Nonparametric estimation of time-varying covariance matrix in a slowly changing vector random walk model

Feng, Yuanhua and Yu, Keming
Heriot-Watt University and Brunel University

2006

Nonparametric Estimation of Time-Varying Covariance Matrix in a Slowly Changing Vector Random Walk Model

Yuanhua Feng
Department of Actuarial Mathematics and Statistics, Heriot-Watt University

and Keming Yu
Department of Mathematical Sciences, Brunel University

Abbreviated Title: Slowly changing vector random walk

Summary. A new multivariate random walk model with slowly changing parameters is introduced and investigated in detail. Nonparametric estimation of local covariance matrix is proposed. The asymptotic distributions, including asymptotic biases, variances and covariances of the proposed estimators are obtained. The properties of the estimated value of a weighted sum of individual nonparametric estimators are also studied in detail. The integrated effect of the estimation errors from the estimation for the difference series to the integrated processes is derived. Practical relevance of the model and estimation is illustrated by application to several foreign exchange rates.

Keywords: Multivariate time series; slowly changing vector random walk; local covariance matrix; kernel estimation; asymptotic properties; forecasting.

AMS 2000 subject classifications: Primary 62G08; Secondary 62M10

1 Introduction

Random walk models, including Weiner processes or Brownian motion arise in many applications, particularly in financial time series, statistical physics, genetics, graphs, modelling of turbulent dispersion within the atmosphere and to geographic distributions of animal; see Kijima (2002), Weiss (1994), Neigel and Avise (1993) and Lovász (1993). In these applications often an analysis of the relationships between more than one variable simultaneously is required. Vector (also called multiple or multivariate) random walk is the simplest multiple integrated process whose first differences form a vector white noise (see e.g. Harvey, 1989). Such a model can be used for modelling financial difference or return series which is also often used to built some more complex models. In Harvey et al. (1994) vector random walk is used to model persistent movements in stochastic volatility models. Due to slow change of economic or environmental situations however the means and variances, also the covariances and correlations between the components of the innovation process may all change slowly over time. In this paper a slowly changing vector random walk model is hence introduced to model the slowly change multiple time series.

The paper is organized as follows. Section 1.1 reviews the existing literature on stochastic time varying process. Section 1.2 presents a motivating example involving exchange rate. The multivariate random walk model with slowly changing parameters is introduced in Section 2. Nonparametric inference methods for time-varying covariance matrix are defined in Section 3. Properties of the estimators are derived in Section 4, including asymptotic mean square errors and optimal bandwidth. An optimal prediction of the future value of a portfolio is developed in Section 5. Practical performance of the proposed methods through an application to four daily foreign exchange rate is discussed in Section 6. Concluding remarks are presented in Section 7 and the proofs can be found in Appendix.

1.1 A brief review of closely related research

Univariate locally stationary processes were first introduced and studied by Dahlhaus (1997, 2000). Nonparametric inference for the mean and variance functions in univariate stochastic process has attracted much attention in statistics literature. We list few of a rich literature using in this context. Kernel regression estimation with time series errors is studied e.g. by Hart (1991) and Csögö and Milnikzug (1995). Beran and Ocker (1999) and Beran and Feng (2002) discussed nonparametric trend estimation in integrated processes. Fan and Yao (1998) proposed nonparametric estimation of conditional variance function following the idea of the ARCH (autoregressive conditional heteroskedasticity, Engle, 1982) model. Aït-Sahalia (1996) studied nonparametric estimation of time-varying drift and diffusion coefficients of a Brownian motion.

In contrast to univariate models, which model each component independently, multivariate model takes account of the covariances and correlation between components. There is a vast literature on parametric multivariate GARCH (generalized ARCH) models for conditional covariance matrix (see e.g. Bollerslev et al., 1988 and Engle, 2002) and on nonparametric estimation of conditional covariance matrix (Härdle et al., 1998, 2003). However in this paper we will focus on the estimation of covariance matrix in a slowly changing model. The slowly change in the current model means some deterministic components which are non-stochastic. The real examples in Section 1.2 clearly illustrate the phenomenon of slow change. Actually, in the new model named SCVRW model the change in the means, variances, covariances and correlations is assumed to be deterministic not conditional. They change smoothly over time and do not depend on the past information. This is what slow change means. On the other hand conditional changes in such components are caused by past observations. Conditional changes and local changes can not explained each other.

Despite a huge number of literature on the estimation of conditional covariance matrix, little research on the estimation of local covariance matrix can be found in literature. Herzel et al (2006) explores some ideas on volatility estimation under a non-stationary multivariate return model, but assumes a constant drift term, lacks theoretical investigation and ignores forecasting in depth. Wu and Pourahmadi (2003) discussed nonparametric estimation of the covariance matrix of the observations of a univariate time series based on Cholesky decomposition, where it is proposed to estimate the diagonal matrix with the variances as its diagonal entries and the unit lower triangle matrix using local polynomial separately.

1.2 Motivation examples

The slow change economics phenomenon can be observed in finance such as the indices of major stock markets and exchange rates. Figure 1 displays the scatter plots of four daily foreign exchange rate series w.r.t. the US Dollar (USD). The data are those of the British Pound (Pound), Japanese Yen (Yen), Euro and Canadian Dollar (CAD), from 4 January 1999 to 4 November 2005. Here the USD price per foreign currency is used. Figure 1 shows that there are clear non-constant drifts in these series which change slowly over time. Figure 2 displays the difference series of the original data from which we can see that the variances of the difference series change clearly over time. What cannot be discovered by eye is that the correlation coefficients also change strongly during the observation period (see Figure 6 in Section 6). This can also be shown graphically by plotting two difference series against each other piece by piece.

For this data and the content in finance, one is concerned

(a) how to model properly the slowly changing drifts, variance and correlation coefficients of underlying stochastic process,

(b) what is the estimation of the mean and variance of a portfolio, and

(c) what is the effect of estimation errors in the short term forecasts.

Other slowly change multivariate time series include ice thickness series. It is widely concerned that global warming results in quickly decreasing ice-thickness. However, the ice thickness series are random walks (the sum of snow fall years after years). There is a nonparametric trend in the differences of these data, because of the different strength of press (or maybe also other reasons), the dependence between ice thickness in different years (after removing the trend) is very weak. In all, different series of ice thickness may also be able to be modelled using the SCVRW.

Global and hemispheric series of temperature anomalies can also be modelled by random walks (Gorden, 1991). Similarly, statistical analysis for satellite-based global daily tropospheric and stratospheric temperature anomaly and solar irradiance data sets shows that the behavior of the series appears to be nonstationary with stationary daily increments (Kärner, 2002). The model proposed in this paper provides a useful tool for modelling global warming as slow change in the increments of those series.

2 The model

Let $\mathbf{Y}_t$ be a k-dimensional stochastic process, and $\mathbf{Y_0}$ be the initial value of $\mathbf{Y_t}$. Let $\mathbf{C}$ be a constant vector, and let $\mathbf{Z}_t$ be the increment of $\mathbf{Y}_t$.

The proposed slowly changing vector random walk (SCVRW) is defined by

$$\begin{cases} \mathbf{Y}_0 &= \mathbf{C}, \\ \mathbf{Y}_t &= \mathbf{Y}_{t-1} + \mathbf{Z}_t \qquad \text{for } t = 1, 2, ..., n, \\ \mathbf{Z}_t &= \boldsymbol{\mu}(x_t) + \boldsymbol{\Sigma}^{1/2}(x_t)\mathbf{E}_t, \end{cases} \tag{1}$$

where

$$\mathbf{Y}_t = \begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{kt} \end{pmatrix}, \qquad \mathbf{Z}_t = \begin{pmatrix} Z_{1t} \\ \vdots \\ Z_{kt} \end{pmatrix}, \qquad \mathbf{E}_t = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{kt} \end{pmatrix}$$

are random vectors, $x_t = t/n$ is the re-scaled time,

$$\boldsymbol{\mu}(x_t) = \begin{pmatrix} \mu_1(x_t) \\ \vdots \\ \mu_k(x_t) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}(x_t) = \begin{pmatrix} \sigma_1^2(x_t) & \cdots & \sigma_{1k}(x_t) \\ \vdots & \ddots & \vdots \\ \sigma_{k1}(x_t) & \cdots & \sigma_k^2(x_t) \end{pmatrix} \tag{2}$$

are the vector of local mean functions and the matrix of local variances or cross-covariance functions, respectively, and $\boldsymbol{\Sigma}^{1/2}$ denotes lower triangular Cholesky factorization of a semi-positive definite matrix so that

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{\Sigma}^{1/2}\right)'.$$

For convenience we will also denote $\sigma_i^2(x_t)$ by $\sigma_{ii}(x_t)$, $i = 1, ..., k$. It is assumed that $\sigma_{ii}(\cdot)$ are strictly positive for all i and that ϵ_{it} given i are i.i.d. $N(0,1)$ random variables for $t = 1, 2, ..., n$, and that ϵ_{it} given t are also mutually independent for $i = 1, 2, ..., k$. Further smoothness conditions on the local mean and the local variances or cross-covariance functions, i.e. μ_i and σ_{ij}, $i, j = 1, ..., k$ will be introduced later.

The difference series $\mathbf{Z}_t$ of the slowly changing vector random walk $\mathbf{Y}_t$ defined in the above is non-jointly stationary, because each of its element is non-stationary. However, it is easy to see that $\mathbf{Z}_t$ is jointly locally stationary in the sense that, in a small interval of x whose length tends to zero as $n \to \infty$, the difference between $\mathbf{Z}_t$ and another jointly stationary process $\mathbf{Z}_t^*$ is negligible in probability.

Remark 1 *When $k = 1$, Model (1) reduces to $\Delta Y_t = \mu(x_t) + \sigma(x_t)\epsilon_t$, which is the discretized form of model (1.1) discussed by Aït-Sahalia (1996). This univariate model is also a fixed design heteroscedastic regression model whose random design type was discussed by Fan and Yao (1998).*

Model (1) can be extended substantially e.g. by introducing a VARMA term into the difference process Z_t or by allowing for continuous time. This will however not be considered *here, because the aim of the current paper is to obtain more detailed results under a basic model.*

3 The estimators

Let $\mathbf{y}_t$, $t = 0, 1, ..., n$ denote the observations and $\mathbf{z}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$, $t = 1, ..., n$. In the following two kinds of estimators, called single and joint estimators, will be introduced, which can be used for forecasting the trend and variance of a single financial series or of a portfolio, respectively. First consider p-th order local polynomial estimation (Fan and Gijbels, 1996) of the mean function in the i-th series. Let $K_i(u)$ denote a weight, i.e. a positive kernel, function and $\mathbf{a}_i = (a_{i0}, ..., a_{ip})'$. Solve the locally weighted least square problem

$$\hat{\mathbf{a}}_i'(x) = \arg\min_{\mathbf{a}_i} \sum_{t=1}^{n} \left[z_{it} - a_{i0} - ...a_{ip}(x_t - x)^p\right]^2 K_i\left(\frac{x_t - x}{h_i}\right), \tag{3}$$

where h_i is the bandwidth used for estimating μ_i. Then the resulting estimator of $\mu_i(x)$ is given by $\hat{\mu}_i(x) = \hat{a}_{i0}(x)$, which is a linear estimator, i.e. a weighted sum of z_{it}. The vector estimator $\hat{\boldsymbol{\mu}}(x) = (\hat{\mu}_1(x), ..., \hat{\mu}_k(x))'$ will be called a joint estimator of all mean functions, if each of its element is estimated following (3) but using the same weight function $K(u)$ and the same bandwidth h.

Let $r_{it} = z_{it} - \hat{\mu}_i(x_t)$, $i = 1, ..., k$, denote the residuals at time t calculated using either the single estimators $\hat{\mu}_i(x_t)$ or the joint estimator $\hat{\boldsymbol{\mu}}_t$. Let $\mathbf{r}_t = (r_{1t}, ..., r_{kt})'$ denote the residual vector. Then the variance and cross-covariance functions can be estimated from the residuals. Consider first the estimation of the variance function in the i-th series by a Nadaraya-Watson kernel (i.e. a local constant). Let $W_i(u)$ denote the weight function and b_i the bandwidth. We have

$$\hat{\sigma}_i^2(x) = \frac{\sum_{t=1}^{n} W_i\left(\frac{x_t - x}{b_i}\right) r_{it}^2}{\sum_{t=1}^{n} W_i\left(\frac{x_t - x}{b_i}\right)}. \tag{4}$$

This type of variance estimators is used, e.g. by Feng (2004). $\hat{\sigma}_i^2$ defined in this way is certainly non-negative.

Note that we do not define single estimators for the cross-covariances functions, because the estimation of $\mathbf{\Sigma}$ obtained as a matrix of separate variance and cross-covariance estimators may not be semi-positive definite. Instead, it is proposed to estimate $\mathbf{\Sigma}$ in the following joint way.

$$\hat{\mathbf{\Sigma}}(x) = \frac{\sum_{t=1}^{n} W\left(\frac{x_t - x}{b}\right) \mathbf{r}_t \mathbf{r}_t'}{\sum_{t=1}^{n} W\left(\frac{x_t - x}{b}\right)}. \tag{5}$$

$\hat{\mathbf{\Sigma}}(x)$ is simply a matrix kernel estimator of all $\sigma_{ij}(x)$, $i, j = 1, ..., k$, with the same kernel function and the same bandwidth. Let $\mathbf{\Gamma}(x) = \Big(\rho_{ij}(x)\Big)$, where $\rho_{ij}(x)$ denotes the local cross-correlation between ϵ_{it} and ϵ_{jt}. Then $\mathbf{\Gamma}(x)$ can be estimated as follows.

$$\hat{\mathbf{\Gamma}}(x) = \left(diag(\hat{\mathbf{\Sigma}}(x))\right)^{-1/2} \hat{\mathbf{\Sigma}}(x) \left(diag(\hat{\mathbf{\Sigma}}(x))\right)^{-1/2}. \tag{6}$$

Proposition 1 $\hat{\mathbf{\Sigma}}(x)$ *defined in (5) is semi-positive definite at any point* $x \in [0, 1]$ *and hence* $\hat{\mathbf{\Gamma}}(x)$ *defined in (6) is a correlation matrix.*

Proof of Proposition 1 is omitted. $\hat{\mathbf{\Sigma}}(x)$ is semi-positive definite, because it is a Gram matrix. Indeed $\hat{\mathbf{\Sigma}}(x)$ is a.s. (almost sure) positive definite, because $\mathbf{r}_t$ are a.s. linear independent of each other. Note that in practice we are mainly interested in estimating $\mathbf{\Sigma}(x)$ at the current end of the time series, i.e. with $x = 1$. If a second order kernel is used, kernel estimator has the so-called boundary effect, i.e. the bias in the interior is of a higher order than that at the boundary point. To avoid this problem we only assume the existence of the first derivatives and will focus on discussing the behaviour of $\mathbf{\Sigma}(x)$ at a boundary point.

4 Main results

For the derivation of the asymptotic results the following assumptions are required.

Assumption A1. The weight function is assumed to be a symmetric density on $[-1, 1]$.

Assumption A2. The local mean functions $\mu_i(\cdot)$, $i = 1, ..., k$, are at least $p + 1$ times continuously differentiable, and the local variance and cross-covariance functions $\sigma_{ij}(\cdot)$, $i, j = 1, ..., k$, are at least continuously differentiable.

Assumption A3. The assumptions on the error structure as described in the context around equations (1) and (2) hold.

Assumption A4. Let h_g denote a generic bandwidth used. It is assumed that $h_g \to 0$ and $nh_g \to \infty$ as $n \to \infty$.

The smoothness conditions required in A2 adapt to the different definitions of the estimators. Lower smoothness is required for $\sigma_{ij}(\cdot)$ due to the use of kernel estimators. For a goodness-of-fit criterion the MSE will be used. The normal assumption on ϵ_{it} is only necessary, if interval prediction for future values is of interest. For the derivation of most of the asymptotic results the existence of fourth moments of ϵ_{it} is enough. If automatic bandwidth selection is considered, higher order of smoothness and the existence of $E(\epsilon_{it}^8)$ are required. A4 is a minimal requirement on the bandwidths in nonparametric regression. Further restrictions on the bandwidths will be introduced later.

In local polynomial regression it is often assumed that p is odd so that the bias of the estimator in the interior and at the boundary is of the same order. For the estimators of the mean functions this restriction will be used. Let $K_i^*(\cdot)$ denote the equivalent kernel of the local polynomial estimator of $\hat{\mu}_i$ (see e.g. Ruppert and Wand, 1994). Then $K_i^*(\cdot)$ is a $(p+1)$-th order kernel. Denote by $\beta_i = \int u^{p+1} K_i^*(u)du$ and $R_i = \int (K_i^*(u))^2 du$. The following theorem is given without proof, which summaries well known asymptotic properties of the single estimator $\hat{\mu}_i$ (see e.g. Wand and Jones, 1995).

Theorem 1 *Let $x \in [0,1]$. Assume that p is odd and that the assumptions A1 to A4 hold. Then we have*

1. *The bias of $\hat{\mu}_i(x)$ is $E[\hat{\mu}_i(x) - \mu_i(x)] = \beta_i \mu_i^{(p+1)}(x) h_i^{p+1}[1 + o(1)]/[(p+1)!]$.*

2. *The variance of $\hat{\mu}_i(x)$ is $var[\hat{\mu}_i(x)] = (nh_i)^{-1} R_i \sigma_i^2(x)[1 + o(1)]$.*

3. *The dominant part of the MSE of $\hat{\mu}_i(x)$ is*

$$MSE(\hat{\mu}_i(x)) \doteq \beta_i^2 (\mu_i^{(p+1)}(x))^2 h_i^{2(p+1)} [(p+1)!]^{-2} + (nh_i)^{-1} R_i \sigma_i^2(x). \qquad (7)$$

4. *The right-hand side of (7) is minimizesd by the (asymptotically) optimal bandwidth*

$$h_i^{opt} = \left(\frac{T_{i2}}{2(p+1)T_{i1}} \right)^{1/(2p+3)} n^{-1/(2p+3)}, \qquad (8)$$

provided that $T_{i1} \neq 0$ *and* $T_{i2} \neq 0$*, where* $T_{i1} = \beta_i^2(\mu_i^{(p+1)}(x))^2[(p+1)!]^{-2}$ *and* $T_{i2} = R_i\sigma_i^2(x)$*.*

Now consider the estimation of the mean in the returns of a given portfolio. Let $S = (S_1, ..., S_k)'$ denote the vector of shares in the portfolio with $S_i \geq 0$ and $\sum S_i > 0$. Then the portfolio values and returns at time t are given by $V_P(x_t) = S'\mathbf{Y}_t$ and $Z_P(x_t) = S'\mathbf{Z}_t$, respectively. The unknown mean of $Z_P(x_t)$ is $S'\boldsymbol{\mu}(x)$ which can be estimated by $S'\hat{\boldsymbol{\mu}}(x)$. Let $K^*(\cdot)$ denote the equivalent kernel of this estimator. Denote by $\beta = \int u^{p+1}K^*(u)du$ and $R = \int(K^*(u))^2du$. The MSE of estimating $S'\boldsymbol{\mu}(x)$ is given by

$$MSE^S_{\hat{\boldsymbol{\mu}}}(x) = E\left\{\sum_{i=1}^{k} S_i[\hat{\mu}_i(x) - \mu_i(x)]\right\}^2. \tag{9}$$

The following theorem quantifies the properties of $MSE^S_{\hat{\boldsymbol{\mu}}}$ using the joint estimator.

Theorem 2 *Under the same conditions of Theorem 1 we have*

1. *Results in 1 to 3 of Theorem 1 hold for each element* $\hat{\mu}_i(x)$ *of* $\hat{\boldsymbol{\mu}}(x)$ *by replacing* β_i*,* R_i *and* h_i *with* β*,* R *and* h *respectively.*

2. *The covariance between* $\hat{\mu}_i(x)$ *and* $\hat{\mu}_j(x)$ *is given by*

$$cov(\hat{\mu}_i(x), \hat{\mu}_j(x)) = R\sigma_{ij}(nh)^{-1}[1 + o(1)]. \tag{10}$$

3. *The* $MSE^S_{\hat{\boldsymbol{\mu}}}$ *is dominated by*

$$MSE^S_{\hat{\boldsymbol{\mu}}} \doteq T_1^S h^{2(p+1)} + T_2^S(nh)^{-1}, \tag{11}$$

where

$$T_1^S = \frac{\beta^2}{[(p+1)!]^2}\sum_{i=1}^{k}\sum_{j=1}^{k} S_iS_j\mu_i^{(p+1)}(x)\mu_j^{(p+1)}(x) \tag{12}$$

and

$$T_2^S = R\sum_{i=1}^{k}\sum_{j=1}^{k} S_iS_j\sigma_{ij}(x). \tag{13}$$

4. The optimal bandwidth which minimizes the asymptotic $MSE^S_{\hat{\boldsymbol{\mu}}}$ *is given by*

$$h_S^{opt} = \left(\frac{T_2^S}{2(p+1)T_1^S} \right)^{1/(2p+3)} n^{-1/(2p+3)}, \tag{14}$$

provided that $T_1^S \neq 0$ *and* $T_2^S \neq 0$.

Let $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\mu}}}(x)$ and $\boldsymbol{\Gamma}_{\hat{\boldsymbol{\mu}}}(x)$ denote the covariance respective correlation matrices of $\hat{\boldsymbol{\mu}}$. Following (10) we have

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\mu}}}(x) \doteq R(nh)^{-1}\boldsymbol{\Sigma}(x) \qquad and \qquad \boldsymbol{\Gamma}_{\hat{\boldsymbol{\mu}}}(x) \doteq \boldsymbol{\Gamma}(x), \tag{15}$$

where $\boldsymbol{\Gamma}$ is the cross-correlation matrix defined before. The result in the second part of (15) shows that $\hat{\boldsymbol{\mu}}$ takes over the correlations of the original data.

Remark 2 *In the special case with* $S_1 = ... = S_k = S_0 > 0$, *Parts 3 and 4 of Theorem 2 show the properties of an unweighted (or equally weighted) portfolio. While in another extreme case with* $S_i > 0$ *and* $S_j = 0$ *for all* $j \neq i$, *these results reduce to those given in Theorem 1. Similar statements apply to the results in Theorem 4 below.*

It is well known that the bias of the kernel estimators $\hat{\sigma}_i^2$ and $\hat{\boldsymbol{\Sigma}}$ in the interior point has different order from the boundary. In the following only asymptotic properties of these estimators at $x = n/n = 1$ will be given, because we are mainly interested in the estimation at the current endpoint. In this case the used weight function is $W_i^r(u) = 2W_i(u)\mathbb{1}_{[-1,0]}$. Let $\alpha_i = \int_{-1}^0 uW_i^r(u)du$ and $V_i = \int_{-1}^0 (W_i^r(u))^2 du$. To simplify the derivation of the asymptotic properties we will introduce the following assumption on $\hat{\mu}_i$ and the bandwidths h_i and b_i.

A4$'$. Assume that $\hat{\mu}_i$ is a local linear estimator obtained with a bandwidth $O(n^{-1/3}) < h_i < O(n^{-1/6})$. The bandwidth b_i satisfies $O(n^{-2/3}) < b_i = o(1)$.

Condition A4$'$ ensures that the errors in $\hat{\mu}_i$ are negligible when discussing the asymptotic properties of $\hat{\sigma}_i^2$ and $\hat{\boldsymbol{\Sigma}}$ (see the proofs of the theorems given in the appendix), which implies A4. Note that $b_i > O(n^{-2/3})$ means that b_i should not be too small. This is a very weak restriction. Results in Theorem 3 below show that the optimal order for b_i is $O(n^{-1/3})$. The bounds for h_i and b_i in A4$'$ are chosen for convenience. Note however that they depend on each other. The optimal order $O(n^{-1/5})$ of h_i lies in between the range given in the first part of A4$'$. If the stronger condition $h_i = O(n^{-1/5})$ is used, then the allowed range of b_i becomes larger. The requirements in A4$'$ also depend on the special case considered. For instance weaker restrictions can be used, if $\hat{\mu}_i$ is a high order local polynomial estimator.

The following results hold for the single estimator $\hat{\sigma}_i^2$ defined in (4).

Theorem 3 *Suppose Assumptions A1 to A3 and A4$'$ hold. Then*

1. *The bias of $\hat{\sigma}_i^2(1)$ is given by*

$$E[\hat{\sigma}_i^2(1) - \sigma_i^2(1)] = \alpha_i (\sigma_i^2)'(1) b_i [1 + o(1)], \tag{16}$$

where $(\sigma_i^2)'(1)$ denotes the first derivative of $\sigma_i^2(x)$ at $x = 1$.

2. *The variance of $\hat{\sigma}_i^2(1)$ is given by*

$$var[\hat{\sigma}_i^2(1)] = (nb_i)^{-1} V_i \gamma_{ii}^2(1)[1 + o(1)], \tag{17}$$

where $\gamma_{ii}^2(1) = var[(Z_i(1) - \mu_i(1))^2] = \sigma_i^4(1) var(\epsilon_{in}^2)$ which equals to $2\sigma_i^4(1)$ for normal ϵ_{it}.

3. *The MSE of $\hat{\sigma}_i^2(1)$ is dominated by*

$$MSE_{\hat{\sigma}_i^2}(1) \doteq D_{i1} h_i^2 + D_{i2}(nb_i)^{-1}, \tag{18}$$

where $D_{i1} = \alpha_i^2[(\sigma_i^2)'(1)]^2$ and $D_{i2} = V_i \gamma_{ii}^2(1)$.

4. *The optimal bandwidth which minimises the right-hand side of (18) is given by*

$$b_i^{opt} = \left(\frac{D_{i2}}{2D_{i1}}\right)^{1/3} n^{-1/3}, \tag{19}$$

provided that $D_{i1} \neq 0$ and $D_{i2} \neq 0$.

Remark 3 *At an interior point $x \in (b_i, 1 - b_i)$ the bias of a kernel estimator is of the order $O(b_i^2)$ and the order of variance stays unchanged. Now, we have $MSE \doteq O(b_i^4) + O[(nh)^{-1}]$ and $b_i^{opt} = O(n^{-1/5})$.*

Remark 4 *Asymptotic results for the left endpoint $x = 0$ are the same as given in Theorem 3. In practice the so called one-side kernel estimators, i.e. estimators defined using only observations on the left hand side, may be of interest. In this case any point is treated as a right endpoint. For one-side kernel estimators results in Theorem 3 hold for all $x \in (b_i, 1]$.*

Consider now the estimation of $\text{var}\,[Z_P(x_t)] = \text{var}\,[S'\mathbf{Z}_t]$. First we have

$$\text{var}\,[Z_P(x_t)] = S'\mathbf{\Sigma}(\mathbf{x_t})S, \qquad \hat{\text{var}}\,[Z_P(x_t)] = S'\hat{\mathbf{\Sigma}}(x_t)S$$

and the total estimated error in $\hat{\text{var}}\,[Z_P(x_t)]$ is

$$\hat{\text{var}}\,[Z_P(x_t)] - \text{var}\,[Z_P(x_t)] = S'\left[\hat{\mathbf{\Sigma}}(x_t) - \mathbf{\Sigma}(\mathbf{x_t})\right]S. \tag{20}$$

Define the $k \times k$ random matrix

$$\mathbf{\Xi}(x) = [\mathbf{Z}(x) - \boldsymbol{\mu}(x)]^{\mathrm{T}}[\mathbf{Z}(x) - \boldsymbol{\mu}(x)] =: (\xi_t^{ij}), \tag{21}$$

$i, j = 1, ..., k$.

Note that $E(\mathbf{\Xi}(x)) = \mathbf{\Sigma}(\mathbf{x})$. Define $(\tilde{\epsilon}_{1t}, ..., \tilde{\epsilon}_{kt})' = \tilde{\mathbf{E}}_t = \mathbf{\Gamma}^{1/2}(x_t)\mathbf{E}_t$, where $\tilde{\epsilon}_{it} \sim N(0, 1)$ with $E(\tilde{\epsilon}_{it}\tilde{\epsilon}_{jt}) = \rho_{ij}(x_t)$, $i, j = 1, .., k$. The *covariance matrix* of $\mathbf{\Xi}(x)$ is given by

$$\Upsilon_{\mathbf{\Xi}}(x) := E\left\{[\mathbf{\Xi} - \mathbf{\Sigma}(x)] \otimes [\mathbf{\Xi} - \mathbf{\Sigma}(x)]\right\} =: (\gamma_{ij,lm}) \tag{22}$$

for $i, j, l, m = 1, ..., k$, where $\otimes$ denotes the Kronecker product, and

$$\begin{aligned} \gamma_{ij,lm}(x_t) &= \text{cov}\,(\xi_t^{ij}, \xi_t^{lm}) \\ &= \sigma_i(x_t)\sigma_j(x_t)\sigma_l(x_t)\sigma_m(x_t)\text{cov}\,(\tilde{\epsilon}_{it}\tilde{\epsilon}_{jt}, \tilde{\epsilon}_{lt}\tilde{\epsilon}_{mt}) \end{aligned} \tag{23}$$

for $i, j, l, m = 1, ..., k$. The MSE of $\hat{\text{var}}\,[Z_P(x_t)]$, denoted by $MSE_V^S(x_t)$, is given by

$$\begin{aligned} MSE_V^S(x_t) &= E\left\{S'\left[\hat{\mathbf{\Sigma}}(x_t) - \mathbf{\Sigma}(x_t)\right]S\right\}^2 \\ &= \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{m=1}^k S_i S_j S_l S_m E\left\{[\hat{\sigma}_{ij}(x_t) - \sigma_{ij}(x_t)][\hat{\sigma}_{lm}(x_t) - \sigma_{lm}(x_t)]\right\} \end{aligned} \tag{24}$$

Finally, let $\alpha = \int_{-1}^0 uW^r(u)du$ and $V = \int_{-1}^0 (W^r(u))^2 du$, where $W^r(u) = 2W(u)\mathbb{1}_{[-1,0]}$. Then the asymptotic biases, variances and covariances of $\hat{\sigma}_{ij}(1)$ and the formula of $MSE_V^S(1)$ (all at $x = 1$) are given in the following theorem.

Theorem 4 *Under the same conditions of Theorem 3 we have*

1. *The bias of* $\hat{\sigma}_{ij}(1)$ *is given by*

$$E[\hat{\sigma}_{ij}(1) - \sigma_{ij}(1)] = \alpha\sigma'_{ij}(1)b[1 + o(1)]. \quad (25)$$

2. *The covariance between* $\hat{\sigma}_{ij}(1)$ *and* $\hat{\sigma}_{lm}(1)$ *is given by*

$$cov\,(\hat{\sigma}_{ij}(1), \hat{\sigma}_{lm}(1)) = V\gamma_{ij,lm}(nb)^{-1}[1 + o(1)]. \quad (26)$$

3. $MSE_V^S(1)$ *is dominated by*

$$MSE_V^S(1) \doteq D_1^S b^2 + D_2^S (nb)^{-1}, \quad (27)$$

where

$$D_1^S = \alpha^2 \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{m=1}^k S_i S_j S_l S_m \sigma'_{ij}(1)\sigma'_{lm}(1) \quad (28)$$

and

$$D_2^S = V \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{m=1}^k S_i S_j S_l S_m \gamma_{ij,\,lm}(1). \quad (29)$$

4. *The optimal bandwidth which minimizes the dominant part of* $MSE_V^S(1)$ *is given by*

$$b_S^{opt} = \left(\frac{D_2^S}{2D_1^S}\right)^{1/3} n^{-1/3}, \quad (30)$$

provided that $D_1^S \neq 0$ *and* $D_2^S \neq 0$.

Now let $\mathbf{\Gamma_\Xi}(x)$ denote the standardized correlation matrix of $\mathbf{\Xi}(x)$, and $\mathbf{\Upsilon}_{\hat{\mathbf{\Sigma}}}(x)$ and $\mathbf{\Gamma}_{\hat{\mathbf{\Sigma}}}(x)$ denote the corresponding matrices of the joint estimator $\hat{\mathbf{\Sigma}}(x)$. Following (26) we have

$$\mathbf{\Upsilon}_{\hat{\mathbf{\Sigma}}}(1) \doteq V(nb)^{-1}\mathbf{\Upsilon_\Xi}(1) \qquad \text{and} \qquad \mathbf{\Gamma}_{\hat{\mathbf{\Sigma}}}(1) \doteq \mathbf{\Gamma_\Xi}(1). \quad (31)$$

We see $\hat{\mathbf{\Sigma}}$ takes over the correlations between the components of $\mathbf{\Xi}$.

Remark 5 *Using Taylor expansion it can be shown that the bias and variance of* $\hat{\rho}_{ij}(x)$, $i \neq j$, *are of the same orders as those of* $\hat{\sigma}_{ij}(x)$.

Remark 6 *Assume that the innovations are normal or satisfy the moment condition mentioned before. Assume further that the proposed estimators are obtained using bandwidths of corresponding optimal orders, then it is easy to show that they are all asymptotically normally distributed.*

5 Forecasting of future mean and standard deviation

We now discuss the forecasting of future mean and standard deviation at time $n+T$ given observations until time n. The most current estimation results, i.e. those with $x=1$ will be used. Under our model the optimal linear prediction for a single series at time $n+T$ is

$$E(Y_{i(n+T)}|Y_{in}) = Y_{in} + \sum_{t=1}^{T} \mu_i(x_{n+t}), \tag{32}$$

which can be estimated by

$$\hat{Y}_{i(n+T)} = Y_{in} + T\hat{\mu}_i(1). \tag{33}$$

The optimal linear prediction for the value of a given portfolio $S = (S_1, ..., S_k)'$ at time $n+T$ is

$$E(P^S_{n+T}|\mathbf{Y}_n) = E(S'\mathbf{Y}_{n+T}|\mathbf{Y}_n) = \sum_{i=1}^{k} S_i Y_{in} + \sum_{i=1}^{k} S_i \sum_{t=1}^{T} \mu_i(x_{n+t}), \tag{34}$$

which can be estimated by

$$\hat{P}^S_{n+T} = \sum_{i=1}^{k} S_i Y_{in} + T\sum_{i=1}^{k} S_i \hat{\mu}_i(1). \tag{35}$$

Clearly, the forecasts should be at least consistent. To this end we need the following assumption on T.

A5. Let h_g denote a generic bandwidth for estimating the mean functions. T satisfies $T^2 h_g^{2(p+1)} \to 0$ and $T^2(nh_g)^{-1} \to 0$ as $n \to \infty$.

The following theorem describes the properties of $\hat{Y}_{i(n+T)}$ and $\hat{P}^S_{n+T}$.

Theorem 5 *Suppose the assumptions of Theorem 2 and A5 hold. Then we have*

1. $\hat{Y}_{i(n+T)}$ *is consistent with* $E[\hat{Y}_{i(n+T)} - E(Y_{i(n+T)}|Y_{in})]^2 \doteq T^2 MSE[\hat{\mu}_i(1)]$*, which is minimized by* $\hat{\mu}_i(1)$ *with* $h_i^{opt}(1)$*.*

2. $\hat{P}^S_{n+T}$ *is consistent with* $E[\hat{P}^S_{n+T} - E(P^S_{n+T}|\mathbf{Y}_n)]^2 \doteq T^2 MSE^S_{\hat{\boldsymbol{\mu}}}$*, which is minimized by* $\hat{\boldsymbol{\mu}}(1)$ *with* $h_S^{opt}(1)$*.*

If a bandwidth of the optimal order $O(n^{-1/(2p+3)})$ is used, then A5 becomes $T = o(n^{(p+1)/(2p+3)})$ which is the biggest allowed order of the forecasting period.

The variance of the future value of a single series is

$$\operatorname{var}[Y_{i(n+T)}|Y_{in}] = \sum_{t=1}^{T} \sigma_i^2(x_{n+t}) \tag{36}$$

which can be estimated by

$$\hat{V}[Y_{i(n+T)}|Y_{in}] = T\hat{\sigma}_i^2(1). \tag{37}$$

The variance of the future value of the portfolio at time $n+T$ is

$$\operatorname{var}(P_{n+T}^S|\mathbf{Y}_n) = S' \left[\sum_{t=1}^{T} \mathbf{\Sigma}(x_{n+t}) \right] S, \tag{38}$$

which can be estimated by

$$\widehat{\operatorname{var}}(P_{n+T}^S) = TS'\hat{\mathbf{\Sigma}}(1)S. \tag{39}$$

Now we need the following assumption on T.

A5′. Let b_g denote a generic bandwidth for estimating the variance-covariance functions. T satisfies $Tb_g \to 0$ and $T(nb_g^2)^{-1} \to 0$ as $n \to \infty$.

Let SD_T denote the standard deviation of $Y_{i(n+T)}|Y_{in}$ with the estimate $\sqrt{T}\hat{\sigma}_i(1)$ and SD_T^S the standard deviation of $\operatorname{var}(P_{n+T}^S|\mathbf{Y}_n)$ with the estimate $\sqrt{T}\sqrt{S'\hat{\mathbf{\Sigma}}(1)S}$. The following holds.

Theorem 6 *Suppose the assumptions of Theorem 3 and A5′ hold. Then we have*

1. $\sqrt{T}\hat{\sigma}_i(1)$ *is a consistent estimator of* SD_T *with* $E\left[\sqrt{T}\hat{\sigma}_i(1) - SD_T\right]^2 = O(T{\cdot}MSE[\hat{\sigma}_i^2(1)])$, *which is minimized by that using* $b_i^{opt}(1)$.
2. $\sqrt{TS'\hat{\mathbf{\Sigma}}(1)S}$ *is a consistent estimator of* SD_T^S *with* $E\left[\sqrt{TS'\hat{\mathbf{\Sigma}}(1)S} - SD_T^S\right]^2 = O[T{\cdot}MSE_{\hat{\mathbf{\Sigma}}}^S(1)]$, *which is minimized by that using* $b_S^{opt}(1)$.

If a bandwidth of the optimal order $O[n^{-1/3}]$ is used, then A5′ becomes $T = o(n^{1/3})$ which is the biggest allowed order of the forecasting period.

Both theorems show that the optimal bandwidths for estimating the local mean and local variance functions respectively are also optimal for carrying future forecasting following the random walks. This also holds for calculating the forecasting intervals, if the innovation distribution is normal. The above results show that forecasts of future values should be carried out for a very short time period T. Otherwise the MSE of the forecasts would be enlarged very strongly so that the forecasts are no longer consistent or only converge very slowly.

6 Data examples

The four daily foreign exchange rate series w.r.t. the US Dollar (USD) outlined in Section 1.3 are chosen to illustrate the practice usefulness of the proposed model. There are 1723 observations in each exchange rate series.

In the following, the mean functions will be estimated by local linear and the variance-covariance functions by local constant regression. Bandwidth-choice methods in this problem would make a particularly interesting topic to study. Cross-validation is an obvious approach, and would require no additional work other than development of theory. However, it would be interesting to develop alternative approaches, based on plug-in methods, since cross-validation often results in highly stochastic variations in the selected bandwidth. Nevertheless, it seems to us that this aspect of the problem is well outside the scope of the present paper.

To this end our experiments found that we can use bandwidth selection rules developed under univariate cases. The bandwidth for the estimation of the mean of a single series can be selected following the iterative plug-in algorithm proposed by Gasser et al. (1991). The bandwidth for estimating the variance by a single series can be selected by adapting the iterative plug-in algorithm in Feng (2004). Hence, the bandwidths used in this paper are selected according to the information obtained by bandwidth selection for each univariate series. The rule gives about $h = 0.125$ for estimating the means and $b = 0.1$ for estimating the variances and covariances. Also the method of k-nearest-neighbours is used at the boundary. This means that the number of observations used at each point is fixed which is the nearest integer to $2nh + 1$ or $2nb + 1$ in the two cases respectively. Through the paper the Epanechnikov kernel is used as the weight function. The estimated local mean functions are displayed in Figure 3 together with a horizontal line $y = 0$. Where negative values of the estimated means correspond to weaker USD periods while positive estimate to strong USD periods. We see at the current end the USD is stronger than all foreign currencies except for the CAD, which reflects the current strong dollar policy of the US government. Figure 4 shows the estimated local standard deviations. We can see that the variance of a difference series also change clearly during this period. The biggest relative change happens by the CAD, the the ratio between the smallest and biggest values is bigger than 2. The standardized difference series are shown in Figure 5, from which we can see that the change in the variances is well fitted and removed. Now each single series looks stationary. However they are still not jointly stationary, because the cross-correlation matrix changes very strongly over this period. This is shown in Figure 6, where all of the correlation coefficients show a clearly increasing pattern. This indicates that the dependence between the main world economies becomes stronger and stronger in these years. Detailed analysis shows that the dependence level between the two European currencies was still very high at the beginning and becomes higher (about 0.8) at the end of the period. The other correlation coefficients have been very slow at the beginning. Those between Pound and CAD, and Euro and CAD are even slightly negative at that time. But they increase very quickly during the observation period until about 0.5 at the current end. At this end the cross-correlations between CAD and the other currencies seem to begin decreasing slightly again, because all other currencies become weaker but the CAD is still stronger (cf Figures 1 and 3).

Results obtained following the proposed model are very useful for risk management and portfolio optimisation. This will be explained briefly in the following. The estimated means, variances, covariances and correlation coefficients at the current end (with $x = 1$) for each currency (standardised for a 100 USD unit to keep them comparable) are shown in Table 1, where the diagonal elements in the second part are the estimated variances, those in the upper triangular part are the estimated covariances while the estimated correlation coefficients are listed in the lower triangular part. Following the mean, the Yen is the

Table 1. $\hat{\mu}_i(1)$, $\hat{\sigma}_{ij}(1)$ and $\hat{\rho}_{ij}(1)$ (italic) (for 100 USD).

		$\hat{\sigma}_{ij}(1)$ $(j \geq i)$, $\hat{\rho}_{ij}(1)$ $(j < i)$			
Names	$\hat{\mu}_i(1)$	Pound	Yen	Euro	CAD
Pound	-0.0598	0.3000	0.2254	0.2574	0.1302
Yen	-0.1087	*0.629*	0.3722	0.2326	0.1311
Euro	-0.0831	*0.796*	*0.618*	0.3485	0.1542
CAD	0.0203	*0.467*	*0.423*	*0.534*	0.2670

weakest and the Euro the second weakest currencies. Following the variance (risk) the Euro is the weakest and the Yen the second weakest currencies. Following both criteria the CAD is the strongest and the Pound the second strongest currencies. This information is very useful for decision making, for instance to calculate an optimal portfolio.

7 Concluding remarks

We see that this article provides a good practical model and estimation methods for slowly changing drifts, variance and correlation coefficients of underlying stochastic process. It is the first multivariate model in the literature for studying slow change stochastic process; it develops effect estimation methods and asymptotic theory; and it assesses the effect of errors in the short term forecasts. Also, any increase in the number of component of the model does not increase the number of unknown parameters, nor the estimation burden, hence the model and fitting method are suitable for jointly modelling any number of assets in practice.

Some important open questions are still there and left for further study. These questions include extension of the model to allow for autocorrelations, development of significant tests and the need of robust estimation under asymmetric distribution and possibly outliers in the data.

Acknowledgments

The data used in this paper are downloaded from the data releases of the US Federal Reserve Bank under the address 'http://www.federalreserve.gov/releases/h10/'. We are grateful to Prof. Serguei Foss and Dr. George Streftaris, Heriot-Watt University, and Prof. Winfried Pohlmeier, University of Konstanz, for useful discussions and comments. We are also grateful to Miss Jiangjiang Yu and Mr Yao-Chih Chen who carried out some empirical analysis using similar idea in their MSc project under the supervision of the first author.

Appendix: Proofs of results

Proof of Theorem 2. 1) The results are just special cases of those given in Theorem 1.

2) Note that $\hat{\mu}_i(x)$ and $\hat{\mu}_j(x)$ are two linear estimators with the same weights generated by the local polynomial approach. Let w_t denote the weights. We have

$$\hat{\mu}_i(x) = \sum_{t=1}^n w_t z_{it} \qquad \text{and} \qquad \hat{\mu}_j(x) = \sum_{t=1}^n w_t z_{jt}.$$

Note that, for $t_1 \neq t_2$, z_{it_1} are independent of z_{it_2} and z_{jt_2}. Hence

$$\begin{aligned}
\mathrm{cov}\left(\hat{\mu}_i(x), \hat{\mu}_j(x)\right) &= \mathrm{cov}\left(\sum_{t=1}^n w_t z_{jt}, \sum_{t=1}^n w_t z_{jt}\right) \\
&= \sum_{t=1}^n \mathrm{cov}\left(w_t z_{jt}, w_t z_{jt}\right) \\
&= \sum_{t=1}^n w_t^2 \mathrm{cov}\left(z_{jt}, z_{jt}\right) \\
&= \sum_{t=1}^n w_t^2 \sigma_{ij}(x_t). \qquad \text{(A.1)}
\end{aligned}$$

Under the conditions of Theorem 2 we have $w_t = 0$ for $|x - x_t| > h$ and

$$\sigma_{ij}(x_t) = \sigma_{ij}(x)[1 + O(h)] = \sigma_{ij}(x)[1 + o(1)]$$

for $|x - x_t| \leq h$. Following known results in nonparametric regression it is easy to show that $\sum w_t^2 = (nh)^{-1} R[1 + o(1)]$ (see e.g. Wand and Jones, 1995 and Fan and Gijbels, 1996). Hence,

$$\mathrm{cov}\left(\hat{\mu}_i(x), \hat{\mu}_j(x)\right) = \sum_{t=1}^n w_t^2 \sigma_{ij}(x_t)$$

$$
\begin{aligned}
&= \sigma_{ij}(x)[1+o(1)]\sum_{t=1}^{n} w_t^2 \\
&= (nh)^{-1}R\sigma_{ij}(x)[1+o(1)]. \qquad \text{(A.2)}
\end{aligned}
$$

3) Results in this part can be shown by straightforward calculations. In the following we will give an indirect proof to show more details. Denote by $B^S_{\hat{\boldsymbol{\mu}}}(x)$ and $V^S_{\hat{\boldsymbol{\mu}}}(x)$ respectively the bias and variance of the estimated portfolio returns. Then it is easy to show that

$$
MSE^S_{\hat{\boldsymbol{\mu}}}(x) = [B^S_{\hat{\boldsymbol{\mu}}}(x)]^2 + V^S_{\hat{\boldsymbol{\mu}}}(x). \qquad \text{(A.3)}
$$

Following results in part 1) we have

$$
\begin{aligned}
B^S_{\hat{\boldsymbol{\mu}}}(x) &= E\left\{\sum_{i=1}^{k} S_i\hat{\mu}_i(x)\right\} - \sum_{i=1}^{k} S_i\mu_i(x) \\
&= E\left\{\sum_{i=1}^{k} S_i[\hat{\mu}_i(x)-\mu_i(x)]\right\} \\
&= \left\{\sum_{i=1}^{k} S_i\mu^{(p+1)}(x)\right\}\frac{\beta h^{p+1}}{(p+1)!}[1+o(1)]. \qquad \text{(A.4)}
\end{aligned}
$$

Furthermore, following (A.2) we have

$$
\begin{aligned}
V^S_{\hat{\boldsymbol{\mu}}}(x) &= \text{var}\left\{\sum_{i=1}^{k} S_i\hat{\mu}_i(x)\right\} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} S_iS_j\text{cov}\left(\hat{\mu}_i(x),\hat{\mu}_j(x)\right) \\
&= \left\{\sum_{j=1}^{k} S_iS_j\sigma_{ij}\right\}\frac{R}{nh}[1+o(1)]. \qquad \text{(A.5)}
\end{aligned}
$$

Equation (11) is proved by inserting those in (A.4) and (A.5) into (A.3).

4) Now, note that both T_1^S and T_2^S are both non-negative. The dominated part of $MSE^S_{\hat{\boldsymbol{\mu}}}(x)$ given in (11) is concave in h, provided that T_1^S and T_2^S are both non-zero, which is minimized by h_S^{opt} given in (14). ◇

Proof of Theorem 3. Let w_{it} denote the weights for $\hat{\sigma}_i^2(1)$ and w_{itk}, $k=1,...,n$, those for $\hat{\mu}_i(x_t)$. Then $w_{it} = O[(nb_i)^{-1}]$ for $1-x_t \le b_i$ and otherwise zero, and $w_{itk} = O[(nh_i)^{-1}]$

for $|x_k - x_t| \leq h_i$ and zero otherwise. Let $\xi_{it} = z_{it} - \mu_i(x_t)$. We have $\xi_{it} = \sigma_i(x_t)\tilde{\epsilon}_{it}$ and $\xi_t^{ij} = \xi_{it}\xi_{jt} = \sigma_i(x_t)\tilde{\epsilon}_{it}\sigma_j(x_t)\tilde{\epsilon}_{jt}$, where ξ_t^{ij} and $\tilde{\epsilon}_{it}$ are as defined in the context of Theorem 4. Define $\eta_{it} = \tilde{\epsilon}_{it}^2 - 1$. Then, for given i, η_{it} are zero mean i.i.d. errors such that

$$\xi_{it}^2 = \sigma_i^2(x_t) + \sigma_i^2(x_t)\eta_{it}, \tag{A.6}$$

which represents a special nonparametric regression model with independent errors and scale change, where the scale function turns to be the same as the trend function. A key point here is that ξ_{it} are unobservable. In the following we will show first Theorem 3 holds, if $\mu_i(x_t)$ is known, i.e. if ξ_{it} are observable. Define

$$\tilde{\sigma}_i^2(1) = \sum_{t=1}^{n} w_{it}\xi_{it}^2, \tag{A.7}$$

which is a kernel estimator of $\sigma_i^2(1)$ but obtained under the assumption that ξ_{it} are observable, i.e. $\mu_i(x_t)$ are known. The asymptotic bias and variance of $\tilde{\sigma}_i^2(1)$ can be obtained by adapting well known results in nonparametric regression with a scale function. See e.g. Fan and Gijbels (1995), Efromovich (1999) and Feng (2004). However, for a kernel estimator at the endpoint, the weights are non-symmetric and the first order term in the Taylor expansion of $\sigma_i^2(x)$ can hence not be cancelled. Denote the bias of $\tilde{\sigma}_i^2(1)$ by $B[\tilde{\sigma}_i^2(1)]$. We have

$$\begin{aligned} B[\tilde{\sigma}_i^2(1)] &= \sum_{t=1}^{n} w_{it}E[\xi_{it}^2] - \sigma_i^2(1) \\ &= \alpha_i(\sigma_i^2)'(1)b_i[1 + o(1)] \end{aligned} \tag{A.8}$$

as given in 1) of Theorem 3.

The variance of $\tilde{\sigma}_i^2(1)$ is given by

$$\begin{aligned} \text{var}\,[\tilde{\sigma}_i^2(1)] &= \sum_{t=1}^{n} w_{it}^2\text{var}\,(\eta_{it}) = \sum_{t=1}^{n} w_{it}^2\text{var}\,(\xi_{it}^2) \\ &\doteq \gamma_{ii}^2(1)\sum_{t=1}^{n} w_{it}^2 = (nb_i)^{-1}V_i\gamma_{ii}^2(1)[1 + o(1)] \end{aligned} \tag{A.9}$$

as given in 2) of Theorem 3, because $\sum_{t=1}^{n} w_{it}^2 \doteq (nb_i)^{-1}V_i$.

Results in 3 and 4 of Theorem 3 follow directly from those in the first two parts.

We now show that the changes of the asymptotic bias and variance of $\hat{\sigma}_i^2(1)$ caused by the error in $\hat{\mu}_i(x_t)$ are both negligible under A4$'$. Note that $r_{it} = z_{it} - \hat{\mu}_i(x_t)$ and

$$\begin{aligned} r_{it}^2 &= z_{it}^2 - 2z_{it}\hat{\mu}_i(x_t) + \hat{\mu}_i^2(x_t) \\ &= \xi_{it}^2 + 2\xi_{it}\Delta_{it} + \Delta_{it}^2, \end{aligned} \tag{A.10}$$

where $\Delta_{it} = \mu_i(x_t) - \hat{\mu}_i(x_t)$ is the estimation error in $\hat{\mu}_i(x_t)$, for which we have

$$\Delta_{it} = O(h_i^2) + O_p[(nh_i)^{-1/2}]. \tag{A.11}$$

Following (A.10) we have

$$E[\hat{\sigma}_i^2(1) - \sigma_i^2(1)] = B[\tilde{\sigma}_i^2(1)] + \sum_{t=1}^{n} w_{it}[2E(\xi_{it}\Delta_{it}) + E(\Delta_{it}^2)]. \tag{A.12}$$

Furthermore,

$$\begin{aligned} E(\xi_{it}\Delta_{it}) &= E\left[\xi_{it}\left(\sum_{k=1}^{n} w_{itk}z_{ik} - \mu_i(x_t)\right)\right] \\ &= E\left[\xi_{it}\left(\sum_{k=1}^{n} w_{itk}[\xi_{ik} + \mu_i(x_k)] - \mu_i(x_t)\right)\right] \\ &= w_{itt}\sigma_i^2(x_t) = O[(nh_i)^{-1}], \end{aligned} \tag{A.13}$$

since ξ_{it} are i.i.d. with $E(\xi_{it}) = 0$ and $E(\xi_{it}^2) = \sigma_i^2(x_t)$, and

$$\begin{aligned} E(\Delta_{it}^2) &= \text{MSE}[\hat{\mu}_i(x_t)] \\ &= O[h_i^4 + (nh_i)^{-1}]. \end{aligned} \tag{A.14}$$

Condition A4$'$ means that $h_i^4 = o(n^{-2/3})$, $(nh_i)^{-1} = o(n^{-2/3})$ and $b_i > O(n^{-2/3})$. This ensures that $O[h_i^4 + (nh_i)^{-1}] = o(b_i) = o\{B[\tilde{\sigma}_i^2(1)]\}$. Observe that $\sum w_{it} = 1$ we obtain

$$\begin{aligned} E[\hat{\sigma}_i^2(1) - \sigma_i^2(1)] &= B[\tilde{\sigma}_i^2(1)] + O[h_i^4 + (nh_i)^{-1}] \\ &= B[\tilde{\sigma}_i^2(1)][1 + o(1)]. \end{aligned} \tag{A.15}$$

Now we will analyze the effect of the error in $\hat{\mu}_i(x_t)$ on $\text{var}\,(\hat{\sigma}_i(1))$. Note first that

$$\text{var}\,[r_{it}^2] = \text{var}\,[\xi_{it}^2][1 + o(1)] = \text{var}\,[z_{it}^2][1 + o(1)]. \tag{A.16}$$

Although $\text{cov}(z_{it_1}^2, z_{it_2}^2) = \text{cov}(\xi_{it_1}^2, \xi_{it_2}^2) = 0$ for $t_1 \neq t_2$, but $\text{cov}(r_{it_1}^2, r_{it_2}^2) \neq 0$. Following the first equation in (A.10) we have

$$\begin{aligned}\text{cov}(r_{it_1}^2, r_{it_2}^2) &= \text{cov}(z_{it_1}^2 - 2z_{it_1}\hat{\mu}_i(x_{t_1}) + \hat{\mu}_i^2(x_{t_1}), z_{it_2}^2 - 2z_{it_2}\hat{\mu}_i(x_{t_2}) + \hat{\mu}_i^2(x_{t_2})) \\ &= C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8 + C_9,\end{aligned}$$

where $C_1 = \text{cov}(z_{it_1}^2, z_{it_2}^2) = 0$, $C_2 = -2\text{cov}(z_{it_1}^2, z_{it_2}\hat{\mu}_i(x_{t_2}))$, $C_3 = \text{cov}(z_{it_1}^2, \hat{\mu}_i^2(x_{t_2}))$,

$$C_4 = -2\text{cov}(z_{it_1}\hat{\mu}_i(x_{t_1}), z_{it_2}^2), \quad C_5 = 4\text{cov}(z_{it_1}\hat{\mu}_i(x_{t_1}), z_{it_2}\hat{\mu}_i(x_{t_2})),$$

$$C_6 = -2\text{cov}(z_{it_1}\hat{\mu}_i(x_{t_1}), \hat{\mu}_i^2(x_{t_2})), \quad C_7 = \text{cov}(\hat{\mu}_i^2(x_{t_1}), z_{it_2}^2),$$

$$C_8 = -2\text{cov}(\hat{\mu}_i^2(x_{t_1}), z_{it_2}\hat{\mu}_i(x_{t_2})), \quad C_9 = \text{cov}(\hat{\mu}_i^2(x_{t_1}), \hat{\mu}_i^2(x_{t_2})).$$

It can be shown that $C_2 = C_4 = C_5 = 0$. This will be shown for C_2.

$$\begin{aligned}C_2 &= -2\text{cov}\left(z_{it_1}^2, z_{it_2}\sum_{k=1}^{n} w_{it_2k}z_{ik}\right) \\ &= -2\sum_{k=1}^{n} w_{it_2k}\text{cov}(z_{it_1}^2, z_{it_2}z_{ik}) = 0, \qquad \text{(A.17)}\end{aligned}$$

because z_{it} are i.i.d., $t_1 \neq t_2$ and hence $\text{cov}(z_{it_1}^2, z_{it_2}z_{ik}) = 0$, for all k. Hence,

$$\text{cov}(r_{it_1}^2, r_{it_2}^2) = C_3 + C_6 + C_7 + C_8 + C_9. \qquad \text{(A.18)}$$

For C_3 we have

$$\begin{aligned}C_3 &= \text{cov}\left(z_{it_1}^2, \sum_{k=1}^{n}\sum_{l=1}^{n} w_{it_2k}w_{it_2l}z_{ik}z_{il}\right) \\ &= \sum_{k=1}^{n}\sum_{l=1}^{n} w_{it_2k}w_{it_2l}\text{cov}(z_{it_1}^2, z_{ik}z_{il}) \\ &= w_{it_2t_1}^2\text{var}(z_{it_1}^2) \qquad \text{(A.19)}\end{aligned}$$

due to the term with $k = l = t_1$, because all other covariances are equal to zero. Similar calculations lead to

$$C_6 = -2w_{it_1t_1}w_{it_2t_1}^2\text{var}(z_{it_1}^2), \qquad C_7 = w_{it_1t_2}^2\text{var}(z_{it_2}^2),$$

$$C_8 = -2w_{it_1t_2}^2w_{it_2t_2}\text{var}(z_{it_2}^2), \qquad C_9 = \sum_{k=1}^{n} w_{it_1k}^2w_{it_2k}^2\text{var}(z_{ik}^2).$$

Observe the properties of the weights it is easy to see that $C_3 = O(C_7) = O[(nh_i)^{-2}]$, $C_6 = O(C_8) = O(C_9) = O[(nh_i)^{-3}] = o(C_3)$. That is

$$\text{cov}\,(r^2_{it_1}, r^2_{it_2}) \doteq w^2_{it_2t_1}\text{var}\,(z^2_{it_1}) + w^2_{it_1t_2}\text{var}\,(z^2_{it_2}) = O[(nh_i)^{-2}]. \tag{A.20}$$

We see $\text{cov}\,(r^2_{it_1}, r^2_{it_2}) = o(n^{-1})$, if $h_i > O(n^{-1/2})$, which is implied by A4′. Hence we have

$$\begin{aligned}\text{var}\,[\hat{\sigma}^2_i(1)] &= \text{var}\left[\sum_{t=1}^n w_{it} r^2_{it}\right] \\ &= \sum_{t=1}^n w^2_{it}\text{var}\,(r^2_{it}) + \sum_{t_1=1}^n \sum_{\substack{t_2=1 \\ t_2\neq t_1}}^n w_{it_1}w_{it_2}\text{cov}\,(r^2_{it_1}, r^2_{it_2}) \\ &\doteq \sum_{t=1}^n w^2_{it}\text{var}\,(r^2_{it}) \\ &\doteq \sum_{t=1}^n w^2_{it}\text{var}\,(\xi^2_{it}) = \text{var}\,[\tilde{\sigma}^2_i(1)]. \end{aligned} \tag{A.21}$$

Theorem 3 is proved. ◇

Proof of Theorem 4. Note that the biases are of the same order as in Theorem 3. And the variances and convariances are of the same order as the variances in Theorem 3. The estimation errors in $\hat{\mu}_i(x_t)$ are also negligible under A4′. In the following proof the residuals r_{it} will hence be simply replaced by ξ_{it}.

By extending the ideas described at the beginning of the proof of Theorem 3 we can obtain

$$\xi^{ij}_t = \sigma_{ij}(x_t) + \sigma_i(x_t)\sigma_j(x_t)\zeta^{ij}_t, \tag{A.22}$$

where $\zeta^{ij}_t = \tilde{\epsilon}_{it}\tilde{\epsilon}_{jt} - \rho_{ij}(x_t)$ are i.i.d. zero mean random variables. This is again a special nonparametric regression model with i.i.d. errors and a scale function, where the trend is the corresponding covariance function. Of course (A.6) is a special case of (A.22) with $i = j$.

1) As mentioned before, the bias of $\hat{\sigma}_{ij}(1)$ is not affected by the scale function and is the same as in common nonparametric regression. The proof of this part is hence omitted.

2) Now let w_t denote the common weights for all estimators at $x = 1$, which are determined

by $W^r(u)$ defined in Theorem 4 and the bandwidth b. We have

$$\hat{\sigma}_{ij}(1) = \sum_{t=1}^{n} w_t \xi_t^{ij}$$

and

$$\begin{aligned} \operatorname{cov}(\hat{\sigma}_{ij}(1), \hat{\sigma}_{lm}(1)) &= \sum_{t_1=1}^{n}\sum_{t_2=1}^{n} w_{t_1} w_{t_2} \operatorname{cov}(\xi_{t_1}^{ij}, \xi_{t_2}^{lm}) \\ &= \sum_{t=1}^{n} w_t^2 \operatorname{cov}(\xi_t^{ij}, \xi_t^{lm}) \\ &= \gamma_{ij,\,lm} \sum_{t=1}^{n} w_t^2 \end{aligned} \tag{A.23}$$

following the definition before, since $\operatorname{cov}(\xi_{t_1}^{ij}, \xi_{t_2}^{lm}) = 0$ for $t_1 \neq t_2$. Furthermore we have

$$\operatorname{cov}(\hat{\sigma}_{ij}(1), \hat{\sigma}_{lm}(1)) = V\gamma_{ij,\,lm}(nb)^{-1}[1 + o(1)], \tag{A.24}$$

because $\sum_{t=1}^{n} w_t^2 = V(nb)^{-1}[1 + o(1)]$, where $V = \int_{-1}^{0}(W^r(u))^2 du$ as defined in Theorem 4.

3) Denote by B the bias. It is easy to show that, for given i, j, l, m,

$$E\{[\hat{\sigma}_{ij}(1) - \sigma_{ij}(1)][\hat{\sigma}_{lm}(1) - \sigma_{lm}(1)]\} = B[\hat{\sigma}_{ij}(1)]B[\hat{\sigma}_{lm}(1)] + \operatorname{cov}[\hat{\sigma}_{ij}(1), \hat{\sigma}_{lm}(1)] \tag{A.25}$$

Using results in 1) we have

$$B[\hat{\sigma}_{ij}(1)]B[\hat{\sigma}_{lm}(1)] = \alpha^2 \sigma'_{ij}(1)\sigma'_{lm}(1) b^2[1 + o(1)]. \tag{A.26}$$

Insert these results and those in 2) into (24) we obtain the results in 3) of Theorem 4.

4) Results in this part follow from those in 3) directly.

Proof of Theorem 5. 1) Under the smoothness assumption on μ_i we have $\mu_i(x_{n+t}) = \mu_i(1) + O(Tn^{-1})$ for any $t \leq T$. Following (32) and (33) we have

$$\begin{aligned} E[\hat{Y}_{i(n+T)} - E(Y_{i(n+T)}|Y_{in})]^2 &= E[T\hat{\mu}_i(1) - \sum_{t=1}^{T} \mu_i(x_{n+t})]^2 \\ &= T^2 E\{[\hat{\mu}_i(1) - \mu_i(1)] + O(Tn^{-1})]\}^2. \end{aligned} \tag{A.27}$$

Note that $\hat{\mu}_i(1) - \mu_i(1) = O(h_i^{(p+1)}) + O_p[(nh_i)^{-1/2}]$. The condition $T^2(nh_i)^{-1} \to 0$ results in $T = o(nh_i^{1/2})$ and $Tn^{-1} = o[n^{-1/2}h_i^{1/2}] = o[\hat{\mu}_i(1) - \mu_i(1)]$. Hence,

$$\begin{aligned} E[\hat{Y}_{i(n+T)} - E(Y_{i(n+T)}|Y_{in})]^2 &\doteq T^2 E[\hat{\mu}_i(1) - \mu_i(1)]^2 \\ &= T^2 MSE[\hat{\mu}_i(1)]. \end{aligned} \tag{A.28}$$

Furthermore, A5 ensures that $T^2 MSE[\hat{\mu}_i(1)] \to 0$, i.e. $\hat{Y}_{i(n+T)}$ is consistent.

2) Similarly, under the conditions of Theorem 5 it can be shown that

$$\begin{aligned} E[\hat{P}^S_{n+T} - E(P^S_{n+T}|\mathbf{Y}_n)]^2 &\doteq T^2 E\left\{\sum_{i=1}^{k} S_i[\hat{\mu}_i(1) - \mu_i(1)]\right\}^2 \\ &= T^2 MSE^S_{\hat{\boldsymbol{\mu}}}, \end{aligned} \tag{A.29}$$

$T^2 MSE^S_{\hat{\boldsymbol{\mu}}} \to 0$ under A5 and $\hat{P}^S_{n+T}$ is consistent. ◇

Proof of Theorem 6. 1) Note that $\hat{\sigma}_i(1) = \sqrt{\hat{\sigma}^2_i(1)}$. Based on Taylor expension of random variables it can be shown that

$$MSE[\hat{\sigma}_i(1)] = O(MSE[\hat{\sigma}^2_i(1)]). \tag{A.30}$$

Analogously to the analysis in 1) of the proof of Theorem 5 we have $\sigma^2_i(x_{n+t}) = \sigma_i(1) + O(Tn^{-1})$ for any $t \le T$ and

$$\begin{aligned} E[\sqrt{T}\hat{\sigma}_i(1) - SD_T]^2 &= E\left\{\sqrt{T}\hat{\sigma}_i(1) - \sqrt{\left[\sum_{t=1}^{T} \sigma^2_i(x_{n+t})\right]}\right\}^2 \\ &\doteq E\left\{\sqrt{T}\left[\hat{\sigma}_i(1) - \sigma_i(1) + O(T^{1/2}n^{-1/2})\right]\right\}^2 \\ &\doteq T \cdot E\left[\hat{\sigma}_i(1) - \sigma_i(1)\right]^2 \\ &= O(T \cdot MSE[\hat{\sigma}^2_i(1)]). \end{aligned} \tag{A.31}$$

Condition A5′ ensures that $T^{1/2}n^{-1/2} = o[\hat{\sigma}_i(1) - \sigma_i(1)]$, $T \cdot MSE[\hat{\sigma}^2_i(1)] \to 0$ and $\sqrt{T}\hat{\sigma}_i(1)$ is a consistent estimator of SD_T.

2) Similarly, it can be shown that

$$MSE\left[\sqrt{S'\hat{\boldsymbol{\Sigma}}(1)S}\right] \doteq MSE^S_{\hat{\Sigma}}(1), \tag{A.32}$$

$$\begin{aligned} E\left[\sqrt{TS'\hat{\Sigma}(1)S} - SD^S_T\right]^2 &\doteq T \cdot MSE\left[\sqrt{S'\hat{\boldsymbol{\Sigma}}(1)S}\right] \\ &= T \cdot MSE^S_{\hat{\Sigma}}(1). \end{aligned} \tag{A.33}$$

And $T \cdot MSE^S_{\hat{\Sigma}}(1) \to 0$ under A5′ so that $\sqrt{TS'\hat{\Sigma}(1)S}$ is consistent. ◇

REFERENCES

Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities, *Econometrica*, **64,** 527–560.

Beran, J. and Y. Feng (2002) SEMIFAR models – a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Comptat. Statist. & Data Anal.*, **40**, 393–419.

Beran, J. and Ocker, D. (1999) SEMIFAR forecasts, with applications to foreign exchange rates. *J. Statistical Planning and Inference*, **80**, 137–153.

Bollerslev, T., R. F. Engle and J. Wooldridge (1988) A capital asset-pricingmodel with time-varying covariances. *Journal of Political Economy*, **96**, 116131.

Csörgö, S. and J. Mielniczuk (1995) Nonparametric regression under long-range dependent normal errors. *Annals of Statistics*, **23**, 1000–1014.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics*, **25**, 1–37.

Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes. *Annals of Statistics*, **28**, 1762–1794.

Efromovich, S. (1999) *Nonparametric curve estimation: Methods, Theory, and Applications.* New York: Springer.

Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, **50**, 987-1008

Engle, R. (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, **20**, 339–350.

Fan, J. and I. Gijbels (1995) Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B*, **57,** 371–394.

Fan, J. and Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika,* **85**, 645–60.

Feng, Y. (2004) Simultaneously modelling conditional heteroskedasticity and scale change. *Econometric Theory*, **20,** 563–596.

Gasser, T., A. Kneip and W. Köhler (1991) A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.,* **86,** 643–652.

Gordon, A.H. (1991) Global warming as a manifestation of a random walk. *Journal of Climate*, **4**, 589-597.

Härdle, W., A.B. Tsybakov and L. Yang (1998) Nonparametric vector autoregression. *J. Statist. Plann. Infer.,* **68,** 221–245.

Härdle, W., H. Herwatz and V. Spokoiny (2003) Time inhomogeneous multiple volatility modelling. *J. Financial Econometrics,* **1,** 55–99.

Hart, J. D. (1991) Kernel regression estimation with time series errors. *J. Roy. Statist. Soc. Ser. B*, **53**, 173–187.

Harvey, A. (1989) *Forecasting structural time series models and the Kalman filter.* Cambridge: Cambridge University Press.

Harvey, A., Ruiz, E. and N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.

Herzel, S., C. Starica and R. Tutuncu (2006) A non-stationary multivariate model for financial returns. Forthcoming in Statistics for dependent data, Ed. Patrice Bertail, and P. Doukhan, Springer.

Kärner, O. (2002) On nonstationarity and antipersistency in global temperature series. *Journal of Geophysical Research*, **107**, doi: 10.1029/2001JD002024.

Kijima, M. (2002) *Stochastic Processes with Applications to Finance*, Cambridge: Chapman and Hall.

Lovász, L. (1993) Random Walks on Graphs: A Survey, *Mathematical Studies*, **2**, 1–46.

Neigel, J. E. and J.C. Avise (1993) Application of a random walk model to geographic distributions of animal mitochondrial DNA variation, *Genetics*, **135**, 1209–20.

Ruppert, D. and M.P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346–1370.

Wand, M.P. and M.C. Jones (1995) *Kernel Smoothing*, London: Chapman & Hall.

Weiss, G.H. (1994) *Aspects and Applications of the Random Walk*, Amsterdam: North Holland Press.

Wu, W. and M. Pourahmadi (2003) Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, **90**, 831–844.

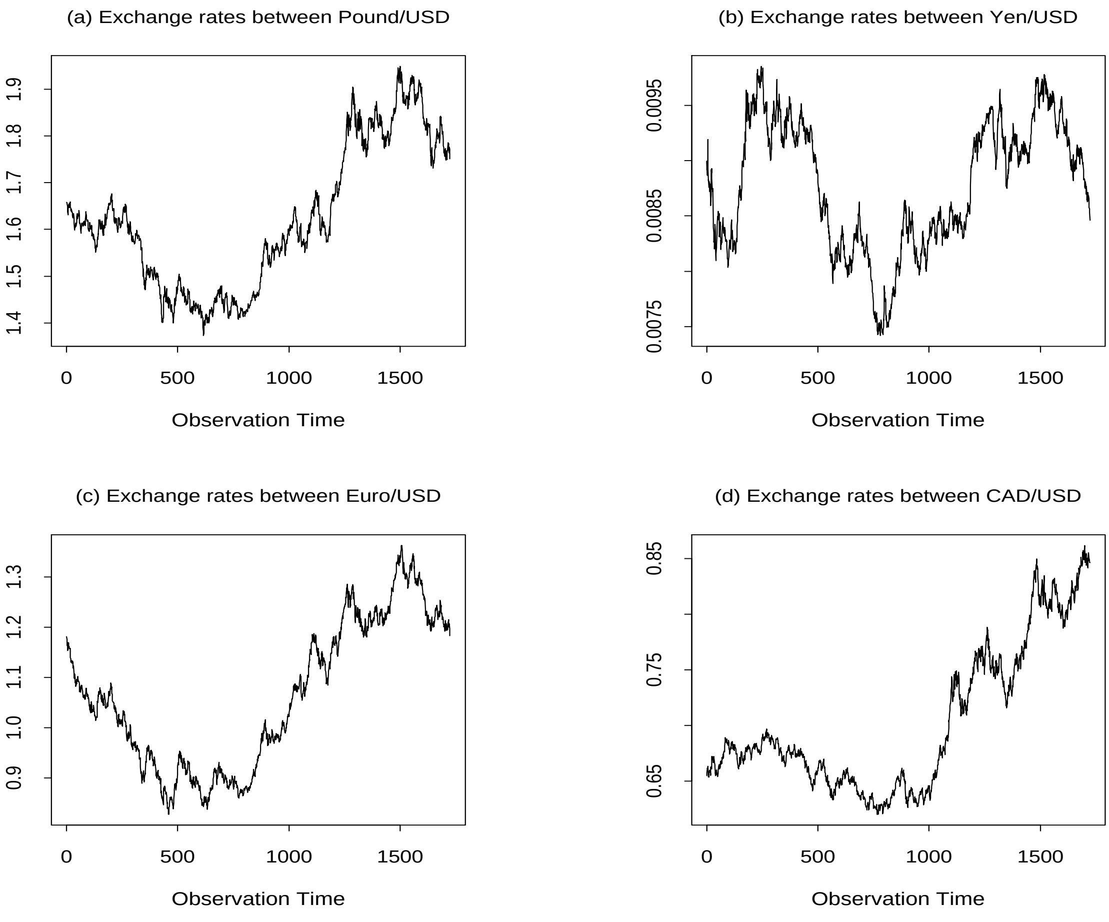

Figure 1: The original data of the selected exchange rate series.

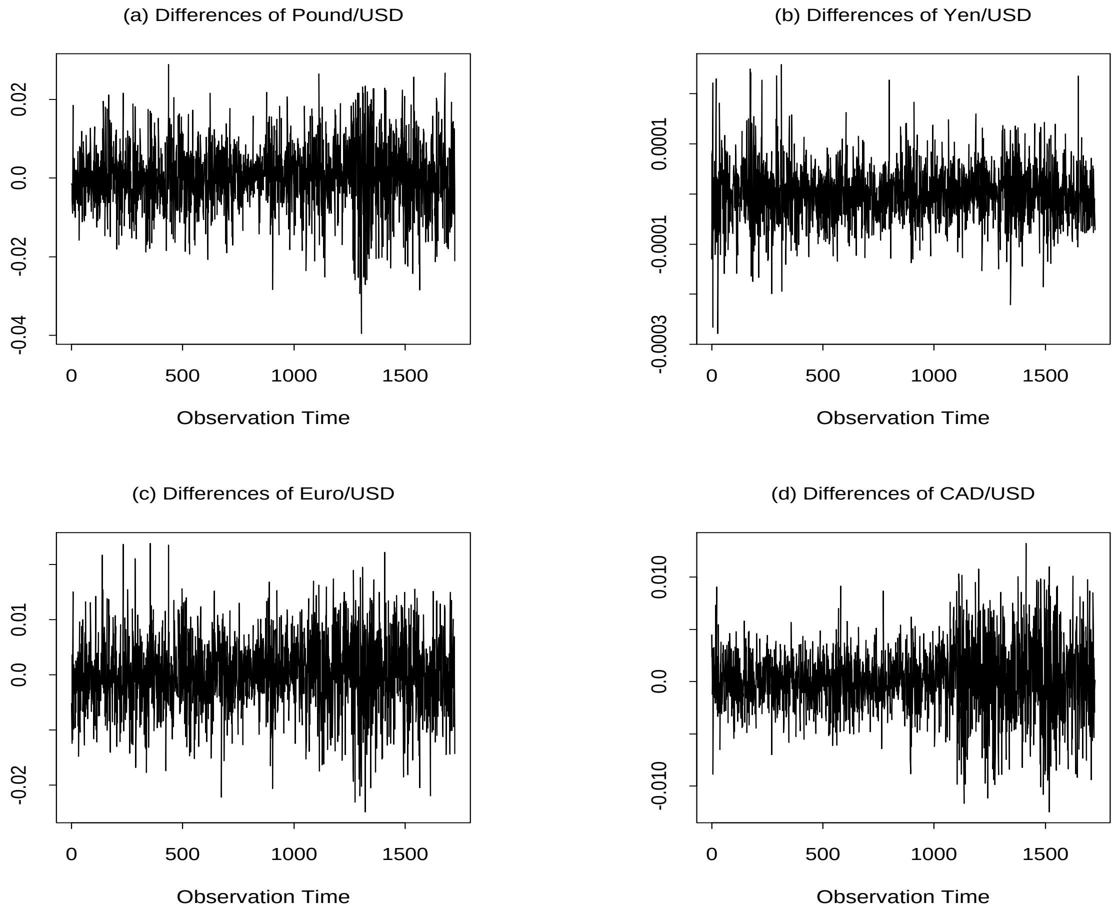

Figure 2: The difference series of the series shown in Figure 1.

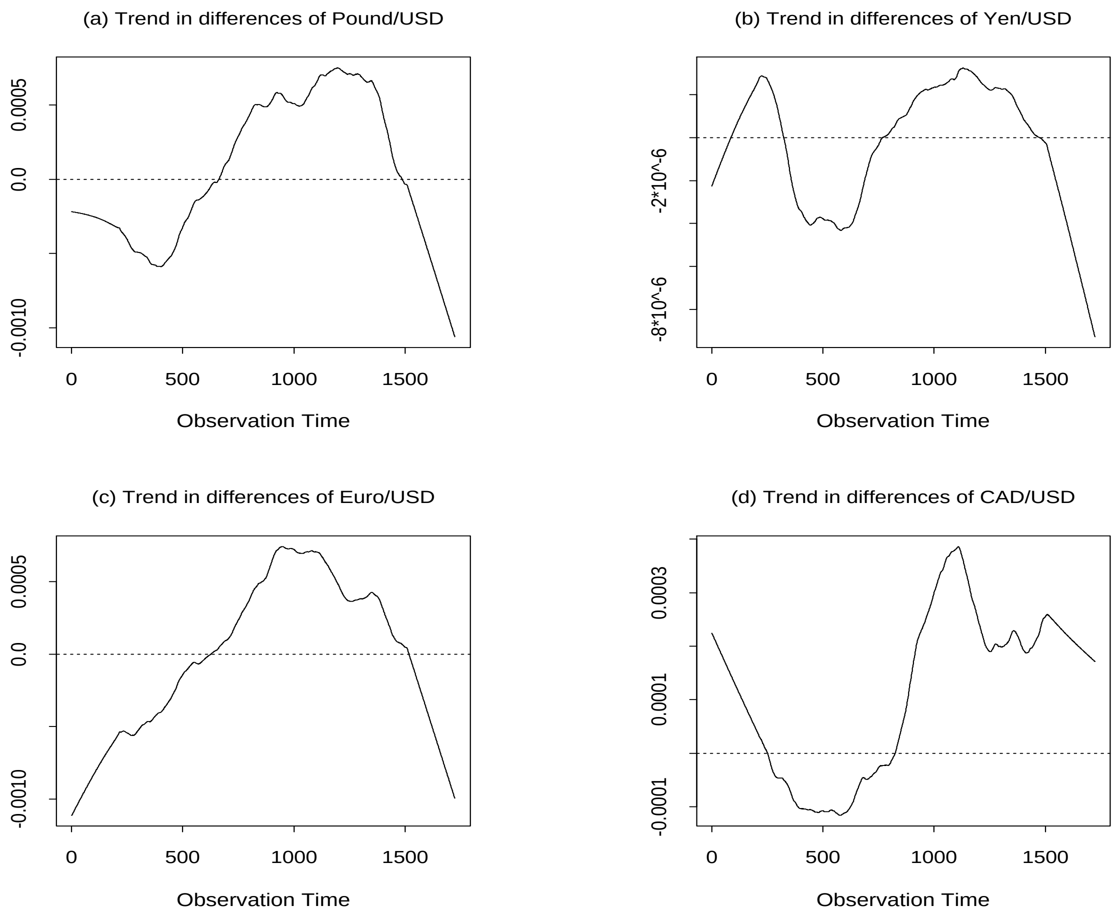

Figure 3: Estimated mean functions from the series in Figure 2 (with a horizontal line at zero).

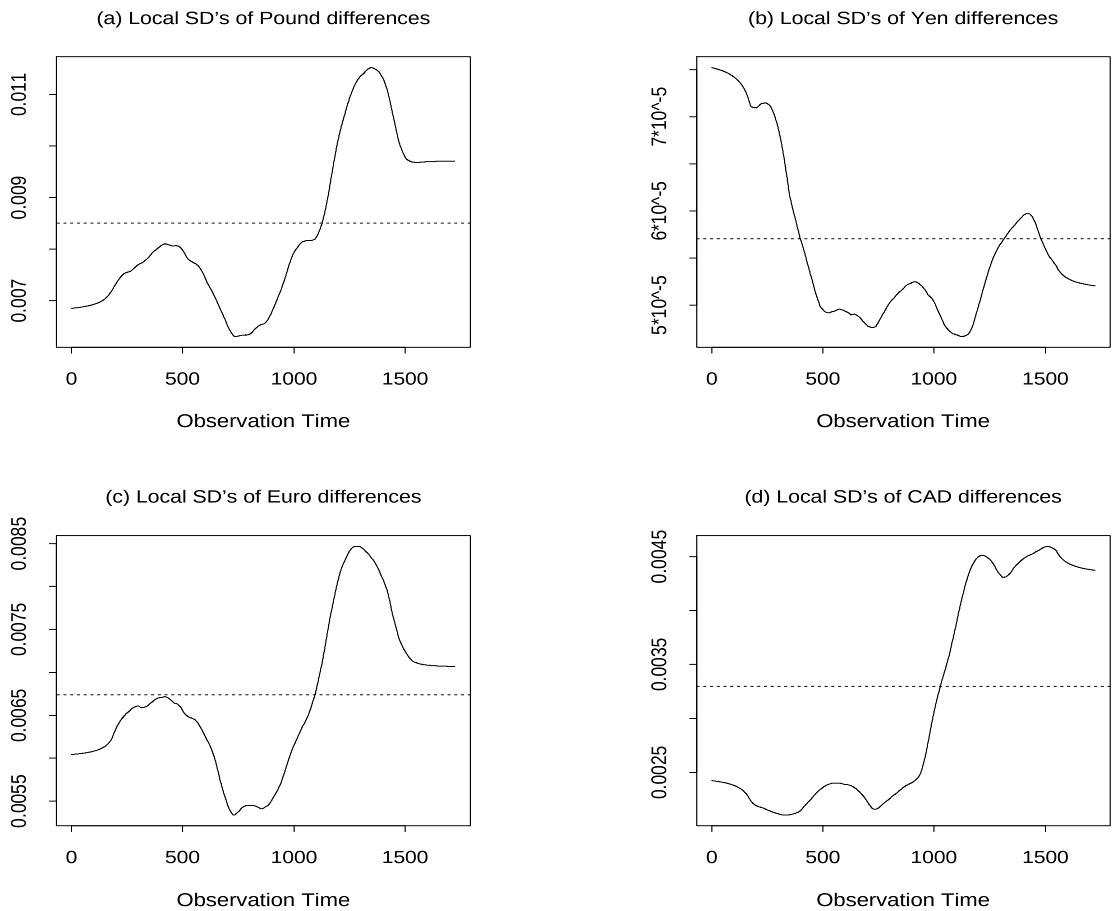

Figure 4: Estimated standard deviations together with the global sample values.

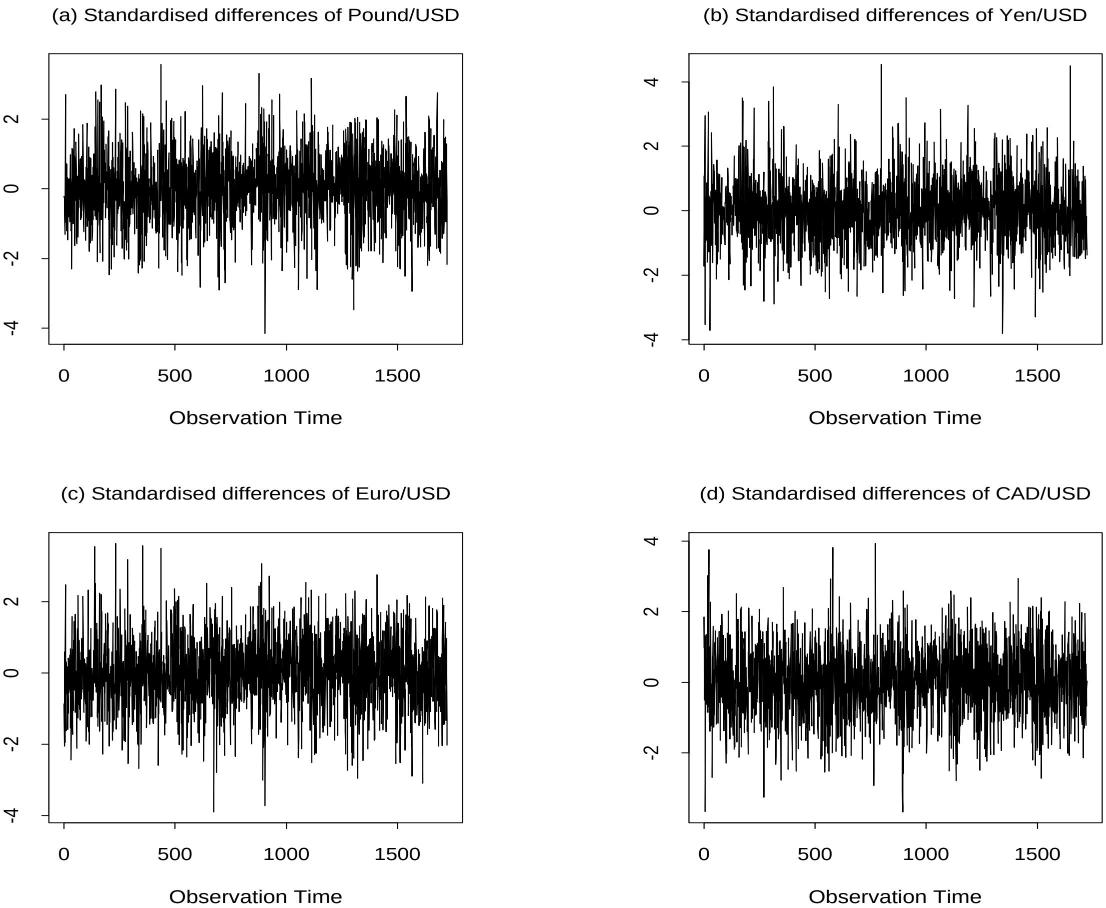

Figure 5: The standardised difference series.

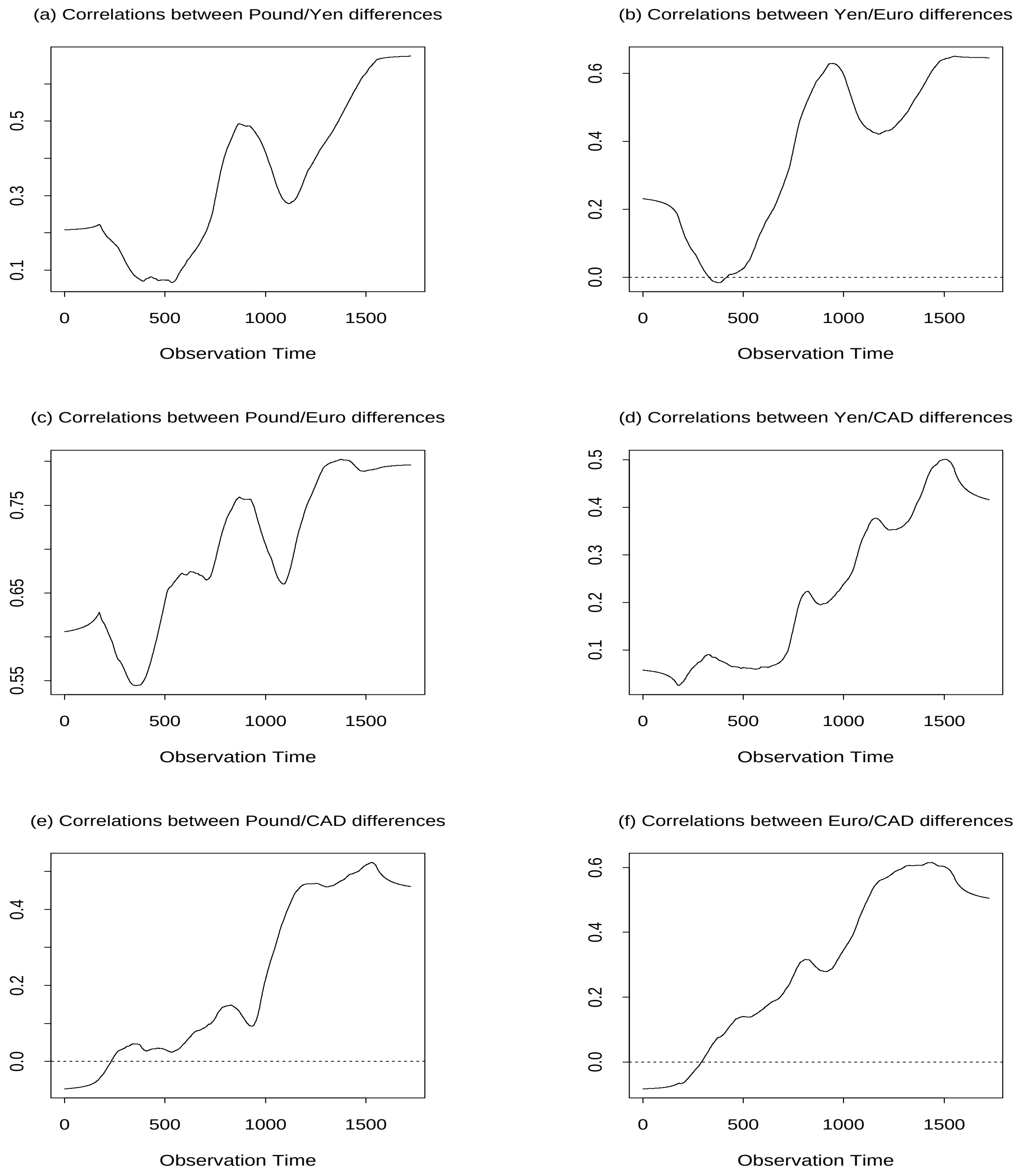

Figure 6: Estimated local correlation coefficients in all cases.