A local dynamic conditional correlation model

Yuanhua Feng

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

2006

A Local Dynamic Conditional Correlation Model

Yuanhua Feng

Department of Actuarial Mathematics and Statistics

Heriot-Watt University

Abstract: This paper introduces the idea that the variances or correlations in financial returns may all change conditionally and slowly over time. A multi-step local dynamic conditional correlation model is proposed for simultaneously modelling these components. In particular, the local and conditional correlations are jointly estimated by multivariate kernel regression. A multivariate k-NN method with variable bandwidths is developed to solve the curse of dimension problem. Asymptotic properties of the estimators are discussed in detail. Practical performance of the model is illustrated by applications to foreign exchange rates.

JEL Classification Codes: C32, G0, G1

Key Words: Local and conditional correlations, multivariate nonparametric ARCH, multivariate kernel regression, multivariate k-NN method.

1 Introduction

Financial econometrics becomes an active new discipline of economics (Engle, 2001) with one of its main focuses on the development of new tools for estimating and forecasting dynamic variances and correlations of financial assets. In the seminal paper Engle (1982) the ARCH (autoregressive conditional heteroskedasticity) model is introduced for modelling conditional variances, which is then generalized by Bollerslev (1986) to the GARCH (generalized ARCH) model. See Bollerslev et al. (1992) for an earlier survey on these models. For modelling dynamic correlations some multivariate GARCH (MGARCH) models with complex parameter specification, e.g. the vec-model (Bollerslev et al., 1988) and the BEKK (Engle and Kroner, 1995) are first introduced. A simple MGARCH, called the constant conditional correlation (CCC) model, was introduced by Bollerslev (1990), where the conditional variances are modelled by separate univariate GARCH models under constant correlation assumption.

*Dr. Yuanhua Feng, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, EH14 4AS, UK. Tel. +44 131 451 4365, fax: +44 131 451 3249. Email: Y.Feng@ma.hw.ac.uk

Recent research on these topics focuses on developing approaches with dynamic correlations and simple parameter specification so that the models are applicable to many assets. The well known proposal is the dynamic conditional correlation (DCC) model introduced by Engle (2002) (see also Engle and Sheppard, 2001). A similar approach is proposed by Tse and Tsui (2002). The DCC model is generalized by Cappiello et al. (2003) and Hafner and Franses (2003) to allow for richer correlation dynamics. Pelletier (2006) extended the CCC model to a regime-switching conditional correlation model, where the correlation matrix is constant within a regime but different across regimes. Silvennoinen and Teräsvirta (2005) extended this idea to a smooth-transition conditional correlation model by allowing for smooth change in the conditional correlations between two extreme states according to some transition function. Most recently, Hafner et al. (2005) introduced a semiparametric generalization of the CCC model, where conditional correlations are estimated by a univariate kernel regression. For more information on the development of MGARCH models we refer the reader to the recent review of Bauwens et al. (2005) and references therein.

In this paper cases with simultaneous smooth local changes and conditional dynamics in variances and correlations are considered are considered. A local dynamic conditional correlation (LDCC) model is introduced for all of these components, where mooth change in the mean of financial returns is also allowed. The variances are decomposed into a conditional and a local (unconditional) parts. The correlation structures are jointly modelled by a multivariate nonparametric ARCH-type approach with the observation time and some other variables based on past observations as regressors. The order of such a model does not depend on the number of assets. The LDCC model is hence applicable to cases with a large number of assets. The proposal is motivated by the observation that financial market conditions often change slowly which in turn causes slowly changing components in asset returns. Proper estimation of these quantities will improve the estimation of further parametric models and then improve the forecasting of future trends, variances and correlations. Modelling of conditional and local variances in univariate financial returns is studied by Feng (2004) under a semiparametric GARCH model. Most recently, Feng and Yu (2005) and Herzel et al. (2006) investigated the slowly changing variances and correlations in financial returns under a multivariate random walk model and a VAR(1) model respectively. To our knowledge multivariate models with both conditional and slowly changing unconditional correlations is not yet studied in the literature.

A multi-step semiparametric procedure is proposed for estimating the LDCC model. The local means and variance are estimated first using standard nonparametric regression approaches. Conditional variances are then estimated from the standardised observations using separate univariate GARCH models. A multivariate kernel regression is proposed for jointly estimating the local and conditional correlations. This idea can also be applied to time series smoothing involving common exogenous variables. A k-NN (k-nearest-neighbours) method is developed to solve the curse of dimension problem in multivariate nonparametric regression, which allows for automatic adaptation of the bandwidth to the design density. Asymptotic properties of the proposed estimators are discussed in detail. The use of causal smoothing technique is also investigated briefly. Practical performance of the proposal is illustrated by applications to several foreign exchange rate series. The idea to estimate and remove the slowly changing variances applies to any MGARCH model. Semiparametric generalizations of the CCC and DCC models are given as examples.

The paper is organized as follows. The model is defined in the next section. Section 3 describes the step-wise semiparametric estimation procedure. Asymptotic properties of the proposed estimators of the local and conditional correlations are investigated in Section 4. In Section 5 the model is applied to data examples. Final remarks in Section 6 concludes the paper. Some auxiliary results and proofs of theorems are given in the appendix.

2 The models

In this section we first define the semiparametric LDCC model. Related semiparametric extensions of the CCC and DCC models are then described.

2.1 The main approach

Let X_t, $t = 1, 2, ..., n$, be a vector return series of d financial assets, which are assumed to follow the nonparametric conditional heteroskedastic model

$$X_t = \mu(\tau_t) + r_t^*, \tag{1}$$

where $\tau_t = t/n$ denotes the re-scaled time, $\mu(\cdot)$ is the nonparametric local mean vector and

$$r_t^* | \mathcal{F}_{t-1} \sim N(0, \Sigma_t), \tag{2}$$

where $\mathcal{F}_{t-1}$ denotes the information set generated by past observations and the location, and Σ_t is the total covariance matrix. It is proposed to decompose Σ_t as follows.

$$\Sigma_t = D_t^L D_t^C R_t D_t^C D_t^L, \tag{3}$$

where $D_t^L = diag(\sigma_i(\tau_t))$, $D_t^C = diag(\sqrt{h_{it}})$ and $R_t = (\rho_{ijt})$, $i, j = 1, ..., d$, where $\sigma_i^2(\cdot)$ are the local variances, h_{it} are the conditional variances and ρ_{ijt} are the dynamic correlations which may depend on both the location and past observations. Furthermore, it is assume that $E(D_t^C) = I_d$ with I_d denoting the identity matrix, so that (3) is uniquely defined. Let $D_t = D_t^L D_t^C = diag(\sigma_i(\tau_t)\sqrt{h_{it}})$ which is the total standard deviation matrix. The total covariance matrix is decomposed into a conditional and an unconditional components. The latter depends on the middle-term market conditions and often changes smoothly. The process defined by (1) to (3) is non-stationary. Following Dahlhaus (1997) it can be shown that such a process is jointly locally stationary under suitable conditions.

The local variance $\sigma_i^2(\cdot)$ can be easily estimated by nonparametric regression. Let $r_t = (D_t^L)^{-1} r_t^*$ be the standardized observations. Following the CCC and DCC, the conditional variances can be modelled by univariate GARCH models,

$$h_{it} = \alpha_{i0} + \sum_{l=1}^{p_i} \alpha_{il} r_{i(t-l)}^2 + \sum_{m=1}^{q_i} \beta_{im} h_{i(t-m)}, \tag{4}$$

where $\alpha_{i0} > 0$, $\alpha_{il}, \beta_{im} \geq 0$ and $\sum_{l=1}^{p_i} \alpha_{il} + \sum_{m=1}^{q_i} \beta_{im} < 1$. The orders p_i and q_i of the GARCH models may different for different assets.

Now let $\epsilon_t = (D_t^C)^{-1} r_t$, where $\epsilon_t \sim N(0, R_t)$. Unlike the diagonal variance matrix, it is not easy to decompose R_t into a local and a conditional parts separately. We will propose to estimate the local and conditional dynamics in R_t jointly in a nonparametric way. Consider the conditional influence of p lagged observations. A nonparametric regression with ϵ_{t-j}, $j = 1, ..., p$, as regressors is not relevant in practice, because such a model cannot be applied to a large number of assets. In this paper some univariate random variables y_{jt} as functions of ϵ_{t-j}, $j = 1, ..., p$, which summarize the effect of $\epsilon_{t-j}\epsilon_{t-j}'$ on the correlation dynamics, will be used as regressors, where $p > 0$ is the order of the model which does not depend on the number of assets. This makes the model applicable to a multivariate time series with many components. Let and $\mathbf{y}_t = (y_{1t}, ..., y_{pt})'$. In the LDCC model defined in the following the local and conditional correlation matrix will be denoted by $R(\tau_t; \mathbf{y}_t) = (\rho_{ij}(\tau_t; \mathbf{y}_t))$ instead of by R_t. The proposed model is defined by

$$R(\tau_t; \mathbf{y}_t) = g(\tau_t; \mathbf{y}_t), \tag{5}$$

where $g(\cdot)$ is a smooth function. A model defined by (1) through (5) will be called a local dynamic conditional correlation (LDCC) model, which extends known models in the literature in different ways. A LDCC(0) model is a semiparametric extension of the CCC model (see later). Another closely related model is the semiparametric dynamic correlation model most recently proposed by Hafner et al. (2005), where $R(\cdot)$ is assumed to be a smooth function of a univariate observable variable. If some exogenous variables, e.g. returns of some other financial index are introduced into the LDCC model, we will obtain local extensions of the model in Hafner et al. (2005). A nonparametric GARCH is proposed by Bühlmann and McNeil (2002). It is worthwhile to introduce the latent variables $R(\tau_{t-i};\cdot)$, $i = 1, ..., q$, into (5) to extend the LDCC to a multivariate nonparametric GARCH-type model.

2.2 Combination with other models

The first part of the proposed model can be used to obtain semiparametric generalizations of well known approaches in the literature. Extensions of the CCC and the DCC models following this idea will be discussed here briefly. If it is assumed that the changes in the correlations only depend on the location but not on the past observations, we will have $R_t = R(\tau_t)$, which can be simply estimated from ϵ_t using nonparametric regression. Now, we obtain a generalized CCC model with slowly changing variances and correlations which is also a LDCC(0) model.

On the other hand, if it is assumed that the unconditional correlation matrix is constant, i.e. R_t only depends on the past observations but not on the location, we will obtain another simplified case. Now, R_t can be modelled by a parametric MGARCH model. For instance, following the DCC model R_t can be modelled by

$$\begin{aligned} R_t &= (diag(Q_t))^{-1/2}\, Q_t\, (diag(Q_t))^{-1/2}, \\ Q_t &= \left(1 - \sum_{l=1}^{L} \alpha_l - \sum_{m=1}^{M} \beta_m\right) \bar{R} + \sum_{l=1}^{L} \alpha_l (\epsilon_{t-l}\epsilon_{t-l}^{'}) + \sum_{m=1}^{M} \beta_m Q_{t-m}, \end{aligned} \tag{6}$$

where $\alpha_l, \beta_m \geq 0$, $\sum_{l=1}^{L} \alpha_l + \sum_{m=1}^{M} \beta_m < 1$ and $\bar{R}$ is the constant correlation matrix of ϵ_t. Equations (1) through (4) and (6) define a DCC model with slowly changing variances which is however not a special case of the LDCC model. Such a model can be estimated by combining the first stage of the algorithm proposed in the next section and the second stage of that in Engle and Sheppard (2001). Moreover, if R_t is assumed to follow (2.4) in Pelletier (2006) or (7) in Silvennoinen and Teräsvirta (2005) we will obtain corresponding extensions of their proposals.

3 The estimation

The LDCC model can be estimated using a step-wise procedure. The first stage of the procedure consists of some common non- and semiparametric estimators which applies to other MGARCH models. The second stage is a multivariate kernel regression for estimating $R(\cdot)$.

3.1 Estimating the means and the variances

Let x_t, $t = 0, 1, ..., n$, denote the observations and x_{it}, $i = 1, ..., d$, the i-th element of x_t. Let $K_{\mu_i}(u)$ be a kernel function and b_{μ_i} the bandwidth. Then $\mu_i(\tau)$, the i-th element of $\mu(\tau)$ can be estimated by solving the local linear regression problem

$$(\hat{a}_0, \hat{a}_1)' = \arg\min_{a_0, a_1} \sum_{t=1}^{n} (x_{it} - a_0 - a_1(\tau_t - \tau)^2 K_{\mu_i}\left(\frac{\tau_t - \tau}{b_{\mu_i}}\right). \tag{7}$$

We have $\hat{\mu}_i(\tau) = \hat{a}_0$. Now let $\hat{\mu}(\tau_t) = (\hat{\mu}_1(\tau_t), ..., \hat{\mu}_d(\tau_t))'$ and $\hat{r}_t^* = x_t - \hat{\mu}(\tau_t)$ denote the residuals. Denote by $V(\tau) = (\sigma_1^2(\tau), ..., \sigma_d^2(\tau))'$ the vector of the local variances. Let $K_{V_i}(u)$ be another kernel and b_{V_i} another bandwidth. It is proposed to estimate the local variances using a kernel estimator to ensure that $\hat{\sigma}_i(\tau) > 0$ a.s. (almost sure). Related proposals may be found e.g. in Fan and Yao (1998), Feng and Heiler (1998) and Härdle et al. (1998). We define

$$\hat{\sigma}_i^2(\tau) = \frac{\sum_{t=1}^{n} K_{V_i}\left(\frac{\tau_t - \tau}{b_{V_i}}\right)(\hat{r}_{it}^*)^2}{\sum_{t=1}^{n} K_{V_i}\left(\frac{\tau_t - \tau}{b_{V_i}}\right)} \tag{8}$$

and set $\hat{V}(\tau) = (\hat{\sigma}_1^2(\tau), ..., \hat{\sigma}_d^2(\tau))'$. If $R(\cdot)$ only depends on the location, the local covariances σ_{ij}, say, can be analogously estimated from $\hat{r}_{it}^*$ and $\hat{r}_{jt}^*$. Then $\rho_{ij}(\tau)$ can be estimated by

$$\hat{\rho}_{ij}(\tau) = \frac{\hat{\sigma}_{ij}(\tau)}{\hat{\sigma}_i(\tau)\hat{\sigma}_j(\tau)}.$$

By means of $\hat{V}$ we obtain the standardized residuals $\hat{r}_{it} = \hat{r}_{it}^*/\hat{\sigma}_i(\tau_t)$. $\hat{r}_{it}$ have asymptotically constant variance. Let $\theta_i = (\alpha_{i0}, \alpha_{i1}, ..., \alpha_{ip_i}, \beta_{i1}, ..., \beta_{iq_i})'$. The unknown parameter vector θ_i can be estimated from $\hat{r}_{it}$, $t = 1, ..., n$, using some standard software for fitting a univariate GARCH model. Then we will obtain $\hat{D}_t^C$. Let $\hat{r}_t = (\hat{r}_{1t}, ..., \hat{r}_{dt})'$ and $\hat{\epsilon}_t = (\hat{D}_t^C)^{-1}\hat{r}_t$. Assuming constant unconditional correlations, R_t can be now estimated from $\hat{\epsilon}_t$ following equations (7) and (8) in Engle and Sheppard (2001). This will not be discussed here in detail.

3.2 A multivariate k-NN kernel approach for $R(\cdot)$

Now, consider the estimation of $R(\cdot)$ in a LDCC(p) model with $p > 0$. In this paper the regressors y_{jt} for $j = 1, ..., p$ and $t = p+1, ..., n$ are defined by

$$y_{jt} = \mathbb{I}'\epsilon_{t-j}\epsilon'_{t-j}\mathbb{I}, \tag{9}$$

where $\mathbb{I}$ is a vector of d ones. Note that $y_{jt} = (\epsilon_{1(t-j)} + ... + \epsilon_{d(t-j)})^2 \geq 0$. It is well known that even in one-dimensional nonparametric regression for financial data we will be faced by the problem of data sparsity. In the current case this problem arises together with the curse of dimension. Now, the use of fixed bandwidths is not suitable. To improve the theoretical and practical performance of the proposed estimator and to ensure that the computer program will run smoothly without any numerical problem, a local multivariate k-NN method is developed. Consider the estimation of $R(\cdot)$ at τ and $\mathbf{y} = (y_1, ..., y_p)'$. Following this algorithm the bandwidth b for y_j, $j = 1, ..., p$, will adapt automatically to the design density. Let $t_0 = [n\tau]$ and assume that $t_0 > p$, where $[x]$ denote the largest integer which is smaller than x. Let k be a chosen integer such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. Let b_0 be the half bandwidth for the re-scaled time such that $b_0 \to 0$, $nb_0 \to \infty$ and $(nb_0)^{-1}k \to 0$, as $n \to \infty$. Here b_0 and k are two smoothing parameters chosen beforehand. Let $k_1 = [nb_0]$ and $k_0 = 2k_1 + 1$. k_0 is the total number of observations involved. The bandwidth used for y_j is defined in the following way.

1. Let $n_1 = t_0 - k_1$ and $n_2 = t_0 + k_1$, if $t_0 \geq k_1 + p$, or $n_1 = p + 1$ and $n_2 = p + k_0$, otherwise.

2. Let $d_t = [(\mathbf{y}_t - \mathbf{y})'(\mathbf{y}_t - \mathbf{y})]^{1/2}$ denote the Euclidean distance between $\mathbf{y}_t$ and $\mathbf{y}$.

3. Order d_t starting with the smallest value and define the bandwidth b for y_j, $j = 1, ..., p$, to be the k-th ordered d_t.

Following this algorithm always k observations will be selected from k_0 observations around t_0, independently of p and the design density. The bandwidth b_0 is chosen separately beforehand, because τ is of a different scale than y_j. To simplify the algorithm, kernel functions for τ and y_j are also chosen separately. Let $K_0(v)$ be a univariate kernel with support $[-1, 1]$ and $\mathbf{u} = (u_1, ..., u_p)'$ a p-dimensional spherical kernel defined on the unit ball. Then the finally used kernel is the product of $K_0(v)$ and $K(\mathbf{u})$. The proposed estimator is defined by

$$\hat{R}(\tau; \mathbf{y}) = \left(diag\left[\hat{Q}(\tau; \mathbf{y})\right]\right)^{-1/2} \hat{Q}(\tau; \mathbf{y}) \left(diag\left[\hat{Q}(\tau; \mathbf{y})\right]\right)^{-1/2} \tag{10}$$

with the matrix-wise multivariate kernel estimator

$$\hat{Q}(\tau;\mathbf{y}) = \frac{\sum_{t=n_1}^{n_2} K_0\left(\frac{\tau_t-\tau}{b_0}\right) K\left(\frac{y_{1t}-y_1}{b}, ..., \frac{y_{pt}-y_p}{b}\right) \hat{\epsilon}_t\hat{\epsilon}_t'}{\sum_{t=n_1}^{n_2} K_0\left(\frac{\tau_t-\tau}{b_0}\right) K\left(\frac{y_{1t}-y_1}{b}, ..., \frac{y_{pt}-y_p}{b}\right)} =: \sum_{t=n_1}^{n_2} w_t\hat{\epsilon}_t\hat{\epsilon}_t', \tag{11}$$

where n_1 and n_2 are defined before. The entries of $\hat{R}(\cdot)$ will be denoted by $\hat{\rho}_{ij}(\cdot)$ and those of $\hat{Q}(\cdot)$ by $\hat{q}_{ij}(\cdot)$. The weights w_t are determined by the kernels, b_0, k as well as the past observations ϵ_{t-1}, ..., ϵ_{t-p}, and are non-zero for the selected observations and zero otherwise. The curse of dimension problem is solved and a LDCC model of higher order can be easily fitted. In the above approach w_t are the same for all entries of $\hat{Q}(\cdot)$. This ensures that $\hat{Q}(\cdot)$ is a.s. positive semidefinite, because now it is a Gram matrix, and $\hat{R}(\cdot)$ is hence a.s. a correlation matrix. The derivation of the covariances between $\hat{q}_{ij}(\cdot)$ and $\hat{q}_{lm}(\cdot)$ is also simplified due to the use of same weights for both.

Remark 1. *In practice a causal smoother involving only past observations might be preferable. Setting $n_1 = t - k_0$ and $n_2 = t - 1$ for $t > k_0 + p$ in Step 1 of the k-NN algorithm we will obtain a causal estimator of $R(\cdot)$ which will also be applied to the data examples in Section 5.*

4 Main results

Both $\hat{Q}(\cdot)$ and $\hat{R}(\cdot)$ are estimators of $R(\cdot)$ which have similar properties. However, $\hat{R}(\cdot)$ is a correlation matrix but $\hat{Q}(\cdot)$ is usually not. In the following the asymptotic properties of $\hat{Q}(\cdot)$ will be first investigated. Properties of $\hat{R}(\cdot)$ are then derived based on those of $\hat{Q}(\cdot)$. Most properties of $\hat{\mu}(\cdot)$, $\hat{V}(\cdot)$ as well as $\hat{\theta}$ are known in the literature. The effect of the errors in these estimators on $\hat{Q}(\cdot)$ will be discussed in the appendix. For the proof of the results the following assumptions are required.

A1. i) The local mean and variance functions $\mu_i(\cdot)$ and $V_i(\cdot)$, $i = 1, ..., d$, are at least twice continuously differentiable. ii) The kernels K_{μ_i} and K_{V_i} for estimating μ and V are all symmetric densities with support $[-1, 1]$. And iii) The estimators $\hat{\mu}$ and $\hat{V}$ are obtained following some consistent data-driven algorithms.

A2. There exists a constant $\delta > 0$ such that $E(r_{it}^{4+\delta}) < \infty$ for each of the GARCH models defined in (4).

A3. The estimation point $(\tau; \mathbf{y})$ is a multivariate interior point with $\tau \in (0, 1)$ and $y_j > 0$ for all $j = 1, ..., p$.

A4. Assume that $\epsilon_t \sim N(0, R(\tau_t; \mathbf{y}_t))$ and ϵ_t and ϵ_s are independent for $t \neq s$.

A5. $R(\tau; \mathbf{y})$ is positive definite, uniformly in τ and $\mathbf{y}$, whose off-diagonal entries are at least twice continuously differentiable with respect to $\tau \in [0, 1]$ and $\mathbf{y}$ on their support.

A6. The kernel K_0 is a symmetric density defined on $[-1, 1]$ with $\int v^2 K_0(v)dv = \mu_2(K_0) > 0$. The spherical kernel K is a density defined on the unit ball such that $\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0$ and $\int u_i u_j K(\mathbf{u})d\mathbf{u} = \delta_{ij}\mu_2(K)$ with $\mu_2(K) > 0$.

A7. i) The bandwidth b_0 is of higher order than $n^{-1/5}$, denoted by $b_0 > O(n^{-1/5})$. ii) b_0 satisfies $b_0 \to 0$, $nb_0^{p+1} \to \infty$ as $n \to \infty$. And iii) k is chosen by $k = C_k nb_0^{p+1}$ with $C_k > 0$.

A1 summarizes the common conditions on the estimators $\hat{\mu}(\cdot)$, $\hat{V}(\cdot)$ as well as $\hat{\theta}$ in the first stage, which together with A7 i) ensures that the effect of the errors in these estimators are asymptotically negligible and hence $\hat{Q}(\cdot)$ and $\hat{R}(\cdot)$ obtained from $\hat{\epsilon}_t$ or from the unobservable ϵ_t have the same asymptotic properties. A2 is required for the asymptotic normality of $\hat{V}_i(\cdot)$ and the resulting $\hat{\theta}_i$. Necessary and sufficient conditions that guarantee this may be found e.g. in Ling and McAleer (2002), and in Bollerslev (1986) for GARCH(1, 1) models. Note that $E(r_{it}^{4+\delta}) < \infty$ implies $\sum_{l=1}^{p_i} \alpha_{il} + \sum_{m=1}^{q_i} \beta_{im} < 1$ as assumed before. A3 is introduced to avoid the boundary effect in the time dimension which simplifies the proofs. In the appendix it will be explained that there is indeed no boundary effect caused by the regressors y_j. A4 is not a necessary condition which can be replaced by suitable assumptions on the distribution of ϵ_t (see Hafner et al., 2005). A5, A6 and A7 ii) are regularity conditions in multivariate nonparametric regression. A5 also ensures that the off-diagonal elements of $R(\cdot)$ are strictly between -1 and 1, uniformly in τ and $\mathbf{y}$.

A7 ii) and iii) ensure that the bandwidth b for the regressors y_j obtained by the k-NN method is of the same order of magnitude as b_0. This is given by the following lemma.

Lemma 1. *Assume that* $\mathbf{y}$ *is observable with continuous density on its support and that* $0 < f(\mathbf{x}) < \infty$ *in a neighbourhood of* $\mathbf{y}$. *Then under A7 ii) and iii) the bandwidth* b *obtained by the k-NN method is given by*

$$b \doteq C_0 b_0 \text{ with } C_0 = \left(\frac{C_k}{2} \frac{\Gamma(p/2+1)}{\pi^{p/2} f(\mathbf{y})} \right)^{1/p}. \tag{12}$$

Lemma 1 is given under common regularity conditions on the design density which provides a basis for extending the main results in this paper to more general cases. The explicit form of $f(\mathbf{y})$ under A4 will be given in the appendix. A3 and A4 together ensure that the conditions of Lemma 1 hold. It is clear that Lemma 1 remains to be true when $\mathbf{y}$ is estimated consistently.

Now, let $\xi_{lmt} = \epsilon_{lt}\epsilon_{mt}$, $l, m = 1, ..., d$. Then we have $\rho_{lm}(\tau_t; \mathbf{y}_t) = E[\xi_{lmt}|\mathbf{y} = \mathbf{y}_t]$. Let γ_{lm}^2 denote the conditional variance of ξ_{lmt} and $\gamma_{lm,rs}$ the conditional covariance between ξ_{lmt} and ξ_{rst}. Let $R(K_0) = \int K_0^2(v)dv$ and $R(K) = \int K^2(\mathbf{u})d\mathbf{u}$. Let $\bigtriangledown_f(\mathbf{y})$ and $\bigtriangledown_{lm}(\mathbf{y})$ denote the two $p \times 1$ vectors of the gradient of $f(\mathbf{y})$ and $\rho_{lm}(\tau; \mathbf{y})$ respectively w.r.t. $\mathbf{y}$, and let $\mathcal{H}_{lm}(\tau; \mathbf{y})$ denote the $(p+1) \times (p+1)$ Hessian matrix of $\rho_{lm}(\tau; \mathbf{y})$. Finally, let $\mathrm{T} = diag(\mu_2(K_0)/\mu_2(K), C_0^2, ..., C_0^2)$, $C_{lm} = \frac{1}{2}\mathrm{tr}\{\mathcal{H}_{lm}(\tau; \mathbf{y})\mathrm{T} + 2C_0^2 \bigtriangledown_f(\mathbf{y}) \bigtriangledown_{lm}(\mathbf{y})'/f(\mathbf{y})\}$ and $C_V = 2 * \pi^{p/2}/\Gamma(p/2+1)R(K_0)R(K)$. Note that C_{lm} depends on τ and $\mathbf{y}$. Then the following holds.

Theorem 1. *Under Assumptions A1 to A7 we have*

i) $Bias[\hat{q}_{lm}(\tau; \mathbf{y})] \doteq C_{lm}\mu_2(K)b_0^2$, *for* $l \neq m$,

ii) $Bias[\hat{q}_{lm}(\tau; \mathbf{y})] = O[b_{V_l}^2 + (nb_{V_l})^{-1}] = o(b_0^2)$, *for* $l = m$,

iii) $var[\hat{q}_{lm}(\tau; \mathbf{y})] \doteq C_V k^{-1}\gamma_{lm}^2$ *and*

iv) $cov[\hat{q}_{lm}(\tau; \mathbf{y}), \hat{q}_{rs}(\tau; \mathbf{y})] \doteq C_V k^{-1}\gamma_{lm,rs}$.

The regressors τ and $\mathbf{y}$ have slightly different effects on the asymptotic results, because the used bandwidths for them are different and that τ_t are equidistant. For fixed k the asymptotic variances and covariances of the proposed estimators do not depend on the design density. But the asymptotic biases depend on $f(\mathbf{y})$ through the constant C_0. If C_k is fixed, the optimal b_0, which minimizes the dominating part of the MSE (mean squared error) of $\hat{q}_{lm}(\tau_t; \mathbf{y})$, is given by

$$b_0^{\mathrm{opt}} = C_{opt} n^{-\frac{1}{p+5}} \text{ with } C_{opt} = \left(\frac{(p+1)C_V\gamma_{lm}^2}{4C_{lm}^2\mu_2^2(K)C_k}\right)^{\frac{1}{p+5}}. \tag{13}$$

The optimal order of the MSE is $O(n^{-\frac{4}{p+5}})$. Note that the use of a fixed C_k is only suboptimal. The optimal choice of C_k so that the constant of the dominating part of the MSE is also minimized and the data-driven selection of b_0^{opt} will be discussed elsewhere.

Similar results hold for $\hat{\rho}_{lm}(\cdot)$. In particular they are also asymptotically normal. Consider the case with $l \neq m$, $r \neq s$ and $\{l, m\} \neq \{r, s\}$. We have

Theorem 2. *Assume that* $b_0 = C_b n^{-\frac{1}{p+5}}$ *and that* $C_b, C_k > 0$ *are two positive constants. Then under the other conditions of Theorem 1 we have*

$$\sqrt{k}\begin{pmatrix} \hat{\rho}_{lm}(\tau;\mathbf{y}) - \rho_{lm}(\tau;\mathbf{y}) \\ \hat{\rho}_{rs}(\tau;\mathbf{y}) - \rho_{rs}(\tau;\mathbf{y}) \end{pmatrix} \xrightarrow{\mathcal{D}} N\left\{C_\mu \begin{pmatrix} C_{lm} \\ C_{rs} \end{pmatrix}, C_V \begin{pmatrix} \gamma_{lm}^2 & \gamma_{lm,rs} \\ \gamma_{lm,rs} & \gamma_{rs}^2 \end{pmatrix}\right\}, \tag{14}$$

where $C_\mu = \mu_2(K) C_k^{1/2} C_b^{(p+5)/2}$ *is an unknown constant and* C_V *is as defined before.*

If a bandwidth $b_0 = o(n^{-1/(p+5)})$ is used then the bias in $\hat{q}_{lm}(\cdot)$ is asymptotically negligible. Now Theorem 2 holds with C_μ being replaced by 0. Results in this case are useful for carrying out confidence intervals of $\rho_{lm}(\cdot)$ (Hafner et al., 2005).

Remark 2. *Similar asymptotic properties of the causal estimator described in Remark 1 can be obtained by setting* $\tau = 1$. *This will not be discussed here in detail.*

5 Applications

In the following the practical performance of the model is shown by applications to the daily foreign exchange rate series of the British Pound (Pound), Euro, Japanese Yen (Yen) and Canadian Dollar (CAD) w.r.t. the US Dollar (USD) from 4 Jan 1999 to 30 December 2005. The log-returns of these series are shown in Figure 1. The programmes are developed in S-Plus, which can also be run under R. In all of the examples corresponding univariate and spherical Epanechnikov kernels are used. The nonparametric trend in the returns are fitted using the SEMIFAR (semiparametric fractional autoregressive) model (Beran and Feng, 2002), where the bandwidths are automatically selected and it is also shown that the returns are about uncorrelated, one of the properties of a GARCH process. The local variances are then estimated from the corresponding residual series using bandwidths selected by the algorithm in Feng (2004). Table 1 lists the bandwidths for estimating the means and the variances selected by these programmes. The estimated local means and local standard deviations are omitted to save space.

GARCH models are then fitted from the standardized residuals using S+GARCH. GARCH(1, 1) models are selected following the BIC for Pound, Yen as well as CAD. The fitted GARCH models for the Euro returns of lower orders have however negative coefficients, for which a GARCH(2, 2) model is hence used. The estimated conditional variances for each series are:

$$\hat{h}_{1t} = 0.0448 + 0.0402\hat{r}_{1(t-1)}^2 + 0.9136\hat{h}_{1(t-1)},$$

Table 1. Selected bandwidths for $\hat{\mu}_i$, $\hat{\sigma}_i$

	Pound	Euro	Yen	CAD
$\hat{b}_{\mu_i}$	0.128	0.103	0.084	0.135
$\hat{b}_{V_i}$	0.144	0.120	0.105	0.132

$$\hat{h}_{2t} = 0.0980 + 0.1000\hat{r}^2_{2(t-1)} + 0.0010\hat{r}^2_{2(t-2)} + 0.7999\hat{h}_{2(t-2)} + 0.0101\hat{h}_{2(t-1)},$$

$$\hat{h}_{3t} = 0.0590 + 0.0139\hat{r}^2_{3(t-1)} + 0.9265\hat{h}_{3(t-1)},$$

$$\hat{h}_{4t} = 0.0360 + 0.0219\hat{r}^2_{4(t-1)} + 0.9415\hat{h}_{4(t-1)}.$$

The estimated total standard variations $\hat{\sigma}_i(\tau_t) \cdot \sqrt{\hat{h}_{it}}$ for the return series are shown in Figure 2.

Selection of the smoothing parameters and the order of the LDCC model is still an open question. For a comparison, results using different parameter combinations will be given. We will see that estimates using b_0 and k within a large range have quite similar conditional patterns. This means that the fitted results discover the nature of the true correlation dynamics and the proposed model works well in practice. Note that for a LDCC(p) model relatively large smoothing parameters should be used, because it is a $(p+1)$-dimensional nonparametric regression. Figure 3 displays the estimated local and conditional correlations of a LDCC(4) model with $b_0 = 0.18$ and $k = 400$, i.e. $Q(\cdot)$ is estimated by 400 observations selected from a total of 36% of all observations around t. Figure 4 shows the estimation results of a LDCC(6) model with the same smoothing parameters.

We see both the conditional and local changes are clear. The conditional effect changes from one point to another. Information in the past observations may cause lower or higher correlations comparing to the average level. Sometimes the conditional effect may cause clear changes in the correlations in both directions. The largest conditional changes in correlations are as high as about 0.2. The average level of the correlations depends on the location, which changes clearly in some periods. This also depends on the two series involved. For instance, the local change of the correlations between Pound and Euro is quite small, because these two series are always highly correlated. An interesting phenomenon is that the correlations between CAD and the other currencies were about zero at the beginning, but increase to about 0.5 at the end of the observation period. Note that the conditional changes estimated by a nonparametric model is non-smooth over time, because the changes of the regressors y_j are non-smooth over time. This fact can be also found from the examples in Hafner et al. (2005). By comparing Figures 3 and 4 we can find that the LDCC(6) model seems to perform better than the LDCC(4). This observation provides evidence for the need of the development of a GARCH-type LDCC model. To show the effect of the smoothing parameters some further estimation results are given in Figure 5. These are correlations between the Euro and the other currencies estimated using a LDCC(6) model with smoothing parameters $b_0 = 0.16$ and $k = 300$ (Figures 5(a) to (c)) and with $b_0 = 0.20$ and $k = 500$ (Figures 5(d) to (f)), corresponding to those given in Figures 4(a), (d) and (e) respectively. Comparing the estimates obtained using different smoothing parameters we can see that they look quite similar to each other. In particular they show the same conditional patterns. The variations of these estimates depend on the used smoothing parameters.

The causal smoother is also applied to these data. Figure 6 shows the estimation results of a causal LDCC(6) with the same smoothing parameters as for Figure 4. Note that causal and non-causal estimates are the same for $t \leq k_0 + p$. Comparing Figures 4 and 6 we can find that, for large t, there are clear differences between the local and conditional patterns obtained using causal and non-causal methods. The conditional changes are more clear by the causal estimates. We can also see that there is a phase-difference between the periods estimated by these two approaches where the local correlations increase quickly. The local changes discovered by the causal estimates seem to be more practically relevant. The phase-difference in the non-causal estimates is caused by future observations. It is worthwhile to carry out further study on the causal smoother.

The proposed model is also applied to other financial data. In particular, applications to some weekly UK equity index returns from 1 Jan 1965 to Nov 1999 (Brooks and Henry, 2002) show that the local and conditional changes in the variances and the conditional changes in the correlations are more clear in those data. But the local changes in the correlations are not so clear.

6 Final remarks

In this paper simultaneous modelling of local and conditional changes in the variances and correlations of financial returns is discussed. A semiparametric approach is introduced to model all of these components. In particular the local and conditional correlations are jointly estimated by a multivariate kernel regression. Asymptotic results and applications confirm the theoretical and practical performance of the proposal. So far as we know, this is the first effort to define and to estimate local and conditional correlations jointly. There are still many open questions in this context, e.g. the development of a suitable bandwidth selection rule, the development of significance test as well as the the model selection. Finally, the extension of the current model to a GARCH-type approach by including some latent variables is also of great interest.

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/'. We would like to thank Prof Chris Brooks, City University London, and Prof Olan Henry, University of Melbourne, for kindly providing us the weekly UK equity index data, to which the model is also applied. We are very grateful to Prof Winfried Pohlmeier, University of Konstanz, and Dr Kevin Sheppard, Oxford University, for useful discussions and comments, which helped to improve the quality of this paper.

Appendix: Auxiliary results and proofs of theorems

Let $\hat{\xi}_{lmt} = \hat{\epsilon}_{lt}\hat{\epsilon}_{mt}$, $l, m = 1, ..., d$, denote estimates of ξ_{lmt}, where ξ_{lmt} are as defined in Section 4. Let $\tilde{q}_{lm}(\cdot)$ denote the same estimator as $\hat{q}_{lm}(\cdot)$, but obtained with $\hat{\xi}_{lmt}$ being replaced by ξ_{lmt}. To simplify the notations let b_μ and b_V denote some generic bandwidths for estimating the mean and variance functions which are of suitable orders depending on the cases under consideration.

The effect of the errors in $\hat{\mu}_i$ on $\hat{\sigma}_i^2$ is discussed in Feng and Yu (2005). For the LDCC model this effect is of the same orders of magnitude. Discussion on the effect of the errors in $\hat{\sigma}_i^2$ on $\hat{\theta}_i$ may be found in Feng (2004). Their results will be adapted to the current case. We now introduce the following variant of Assumption A1 iii).

A1 iii)$'$. b_μ and b_V satisfy $O(n^{-1/3}) < b_\mu < O(n^{-1/6})$ and $O(n^{-1/3}) < b_V = o(1)$.

Here '$<$' or '$>$' are applied to the orders. Condition A3$'$ is sufficient but not necessary.

Lemma A.1. *Consider* $0 < \tau < 1$. *Under the conditions A1 i) and ii), and A1 iii)$'$ the error in* $\hat{\mu}(\tau)$ *does not affect the asymptotic properties of* $\hat{V}(\tau)$.

This means that $\hat{V}(\tau)$ obtained from the residuals $\hat{r}_t^*$ has the same asymptotic properties as that obtained from the unobservable r_t^*.

A sketched proof of Lemma A.1. Results for $\tau = 1$ are discussed in detail in the proof of Theorem 3 of Feng and Yu (2005). In our case the corresponding results are

1. The additional bias in $\hat{\sigma}_i^2(\tau)$ caused by the error in $\hat{\mu}_i(\tau)$ is of the order $O[b_\mu^4 + (nb_\mu)^{-1}]$.

2. The additional variance in $\hat{\sigma}_i^2(\tau)$ caused by the error in $\hat{\mu}_i(\tau)$ is negligible, if $b_\mu > O(n^{-1/2})$.

For more details we refer the reader to the corresponding proofs given there. Note that there are some differences between the above results and those given there, because now $0 < \tau < 1$ is considered. The bias of $\hat{\sigma}_i^2(\cdot)$ under the assumption that $\mu_i(\cdot)$ were known is of the order $O(b_V^2)$. Under A1 iii)$'$ we have $O[b_\mu^4 + (nb_\mu)^{-1}] = o(b_V^2)$. Hence, the asymptotic bias of $\hat{\sigma}_i^2(\cdot)$ will not be affected by the error in $\hat{\mu}_i(\cdot)$. Furthermore, the asymptotic variance of $\hat{\sigma}_i^2(\cdot)$ will also not be affected by the error in $\hat{\mu}_i(\cdot)$, because $b_\mu > O(n^{-1/2})$. $\diamond$

If $\hat{\mu}_i$ is estimated data-drivenly as assumed in A1 iii), we have $b_{\mu_i} = O(n^{-1/5})$. Now A1 iii)$'$ is fulfilled and the error in $\hat{\mu}_i$ on $\hat{\sigma}_i^2$ is asymptotically negligible. The effect of the error in $\hat{\sigma}_i^2$ on $\hat{\theta}_i$ is shown by the following lemma which follows from Theorem 3 in Feng (2004).

Lemma A.2. *Under the same conditions of Lemma A.1 $\hat{\theta}_i$ is root-n consistent and asymptotically normal as in the parametric case except for a bias term of the order $O[b_V^2+(nb_V)^{-1}]$.*

The following lemma quantifies the difference between $\tilde{q}_{lm}(\cdot)$ and $\hat{q}_{lm}(\cdot)$ caused by the error in $\hat{\sigma}_i^2(\cdot)$ and that in $\hat{\theta}_i$. Note that the latter is resulted by the former.

Lemma A.3. *Under the assumptions of Theorem 1 we have*

i) $E[\hat{q}_{lm}(\tau;\mathbf{y})-\rho_{lm}(\tau;\mathbf{y})] = E(\tilde{q}_{lm}(\tau;\mathbf{y})-\rho_{lm}(\tau;\mathbf{y}))+O[b_V^2+(nb_V)^{-1}]$,

ii) $var[\hat{q}_{lm}(\tau;\mathbf{y})] = var(\tilde{q}_{lm}(\tau;\mathbf{y}))[1+o(1)]+O[(nb_V)^{-1}]$ *and*

iii) $cov[\hat{q}_{lm}(\tau;\mathbf{y}),\hat{q}_{rs}(\tau;\mathbf{y})] = cov[\tilde{q}_{lm}(\tau;\mathbf{y}),\tilde{q}_{rs}(\tau;\mathbf{y})][1+o(1)]+O[(nb_V)^{-1}]$.

The additional errors caused by the error in $\hat{\sigma}_i^2(\cdot)$ and that in $\hat{\theta}_i$ are asymptotically negligible.

Proof of Lemma A.3. Observe that

$$\hat{\xi}_{lmi} = \hat{\epsilon}_{li}\hat{\epsilon}_{mi} \doteq \frac{\sigma_l(\tau_i)\sigma_m(\tau_i)\sqrt{h_{li}}\sqrt{h_{mi}}}{\hat{\sigma}_l(\tau_i)\hat{\sigma}_m(\tau_i)\sqrt{\hat{h}_{li}}\sqrt{\hat{h}_{mi}}}\xi_{lmi}.$$

We have

$$\begin{aligned}
\hat{\xi}_{lmi}-\xi_{lmi} &\doteq \frac{\sigma_l(\tau_i)\sigma_m(\tau_i)\sqrt{h_{li}}\sqrt{h_{mi}}-\hat{\sigma}_l(\tau_i)\hat{\sigma}_m(\tau_i)\sqrt{\hat{h}_{li}}\sqrt{\hat{h}_{mi}}}{\hat{\sigma}_l(\tau_i)\hat{\sigma}_m(\tau_i)\sqrt{\hat{h}_{li}}\sqrt{\hat{h}_{mi}}}\xi_{lmi} \\
&= O_p\left[\sigma_l(\tau_i)\sigma_m(\tau_i)\sqrt{h_{li}}\sqrt{h_{mi}}-\hat{\sigma}_l(\tau_i)\hat{\sigma}_m(\tau_i)\sqrt{\hat{h}_{li}}\sqrt{\hat{h}_{mi}}\right]\xi_{lmi} \\
&= \left[O(b_V^2)+O_p(nb_V)^{-1}\right]\xi_{lmi}, \qquad \text{(A.1)}
\end{aligned}$$

where the first step is because of $\hat{\sigma}_l(\tau_i)\hat{\sigma}_m(\tau_i)\sqrt{\hat{h}_{li}}\sqrt{\hat{h}_{mi}} = O_p(1)$ and the last equation is obtained by means of Taylor expansion.

Although ξ_{lmi} are iid random variables, $\hat{\xi}_{lmi}$ correlate to each other. From (A.1) we have

$$\hat{\xi}_{lmi} = \left[1+O(b_V^2)+O_p(nb_V)^{-1}\right]\xi_{lmi}.$$

Straightforward analysis shows that, for $i \neq j$,

$$\mathrm{cor}\,[\hat{\xi}_{lmi},\hat{\xi}_{lmj}] = O[(nb_V)^{-1}].$$

i) The difference in the biases of $\hat{q}_{lm}(\cdot)$ and $\tilde{q}_{lm}(\cdot)$ caused by that between $\hat{\xi}_{lmi}$ and ξ_{lmi} is of the same order of magnitude as $E[\hat{\xi}_{lmi}-\xi_{lmi}]$, provided that $E[\xi_{lmi}|\mathbf{y}=\mathbf{y}_i] = \rho_{lm}(\tau_i;\mathbf{y}_i) \neq 0$ for some i such that $\sum_i w_i = O(1)$, where w_i are the weights used in $\hat{q}_{lm}(\cdot)$. Conditioning on past observations, we have

$$
\begin{aligned}
E[\hat{\xi}_{lmi} - \xi_{lmi}] &= E\left\{\left[O(b_V^2) + O_p(nb_V)^{-1}\right]\xi_{lmi}\right\} \\
&= \left[O(b_V^2) + O_p(nb_V)^{-1}\right]\rho_{lm}(\tau_i; \mathbf{y}_i). \qquad \text{(A.2)}
\end{aligned}
$$

ii) Observe that $\sum_{i=1}^n \sum_{j\neq i} w_i w_j \doteq 1$. We have

$$
\begin{aligned}
\mathrm{var}\,[\hat{q}_{lm}(\tau;\mathbf{y})] &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \mathrm{cov}\,[\hat{\xi}_{lmi}, \hat{\xi}_{lmi}] \\
&= \sum_{i=1}^n w_i^2 \mathrm{var}\,[\hat{\xi}_{lmi}] + \sum_{i=1}^n \sum_{j\neq i} w_i w_j \,\mathrm{cov}\,[\hat{\xi}_{lmi}, \hat{\xi}_{lmj}] \\
&= \mathrm{var}\,[\tilde{q}_{lm}(\tau;\mathbf{y})][1 + o(1)] + O[(nb_V)^{-1}]. \qquad \text{(A.3)}
\end{aligned}
$$

iii) From the calculation of $\mathrm{cov}\,[\tilde{q}_{lm}(\cdot), \tilde{q}_{rs}(\cdot)]$ given later we can see that the proof of the result in this part is similar to that in ii). Lemma A.3 is proved. $\diamond$

Remark A.1. *If our aim is to test the hypothesis* $\mathrm{H}_0 : R_t \equiv \mathbf{I}_d$, *then the* b_V^2 *term in the additional bias of* $\hat{q}_{lm}(\cdot)$ *caused by the errors in* $\hat{\sigma}_i^2$ *will vanish.*

Proof of Lemma 1. Note that $b \to 0$ as $n \to \infty$. Under the assumptions of Lemma 1 there exists a p-ball around $\mathbf{y}$ with radius b for large n and the density on this ball is approximately $0 < f(\mathbf{y}) < \infty$. Following the k-NN method the bandwidth b is selected so that the probability for $\mathbf{y}$ within this p-ball is approximately k/k_0. Note that $k_0 \doteq 2nb_0$. Under A7 iii) we have $k = C_k n b_0^{p+1}$. This leads to

$$
\frac{\pi^{p/2} b^p}{\Gamma(p/2+1)} f(\mathbf{y}) \doteq \frac{k}{k_0} \doteq \frac{C_k n b_0^{p+1}}{2nb_0} \qquad \text{(A.4)}
$$

and

$$
b \doteq C_0 b_0 \quad \text{with} \quad C_0 = \left(\frac{C_k}{2}\frac{\Gamma(p/2+1)}{\pi^{p/2} f(\mathbf{y})}\right)^{1/p}
$$

as given in (12). Lemma 1 is proved. $\diamond$

Lemma A.4. *Under the assumptions of Theorem 1 the design density* $f(\mathbf{y})$ *of* $\mathbf{y}$ *satisfies the assumptions of Lemma 1.*

Proof of Lemma A.4. For simplicity let $\mathbf{y} = \mathbf{y}_t$ be an observation point. Assume first that ϵ_t are observable. Define $S_{jt}^2 = \mathbb{1}' R_t \mathbb{1}$, $j = 1, ..., p$, where $\mathbb{1}$ is a vector of ones as defined before. S_{jt}^2 is the variance of the normal random variable $\sum_{j=1}^{p} \epsilon_{i(t-j)}$. Following the construction of y_{jt} we have $y_{jt} = S_{jt}^2 Z_{jt}^2$, where $Z_{jt} \sim N(0,1)$. Under the assumptions of Theorem 1 Z_{jt}, $j = 1, ..., p$, are independent of each other. Therefore $\mathbf{y}_t$ is a vector of independent non-standardized χ_1^2 random variables with the joint density given by

$$f(\mathbf{y}_t) = \left(\sqrt{2}\Gamma(1/2)\right)^{-p} \prod_{j=1}^{p} y_{jt}^{-1/2} \exp\{-y_{jt} S_{jt}^{-2}/2\} S_{jt}^{-3}. \tag{A.5}$$

If $\mathbf{y}_t$ is an interior point, we have $0 < f(\mathbf{y}) < \infty$ in a neighbourhood of $\mathbf{y}_t$. Furthermore, it is easy to see that the above approximation holds, if $\mathbf{y}_t$ are calculated from some consistent estimates $\hat{\epsilon}_t$ of ϵ_t. The consistency of $\hat{\epsilon}_t$ is ensured by the conditions of Theorem 1. $\diamond$

Remark A.2. *The regressors y_{jt} are squared values of some normal random variables. There is no boundary effect at a point with $y_j = 0$ for some j. This can be seen from the form of $f(\mathbf{y})$ given in (A.5). If $y_j = 0$ for some j, we only have observations with $y_{jt} \geq y_j$. But the marginal density for y_j tends to infinite in a power rate $y_{jt}^{-1/2}$ for y_{jt} near the origin. Choosing $b_j = O(b_0^2)$, the resulting bias will be till of the order $O(b_0^2)$.*

Proof of Theorem 1. Define $\eta_{lmt} = \xi_{lmt} - \rho_{lm}(\tau_t; \mathbf{y}_t)$. Then we have

$$E[\eta_{lmt}|\mathbf{y}_t] = 0 \text{ and } \mathrm{var}\,[\eta_{lmt}|\mathbf{y}_t] = \gamma_{lm}^2.$$

Under the assumptions η_{lmt} and η_{lms} are conditionally independent for $t \neq s$. The fact that $\rho_{lm}(\cdot)$ can be estimated from ξ_{lmt} using kernel method is based on the following special conditional nonparametric regression model

$$\xi_{lmt}|\mathbf{y}_t = \rho_{lm}(\tau_t; \mathbf{y}_t) + \eta_{lmt}. \tag{A.6}$$

Consider first the case with $l \neq m$. Model (A.6) ensures that $\hat{q}_{lm}(\cdot)$ has the same asymptotic properties as in multivariate kernel regression. Let $\mathrm{U} = \int (v; \mathbf{u})(v; \mathbf{u})' K_0(\tau) K(\mathbf{u}) dv d\mathbf{u}$. Under A6 we have $\mathrm{U} = diag(\mu_2(K_0), \mu_2(K), ..., \mu_2(K))$. Let $\mathrm{H} = diag(b_0, b, ..., b)$ denote the bandwidth matrix. We have $\mathrm{HH'U} = \mu_2(K)\mathrm{T}$, where T is as defined in Theorem 1. Let $h(\tau; \mathbf{y})$ denote the joint density of τ and $\mathbf{y}$. Note that $h'_\tau(\tau; \mathbf{y}) \equiv 0$ and $h(\tau; \mathbf{y}) \equiv f(\mathbf{y})$, because τ_t are equidistant. Following Lemma A.3 and well known results in multivariate kernel regression we have:

$$\begin{aligned}
Bias[\hat{q}_{lm}(\tau; \mathbf{y})] &\doteq Bias[\tilde{q}_{lm}(\tau; \mathbf{y})] \\
&\doteq \mathrm{tr}\left\{\frac{1}{2}\mathcal{H}_{lm}(\tau; \mathbf{y})\mathrm{HH'U} + \frac{\nabla_h(\tau; \mathbf{y})\, \nabla'_{lm}(\tau; \mathbf{y})\mathrm{HH'U}}{h(\tau; \mathbf{y})}\right\} \\
&\doteq \frac{1}{2}\mu_2(K) b_0^2 \mathrm{tr}\left\{\mathcal{H}_{lm}(\tau; \mathbf{y})\mathrm{T} + 2C_0^2 \frac{\nabla_f(\mathbf{y})\, \nabla'_{lm}(\mathbf{y})}{f(\mathbf{y})}\right\},
\end{aligned} \tag{A.7}$$

which is $C_{lm}\mu_2(K)b_0^2$ as defined in Theorem 1, where the second term in the brackets is not shared by the local linear approach (see Ruppert and Wand, 1994).

For $l = m$, we have $\rho_{ll}(\cdot) \equiv 1$. Now $\tilde{q}_{ll}(\cdot)$ is unbiased and we have

$$\begin{aligned} E[\hat{q}_{ll}(\tau;\mathbf{y}) - \rho_{ll}(\tau;\mathbf{y})] &= O[b_V^2 + (nb_V)^{-1}] \\ &= o(b_0^2), \end{aligned} \tag{A.8}$$

which is caused by the error in the estimated variances.

For any l and m the variance of $\hat{q}_{lm}(\cdot)$ is of the same order. In the current case the asymptotic variance of $\hat{q}_{lm}(\cdot)$ is the same as for a local linear estimator. By adapting the result in (2.4) of Ruppert and Wand (1994) we have

$$\begin{aligned} \text{var}\,[\hat{q}_{lm}(\tau;\mathbf{y})] &\doteq \text{var}\,[\tilde{q}_{lm}(\tau;\mathbf{y})] \\ &\doteq \frac{1}{nb_0b^p}R(K_0)R(K)\frac{\gamma_{lm}^2}{h(\tau;\mathbf{y})}, \end{aligned} \tag{A.9}$$

where $h(\tau;\mathbf{y}) \equiv f(\mathbf{y})$. Observe that

$$\begin{aligned} b_0b^p &= C_0^p b_0^{p+1} \\ &= \frac{C_k}{2}\frac{\Gamma(p/2+1)}{\pi^{p/2}f(\mathbf{y})} \\ &= kn^{-1}C_V^{-1}R(K_0)R(K)f^{-1}(\mathbf{y}), \end{aligned}$$

where C_V is as defined in Theorem 1. Insert this into (A.9) we have

$$\text{var}\,[\hat{q}_{lm}(\tau;\mathbf{y})] \doteq C_V k^{-1}\gamma_{lm}^2. \tag{A.10}$$

Now consider the covariance between $\hat{q}_{lm}(\cdot)$ and $\hat{q}_{rs}(\cdot)$ for any $\{l,m\} \neq \{r,s\}$. Note that these two estimators are obtained using the same kernel weights. This results in a close relationship between $\text{cov}\,[\hat{q}_{lm}(\cdot), \hat{q}_{rs}(\cdot)]$ and $\text{var}\,[\hat{q}_{lm}(\cdot)]$ (or $\text{var}\,[\hat{q}_{rs}(\cdot)]$). Observe that ξ_{lmi} is independent of $\hat{\xi}_{lmj}$ for $i \neq j$. Following Lemma A.3 we have,

$$\begin{aligned} \text{cov}\,[\hat{q}_{lm}(\tau;\mathbf{y}), \hat{q}_{rs}(\tau;\mathbf{y})] &\doteq \text{cov}\,[\tilde{q}_{lm}(\tau;\mathbf{y}), \tilde{q}_{rs}(\tau;\mathbf{y})] \\ &\doteq \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \text{cov}\,[\xi_{lmi}, \xi_{rsj}] \\ &\doteq \sum_{i=1}^{n} w_i^2 \text{cov}\,[\xi_{lmi}, \xi_{rsj}] \\ &= \sum_{i=1}^{n} w_i^2 \gamma_{lm,rs} \\ &= \text{var}\,[\tilde{q}_{lm}(\tau;\mathbf{y})]\gamma_{lm,rs}/\gamma_{lm}^2 \\ &\doteq C_V k^{-1}\gamma_{lm,rs}. \end{aligned} \tag{A.11}$$

$\diamond$

Proof of (13). Note that $k = C_k n b_0^{p+1}$. The dominating part of the MSE of $\hat{q}_{lm}(\cdot)$ is

$$\text{MSE} \doteq C_k^{-1} C_V \gamma_{lm}^2 \left(n b_0^{p+1}\right)^{-1} + C_{lm}^2 \mu_2^2(K) b_0^4$$

and

$$\text{MSE}'_{b_0} \doteq -(p+1) C_k^{-1} C_V \gamma_{lm}^2 \left(n b_0^{p+2}\right)^{-1} + 4 C_{lm}^2 \mu_2^2(K) b_0^3.$$

Set $\text{MSE}'_{b_0} = 0$ we obtain

$$b_0^{opt} = \left(\frac{(p+1) C_V \gamma_{lm}^2}{4 C_{lm}^2 \mu_2^2(K) C_k} \right)^{\frac{1}{p+5}} n^{-\frac{1}{p+5}},$$

as given in (13). $\diamond$

The results in Theorem 2 are obtained based on the same results on $\hat{q}_{lm}(\cdot)$ and $\hat{q}_{rs}(\cdot)$ given in the following lemma, where it is assumed that $l \neq m$, $r \neq s$ and $\{l, m\} \neq \{r, s\}$.

Lemma A.5. *Under the same conditions of Theorem 2 we have*

$$\sqrt{k} \begin{pmatrix} \hat{q}_{lm}(\tau; \mathbf{y}) - \rho_{lm}(\tau; \mathbf{y}) \\ \hat{q}_{rs}(\tau; \mathbf{y}) - \rho_{rs}(\tau; \mathbf{y}) \end{pmatrix} \xrightarrow{\mathcal{D}} N \left\{ C_\mu \begin{pmatrix} C_{lm} \\ C_{rs} \end{pmatrix}, C_V \begin{pmatrix} \gamma_{lm}^2 & \gamma_{lm,rs} \\ \gamma_{lm,rs} & \gamma_{rs}^2. \end{pmatrix} \right\}, \quad \text{(A.12)}$$

where the constants are the same as defined in Theorem 2.

Proof of Lemma A.5. Following Theorem 1, the squared asymptotic bias and the asymptotic variance of $\sqrt{k}\hat{q}_{lm}(\cdot)$ are both constants, if $b_0 = C_b n^{-\frac{1}{p+5}}$ is used. Following Theorem iii) and iv), it is easy to see that the variances of $\sqrt{k}\hat{q}_{lm}(\cdot)$ and $\sqrt{k}\hat{q}_{rs}(\cdot)$, and the covariance between them are those given in Theorem 2. Note that $k = C_k n b_0^{p+1}$. Following Theorem 1 i) we have

$$\begin{aligned} Bias[\sqrt{k}\hat{q}_{lm}(\tau; \mathbf{y})] &\doteq C_{lm} \mu_2(K) C_k^{1/2} C_b^{(p+1)/2} C_b^2 \\ &= C_\mu C_{lm}, \end{aligned}$$

where $C_\mu = \mu_2(K) C_k^{1/2} C_b^{(p+5)/2}$ as defined in Theorem 2. Similarly, we have

$$Bias[\sqrt{k}\hat{q}_{rs}(\tau; \mathbf{y})] \doteq C_\mu C_{rs}.$$

Following Lemma A.3 we have

$$\sqrt{k}\{\hat{q}_{lm}(\tau; \mathbf{y}) - \rho_{lm}(\tau; \mathbf{y})\} = \sqrt{k}\{\tilde{q}_{lm}(\tau; \mathbf{y}) - \rho_{lm}(\tau; \mathbf{y})\}[1 + o_p(1)],$$

$$\sqrt{k}\{\hat{q}_{rs}(\tau; \mathbf{y}) - \rho_{rs}(\tau; \mathbf{y})\} = \sqrt{k}\{\tilde{q}_{rs}(\tau; \mathbf{y}) - \rho_{sr}(\tau; \mathbf{y})\}[1 + o_p(1)].$$

We will prove the asymptotic normality with $\hat{q}_{lm}(\cdot)$ and $\hat{q}_{rs}(\cdot)$ being replaced by $\tilde{q}_{lm}(\cdot)$ and $\tilde{q}_{rs}(\cdot)$ respectively. Note that η_{lmt} are normally distributed, and η_{lmi} and η_{lmj} are independent for $i \neq j$. This together with the condition on the bandwidth ensures that $\sqrt{k}\{\tilde{q}_{lm}(\cdot)-\rho_{lm}(\cdot)\}$ is asymptotically normal. Analogously, $\sqrt{k}\{\tilde{q}_{rs}(\cdot) - \rho_{rs}(\cdot)\}$ is also asymptotically normal.

Using the Cramér-Wold device, the joint asymptotic normality of $\sqrt{k}\{\tilde{q}_{lm}(\cdot) - \rho_{lm}(\cdot)\}$ and $\sqrt{k}\{\tilde{q}_{rs}(\cdot) - \rho_{rs}(\cdot)\}$ is proved, if we can show that, for any $\alpha_1, \alpha_2 \in \Re$,

$$\sqrt{k}\{\alpha_1[\tilde{q}_{lm}(\tau;\mathbf{y}) - \rho_{lm}(\tau;\mathbf{y})] + \alpha_2[\tilde{q}_{rs}(\tau;\mathbf{y}) - \rho_{rs}(\tau;\mathbf{y})]\} \xrightarrow{\mathcal{D}} N(\mu(\alpha_1,\alpha_2), V(\alpha_1,\alpha_2)),$$

where $\mu(\alpha_1,\alpha_2) = C_\mu(\alpha_1 C_{lm} + \alpha_2 C_{rs})$ and $V(\alpha_1,\alpha_2) = C_V(\alpha_1^2\gamma_{lm}^2 + \alpha_2^2\gamma_{rs}^2 + 2\alpha_1\alpha_2\gamma_{lm,rs})$. In the following we will explained briefly that this holds. Let

$$\begin{aligned} Z &= \alpha_1\tilde{q}_{lm}(\tau;\mathbf{y}) + \alpha_2\tilde{q}_{rs}(\tau;\mathbf{y}) \\ &= \alpha_1\sum_{i=1}^{n} w_i\xi_{lmi} + \alpha_2\sum_{i=1}^{n} w_i\xi_{rsi} \\ &= \sum_{i=1}^{n} w_i[\alpha_1\xi_{lmi} + \alpha_2\xi_{rsi}]. \end{aligned} \tag{A.13}$$

We see that $\sqrt{k}Z$ is asymptotically normal with the given bias and variance. This finishes the proof of Lemma A.5. ◇

Proof of Theorem 2. Straightforward calculations using Taylor lead to

$$\begin{aligned} \hat{r}_{lm}(\tau;\mathbf{y}) - \hat{q}_{lm}(\tau;\mathbf{y}) &= \frac{\hat{q}_{lm}(\tau;\mathbf{y})}{\sqrt{\hat{q}_{ll}\hat{q}_{mm}}} - \hat{q}_{lm}(\tau;\mathbf{y}) \\ &= T_1 + T_2, \end{aligned} \tag{A.14}$$

where $T_1 = O[b_V^2 + (nb_V)^{-1/2}] = o(k^{-1/2})$ is the bias term caused by those in $\hat{q}_{ll}$ and $\hat{q}_{mm}$, and $T_2 = \hat{q}_{lm}(\cdot)O_p(k^{-1/2})$, where the $O_p(k^{-1/2})$ term is a random variable with zero mean caused by the stochastic part of $\hat{q}_{ll}$ and $\hat{q}_{mm}$. It is clear that var $(T_2) = o(k^{-1})$. Using similar idea as in the proof of Lemma A.3, it can be shown that $E(T_2) = O(k^{-1})$. In summary we have

$$\sqrt{k}\{\hat{r}_{lm}(\tau;\mathbf{y}) - \rho_{lm}(\tau;\mathbf{y})\} = \sqrt{k}\{\hat{r}_{lm}(\tau;\mathbf{y}) - \rho_{lm}(\tau;\mathbf{y})\}[1 + o(1)].$$

This holds of course for $\hat{r}_{rs}(\cdot)$, too. Analogously, it can be shown that the above approximation also holds for the difference between $\alpha_1\hat{r}_{lm}(\cdot) + \alpha_2\hat{r}_{rs}(\cdot)$ and $\alpha_1\hat{q}_{lm}(\cdot) + \alpha_2\hat{q}_{rs}(\cdot)$. The results of Theorem 2 hence follow from those of Lemma A.5. ◇

References

Bauwens, L., S. Laurent and J. Rombouts, 2005, Multivariate GARCH models: A survey. CORE discussion paper, 2003/31 (revised 2005).

Beran, J. and Y. Feng, 2002, SEMIFAR models - A semiparametric framework for modelling trends, long-range dependence and nonstationarity. Computational Statistics & Data Analysis, 27, 393-419.

Bollerslev, T, 1986, Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, 31, 307-327.

Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. Review of Economics and Statistics, 72, 498-505.

Bollerslev, T., R. Chou and K. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics, 52, 5-59.

Bollerslev, T., R.F. Engle and J.M. Wooldridge, 1988, A Capital Asset Pricing Model with Time-Varying Covariances. Journal of Political Economy, 96, 116-131.

Brooks, C. and O. Henry, 2002, The impact of news on measures of undiversifiable risk: Evidence from the UK stock markets. Oxford Bulletin of Econonics and Statistics, 64, 487-507.

Bühlmann, P. and A.J. McNeil, 2002, An algorithm for nonparametric GARCH modelling. Computational Statistics & Data Analysis, 40, 665-683.

Cappiello, L., R.F. Engle and K. Sheppard, 2003, Evidence of Asymmetric Effects in the Dynamics of International Equity and Bond Return Covariance. Preprint, European Central Bank.

Dahlhaus, R., 1997, Fitting time series models to nonstationary processes. Annals of Statistics, 25, 1-37.

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

Engle, R.F., 2001, Financial econometrics - A new discipline with new methods. Journal of Econometrics, 100, 53-56.

Engle, R.F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20, 339-350.

Engle, R.F. and K.F. Kroner, 1995, Multivariate simultaneous GARCH. Econometric Theory, 11, 122-150.

Engle, R.F. and K. Sheppard, 2001, Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Preprint, National Bureau of Economic Research.

Fan, J. Q. and 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.

Feng, Y. and S. Heiler, 1998, Locally weighted autoregression, in: R. Galata and H. Küchenhoff (Eds.) Econometrics in theory and practice, Physica-Verlag, Heidelberg, pp. 101-117.

Feng, Y. and K. Yu, 2005, A Slowly Changing Vector Random Walk Model. Preprint, Heriot-Watt University and Brunel University.

Härdle, W., A.B. Tsybakov and L. Yang, 1998, Nonparametric vector autoregression. Journal of Statistical Planning and Inference, 68, 221-245.

Hafner, C.M., D. van Dijk and P.H. Franses, 2005, Semi-parametric modelling of correlation dynamics. Research Report, Erasmus University Rotterdam.

Hafner, C.M. and P.H. Franses, 2003, A generalized dynamic conditional correlation model for many asset returns. Research Report, Erasmus University Rotterdam.

Herzel, S. C. Starica and R. Tutungu, 2006, A non-stationary paradigm for the dynamics of multivariate model for financial returns. Preprints, University of Perugia.

Ling, S. and M. McAleer, 2002, Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. Econometric Theory, 18, 722-729.

Pelletier, D., 2006, Regime Switching for Dynamic Correlations. *Journal of Econometrics*, 131, 445-473.

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

Silvennoinen, A. and T. Teräsvirta, 2005, Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations. Preprint, Stockholm School of Economics.

Tse, Y. and A. Tsui, 2002, A multivariate GARCH model with time varying correlations. Journal of Business & Economic Statistics, 20, 352-362.

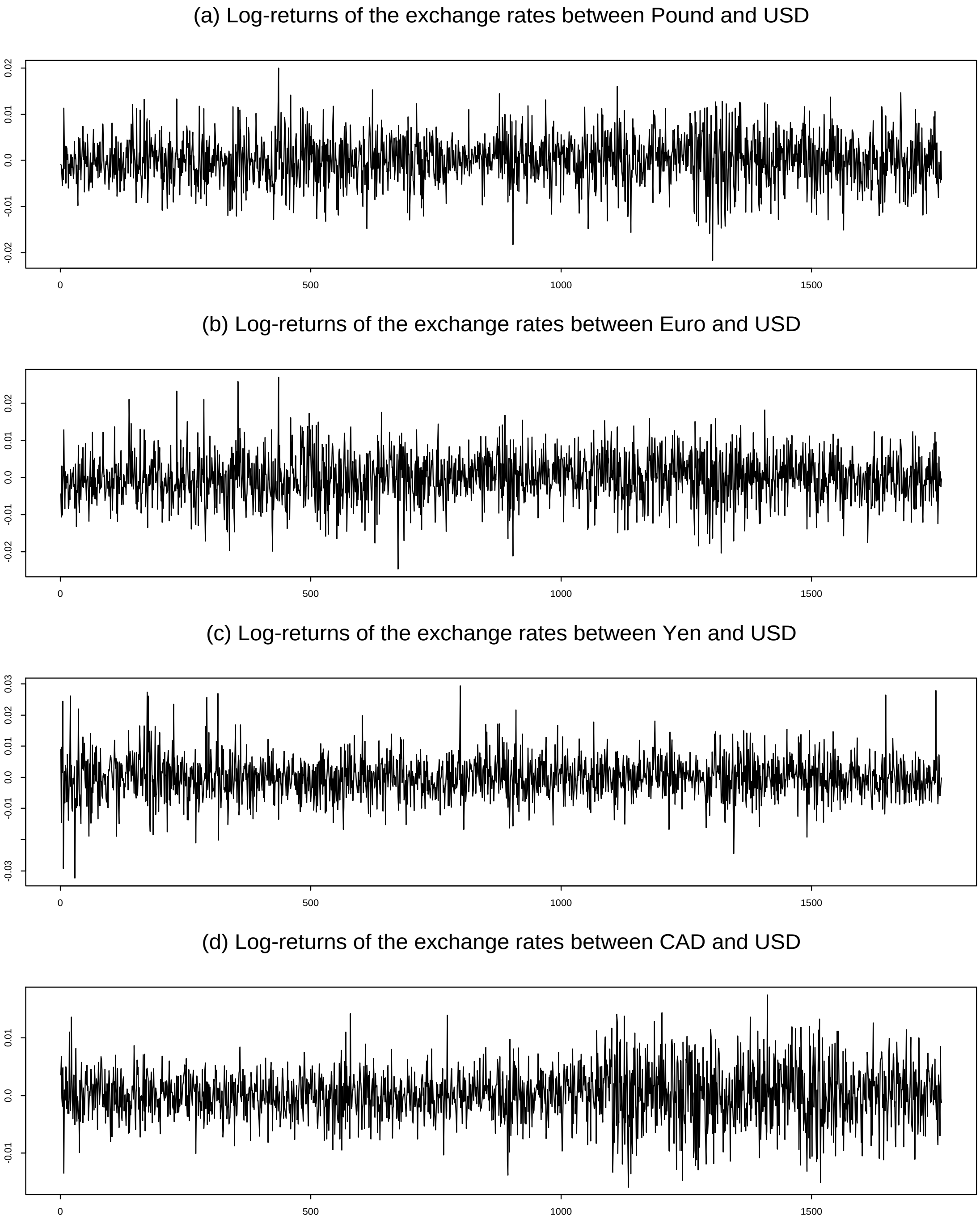

Figure 1: Return series of the selected exchange rates between 4 Jan 1999 to 30 Dec 2005.

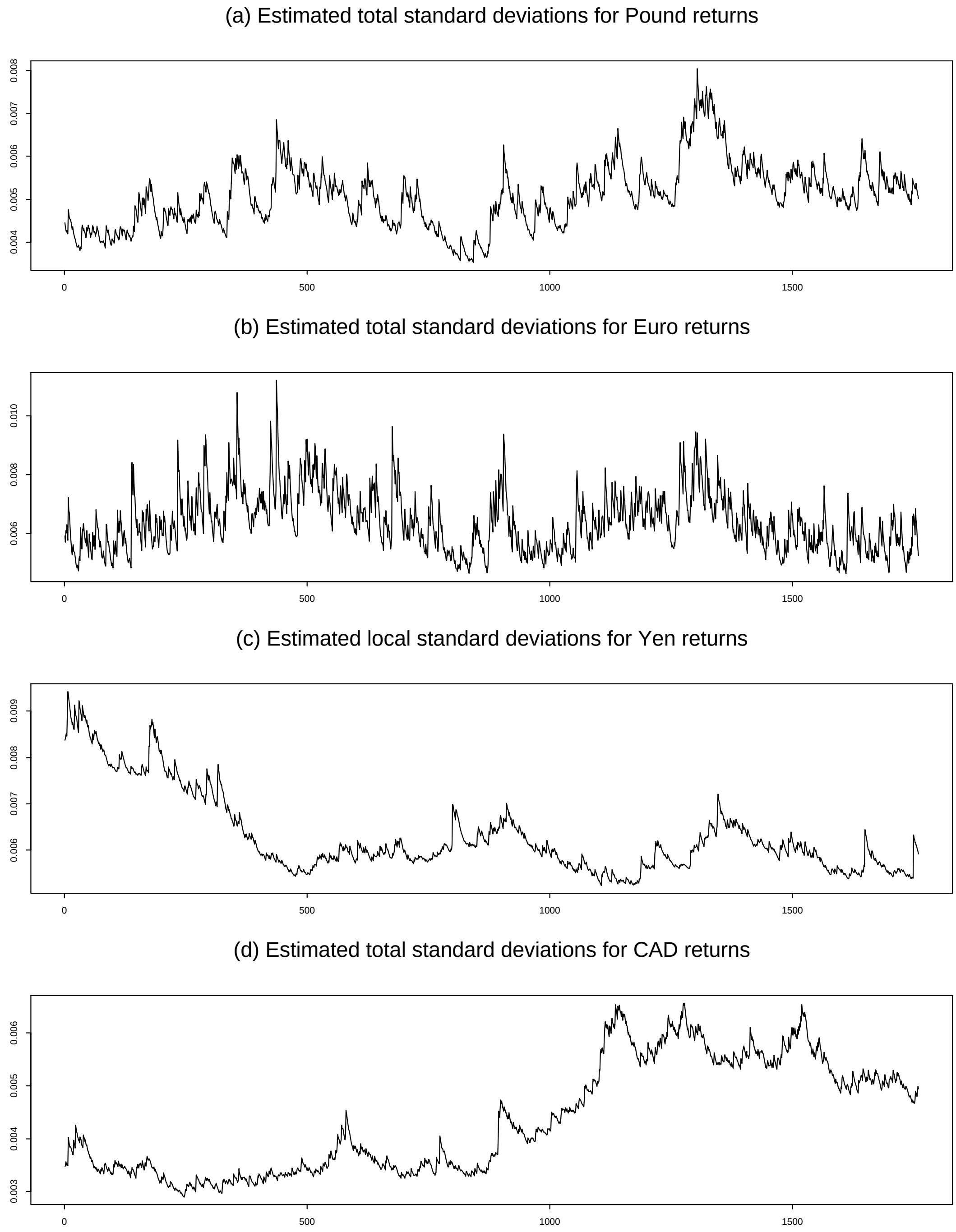

Figure 2: Estimated total standard deviations $\hat{\sigma}_i(\tau_t)\sqrt{\hat{h}_{it}}$ for the four return series.

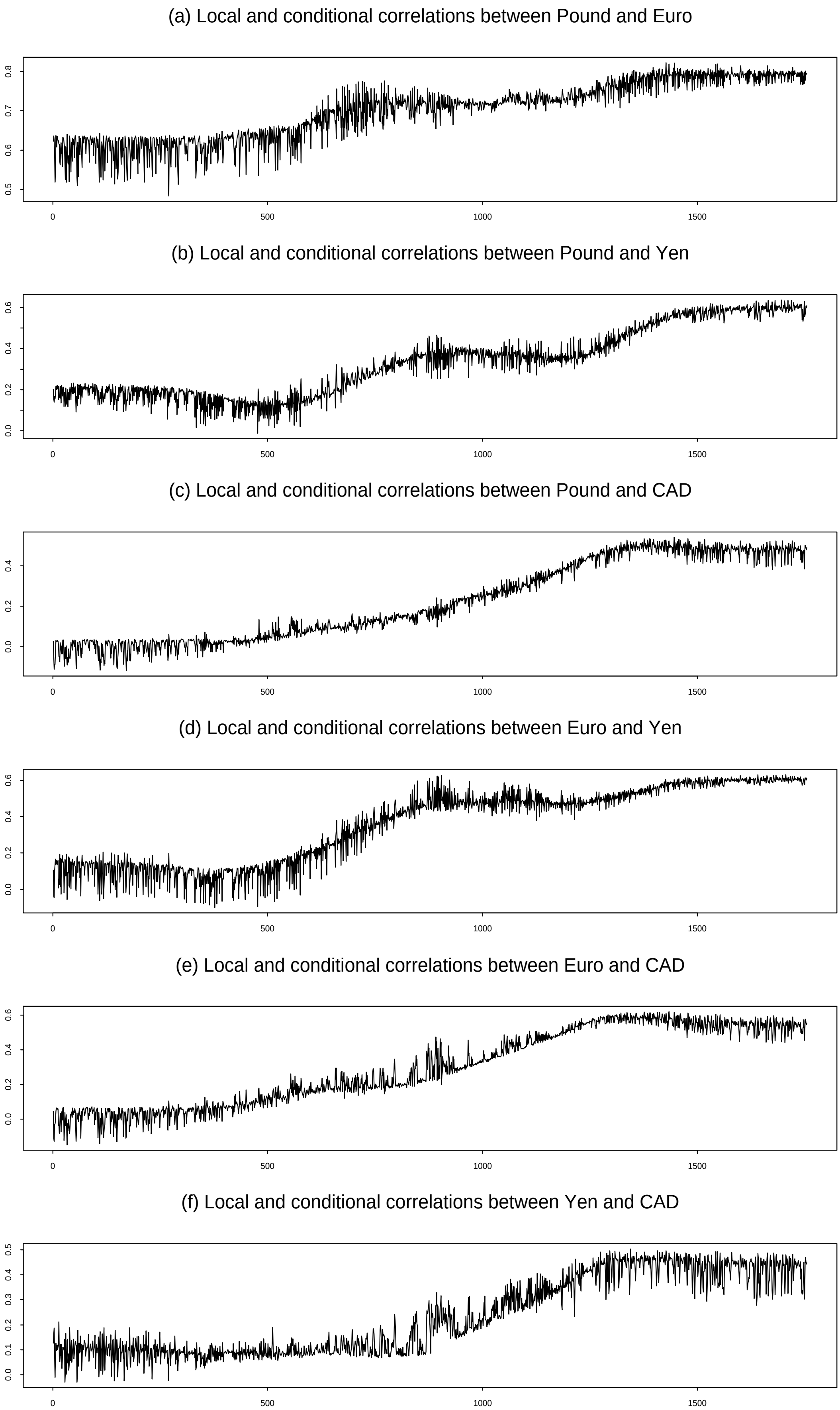

Figure 3: Local and conditional correlations estimated by a LDCC(4) model with $b_0 = 0.18$ and $k = 400$.

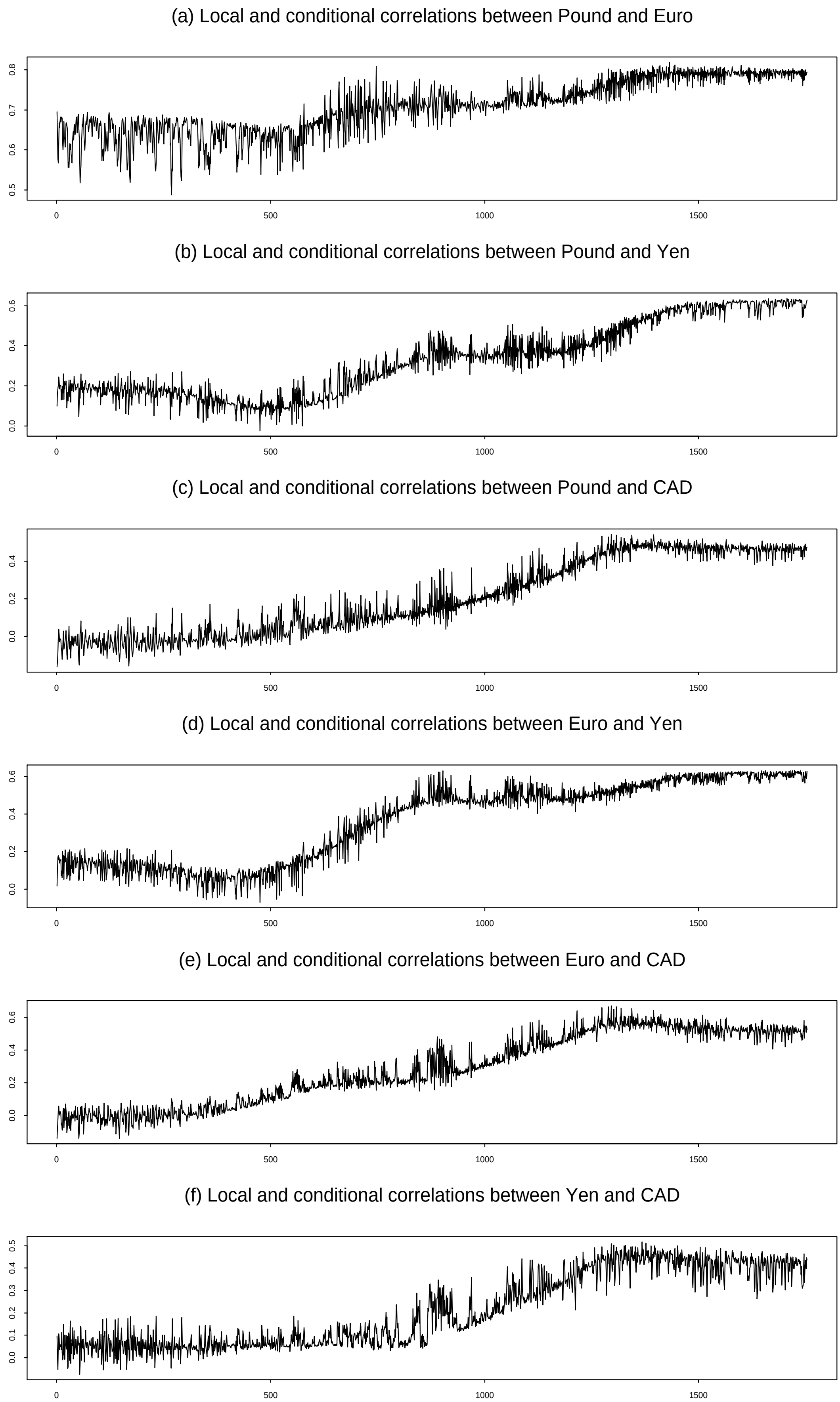

Figure 4: Local and conditional correlations estimated by a LDCC(6) model with $b_0 = 0.18$ and $k = 400$.

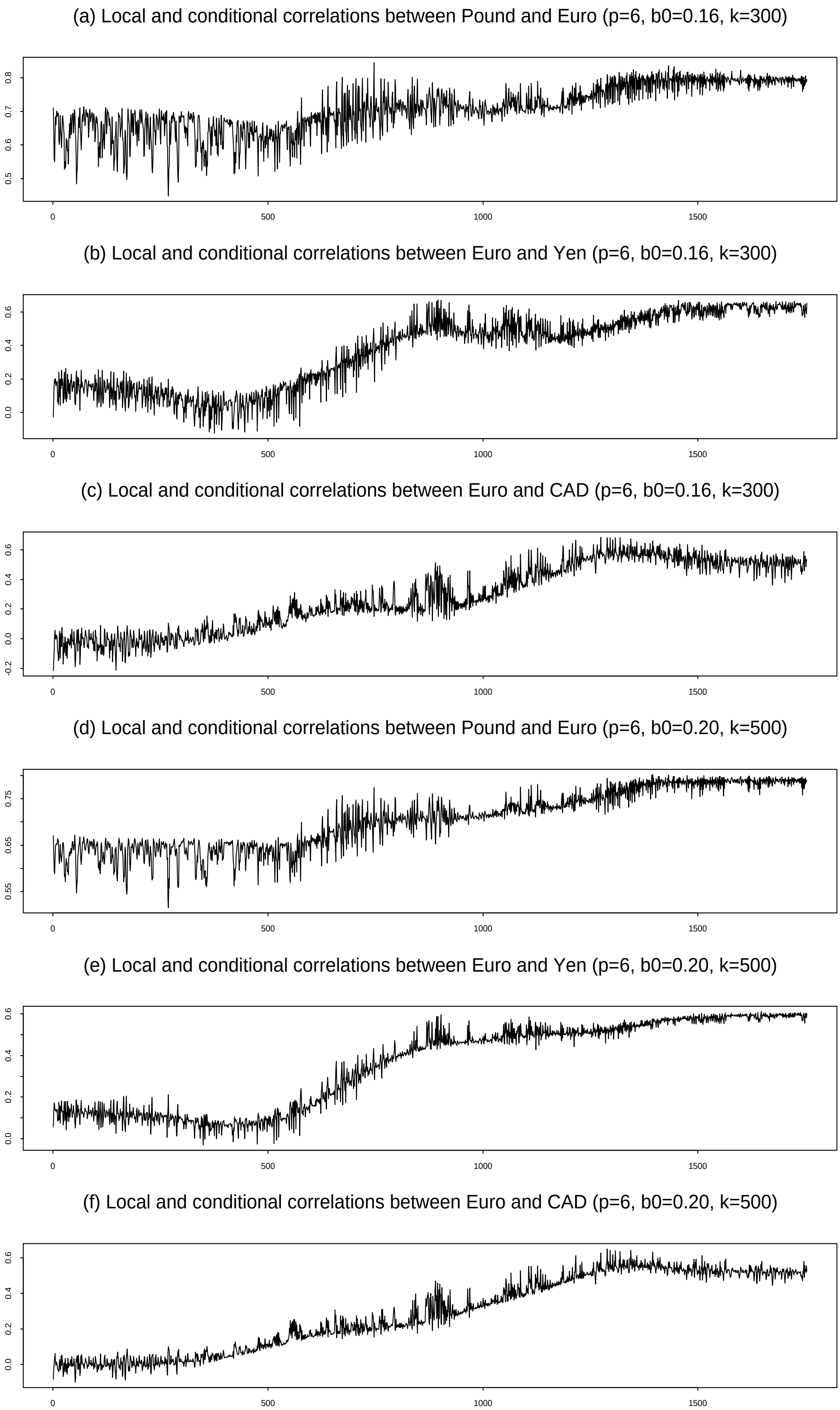

Figure 5: Local and conditional correlations estimated by a LDCC(6) model with different smoothing parameters.

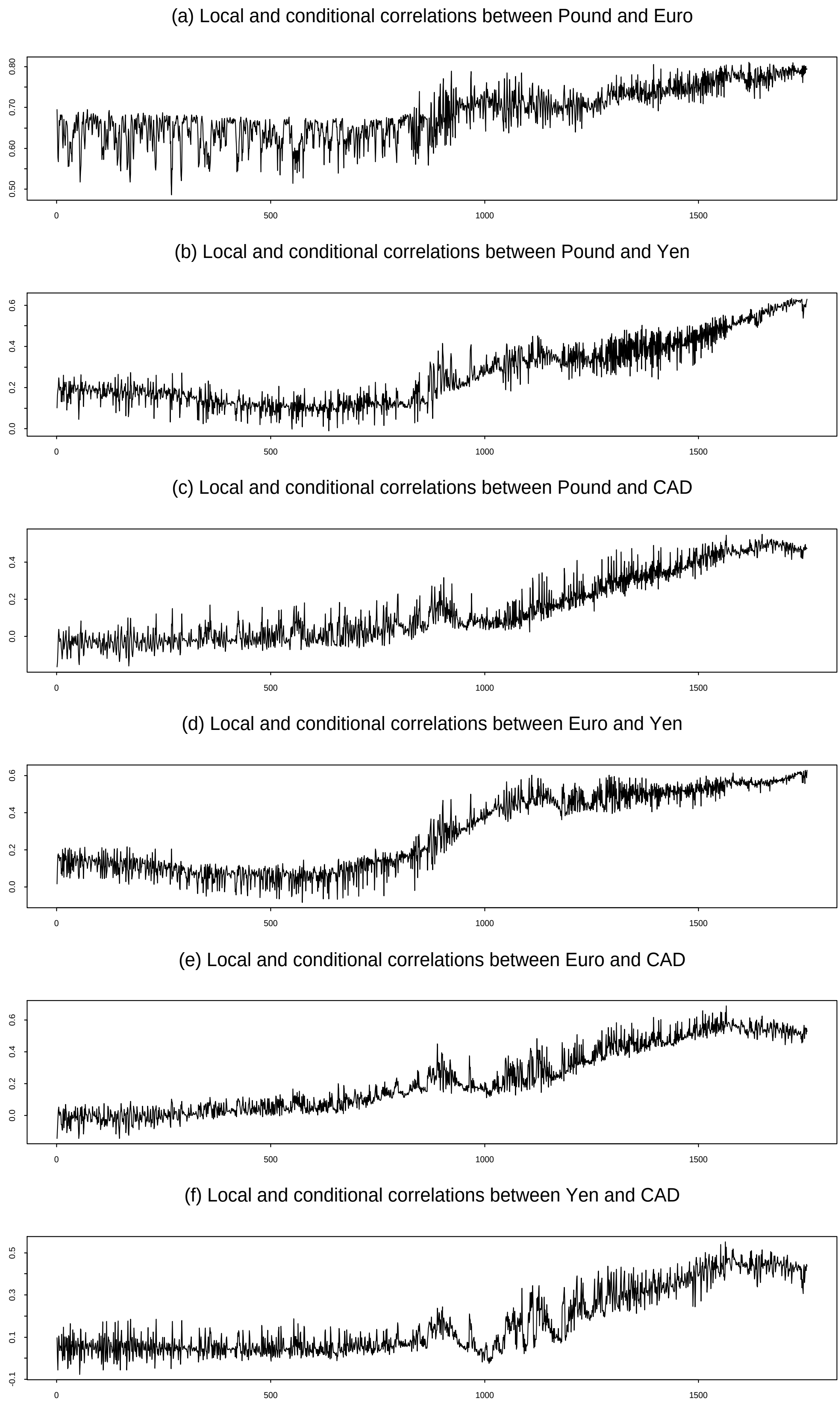

Figure 6: Local and conditional correlations estimated by a LDCC(6) model using causal smoothing, where the smoothing parameters are the same as in Figure 4.