$$

which hold for any bounded continuous functions g. Note that the above integrals are already in state space, not in time domain. These results will be useful when we integrate along the maturity dimension. We also use the following results, with the understanding that the initial integrals are in the time domain:

$$\begin{aligned}\int g(.,s)K_{h_r}(r-r(s))ds &\rightarrow \int_{-\infty}^{\infty} g\left(.,a\right)K_{h_r}\left(\frac{a-r}{h_r}\right)\overline{L}\left(t_n,a\right)da \\ &\rightarrow g\left(.,r\right)\overline{L}\left(t_n,r\right)\end{aligned}$$

$$\begin{aligned}h_r\int g(.,s)K^2_{h_r}(r-r(s))ds &\rightarrow h_r\int_{-\infty}^{\infty} g\left(.,a\right)K^2_{h_r}\left(\frac{a-r}{h_r}\right)\overline{L}\left(t_n,a\right)da \\ &\rightarrow g\left(.,r\right)\overline{L}\left(t_n,r\right)\left\|K\right\|^2.\end{aligned}$$

We first analyze the quantity $\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)$, which turns out to be central to the fundamental source of uncertainty driving the asymptotics. We have

$$\begin{aligned}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) &= \left[y(t_{i+1},\tau_j)-y(t_i,\tau_j)\right]\left[y\left(t_{i+1},\tau_k\right)-y\left(t_i,\tau_k\right)\right] \\ &= \left[y(t_{i+1},\tau_j)y\left(t_{i+1},\tau_k\right)-y(t_i,\tau_j)y\left(t_i,\tau_k\right)\right]- \\ &\quad y\left(t_i,\tau_k\right)\left[y(t_{i+1},\tau_j)-y(t_i,\tau_j)\right]-y(t_i,\tau_j)\left[y\left(t_{i+1},\tau_k\right)-y\left(t_i,\tau_k\right)\right],\end{aligned}$$

where

$$\begin{aligned}y(t_{i+1},\tau_j)-y(t_i,\tau_j) &= \int_{t_i}^{t_i+\Delta_n}\alpha(s,\tau_j)ds+\int_{t_i}^{t_i+\Delta_n}\gamma(r(s),\tau_j)dW(s)+\sigma^2_\varepsilon\int_{t_i}^{t_i+\Delta_n}dW_j(s) \\ y\left(t_{i+1},\tau_k\right)-y\left(t_i,\tau_k\right) &= \int_{t_i}^{t_i+\Delta_n}\alpha(s,\tau_k)ds+\int_{t_i}^{t_i+\Delta_n}\gamma(r(s),\tau_k)dW(s)+\sigma^2_\varepsilon\int_{t_i}^{t_i+\Delta_n}dW_j(s).\end{aligned}$$

By the Multivariate Itô lemma [Theorem 4.2.1 of Øksendal (1998)], we have

$$\begin{aligned}d\left(y(t,\tau_j)\cdot y\left(t,\tau_k\right)\right) &= y(t,\tau_j)dy\left(t,\tau_k\right)+y\left(t,\tau_k\right)dy(t,\tau_j)+dy\left(t,\tau_k\right)dy(t,\tau_j) \\ &= y(t,\tau_j)\left[\alpha(t,\tau_k)dt+\gamma(r\left(t\right),\tau_k)dW\left(t\right)+\sigma^2_\varepsilon dW_k(t)\right] \\ &\quad +y\left(t,\tau_k\right)\left[\alpha(t,\tau_j)dt+\gamma(r\left(t\right),\tau_j)dW\left(t\right)+\sigma^2_\varepsilon dW_j(t)\right] \\ &\quad +\gamma(r\left(t\right),\tau_j)\gamma(r\left(t\right),\tau_k)dt,\end{aligned}$$

because the processes $dW_j(t)$ and $dW_k(t)$ are orthogonal. This implies that

$$\begin{aligned}
& [y(t_{i+1},\tau_j)y\left(t_{i+1},\tau_k\right) - y\left(t_i,\tau_k\right)y\left(t_i,\tau_k\right)] \\
= & \int_{t_i}^{t_i+\Delta_n} y(s,\tau_j)\alpha(s,\tau_k)ds + \int_{t_i}^{t_i+\Delta_n} y(s,\tau_j)\gamma(r(s),\tau_k)dW(s) \\
& +\sigma_\varepsilon^2 \int_{t_i}^{t_i+\Delta_n} y(s,\tau_j)dW_k(s) + \int_{t_i}^{t_i+\Delta_n} y\left(s,\tau_k\right)\alpha(s,\tau_j)ds \\
& + \int_{t_i}^{t_i+\Delta_n} y\left(s,\tau_k\right)\gamma(r(s),\tau_j)dW(s) + \sigma_\varepsilon^2 \int_{t_i}^{t_i+\Delta_n} y(s,\tau_k)dW_j(s) \\
& + \int_{t_i}^{t_i+\Delta_n} \gamma(r(s),\tau_j)\gamma(r(s),\tau_k)ds.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) = & \int_{t_i}^{t_i+\Delta_n} \left[y(s,\tau_j) - y(t_i,\tau_j)\right]\alpha(s,\tau_k)ds \\
& + \int_{t_i}^{t_i+\Delta_n} \left[y(s,\tau_j) - y(t_i,\tau_j)\right]\gamma(r(s),\tau_k)dW(s) \\
& +\sigma_\varepsilon^2 \int_{t_i}^{t_i+\Delta_n} \left[y(s,\tau_j) - y(t_i,\tau_j)\right]dW_k(s) \\
& + \int_{t_i}^{t_i+\Delta_n} \left[y\left(s,\tau_k\right) - y\left(t_i,\tau_k\right)\right]\alpha(s,\tau_j)ds \\
& + \int_{t_i}^{t_i+\Delta_n} \left[y\left(s,\tau_k\right) - y\left(t_i,\tau_k\right)\right]\gamma(r(s),\tau_j)dW(s) \\
& +\sigma_\varepsilon^2 \int_{t_i}^{t_i+\Delta_n} \left[y(s,\tau_k) - y(t_i,\tau_k)\right]dW_j(s) \\
& + \int_{t_i}^{t_i+\Delta_n} \gamma(r(s),\tau_j)\gamma(r(s),\tau_k)ds.
\end{aligned} \tag{14}$$

We have

$$E\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)|\mathcal{F}_{t_i}\right] \simeq 2\gamma(r_i,\tau_j)\gamma(r_i,\tau_k)\Delta_n,$$

where $r_i = r(t_i)$, and

$$E\left[\Delta y(t_i,\tau_j)^2\Delta y(t_i,\tau_k)^2|\mathcal{F}_{t_i}\right] \simeq 4\gamma(r_i,\tau_j)^2\gamma(r_i,\tau_k)^2\Delta_n^2. \tag{15}$$

We next establish the asymptotic properties of our estimators.

8.1.1 Consistency

We will restrict our attention to the class of Sobolev functions which we now define. For any vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)'$ and function $g : \mathbb{R}^d \to \mathbb{R}$, let

$$D^{\boldsymbol{\alpha}} g(x) = \frac{\partial^{\sum_{i=1}^{d} \alpha_i} g(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

Define the following seminorm on the class of real-valued functions with domain $\mathcal{X}$:

$$\|g\|_{q,2}^2 = \sum_{|\alpha| \leq q} \int_{\mathcal{X}_0} \left(D^{\boldsymbol{\alpha}} g(x)\right)^2 dx,$$

where q is an integer and $\mathcal{X}_0 \subset \mathcal{X}$ is a compact set, and let $||g||_2^2 = ||g||_{0,2}^2$ denote the usual L_2 norm. Finally, define $\Gamma = \{\gamma : \mathbb{R}^d \to \mathbb{R} : ||\gamma||_{q,2}^2 \leq C\}$ for some large $C < \infty$, and let Γ_0 denote the subset of Γ that excludes an ϵ-neighborhood of the zero function. We take $q = 1$ and $d = 2$.

Let

$$G_n(\gamma)(r, \tau) = \gamma(r, \tau) \int \widehat{H}_2(r, \tau, s)\, \gamma^2(r, s) ds - \int \widehat{H}_1(r, \tau, s)\, \gamma(r, s) ds$$

for any function $\gamma(.)$, and define

$$Q_n(\gamma) = ||G_n(\gamma)||_2.$$

We use the following lemma of Newey and Powell (2000, Lemma A2).

Lemma. *Suppose that (i) There exists some deterministic function $Q(\gamma)$ with a unique minimum on the parameter space Γ_0; (ii) Q_n and Q are continuous, Γ_0 is compact, and* $\sup_{\gamma \in \Gamma} |Q_n(\gamma) - Q(\gamma)| = o_p(1)$; *(iii) $\widetilde{\Gamma}$ are subsets of Γ_0 such that for any $\gamma \in \Gamma_0$ there exists $\widetilde{\gamma} \in \widetilde{\Gamma}$ such that $\widetilde{\gamma} \to \gamma$. Then* $\widehat{\gamma} = \arg\min_{\gamma \in \Gamma} Q_n(\gamma) \to^p \gamma_0$.

The proof of this lemma is given in Newey and Powell (2000). We next verify its main conditions in our case. We first calculate the limit function $G(\gamma)$ and hence $Q(\gamma) = ||G(\gamma)||_2$. In the limit, when the bandwidth of all kernels approach zero, $h \to 0$, and number of observations approach infinity, $n \to \infty$ and $J \to \infty$, observing that for any twice continuously differentiable function γ

$$\int \gamma(r, s) K_\tau(s - \tau_k) ds = \gamma(r, \tau_k) + O(h^2),$$

and scaling the whole equation by J^{-2}, we have the first term's first component

$$\frac{1}{J^2} \sum_{i=1}^{n-1} \sum_{\substack{j=1 \\ k \neq j}}^{J} \sum_{k=1}^{J} \Delta y(t_i, \tau_j) \Delta y(t_i, \tau_k) K_{h_r}(r_i - r)$$
$$\times \left[K_{h_\tau}(\tau - \tau_j) \int \gamma(r, s) K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \gamma(r, s) K_{h_\tau}(s - \tau_j) ds \right]$$

$$
\begin{aligned}
&\simeq \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right] \\
&= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}K_{h_r}(r_i-r)\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right] \\
&\quad\times\left\{\int_{t_i}^{t_i+\Delta_n}\left[y(s,\tau_j)-y(t_i,\tau_j)\right]\alpha_0(s,\tau_k)ds+\int_{t_i}^{t_i+\Delta_n}\left[y(s,\tau_j)-y(t_i,\tau_j)\right]\gamma_0(r(s),\tau_k)dW(s)\right. \\
&\quad+\sigma_\varepsilon^2\int_{t_i}^{t_i+\Delta_n}\left[y(s,\tau_j)-y(t_i,\tau_j)\right]dW_k(s)+\sigma_\varepsilon^2\int_{t_i}^{t_i+\Delta_n}\left[y(s,\tau_k)-y(t_i,\tau_k)\right]dW_j(s) \\
&\quad+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\alpha_0(s,\tau_j)ds+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\gamma_0(r(s),\tau_j)dW(s) \\
&\quad\left.+\int_{t_i}^{t_i+\Delta_n}\gamma_0(r(s),\tau_j)\gamma_0(r(s),\tau_k)ds\right\}.
\end{aligned}
$$

The first six terms are of smaller order in probability and by a law of large numbers, we have

$$
\begin{aligned}
&\simeq \frac{\Delta_n}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma_0(r_i,\tau_j)\gamma_0(r_i,\tau_k)K_{h_r}(r_i-r)\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right] \\
&\simeq \frac{\Delta_n}{J}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\gamma_0(r_i,\tau_j)K_{h_r}(r_i-r)K_{h_\tau}(\tau-\tau_j)\frac{1}{J}\sum_{k,k\neq j}^{J}\gamma_0(r_i,\tau_k)\gamma(r,\tau_k) \\
&\quad+\frac{\Delta_n}{J}\sum_{i=1}^{n-1}\sum_{k,k\neq j}^{J}\gamma_0(r_i,\tau_k)K_{h_r}(r_i-r)K_{h_\tau}(\tau-\tau_k)\frac{1}{J}\sum_{j}^{J}\gamma_0(r_i,\tau_j)\gamma(r,\tau_j) \\
&\simeq \frac{\Delta_n}{J}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\gamma_0(r_i,\tau_j)K_{h_r}(r_i-r)K_{h_\tau}(\tau-\tau_j)\int\gamma_0(r_i,s)\gamma(r,s)p\left(s\right)ds \\
&\quad+\frac{\Delta_n}{J}\sum_{i=1}^{n-1}\sum_{k,k\neq j}^{J}\gamma_0(r_i,\tau_k)K_{h_r}(r_i-r)K_{h_\tau}(\tau-\tau_k)\int\gamma_0(r_i,s)\gamma(r,s)p\left(s\right)ds \\
&\simeq 2\Delta_n\sum_{i=1}^{n-1}\gamma_0(r_i,\tau)K_{h_r}(r_i-r)p\left(\tau\right)\int\gamma_0(r_i,s)\gamma(r,s)p\left(s\right)ds \\
&\simeq 2\overline{L}\left(t_n,r\right)p\left(\tau\right)\gamma_0(r,\tau)\int\gamma_0(r,s)\gamma(r,s)p\left(s\right)ds,
\end{aligned}
$$

and the second term

$$
\frac{1}{J^2}\Delta_n\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma(r,\tau)K_{h_r}(r-r_i)
$$

$$
\times \left[K_{h_\tau}(\tau - \tau_j) \int \gamma(r,s)^2 K_\tau(s-\tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \gamma(r,s)^2 K_\tau(s-\tau_j) ds \right]
$$

$$
\simeq \gamma(r,\tau) \frac{1}{J^2} \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_r}(r - r_i) \left[K_{h_\tau}(\tau - \tau_j)\gamma(r,\tau_k)^2 + K_{h_\tau}(\tau - \tau_k)\gamma(r,\tau_j)^2 \right] \Delta_n
$$

$$
\simeq 2\overline{L}\,(t_n, r)\, p\,(\tau)\, \gamma(r,\tau) \int \gamma(r,s)^2 p\,(s)\, ds.
$$

In conclusion, $G_n(\gamma) \to^p G(\gamma)$, where

$$
G(\gamma)(r,\tau) = \gamma_0(r,\tau) \int \gamma_0(r,s)\gamma(r,s) p\,(s)\, ds - \gamma(r,\tau) \int \gamma(r,s)^2 p\,(s)\, ds.
$$

The convergence of G_n to G is uniform in $\gamma \in \Gamma_0$ because of the polynomial nature of the criterion function and the boundedness assumptions.

The equation $G(\gamma)(r,\tau) = 0$ has solution $\gamma(r,\tau) = \gamma_0(r,\tau)$ on Γ_0. For uniqueness, we will examine the Gateaux derivatives of both sides of the following equation

$$
\gamma(r,\tau) = \frac{\gamma_0(r,\tau) \int \gamma_0(r,s)\gamma(r,s) p\,(s)\, ds}{\int \gamma(r,s)^2 p\,(s)\, ds},
$$

which is from the first order condition above. Expand $\gamma(r,s)$ around the above solution $\gamma_0(r,s)$ as

$$
\gamma(r,s) = \gamma_0(r,s) + \epsilon g\,(s)\,.
$$

The above equation can then be written as

$$
\begin{aligned}
\gamma_0(r,\tau) + \epsilon g\,(\tau) &= \frac{[\gamma_0(r,\tau) + \epsilon g\,(\tau)] \int \gamma_0(r,s)\,[\gamma_0(r,s) + \epsilon g\,(s)]\, p\,(s)\, ds}{\int [\gamma_0(r,\tau) + \epsilon g\,(\tau)]^2 f\,(s)\, ds} \\
&= \frac{\gamma_0(r,\tau) \int \gamma_0(r,s)^2 p\,(s)\, ds + \varepsilon \gamma_0(r,\tau) \int \gamma_0(r,s) g\,(s)\, p\,(s)\, ds}{\int \gamma_0(r,s)^2 p\,(s)\, ds \left[1 + 2\epsilon \frac{\int \gamma_0(r,s) g(s) p(s) ds}{\int \gamma_0(r,s)^2 p(s) ds} \right]} + o\,(\varepsilon) \\
&= \gamma_0(r,\tau) - \varepsilon \frac{\gamma_0(r,\tau) \int \gamma_0(r,s) g\,(s)\, p\,(s)\, ds}{\int \gamma_0(r,s)^2 p\,(s)\, ds} + o\,(\varepsilon)\,.
\end{aligned}
$$

Differentiating both sides with respect to ε, and evaluating them at $\varepsilon = 0$ yields the following condition for $g\,(.)$

$$
g\,(\tau) = -\frac{\gamma_0(r,\tau) \int \gamma_0(r,s) g(r,s) p(s) ds}{\int \gamma_0(r,s)^2 p(s) ds},
$$

which has only one solution that $g(r,s) = 0$. Uniqueness is thus proved with $\gamma(r,s) = \gamma_0(r,s)$ as the only solution. Consistency is therefore achieved. ■

8.1.2 Asymptotic Distribution

It will be convenient to use the representation in (7). First order Taylor expansion of the right hand side with respect to $\widehat{\gamma}(.)$ around $\gamma_0(.)$ yields

$$\begin{aligned}
\widehat{\gamma}(r,\tau)-\gamma_0(r,\tau) &= \frac{\int \widehat{H}_1(r,\tau,s)\,\widehat{\gamma}(r,s)ds}{\int \widehat{H}_2(r,\tau,s)\,\widehat{\gamma}^2(r,s)ds}-\frac{\int H_1(r,\tau,s)\,\gamma_0(r,s)ds}{\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds} \\
&= \frac{\int[\widehat{H}_1(r,\tau,s)-H_1(r,\tau,s)]\gamma_0(r,s)ds}{\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds} \\
&\quad -\gamma_0(r,\tau)\frac{\int[\widehat{H}_2(r,\tau,s)-H_2(r,\tau,s)]\gamma_0(r,s)ds}{\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds} \\
&\quad +\frac{\int H_1(r,\tau,s)\,[\widehat{\gamma}(r,s)-\gamma_0(r,s)]\,ds\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds}{\left[\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds\right]^2} \\
&\quad -\frac{\int H_1(r,\tau,s)\,\gamma_0(r,s)ds\int H_2(r,\tau,s)\,2\gamma_0(r,s)\,[\widehat{\gamma}(r,s)-\gamma_0(r,s)]\,ds}{\left[\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds\right]^2} \\
&\quad +O\left(||\widehat{H}_1-H_1||^2+||\widehat{H}_2-H_2||^2+||\widehat{\gamma}-\gamma||^2\right),
\end{aligned}$$

where

$$\gamma_0(r,\tau)=\frac{\int H_1(r,\tau,s)\,\gamma_0(r,s)ds}{\int H_2(r,\tau,s)\,\gamma_0^2(r,s)ds}.$$

Let

$$\begin{aligned}
\Psi(\gamma(.)) &= \int \widehat{H}_1(r,\tau,s)\,\gamma_0(r,s)ds \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r) \\
&\quad \times\left[K_{h_\tau}(\tau-\tau_j)\int\gamma_0(r,s)K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\gamma_0(r,s)K_{h_\tau}(s-\tau_j)ds\right],
\end{aligned}$$

$$\begin{aligned}
\Phi(\gamma(.)) &= \int \widehat{H}_2(r,\tau,s)\,\gamma_0^2(r,s)ds \\
&= \Delta_n\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r-r_i) \\
&\quad \times\left[K_{h_\tau}(\tau-\tau_j)\int\gamma_0^2(r,s)K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\gamma_0^2(r,s)K_{h_\tau}(s-\tau_j)ds\right]
\end{aligned}$$

and

$$
\begin{aligned}
\Xi\left(\gamma\left(.\right),\widehat{\gamma}\left(.\right)\right) &= \int\widehat{H}_1\left(r,\tau,s\right)\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]ds\int\widehat{H}_2\left(r,\tau,s\right)\gamma_0^2(r,s)ds \\
&\quad -\int\widehat{H}_1\left(r,\tau,s\right)\gamma_0(r,s)ds\int\widehat{H}_2\left(r,\tau,s\right)2\gamma_0(r,s)\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]ds \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r) \\
&\quad \times\{K_{h_\tau}(\tau-\tau_j)\int\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma_0(r,s)\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]K_{h_\tau}(s-\tau_j)ds\}\times\Phi\left(\gamma_0\left(.\right)\right) \\
&\quad -\Psi\left(\gamma_0\left(.\right)\right)\{\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r-r_i)[K_{h_\tau}(\tau-\tau_j) \\
&\quad \times\int 2\gamma_0(r,s)\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma_0(r,s)\left[\widehat{\gamma}(r,s)-\gamma_0(r,s)\right]K_{h_\tau}(s-\tau_j)ds\},
\end{aligned}
$$

which then allow us to rewrite the above Taylor expansion as

$$
\frac{\int\widehat{H}_1\left(r,\tau,s\right)\widehat{\gamma}(r,s)ds}{\int\widehat{H}_2\left(r,\tau,s\right)\widehat{\gamma}^2(r,s)ds}=\frac{\Psi(\gamma)}{\Phi(\gamma)}+\frac{\Xi(\widehat{\gamma},\gamma)}{\Phi(\gamma)^2}. \tag{16}
$$

First consider $\Psi(\gamma(.))$. As shown in the consistency part above, when scaled by J^{-2}, it converges to

$$
2\overline{L}\left(t_n,r\right)p(\tau)\gamma(r,\tau)\int\gamma(r,s)^2p\left(s\right)ds.
$$

Its stochastic part, which can be written as follows, due to (14)

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r_i-r)\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right] \\
&\times\left[\int_{t_i}^{t_i+\Delta_n}[y(s,\tau_j)-y(t_i,\tau_j)]\gamma(r(s),\tau_k)dW(s)+\int_{t_i}^{t_i+\Delta_n}[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)]\gamma(r(s),\tau_j)dW(s)\right].
\end{aligned}
$$

Denote

$$
U_n\left(x\right) = \frac{1}{J}\sqrt{h_rh_\tau}\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{[nx]-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r_i-r)[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)]
$$

$$\times\{\int_{t_i}^{t_i+\Delta_n}[y(s,\tau_j)-y(t_i,\tau_j)]\gamma(r(s),\tau_k)dW(s)$$
$$+\int_{t_i}^{t_i+\Delta_n}[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)]\gamma(r(s),\tau_j)dW(s)\}.$$

Since $U_n(x)$ is a continuous martingale, by the strong law of large numbers for martingale difference sequences with finite second moment (Hall and Heyde (1986)), $U_n(x)$ converges to zero almost surely, with the rate of convergence can be found from Knight's embedding theorem. Its quadratic variation process $[U_n]_x$ is

$$\begin{aligned}
[U_n]_x &= \frac{1}{J^2}h_r h_\tau \frac{1}{\Delta_n}\sum_{i=1}^{[nx]-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\int_{t_i}^{t_i+\Delta_n}\{[y(s,\tau_j)-y(t_i,\tau_j)]^2\gamma(r(s),\tau_k)^2ds+ \\
&\quad [y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)]^2\gamma(r(s),\tau_j)^2ds \\
&\quad +2\left[y(s,\tau_j)-y(t_i,\tau_j)\right]\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\gamma(r(s),\tau_j)\gamma(r(s),\tau_k)ds \\
&\quad \times K_{h_r}(r_i-r)^2\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right]^2 \\
&= \frac{1}{J^2}4\Delta_n h_r h_\tau\sum_{i=1}^{[nx]-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma(r_i,\tau_j)^2\gamma(r_i,\tau_k)^2 \\
&\quad \times K_{h_r}(r_i-r)^2\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right]^2 \\
&= \frac{1}{J^2}4h_r h_\tau\int_0^{xt_n}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma(r(s),\tau_j)^2\gamma(r(s),\tau_k)^2 \\
&\quad \times K_{h_r}(r_i-r)^2\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right]^2 ds \\
&= \frac{1}{J^2}4h_r h_\tau\int_{-\infty}^{+\infty}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma(a,\tau_j)^2\gamma(a,\tau_k)^2\overline{L}\left(xt_n,a\right) \\
&\quad \times K_{h_r}(a-r)^2\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right]^2 da \\
&= \frac{1}{J^2}4h_\tau\int_{-\infty}^{+\infty}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\gamma(r+h_r q,\tau_j)^2\gamma(r+h_r q,\tau_k)^2\overline{L}\left(xt_n,r+h_r q\right) \\
&\quad \times K(q)^2\left[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\right]^2 dq
\end{aligned}$$

This is

$$
\begin{aligned}
&\simeq 4h_\tau \left\|K\right\|^2 \frac{1}{J^2} \sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} \gamma(r,\tau_j)^2\gamma(r,\tau_k)^2\overline{L}\left(xt_n,r\right) \times [K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)]^2 \\
&= 4\left\|K\right\|^2 \overline{L}\left(xt_n,r\right)\{\frac{1}{J^2}h_\tau \sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} \gamma(r,\tau_j)^2\gamma(r,\tau_k)^2 K_{h_\tau}(\tau-\tau_j)^2\gamma(r,\tau_k)^2 \\
&\quad +\frac{1}{J^2}h_\tau \sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} \gamma(r,\tau_j)^2\gamma(r,\tau_k)^2 K_{h_\tau}(\tau-\tau_k)^2\gamma(r,\tau_j)^2 \\
&\quad +\frac{1}{J^2}h_\tau \sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} \gamma(r,\tau_j)^3\gamma(r,\tau_k)^3 K_{h_\tau}(\tau-\tau_j)K_{h_\tau}(\tau-\tau_k)\} \\
&\simeq 8\left\|K\right\|^2 \overline{L}\left(xt_n,r\right)\left\{\gamma(r,\tau)^2 p\left(\tau\right)\int \gamma(r,s)^4 p\left(s\right)ds\left\|K\right\|^2 + h_\tau p(\tau)\gamma(r,\tau)^6\right\} \\
&\simeq 8\left\|K\right\|^4 \overline{L}\left(xt_n,r\right)p\left(\tau\right)\gamma(r,\tau)^2 \int \gamma(r,s)^4 p\left(s\right)ds.
\end{aligned}
$$

We thus have

$$
U_n\left(1\right) \xrightarrow{d} N\left(0, 8\left\|K\right\|^4 \overline{L}\left(t_n,r\right)p\left(\tau\right)\gamma(r,\tau)^2 \int \gamma(r,s)^4 p\left(s\right)ds\right). \tag{17}
$$

Next consider $\Phi(\gamma(.))$. After scaling by J^{-2}

$$
\begin{aligned}
\frac{1}{J^2}\Phi(\gamma) &= \frac{\Delta_n}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} K_{h_r}(r-r_i) \\
&\quad \times[K_{h_\tau}(\tau-\tau_j)\int \gamma(r,s)^2 K_{h_\tau}(s-\tau_k)ds + K_{h_\tau}(\tau-\tau_k)\int \gamma(r,s)^2 K_{h_\tau}(s-\tau_j)ds] \qquad (18) \\
&\simeq \frac{\Delta_n}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} K_{h_r}(r-r_i[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)^2 + K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)^2] \\
&\simeq \frac{\Delta_n}{J}\sum_{i=1}^{n-1} K_{h_r}(r-r_i)\sum_{j=1}^{J} K_{h_\tau}(\tau-\tau_j)\frac{1}{J}\sum_{k,k\neq j}^{J}\gamma(r,\tau_k)^2 \\
&\quad +\frac{\Delta_n}{J}\sum_{i=1}^{n-1} K_{h_r}(r-r_i)\sum_{k,k\neq j}^{J} K_{h_\tau}(\tau-\tau_k)\frac{1}{J}\sum_{j=1}^{J}\gamma(r,\tau_j)^2
\end{aligned}
$$

$$
\begin{aligned}
&\simeq \frac{\Delta_n}{J}\sum_{i=1}^{n-1} K_{h_r}(r-r_i)\sum_{j=1}^{J} K_{h_\tau}(\tau-\tau_j)\int \gamma(r,s)^2 p\,(s)\,ds \\
&\quad +\frac{\Delta_n}{J}\sum_{i=1}^{n-1} K_{h_r}(r-r_i)\sum_{k,k\neq j}^{J} K_{h_\tau}(\tau-\tau_k)\int \gamma(r,s)^2 p\,(s)\,ds \\
&\simeq 2\overline{L}\,(t_n,r)\,p\,(\tau)\int \gamma(r,s)^2 p\,(s)\,ds.
\end{aligned}
$$

Now consider $\Xi\,(.)$. We have

$$
\begin{aligned}
\frac{1}{J^4}\Xi\,(.) &= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J} \Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)\frac{1}{J^2}\Phi(\gamma)\times \\
&\quad \{K_{h_\tau}(\tau-\tau_j)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_j)ds\} \\
&\quad -\frac{1}{J^2}\Delta_n\frac{1}{J^2}\Psi(\gamma(.))\{\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J} K_{h_r}(r-r_i) \\
&\quad \times[K_{h_\tau}(\tau-\tau_j)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_j)ds] \\
&= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J} 2\gamma(r,\tau_j)\gamma(r,\tau_k)\Delta_n K_{h_r}(r_i-r)\}\frac{1}{J^2}\Phi(\gamma(.)) \\
&\quad \times\{K_{h_\tau}(\tau-\tau_j)\int 2\gamma(r,s)[\widehat{\gamma}(r,s)-\gamma(r,s)]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_j)ds\} \\
&\quad -\frac{1}{J^2}\Delta_n\frac{1}{J^2}\Psi(\gamma(.))\{\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J} K_{h_r}(r-r_i)[K_{h_\tau}(\tau-\tau_j) \\
&\quad \times\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int 2\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]K_{h_\tau}(s-\tau_j)ds]\},
\end{aligned}
$$

where the second equation follows (14). It is apparent that the two components are perfectly correlated, where the variation comes from $\widehat{\gamma}(r,s)-\gamma(r,s)$.

Its first component approaches

$$\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J} 2\gamma(r,\tau_j)\gamma(r,\tau_k)\Delta_n K_{h_r}(r_i-r)\frac{1}{J^2}\Phi(\gamma(.))$$
$$\{K_{h_\tau}(\tau-\tau_j)\left[\widehat{\gamma}(r,\tau_k)-\gamma(r,\tau_k)\right]+K_{h_\tau}(\tau-\tau_k)\left[\widehat{\gamma}(r,\tau_j)-\gamma(r,\tau_j)\right]\}$$

$$\begin{aligned}
&= 2\gamma(r,\tau)\overline{L}\left(t_n,r\right)p\left(\tau\right)\int\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]p\left(s\right)ds\\
&\quad\times 2\overline{L}\left(t_n,r\right)p\left(\tau\right)\gamma(r,\tau)\int\gamma(r,s)^2p\left(s\right)ds\\
&= 4\overline{L}\left(t_n,r\right)^2p\left(\tau\right)^2\gamma(r,\tau)^2\int\gamma(r,s)^2p\left(s\right)ds\int\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]p\left(s\right)ds.
\end{aligned}$$

The second component also approaches

$$\frac{1}{J^2}2\gamma(r,\tau)\Delta_n\{\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r-r_i)\frac{1}{J^2}\Psi(\gamma(.))$$
$$\times[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)\left[\widehat{\gamma}(r,\tau_k)-\gamma(r,\tau_k)\right]+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\left[\widehat{\gamma}(r,\tau_j)-\gamma(r,\tau_j)\right]]\}$$

$$\begin{aligned}
&= 2\gamma(r,\tau)\overline{L}\left(t_n,r\right)p\left(\tau\right)\int\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]p\left(s\right)ds\times 2\overline{L}\left(t_n,r\right)p\left(\tau\right)\gamma(r,\tau)\int\gamma(r,s)^2p\left(s\right)ds\\
&= 4\overline{L}\left(t_n,r\right)^2p\left(\tau\right)^2\gamma(r,\tau)^2\int\gamma(r,s)^2f\left(s\right)ds\int\gamma(r,s)\left[\widehat{\gamma}(r,s)-\gamma(r,s)\right]p\left(s\right)ds
\end{aligned}$$

so they are cancelled out, which means that the variation of the right hand side will be driven by $\Psi(\gamma)/\Phi(\gamma)$.

From (16), (17) and (18), we obtain

$$\begin{aligned}
&\frac{1}{J}\sqrt{h_rh_\tau}\frac{1}{\frac{1}{J^2}}\frac{1}{\sqrt{\Delta_n}}\left[\widehat{\gamma}(r,\tau)-\gamma(r,\tau)\right]\\
&= \frac{\frac{1}{J}\sqrt{h_rh_\tau}\frac{1}{\sqrt{\Delta_n}}}{\frac{1}{J^2}}\left[\frac{\Psi\left(\gamma\right)}{\Phi\left(\gamma\right)}-\gamma(r,\tau)\right]\\
&\xrightarrow{d}\frac{1}{2\overline{L}\left(t_n,r\right)p\left(\tau\right)\int\gamma(r,s)^2p\left(\tau\right)ds}N\left(0,8\left\|K\right\|^4\overline{L}\left(t_n,r\right)p\left(\tau\right)\gamma(r,\tau)^2\int\gamma(r,s)^4p\left(s\right)ds\right),
\end{aligned}$$

or

$$J\sqrt{\frac{h_rh_\tau\overline{L}\left(t_n,r\right)}{\Delta_n}}\left[\widehat{\gamma}(r,\tau)-\gamma(r,\tau)\right]\xrightarrow{d}N\left(0,2\left\|K\right\|^4\frac{\gamma(r,\tau)^2\int\gamma(r,s)^4p\left(s\right)ds}{p\left(\tau\right)\left(\int\gamma(r,s)^2p\left(s\right)ds\right)^2}\right)$$

the results stated in Theorem 1. ■

8.2 Proof of Theorem 2

The multiplicative separability assumptions that we invoke for the volatility structures and the local time allow us to have the following formula, which is central to our derivations, as in the one-factor case. For the two-variable case, for any multiplicatively separable kernel function $K(\cdot)$ and continuous bounded multiplicatively separable function $f(\cdot)$,

$$\begin{aligned}
&\lim_{h\downarrow 0}\frac{1}{h}\int_0^t K\left(\frac{(r,l)-(r(s),l(s))}{h}\right) f\left(r(s),l(s)\right)ds \\
=&\left[\lim_{h_r\downarrow 0}\frac{1}{h_r}\int_{-\infty}^{+\infty}K\left(\frac{r-a}{h_r}\right)f_r(a)\overline{L}_r(t,a)da\right]\left[\lim_{h_l\downarrow 0}\frac{1}{h_l}\int_{-\infty}^{+\infty}K\left(\frac{x-a}{h_l}\right)f_l(a)\overline{L}_l(t,a)da\right] \\
=&\left[\lim_{h_r\downarrow 0}\int_{-\infty}^{+\infty}K(q)f_r(x+h_rq)\overline{L}_r(t,x+h_rq)dq\right]\left[\lim_{h_l\downarrow 0}\int_{-\infty}^{+\infty}K(q)f_l(x+h_lq)\overline{L}_l(t,x+h_lq)dq\right] \\
\rightarrow\;&\overline{L}_r(t,r)f_r(r)\overline{L}_l(t,r)f_l(l)=\overline{L}(t,(r,l))f(r,l).
\end{aligned}$$

The result is essentially identical to the one dimensional case.

We first derive the first order condition for the two factor model. Differentiate the sample criteria function with respect to $\eta(.)$ we have

$$\begin{aligned}
0 =& \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\int_{\widetilde{r}_i,\widetilde{l}_i}\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)-\left(\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)+\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)\right)\Delta_n\right] \\
&\times\left(\delta\left(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j\right)\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)+\delta\left(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k\right)\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\right) \\
&\times\Delta_n K_{h_r}(r_i-\widetilde{r}_i)K_{h_r}(l_i-\widetilde{l}_i)K_{\tau}(\tau_j-\widetilde{\tau}_j)K_{h_\tau}(\tau_k-\widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j d\widetilde{\tau}_k
\end{aligned}$$

and with respect to $\gamma(.)$

$$\begin{aligned}
0 =& \sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\int\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)-\left(\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)+\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)\right)\Delta_n\right] \\
&\times\left(\delta\left(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j\right)\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)+\delta\left(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k\right)\widehat{\eta}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\right) \\
&\times\Delta_n K_{h_r}(r_i-\widetilde{r}_i)K_{h_r}(l_i-\widetilde{l}_i)K_{h_\tau}(\tau_j-\widetilde{\tau}_j)K_{h_\tau}(\tau_k-\widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j d\widetilde{\tau}_k.
\end{aligned}$$

The first order condition stated in the body of the paper is obtained through the following manipulation, for representative terms, which are

$$\begin{aligned}
&\int\delta(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_j)\widehat{\gamma}(\widetilde{r}_i,\widetilde{l}_i,\widetilde{\tau}_k)K_{h_r}(r_i-\widetilde{r}_i)K_{h_r}(l_i-\widetilde{l}_i)K_{h_\tau}(\tau_j-\widetilde{\tau}_j)K_{h_\tau}(\tau_k-\widetilde{\tau}_k)d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j d\widetilde{\tau}_k \\
=\;& K_{h_r}(r-r_i)K_{h_r}(l-l_i)[K_{h_\tau}(\tau-\tau_j)\int\widehat{\gamma}(r,l,s)K_{h_\tau}(s-\tau_k)ds \\
&+K_{h_\tau}(\tau-\tau_k)\int\widehat{\gamma}(r,l,s)K_{h_\tau}(s-\tau_j)ds],
\end{aligned}$$

$$
\begin{aligned}
&\int \delta\left(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_j\right) \widehat{\gamma}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_k) \widehat{\gamma}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_j) \widehat{\gamma}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_k) \\
&\times K_{h_r}(r_i - \widetilde{r}_i) K_{h_r}(l_i - \widetilde{l}_i) K_\tau(\tau_j - \widetilde{\tau}_j) K_{h_\tau}(\tau_k - \widetilde{\tau}_k) d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j \\
= \;& \widehat{\gamma}(r, l, \tau) K_{h_r}(r - r_i) K_{h_r}(l - l_i) \\
&\times [K_{h_\tau}(\tau - \tau_j) \int \widehat{\gamma}(r, l, s)^2 K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \widehat{\gamma}(r, l, s)^2 K_{h_\tau}(s - \tau_j) ds]
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{\widetilde{r}_i} \int_{\widetilde{\tau}_j} \int_{\widetilde{\tau}_k} \delta\left(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_j\right) \widehat{\gamma}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_k) \widehat{\eta}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_j) \widehat{\eta}(\widetilde{r}_i, \widetilde{l}_i, \widetilde{\tau}_k) \\
&\times K_{h_r}(r_i - \widetilde{r}_i) K_{h_r}(l_i - \widetilde{l}_i) K_{h_\tau}(\tau_j - \widetilde{\tau}_j) K_{h_\tau}(\tau_k - \widetilde{\tau}_k) d\widetilde{r}_i d\widetilde{l}_i d\widetilde{\tau}_j \\
= \;& \widehat{\eta}(r, l, \tau) K_{h_r}(r - r_i) K_{h_r}(l - l_i) \\
&\times \left[K_{h_\tau}(\tau - \tau_j) \int \widehat{\eta}(r, l, s) \widehat{\gamma}(r, l, t) K_{h_\tau}(s - \tau_k) dt + K_{h_\tau}(\tau - \tau_k) \int \widehat{\eta}(r, l, s) \widehat{\gamma}(r, l, s) K_{h_\tau}(s - \tau_j) dt\right].
\end{aligned}
$$

Therefore, we have for $\widehat{\gamma}(.)$

$$
\begin{aligned}
0 \;=\; & \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \Delta y(t_i, \tau_j) \Delta y(t_i, \tau_k) K_{h_r}(r_i - r) K_{h_r}(l_i - l) \times \\
& \left[K_{h_\tau}(\tau - \tau_j) \int \widehat{\gamma}(r, l, s) K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \widehat{\gamma}(r, l, s) K_{h_\tau}(s - \tau_j) ds\right] \\
& - \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \widehat{\gamma}(r, l, \tau) K_{h_r}(r - r_i) K_{h_r}(l - l_i) \Delta_n \\
& \times \left[K_{h_\tau}(\tau - \tau_j) \int \widehat{\gamma}(r, l, s)^2 K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \widehat{\gamma}(r, l, s)^2 K_{h_\tau}(s - \tau_j) ds\right] \\
& - \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \widehat{\eta}(r, l, \tau) K_{h_r}(r - r_i) K_{h_r}(l - l_i) \Delta_n \\
& \times \left[K_{h_\tau}(\tau - \tau_j) \int \widehat{\eta}(r, l, s) \widehat{\gamma}(r, l, s) K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \widehat{\eta}(r, l, s) \widehat{\gamma}(r, l, s) K_{h_\tau}(t - \tau_j) ds\right],
\end{aligned}
$$

and similarly for $\widehat{\eta}(.)$

$$
\begin{aligned}
0 \;=\; & \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \Delta y(t_i, \tau_j) \Delta y(t_i, \tau_k) K_{h_r}(r_i - r) K_{h_r}(l_i - l) \\
& \times \left[K_\tau(\tau - \tau_j) \int \widehat{\eta}(r, l, s) K_{h_\tau}(s - \tau_k) ds + K_{h_\tau}(\tau - \tau_k) \int \widehat{\eta}(r, l, s) K_{h_\tau}(s - \tau_j) ds\right]
\end{aligned}
$$

$$
-\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\widehat{\eta}(r,l,\tau)K_{h_r}(r-r_i)K_{h_r}(l-l_i)\Delta_n
$$
$$
\times\left[K_{h_\tau}(\tau-\tau_j)\int\widehat{\eta}(r,l,s)^2K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\widehat{\eta}(r,l,s)^2K_{h_\tau}(s-\tau_j)ds\right]
$$
$$
-\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\widehat{\gamma}(r,l,\tau)K_{h_r}(r-r_i)K_{h_r}(l-l_i)\Delta_n
$$
$$
\times\left[K_{h_\tau}(\tau-\tau_j)\int\widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s)K_{h_\tau}(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s)K_{h_\tau}(s-\tau_j)ds\right].
$$

Similar to the one-factor case above, we will derive the asymptotics for our estimate by examining the first order condition in the limit. For the interest of brevity, we will skip steps that should be familiar now via the proof of the one-factor case done in details earlier. When $h\to 0$, $n\to\infty$, $J\to\infty$, observing that

$$
\begin{aligned}
\int\widehat{\eta}(r,l,s)K_{h_\tau}(t-\tau_j)ds &\to \widehat{\eta}(r,l,\tau_j)\\
\int\widehat{\eta}(r,l,s)^2K_{h_\tau}(t-\tau_j)ds &\to \widehat{\eta}(r,l,\tau_j)^2\\
\int_s\widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s)K_{h_\tau}(t-\tau_j)ds &\to \widehat{\eta}(r,l,\tau_j)\widehat{\gamma}(r,l,\tau_j)
\end{aligned}
$$

and similarly for the alike terms. The analogue to (14) is

$$
\begin{aligned}
\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) =& \int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_j\right)-y\left(t_i,\tau_j\right)\right]\alpha(s,\tau_k)ds\\
&+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_j\right)-y\left(t_i,\tau_j\right)\right]\gamma(r(s),l\left(s\right),\tau_k)dW_1\left(s\right)\\
&+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_j\right)-y\left(t_i,\tau_j\right)\right]\eta(r(s),l\left(s\right),\tau_k)dW_2\left(s\right) \qquad (19)\\
&+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\alpha(s,\tau_j)ds\\
&+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\gamma(r(s),l\left(s\right),\tau_j)dW_1\left(s\right)\\
&+\int_{t_i}^{t_i+\Delta_n}\left[y\left(s,\tau_k\right)-y\left(t_i,\tau_k\right)\right]\eta(r(s),l\left(s\right),\tau_j)dW_2\left(s\right)\\
&+\int_{t_i}^{t_i+\Delta_n}\left[\gamma(r(s),l\left(s\right),\tau_j)\gamma(r(s),\tau_k)+\eta(r(s),l\left(s\right),\tau_j)\eta(r(s),l\left(s\right),\tau_k)\right]ds.
\end{aligned}
$$

The above term has quadratic variation as

$$
\begin{aligned}
\left[\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)\right]=2\quad&\left[\gamma(r(t_i),l(t_i),\tau_k)^2+\eta(r(t_i),l(t_i),\tau_k)^2\right]\Delta_n^2\\
&+2\left[\gamma(r(t_i),l(t_i),\tau_j)\gamma(r(t_i),l(t_i),\tau_k)+\eta(r(t_i),l(t_i),\tau_j)\eta(r(t_i),l(t_i),\tau_k)\right]^2\Delta_n^2
\end{aligned}
$$

Another useful approximation can also be derived from (19)

$$\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) \simeq 2\left[\gamma(r_i,l_i,\tau_j)\gamma(r_i,l_i,\tau_k)+\eta(r_i,l_i,\tau_j)\eta(r_i,l_i,\tau_k)\right]\Delta_n.$$

For the first term in the first order condition , using (19) and ignoring small terms (terms bounded by assumption 2) as done in the one-factor case, we have (scaled by J^{-2})

$$\begin{aligned}
&= \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)K_{h_r}(l_i-l)\\
&\quad\times\left[K_\tau(\tau-\tau_j)\int\widehat{\gamma}(r,l,s)K_\tau(s-\tau_k)ds+K_{h_\tau}(\tau-\tau_k)\int\widehat{\gamma}(r,l,t)K_{h_\tau}(s-\tau_j)ds\right]\\
&\simeq \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)K_{h_r}(l_i-l)\\
&\quad\times\left[K_{h_\tau}(\tau-\tau_j)\widehat{\gamma}(r,l,\tau_k)+K_{h_\tau}(\tau-\tau_k)\widehat{\gamma}(r,l,\tau_j)\right]\\
&\simeq \frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\int_{t_i}^{t_i+\Delta_n}\left[\gamma(r(s),l\left(s\right),\tau_j)\gamma(r(s),\tau_k)+\eta(r(s),l\left(s\right),\tau_j)\eta(r(s),l\left(s\right),\tau_k)\right]ds\\
&\quad\times K_{h_r}(r_i-r)K_{h_r}(l_i-l)\left[K_{h_\tau}(\tau-\tau_j)\widehat{\gamma}(r,l,\tau_k)+K_{h_\tau}(\tau-\tau_k)\widehat{\gamma}(r,l,\tau_j)\right]\\
&\simeq \frac{\Delta_n}{J^2}\sum_{i=1}^{n-1}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\left[\gamma(r_i,l_i,\tau_j)\gamma(r_i,l_i,\tau_k)+\eta(r_i,l_i,\tau_j)\eta(r_i,l_i,\tau_k)\right]\\
&\quad\times K_{h_r}(r_i-r)K_{h_r}(l_i-l)\left[K_{h_\tau}(\tau-\tau_j)\widehat{\gamma}(r,l,\tau_k)+K_{h_\tau}(\tau-\tau_k)\widehat{\gamma}(r,l,\tau_j)\right]\\
&\simeq 2\overline{L}\left(t_n,(r,l)\right)p\left(\tau\right)\left[\gamma(r,l,\tau)\int\gamma(r,l,s)\widehat{\gamma}(r,l,s)p\left(s\right)ds+\eta(r,l,\tau)\int\eta(r,l,s)\widehat{\gamma}(r,l,s)p\left(s\right)ds\right].
\end{aligned}$$

For the second term

$$\begin{aligned}
&\frac{\Delta_n}{J^2}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}\widehat{\gamma}(r,l,\tau)K_{h_r}(r-r_i)K_{h_r}(l-l_i)\\
&\times\left[K_{h_\tau}(\tau-\tau_j)\int\widehat{\gamma}(r,l,s)^2K_{h_\tau}(s-\tau_k)dt+K_{h_\tau}(\tau-\tau_k)\int\widehat{\gamma}(r,l,s)^2K_{h_\tau}(s-\tau_j)dt\right]\\
\simeq\ &\frac{\Delta_n}{J^2}\widehat{\gamma}(r,l,\tau)\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{\substack{k=1\\k\neq j}}^{J}K_{h_r}(r-r_i)K_{h_r}(l-l_i)\\
&\times\left[K_{h_\tau}(\tau-\tau_j)\widehat{\gamma}(r,l,\tau_k)^2+K_{h_\tau}(\tau-\tau_k)\widehat{\gamma}(r,l,\tau_j)^2\right]\\
\simeq\ &2\overline{L}\left(t_n,r\right)p\left(\tau\right)\widehat{\gamma}(r,l,\tau)\int\widehat{\gamma}(r,l,s)^2p\left(s\right)ds.
\end{aligned}$$

And for the last term, similarly

$$
\begin{aligned}
& \frac{\Delta_n}{J^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ k \neq j}}^{J} \sum_{k=1}^{J} \widehat{\eta}(r,l,\tau) K_{h_r}(r-r_i) K_{h_r}(l-l_i) \\
& \times \left[K_{h_\tau}(\tau-\tau_j) \int \widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s) K_{h_\tau}(s-\tau_k) ds + K_{h_\tau}(\tau-\tau_k) \int \widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s) K_{h_\tau}(s-\tau_j) ds \right] \\
\simeq \; & \frac{\Delta_n}{J^2} \widehat{\eta}(r,l,\tau) \sum_{i=1}^{n} \sum_{\substack{j=1 \\ k \neq j}}^{J} \sum_{k=1}^{J} K_{h_r}(r-r_i) K_{h_r}(l-l_i) \\
& \times \left[K_{h_\tau}(\tau-\tau_j) \widehat{\eta}(r,l,\tau_k)\widehat{\gamma}(r,l,\tau_k) + K_{h_\tau}(\tau-\tau_k) \widehat{\eta}(r,l,\tau_j)\widehat{\gamma}(r,l,\tau_j) \right] \\
\simeq \; & 2\overline{L}\left(t_n, r\right) p(\tau) \widehat{\eta}(r,l,\tau) \int \widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s) p\left(s\right) ds.
\end{aligned}
$$

So in the limit, the first order condition is

$$
\begin{aligned}
0 \;=\; & \left[\gamma(r,l,\tau) \int \gamma(r,l,s)\widehat{\gamma}(r,l,s) p\left(s\right) ds + \eta(r,l,\tau) \int \eta(r,l,s)\widehat{\gamma}(r,l,s) p\left(s\right) ds \right] \\
& - \widehat{\gamma}(r,l,\tau) \int \widehat{\gamma}(r,l,s)^2 p\left(s\right) ds - \widehat{\eta}(r,l,\tau) \int \widehat{\eta}(r,l,s)\widehat{\gamma}(r,l,s) p\left(s\right) ds,
\end{aligned}
$$

and similarly, the second first order condition is

$$
\begin{aligned}
0 \;=\; & \left[\eta(r,l,\tau) \int \eta(r,l,s)\widehat{\eta}(r,l,s) p\left(s\right) ds + \gamma(r,l,\tau) \int \gamma(r,l,s)\widehat{\eta}(r,l,s) p\left(s\right) ds \right] \\
& - \widehat{\eta}(r,l,\tau) \int \widehat{\eta}(r,l,s)^2 p\left(s\right) ds - \widehat{\gamma}(r,l,\tau) \int \widehat{\gamma}(r,l,s)\widehat{\eta}(r,l,s) p\left(s\right) ds,
\end{aligned}
$$

which has a solution $\widehat{\eta}(r,l,\tau) = \eta(r,l,\tau)$ and $\widehat{\gamma}(r,l,\tau) = \gamma(r,l,\tau)$. Uniqueness can be proved using arguments as in the one-factor case.

We now discuss the asymptotic distribution. Re-write the first order condition as

$$
\begin{aligned}
\begin{bmatrix} \widehat{\gamma}(r,l,\tau) \\ \widehat{\eta}(r,l,\tau) \end{bmatrix} \;=\; & \frac{1}{\int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}(r,s)^2 ds \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\eta}(r,s)^2 ds - \left[\int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}(r,s)\widehat{\eta}(r,s) ds \right]^2} \\
& \times \begin{bmatrix} \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\eta}(r,s)^2 ds & -\int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}(r,s)\widehat{\eta}(r,s) ds \\ -\int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}(r,s)\widehat{\eta}(r,s) ds & \int \widehat{H}_2\left(r,\tau,s\right) \widehat{\gamma}(r,s)^2 ds \end{bmatrix} \\
& \times \begin{bmatrix} \int \widehat{H}_1\left(r,\tau,s\right) \widehat{\gamma}(r,s) ds \\ \int \widehat{H}_1\left(r,\tau,s\right) \widehat{\eta}(r,s) ds \end{bmatrix}.
\end{aligned}
$$

Using first-order Taylor expansion, we can show that the variation, similar to the 1-factor case, comes only from the following term

$$
\frac{1}{\int \widehat{H}_2\left(r,\tau,s\right) \gamma(r,s)^2 ds \int \widehat{H}_2\left(r,\tau,s\right) \eta(r,s)^2 ds - \left[\int \widehat{H}_2\left(r,\tau,s\right) \gamma(r,s)\eta(r,s) ds \right]^2}
$$

$$\times \begin{bmatrix} \int \widehat{H}_2(r,\tau,s)\,\eta(r,s)^2 ds & -\int \widehat{H}_2(r,\tau,s)\,\gamma(r,s)\eta(r,s)ds \\ -\int \widehat{H}_2(r,\tau,s)\,\gamma(r,s)\eta(r,s)ds & \int \widehat{H}_2(r,\tau,s)\,\gamma(r,s)^2 ds \end{bmatrix} \begin{bmatrix} \int \widehat{H}_1(r,\tau,s)\,\gamma(r,s)ds \\ \int \widehat{H}_1(r,\tau,s)\,\eta(r,s)ds \end{bmatrix}.$$

For the rate of convergence of $\widehat{\gamma}(r,l,\tau)$, consider the following quantity

$$\begin{aligned} \Theta_n &= \frac{1}{J}\sqrt{h_r h_\tau}\frac{1}{\sqrt{\Delta_n}}\frac{1}{2\overline{L}(t_n,(r,l))\,p(\tau)} \times [\int \widehat{H}_2(r,\tau,s)\,\eta(r,s)^2 ds \int \widehat{H}_1(r,\tau,s)\,\gamma(r,s)ds \\ &\quad - \int \widehat{H}_2(r,\tau,s)\,\gamma(r,s)\eta(r,s)ds \int \widehat{H}_1(r,\tau,s)\,\eta(r,s)ds], \end{aligned}$$

which is

$$\begin{aligned} & \frac{1}{J}\sqrt{h_r h_l h_\tau}\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n-1}\sum_{\substack{j=1 \\ k\neq j}}^{J}\sum_{k=1}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)K_{h_r}(l_i-l) \\ & \times\{\int \eta(r,s)^2 p(s)\,ds[K_{h_\tau}(\tau-\tau_j)\int \gamma(r,l,s)K_{h_\tau}(s-\tau_k)ds + K_{h_\tau}(\tau-\tau_k)\int \gamma(r,s)K_{h_\tau}(s-\tau_j)ds] \\ & -\int \gamma(r,s)\eta(r,s)p(s)\,ds[K_{h_\tau}(\tau-\tau_j)\int \eta(r,l,s)K_{h_\tau}(s-\tau_k)ds \\ & +K_{h_\tau}(\tau-\tau_k)\int \eta(r,s)K_{h_\tau}(s-\tau_j)ds]\} \\ = & \frac{1}{J}\sqrt{h_r h_l h_\tau}\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n-1}\sum_{\substack{j=1 \\ k\neq j}}^{J}\sum_{k=1}^{J}\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)K_{h_r}(r_i-r)K_{h_r}(l_i-l) \\ & \times\{K_{h_\tau}(\tau-\tau_j)\left[\gamma(r,l,\tau_k)\int \eta(r,s)^2 p(s)\,ds - \eta(r,l,\tau_k)\int \gamma(r,l,s)\eta(r,s)p(s)\,ds\right] \\ & +K_{h_\tau}(\tau-\tau_k)\left[\gamma(r,l,\tau_j)\int \eta(r,s)^2 p(s)\,ds - \eta(r,l,\tau_j)\int \gamma(r,l,s)\eta(r,s)p(s)\,ds\right]\}, \end{aligned}$$

where we have used the fact that

$$\begin{aligned} \frac{1}{J^2}\int \widehat{H}_2(r,\tau,s)\,\eta(r,s)^2 ds &= \Delta_n\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1 \\ k\neq j}}^{J}\sum_{k=1}^{J}K_{h_r}(r_i-r)K_{h_r}(l_i-l)\times \\ &\quad [K_{h_\tau}(\tau-\tau_j)\int \eta(r,l,s)^2 K_{h_\tau}(s-\tau_k)ds \\ &\quad +K_{h_\tau}(\tau-\tau_k)\int \eta(r,s)^2 K_{h_\tau}(s-\tau_j)ds] \\ &= \Delta_n\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1 \\ k\neq j}}^{J}\sum_{k=1}^{J}K_{h_r}(r_i-r)K_{h_r}(l_i-l) \\ &\quad \times\left[K_{h_\tau}(\tau-\tau_j)\eta(r,l,\tau_k)^2 + K_{h_\tau}(\tau-\tau_k)\eta(r,\tau_j)^2\right] \\ &\simeq 2\overline{L}(t_n,(r,l))\,p(\tau)\int \eta(r,s)^2 p(s)\,ds, \end{aligned}$$

and

$$
\begin{aligned}
\frac{1}{J^2}\int \widehat{H}_2(r,\tau,s)\,\gamma(r,s)\eta(r,s)ds &= \Delta_n\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} K_{h_r}(r_i-r)K_{h_r}(l_i-l) \\
&\quad \times[K_{h_\tau}(\tau-\tau_j)\int \gamma(r,s)\eta(r,s)K_{h_\tau}(s-\tau_k)ds \\
&\quad +K_{h_\tau}(\tau-\tau_k)\int \gamma(r,s)\eta(r,s)K_{h_\tau}(s-\tau_j)ds] \\
&= \Delta_n\frac{1}{J^2}\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J} K_{h_r}(r_i-r)K_{h_r}(l_i-l)\times \\
&\quad \times[K_{h_\tau}(\tau-\tau_j)\gamma(r,\tau_k)\eta(r,\tau_k)+K_{h_\tau}(\tau-\tau_k)\gamma(r,\tau_j)\eta(r,\tau_j)] \\
&\simeq 2\overline{L}\,(t_n,(r,l))\,p\,(\tau)\int \gamma(r,s)\eta(r,s)p\,(s)\,ds.
\end{aligned}
$$

For the stochastic part of Θ, note that it is driven by the following terms belong to $\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k)$

$$
\begin{aligned}
\Delta y(t_i,\tau_j)\Delta y(t_i,\tau_k) &= +\int_{t_i}^{t_i+\Delta_n}[y\,(s,\tau_j)-y\,(t_i,\tau_j)]\,\gamma(r(s),l\,(s)\,,\tau_k)dW_1\,(s) \\
&\quad +\int_{t_i}^{t_i+\Delta_n}[y\,(s,\tau_j)-y\,(t_i,\tau_j)]\,\eta(r(s),l\,(s)\,,\tau_k)dW_2\,(s) \\
&\quad +\int_{t_i}^{t_i+\Delta_n}[y\,(s,\tau_k)-y\,(t_i,\tau_k)]\,\gamma(r(s),l\,(s)\,,\tau_j)dW_1\,(s) \\
&\quad +\int_{t_i}^{t_i+\Delta_n}[y\,(s,\tau_k)-y\,(t_i,\tau_k)]\,\eta(r(s),l\,(s)\,,\tau_j)dW_2(s).
\end{aligned}
$$

The quadratic variation of Θ_n is

$$
\begin{aligned}
[\Theta_n] &= \frac{1}{J^2}h_rh_lh_\tau\Delta_n\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\{2[\gamma(r(t_i),l(t_i),\tau_j)^2+\eta(r(t_i),l(t_i),\tau_j)^2] \\
&\quad \times[\gamma(r(t_i),l(t_i),\tau_k)^2+\eta(r(t_i),l(t_i),\tau_k)^2] \\
&\quad +2[\gamma(r(t_i),l(t_i),\tau_j)\gamma(r(t_i),l(t_i),\tau_k)+\eta(r(t_i),l(t_i),\tau_j)\eta(r(t_i),l(t_i),\tau_k)]^2\} \\
&\quad \times K_{h_r}(r_i-r)^2K_{h_r}(l_i-l)^2 \\
&\quad \times\{K_{h_\tau}(\tau-\tau_j)\left[\gamma(r,l,\tau_k)\int\eta(r,s)^2p\,(s)\,ds-\eta(r,l,\tau_k)\int\gamma(r,l,s)\eta(r,s)p\,(s)\,ds\right] \\
&\quad +K_{h_\tau}(\tau-\tau_k)\left[\gamma(r,l,\tau_j)\int\eta(r,s)^2p\,(s)\,ds-\eta(r,l,\tau_j)\int\gamma(r,l,s)\eta(r,s)p\,(s)\,ds\right]\}^2 \\
&\simeq \frac{1}{J^2}h_rh_lh_\tau\Delta_n\sum_{i=1}^{n-1}\sum_{\substack{j=1\\k\neq j}}^{J}\sum_{k=1}^{J}\{2[\gamma(r(t_i),l(t_i),\tau_j)^2+\eta(r(t_i),l(t_i),\tau_j)^2]
\end{aligned}
$$

$$\times[\gamma(r(t_i),l(t_i),\tau_k)^2+\eta(r(t_i),l(t_i),\tau_k)^2]$$
$$+2[\gamma(r(t_i),l(t_i),\tau_j)\gamma(r(t_i),l(t_i),\tau_k)+\eta(r(t_i),l(t_i),\tau_j)\eta(r(t_i),l(t_i),\tau_k)]^2\}$$
$$\times\{K_{h_\tau}(\tau-\tau_j)[\gamma(r,l,\tau_k)\int\eta(r,s)^2p(s)\,ds-\eta(r,l,\tau_k)\int\gamma(r,l,s)\eta(r,s)p(s)\,ds]$$
$$+K_{h_\tau}(\tau-\tau_k)[\gamma(r,l,\tau_j)\int\eta(r,s)^2p(s)\,ds-\eta(r,l,\tau_j)\int\gamma(r,l,s)\eta(r,s)p(s)\,ds]\}^2$$

Let $G_1=\int\gamma(r,s)^2p(s)\,ds$, $G_2=\int\eta(r,s)^2p(s)\,ds$, $G_3=\int\gamma(r,s)\eta(r,s)p(s)\,ds$. Then the quadratic variation of Θ_n is

$$\begin{aligned}
[\Theta_n] &= \|K\|^6\,\overline{L}(t_n,r)\,p(\tau)\\
&\times\{4\left[\gamma(r,l,\tau)^2+\eta(r,l,\tau)^2\right]\int\left[\gamma(r,l,s)^2+\eta(r,l,s)^2\right]\left[\gamma(r,l,s)G_2-\eta(r,l,s)G_3\right]p(s)\,ds\\
&+2\gamma(r,l,\tau)^2\int\gamma(r,l,s)^2\left[\gamma(r,l,s)G_2-\eta(r,l,s)G_3\right]p(s)\,ds\\
&+2\eta(r,l,\tau)^2\int\eta(r,l,s)^2\left[\gamma(r,l,s)G_2-\eta(r,l,s)G_3\right]p(s)\,ds\\
&+2\gamma(r,l,\tau)\eta(r,l,\tau)\int\gamma(r,l,s)\eta(r,l,\tau)\left[\gamma(r,l,s)G_2-\eta(r,l,s)G_3\right]p(s)\,ds\}
\end{aligned}$$

The last term we need to calculate to find the asymptotic distribution of our estimators is

$$\frac{1}{\int\widehat{H}_2(r,\tau,s)\,\gamma(r,s)^2ds\int\widehat{H}_2(r,\tau,s)\,\eta(r,s)^2ds-\left[\int\widehat{H}_2(r,\tau,s)\,\gamma(r,s)\eta(r,s)ds\right]^2}.$$

The denominator, scaled by J^{-2} , as shown in earlier approach, is

$$\begin{aligned}
&4\left[\overline{L}(t_n,(r,l))\,p(\tau)\int\gamma(r,s)^2p(s)\,ds\right]\left[\overline{L}(t_n,(r,l))\,p(\tau)\int\eta(r,s)^2p(s)\,ds\right]\\
&-4\left[\overline{L}(t_n,(r,l))\,p(\tau)\int\gamma(r,s)\eta(r,s)p(s)\,ds\right]^2\\
=\;&4\left[\overline{L}(t_n,(r,l))\,p(\tau)\right]^2\left[\int\gamma(r,s)^2p(s)\,ds\int\eta(r,s)^2p(s)\,ds-\left[\int\gamma(r,s)\eta(r,s)p(s)\,ds\right]^2\right],
\end{aligned}$$

giving the results reported in Theorem 2. ■

References

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

[2] Balduzzi, P., S.Das and S. Foresi (1998). The Central Tendency: A Second Factor in Bond Yields. Review of Economics and Statistics, 80, 62-72

[3] Balduzzi, P., S. Das, S. Foresi and R. Sundaram (1996). A simple Approach to three-factor affine term structure models. Journal of Fixed Income, 6, 43-53

[4] Balduzzi, P., S. Das, S. Foresi and R. Sundaram (1998). Stochastic Mean Models of the Term Structure of Interest Rates. Working Paper.

[5] Backus, D., S. Foresi, C. Telmer (1998). Discrete-Time Models of Bond Pricing. NBER Working Paper 6736.

[6] 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.

[7] Bosq, D. (1999). *Nonparametric Statistics for Stochastic Processes.* Lecture Notes in Statistics. Springer.

[8] Boudoukh, J., M. Richardson, R. Stanton and R. Whitelaw (1998). Pricing Mortgage-backed Securities in a Multifactor Interest Rate Environment: A Multivariate Density Estimation Approach. Working Paper, Stern School of Business, New York University.

[9] Brennan, M. and E. Schwartz (1979). A Continuous Time Approach to the Pricing of Bonds. Journal of Banking and Finance 3, 133-155.

[10] Brugiere, P. (1991). Estimation of the Coefficient of a Diffusion Process in the Multidimensional Case. C. R. Acad. Sci., Paris, t.312 (Serie I), 999-1004.

[11] 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.

[12] Buser, Hendershott and Sanders (1990).

[13] Chan, K.C., G. Karolyi, F. Longstaff and A. Saunder (1992). An Empirical Comparison of Alternative Models of Short-Term Interest Rate. Journal of Finance 47, 1209-1227.

[14] Chen, R.R. and L. Scott (1993). Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. Journal of Fixed Income 3 (December), 14-31.

[15] Chen, R.R. and L. Scott (1992). Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross models of the Short-Term Interest Rates. Review of Financial Studies, 5(4), 613-636.

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

[17] Dai, Q. and K. Singleton (1997). Specification Analysis of Affine Term Structure Models. Working Paper, Stanford University.

[18] Duffie, D. and R. Kan (1996). A Yield Factor Model of Interest Rate. Mathematical Finance, 6(4), 379-406.

[19] Duffiie, D. and K. Singleton (1997). An Econometric Model of the Term Structure of Interest Rate Swap Yields. Journal of Finance 52, 1287-1321.

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

[21] 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.

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

[23] Inui, K and M. Kijima (1998). A Markovian Framework in Multi-Factor Heath-Jarrow-Morton Models. Journal of Financial and Quantitative Analysis, 34 (1), 131-157.

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

[25] Jeffrey, A., O. Linton and T. Nguyen (1999b). Nonparametric Estimation of Heath-Jarrow-Morton models and a Test for Path Independence. Unpublished paper, Yale University, 1999

[26] Jong, de J. and P. Santa-Clara (1999). The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables. Journal of Financial and Quantitative Analysis, 33 (3), 423-440.

[27] Knight, J., F. Li and M. Yuan (1999). Pricing Interest Rate Derivatives in a Non-Parametric Two-Factor Term Structure Model. Working paper 99-19, Bank of Canada.

[28] Linton, O., E. Mammen, J. Nielsen and C. Tanggard (1998). Estimating Yield Curves by Kernel Smoothing Methods. Forthcoming in the Journal of Econometrics.

[29] Litterman, R., and J. Scheinkman (1991). Common Factors Affecting Bond Returns. Journal of Fixed Income, June 1991, 54-61.

[30] Longstaff, F. A. and E. S. Schwartz (1992). Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model. Journal of Finance, 47(4), 1259-1282.

[31] Luenberger, D. G. (1969) *Optimization by Vector Space Methods*. New York: Wiley.

[32] McCulloch, J. H. (1971). Measuring the Term Structure of Interest Rates. Journal of Business 44, 19-31.

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

[34] Newey, W.K., and J.L. Powell (2000). Instrumental Variables Estimation for Nonparametric Models. Manuscript, MIT.

[35] Øksendal, B. (1998). *Stochastic Differential Equations*. Fifth Edition. Springer: Berlin.

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

[37] Pearson, N. and T. Sun (1994). Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model. Journal of Finance 49, 1279-1304.

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

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

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

[41] Santa-Clara, P. and D. Sornette (2000). The Dynamics of the Forward Interest Rate Curve with Stochastic String Schocks. Review of Financial Studies, forthcoming.

[42] Schaefer, S.M. and E.S. Schwartz (1984). A Two Factor Model of the Term Structure: An Approximate Analytical Solution. Journal of Financial and Quantitative Analysis, 19(4), 413-424.

[43] 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.

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

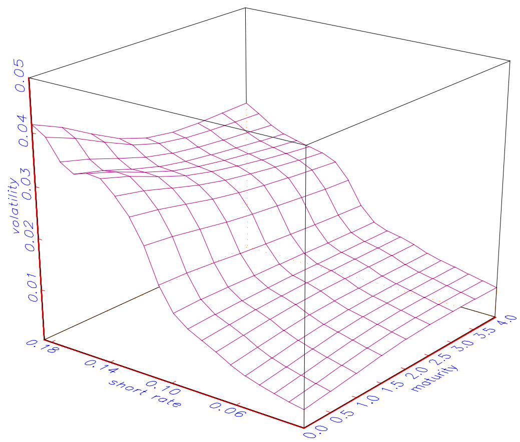

Figure 3.1: Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from Jan 1961 to December 1998. Maturity ranges from 0 to 4 years, and the short rate is from 0% to 18%.

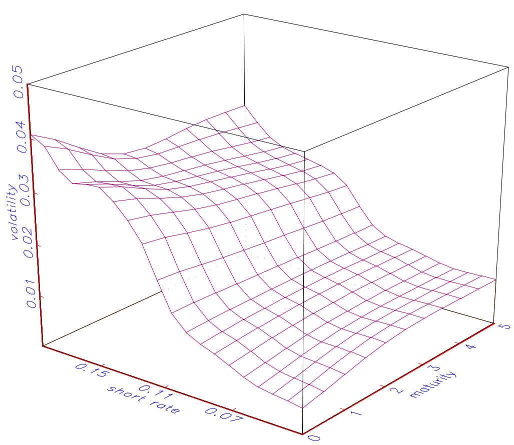

Figure 3.2: Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from Jan 1970 to December 1998. Maturity ranges from 0 to 5 years, and the short rate is from 0% to 18%.

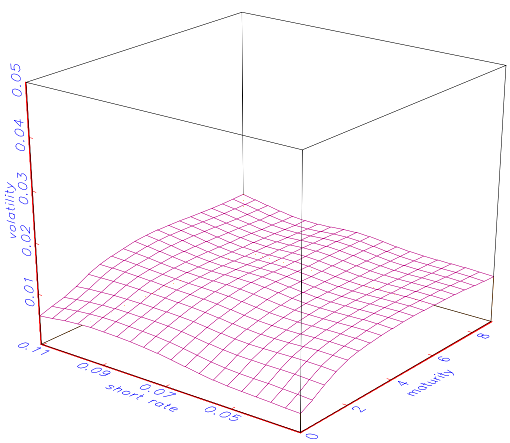

Figure 3.3: Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from Jan 1983 to December 1998. Maturity ranges from 0 to 9 years, and the short rate is from 0% to 10%.

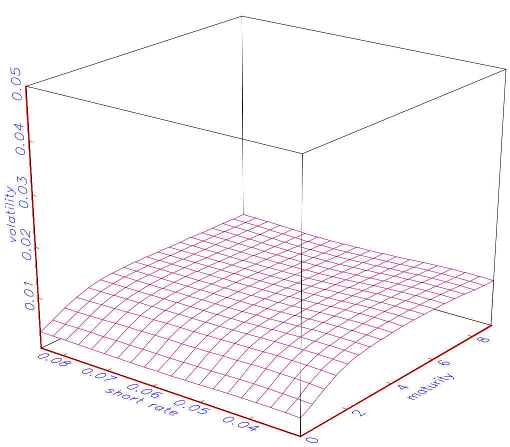

Figure 3.4: Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from Jan 1990 to December 1998. Maturity ranges from 0 to 9 years, and the short rate is from 0% to 8%.

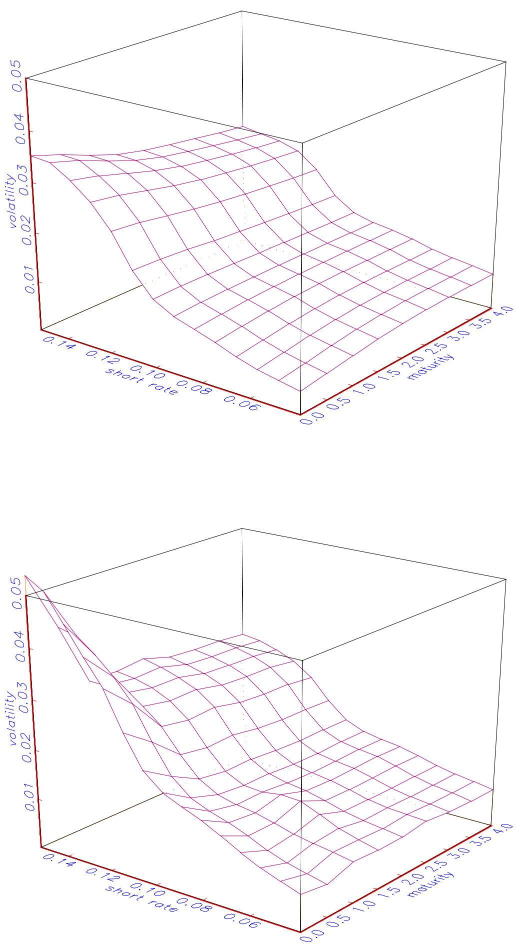

Figure 3.5: Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from Jan 1983 to December 1998 where the yield curve is extracted by the Linton et al. (1998) kernel smoothing-based method. The graph in the top is the volatility structure of the yield curves estimated nonparametrically by the method proposed in this paper. The graph in the bottom is obtained by the method proposed in Jeffrey-Linton and Nguyen (1999b). Maturity ranges from 0 to 4 years, and the short rate is from 0% to 15%.

MAT. *L.R.(%)*	*0*	*0.5*	*1.0*	*1.5*	*2.0*	*2.5*	*3.0*	*3.5*	*4.0*
2.51	7.54	7.78	7.97	8.12	8.25	8.38	8.46	8.48	8.41
3.34	8.52	8.79	8.98	9.14	9.31	9.46	9.52	9.47	9.35
4.18	10.11	10.36	10.53	10.71	10.94	11.13	11.15	11.02	10.85
5.01	12.47	12.57	12.58	12.65	12.85	13.05	13.05	12.87	12.67
5.84	14.84	14.70	14.43	14.25	14.31	14.43	14.42	14.24	14.07
6.67	16.58	16.23	15.69	15.23	15.06	15.07	15.02	14.88	14.75
7.51	17.82	17.27	16.51	15.81	15.42	15.30	15.22	15.11	15.03
8.34	18.87	18.13	17.14	16.22	15.65	15.41	15.30	15.21	15.16
9.17	19.91	18.92	17.67	16.52	15.79	15.46	15.32	15.26	15.22
10.01	20.92	19.63	18.06	16.66	15.78	15.39	15.24	15.19	15.17
10.84	21.79	20.17	18.26	16.60	15.60	15.16	15.02	14.99	14.98

Table 3.1: Volatility structure associated with the second Brownian motion from the 2-factor HJM model is estimated nonparametrically from CRSP daily bond data from Jan 1990 to December 1998 where the yield curve is extracted by the Linton et al. (1998) kernel smoothing-based method. The bandwidth chosen for the volatility structure estimation is 1%. Maturity (τ) ranges from 0 to 4 years, short rate is fixed at 6.67% and the long rate (LR) is from 2.51% to 10.84%. Volatility is reported in 1/1000.

MAT. *L. R(%)*	*0*	*0.5*	*1.0*	*1.5*	*2.0*	*2.5*	*3.0*	*3.5*	*4.0*
2.51	7.55	7.82	8.04	8.19	8.32	8.44	8.54	8.55	8.49
3.34	8.54	8.83	9.05	9.22	9.39	9.53	9.60	9.56	9.44
4.18	10.14	10.42	10.62	10.80	11.03	11.21	11.24	11.12	10.94
5.01	12.52	12.65	12.68	12.75	12.95	13.15	13.15	12.98	12.78
5.84	14.91	14.79	14.54	14.37	14.42	14.54	14.53	14.36	14.18
6.67	16.67	16.33	15.81	15.35	15.18	15.18	15.14	15.00	14.87
7.51	17.91	17.38	16.64	15.93	15.54	15.41	15.33	15.23	15.14
8.34	18.97	18.25	17.27	16.35	15.77	15.52	15.41	15.33	15.28
9.17	20.02	19.04	17.80	16.65	15.91	15.57	15.44	15.38	15.34
10.01	21.04	19.75	18.19	16.79	15.90	15.50	15.36	15.31	15.29
10.84	21.91	20.30	18.40	16.73	15.72	15.27	15.13	15.11	15.10

Table 3.2: Volatility structure associated with the second Brownian motion from the 2-factor HJM model is estimated nonparametrically from CRSP daily bond data from Jan 1990 to December 1998 where the yield curve is extracted by the Linton et al. (1998) kernel smoothing-based method.

The bandwidth chosen for the volatility structure estimation is 1%. Maturity (τ) ranges from 0 to 4 years, short rate is fixed at 6.67% and the long rate (LR) is from 2.51% to 10.84%. Volatility is reported in 1/1000.