

Cai, Biqing; Tjøstheim, Dag

Article

Nonparametric regression estimation for multivariate null recurrent processes

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Cai, Biqing; Tjøstheim, Dag (2015) : Nonparametric regression estimation for multivariate null recurrent processes, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 3, Iss. 2, pp. 265-288,
<https://doi.org/10.3390/econometrics3020265>

This Version is available at:

<https://hdl.handle.net/10419/171827>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Article

Nonparametric Regression Estimation for Multivariate Null Recurrent Processes

Biqing Cai and Dag Tjøstheim *

Department of Mathematics, University of Bergen, 5020 Bergen, Norway;

E-Mail: Biqing.Cai@math.uib.no

* Author to whom correspondence should be addressed; E-Mail: Dag.Tjostheim@math.uib.no;
Tel.: +47-55582885.

Academic Editor: Timo Teräsvirta

Received: 26 November 2014 / Accepted: 2 April 2015 / Published: 14 April 2015

Abstract: This paper discusses nonparametric kernel regression with the regressor being a d -dimensional β -null recurrent process in presence of conditional heteroscedasticity. We show that the mean function estimator is consistent with convergence rate $\sqrt{n(T)h^d}$, where $n(T)$ is the number of regenerations for a β -null recurrent process and the limiting distribution (with proper normalization) is normal. Furthermore, we show that the two-step estimator for the volatility function is consistent. The finite sample performance of the estimate is quite reasonable when the leave-one-out cross validation method is used for bandwidth selection. We apply the proposed method to study the relationship of Federal funds rate with 3-month and 5-year T-bill rates and discover the existence of nonlinearity of the relationship. Furthermore, the in-sample and out-of-sample performance of the nonparametric model is far better than the linear model.

Keywords: β -null recurrent; cointegration; conditional heteroscedasticity; Markov chain; nonparametric regression

JEL classifications: C13; C14; C22; E43

1. Introduction

The interplay of nonlinearity and nonstationarity has been an important topic in recent developments of econometrics. Karlsen and Tjøstheim [1] and Karlsen *et al.* [2], respectively, discuss the asymptotics for nonparametric estimation of autoregression and cointegrating regression when the regressor is a β -null recurrent Markov process. Using different data generating assumptions (*i.e.*, the regressor is a unit root process with innovations being a linear process), Wang and Phillips [3,4] discuss asymptotics for nonparametric estimation for nonlinear cointegrating regression models. The two frameworks have their own advantages and drawbacks. The β -null recurrence framework generalizes the unit root framework by incorporating more kinds of processes than the unit root process although it encounters some other restrictions. For example, the processes need to be Markov. For more discussion of linkage and difference of these two frameworks, we refer to [5].

The papers [1–4] focus on nonparametric estimation when the regressor is a univariate process. The issue of estimation when the regressor is a multivariate process has received less attention. As argued by Park and Phillips [6], the difficulty of extending the theory from a univariate regressor to a multivariate regressor is due to the fact that the recurrence property of a higher dimensional random walk process is different from the one dimensional random walk. Dong *et al.* [7] provide an intuitive example showing that the nonparametric estimate, when the regressor is a bivariate independent random walk, is not consistent. One way to avoid this problem is to use semi-parametric models rather than nonparametric models. For example, Schienle [8] considers an additive model rather than a pure nonparametric model to avoid this problem while in [9] a partial linear model is considered. However, nonparametric estimation is still possible in the multivariate case when the regressors are not independent random walks. Gao and Phillips [10] provide the theory of nonparametric estimation for multivariate regressors when one regressor is a unit root process and the other regressors are stationary processes. The reason why the model setup of [10] works in nonparametric estimation while the two-dimensional random walk does not is due to the fact that a one-dimensional random walk together with a multi-dimensional positive recurrent process form a $1/2$ -null recurrent system while the two dimensional random walk process is null recurrent but not β -null recurrent for any $\beta \in (0, 1)$. As discussed in [1], the β -null recurrence property plays a vital role to guarantee validity of nonparametric estimates.

In this paper, we introduce the theory of nonparametric estimation for a multivariate β -null recurrent system. The multivariate β -null recurrent processes include but are not restricted to the case of [10]. For example, our theory can cover the case where two regressors are both random walks but at the same time are cointegrated which is not covered by [10]. This will be discussed in more detail in Section 2. The cointegrated case is of importance in economics because it is well known that many macroeconomic time series are nonstationary but cointegrated such that they are driven by a common stochastic trend. Furthermore, in this paper, we use different mathematical techniques compared to [10]. In their paper, the technique of local time approximation for partial sums of functionals of unit root process is used, while in our paper, we use the Markov chain null recurrence framework.

It is well known that for the nonparametric kernel estimation for a d -dimensional stationary process, the convergence rate is $\sqrt{Th^d}$ with h being the bandwidth. In our model, the convergence rate is $\sqrt{n(T)h^d}$, where $n(T)$ is the number of regenerations of the β -null recurrent process which

is discussed in more detail in [1] (see also Appendix A). The difference of these two rates is due to the fact that for β -null recurrent processes, the number of observations in a small set (we refer to p. 376 in [1] for the definition and discussion of the small set) is $O_P(T^\beta L_s(T))$ rather than $O_P(T)$ (in the stationary case), where $L_s(\cdot)$ denotes a function slowly varying at infinity (cf. p. 6, [11]).

Furthermore, unlike [2], which assumes the regression error has constant variance, we allow existence of conditional heteroscedasticity. This is important for economic or financial time series modeling because many of these series are regarded to contain conditional heteroscedasticity (cf. [12–14]). In [15,16], estimation of conditional variance functions in autoregressive and regression models are discussed when the data are stationary. Wang and Wang [17] discuss nonparametric estimation of conditional mean and variance function when the regressor is a unit root process. Our paper is different from [17] in two ways: first, in our model, the regressor is multivariate rather than univariate; second, we employ the Markov β -null recurrence technique which is different from the local time approximation technique used in [17].

The rest of this paper is organized as follows. In Section 2, we introduce the model and the nonparametric estimate; in Section 3, the asymptotic properties for the estimator will be provided; in Section 4, Monte Carlo simulations will be conducted to examine the finite sample performance of the estimator; in Section 5, we apply the method to estimate the relationship of Federal funds rate with short term and long term T-bill rates; in Section 6, concluding remarks are made. To make this paper self-contained, we provide some basic notations and theory of Markov processes, especially the β -null recurrent processes in Appendix A. The mathematical proofs are contained in Appendix B.

Throughout this paper, all limits are taken “as $T \rightarrow \infty$ ” where T is the sample size, \rightarrow_d denotes weak convergence, \rightarrow_p denotes convergence in probability. $O_P(\cdot)$ means stochastic order same as, $o_P(\cdot)$ means stochastic order less than.

2. Model and Estimation

We are going to discuss estimation for the following model

$$y_t = g(x_t) + \varepsilon_t, \quad (1)$$

where $t = 1, \dots, T$, $\{x_t\} = \{x_{1,t}, \dots, x_{d,t}\}^\tau$ is a d -dimensional β -null recurrent process (see Appendix A for a precise definition), $\varepsilon_t = \sigma(x_t)e_t$ with $\{e_t\}$ being a positive recurrent process and $\sigma(\cdot)$ being a positive function. When the data are stationary, the model (1) has been widely studied, see, e.g., [15,16] (with univariate regressor) and [18] (with multivariate regressor). Recently, Wang and Wang [17] study the estimation of model (1) with $\{x_t\}$ being a univariate unit root process. As we have mentioned in the introduction, our paper has important differences from their paper.

The examples of univariate β -null recurrent processes with $\beta = 1/2$ include the random walk process with the innovation having second moment (cf. [19]); the threshold unit root model (cf. [20]) with arbitrary behavior in a compact set and unit root behavior outside the compact set. Moreover, under some regularity conditions, it has been shown that several multivariate Markov processes are β -null recurrent. For example, when $d = 2$, the following models of $\{x_{1,t}, x_{2,t}\}$ are $1/2$ -null recurrent:

- **Example 1:** $\{x_{1,t}\}$ is a 1/2-null recurrent process and $\{x_{2,t}\}$ is a positive recurrent process independent of $\{x_{1,t}\}$. This is proved by Lemma 3.1 of [2]. In fact, the independence assumption of $\{x_{1,t}\}$ and $\{x_{2,t}\}$ can be relaxed to asymptotic independence. We refer to Example 4.1 of [2] for more discussion of this.
- **Example 2:** $\{x_{1,t}\}$ and $\{x_{2,t}\}$ are both unit root processes and cointegrated. More specifically

$$x_t = Ax_{t-1} + e_t,$$

where $\{e_t\}$ is a bivariate i.i.d. process satisfying some regularity conditions, A is a 2×2 matrix having one eigenvalue equal to 1 and the other eigenvalue with absolute value less than 1. Myklebust *et al.* [5] show that in this model, $\{x_t\}$ is 1/2-null recurrent.

- **Example 3:** $\{x_t\}$ can be threshold cointegrated process. Consider the model

$$x_t = Ax_{t-1}I(x_{t-1} \in C^c) + Bx_{t-1}I(x_{t-1} \in C) + e_t,$$

where C is a compact set in \mathbb{R}^2 , A is a 2×2 matrix having one eigenvalue equal to 1 and the other eigenvalue with absolute value less than 1, B is an arbitrary matrix and $\{e_t\}$ is a bivariate i.i.d. process satisfying some regularity conditions. Cai *et al.* [21] prove that $\{x_t\}$ is 1/2-null recurrent under this model setup.

- **Example 4:** $\{x_t\}$ is generated from $x_{2,t} = f(x_{1,t}) + u_t$, with $\{x_{1,t}\}$ being a 1/2-recurrent process and $\{u_t\}$ being an i.i.d. sequence and independent of $\{x_{1,t}\}$. This is the nonlinear cointegration type model of [2].

Remark 1: The cases discussed can be extended to dimension higher than 2. For example, Myklebust *et al.* [5] show that a d -dimensional VAR(1) model is 1/2-null recurrent if the autoregressive matrix has one eigenvalue equal to 1 and the other eigenvalues with absolute values less than 1. By Theorem 2 of [5], it is shown that one-to-one transformation of a β -null recurrent process is also a β -null recurrent process.

Remark 2: We can see that the model of [10] is related to Example 1. Our methodology can be applied to other models listed above.

We propose to estimate the functional form $g(x)$ at (x_1, \dots, x_d) by the conventional local constant method through minimizing

$$\frac{1}{T} \sum_{t=1}^T (y_t - \alpha)^2 K\left(\frac{x_t - x}{h}\right) \text{ over } \alpha = g(x_1, \dots, x_d), \quad (2)$$

where $K\left(\frac{x_t - x}{h}\right) = \prod_{i=1}^d k_i\left(\frac{x_{i,t} - x_i}{h}\right)$ with $k_i(\cdot)$ being univariate kernel functions and h being the bandwidth parameter¹.

¹ In this paper, we use same bandwidth for different regressors. This is a little bit restrictive in practice. We leave the case of different choices of bandwidths for different regressors to future research.

Equation (2) implies the resulting estimate is given by

$$\hat{g}(x) = \frac{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)y_t}{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)}. \quad (3)$$

So that we have

$$\begin{aligned} \hat{g}(x) - g(x) &= \frac{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)(g(x_t) - g(x))}{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)} + \frac{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)\varepsilon_t}{\sum_{t=1}^T K\left(\frac{x_t-x}{h}\right)} \\ &\equiv I_1 + I_2, \end{aligned} \quad (4)$$

where I_1 and I_2 are respectively the bias and variance terms for the nonparametric estimate.

3. Asymptotic Theory

To study the asymptotics for the estimate (3), we need to introduce some technical assumptions.

A.1. $\{x_t\} = \{x_{1,t}, \dots, x_{d,t}\}$ is a d -dimensional Harris β -null recurrent Markov chain. Let $\pi_s(\cdot)$ be an invariant measure of the recurrent process admitting a locally Lipschitz continuous density $p_s(\cdot)$ which is locally bounded. $\sigma(\cdot)$ is a locally bounded, positive and Lipschitz continuous function such that for a vector x , there exists a constant C such that when y is in a neighbourhood of x , we have $|\sigma(y) - \sigma(x)| < C \|x - y\|$, where $\|\cdot\|$ is the Euclidean norm. $\{e_t\}$ is an i.i.d. sequence independent of $\{x_t\}$ with $E(e_1) = 0$, $E(e_1^2) = 1$ and $E(|e_1|^{4+\delta}) < \infty$ for some $\delta > 0$.

A.2. For any given x , $g(x)$ is twice continuously differentiable and the second order partial derivatives are locally bounded and Lipschitz continuous, i.e., $|\frac{\partial^2 g(x)}{\partial x_i \partial x_j} - \frac{\partial^2 g(y)}{\partial x_i \partial x_j}| < C \|x - y\|$ when y is in a neighbourhood of x , $i = 1, \dots, d$ and $j = 1, \dots, d$.

A.3. For $i = 1, \dots, d$, each k_i is a symmetric, nonnegative and bounded probability density function with compact support. Furthermore, the support of the kernel functions are small sets.

A.4. $h \rightarrow 0$ as $T \rightarrow \infty$, $u(T)h^d \rightarrow \infty$ as $T \rightarrow \infty$, and $h^2(\sqrt{u(T)h^d}) \rightarrow 0$ as $T \rightarrow \infty$, where $u(T) = T^\beta L_s(T)$, see Appendix A (cf. [1]), which is associated with $n(T)$, where $n(T)$ is the number of regenerations for the null recurrent Markov chain.

Remark: Assumption A.1 restricts the regressors to be a β -null recurrent system with some of the examples having been given in Section 2. The assumption on the error term is quite common in the literature, see, e.g., [18]. As shown in Lemma 3.1 of [2], the compound process $\{x_t, e_t\}$ is also a β -null recurrent process. It is possible to relax the assumption on $\{e_t\}$ such that endogeneity and autocorrelation are involved by applying some results in [2]. For the current paper, we use this assumption for illustrative purpose. Assumptions A.2 and A.3 are often used in nonparametric kernel estimation problems. We assume the support of the kernel function to be compact and a small set as in [1]. The bandwidth restriction in Assumption A.4 ensures that the nonparametric estimator is consistent and the estimation bias converges to 0 in probability.

To derive the asymptotic theory for the nonparametric estimator, we need the following three lemmas.

Lemma 1. Under assumptions A.1–A.4,

$$\frac{1}{n(T)h^d} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) \rightarrow_p p_s(x).$$

Remark. Lemma 1 shows that the denominator of the nonparametric estimator converges to an invariant density of the null recurrent process which is different from the positive recurrent case where the denominator converges to the probability density of the process. In Example 1 of Section 2, $p_s(x) = p_{s_1}(x_1) \times p_{s_2}(x_2)$, where $p_{s_1}(x_1)$ is an invariant density of $\{x_{1,t}\}$ and $p_{s_2}(x_2)$ is the (unique) invariant stationary density of $\{x_{2,t}\}$.

Lemma 2. Under assumptions A.1–A.4,

$$\frac{1}{\sqrt{n(T)h^d}} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) \varepsilon_t \rightarrow_d N(0, p_s(x)\sigma^2(x) \int K^2(u)du),$$

where $N(\cdot, \cdot)$ denotes a normal variable and $\int K^2(u)du = \Pi_i \int k_i^2(v)dv$.

Lemma 3. Under assumptions A.1–A.4,

$$\frac{1}{n(T)h^d} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) = O_P(h^2).$$

After proving Lemmas 1–3, we can derive the asymptotic distribution for $\widehat{g}(x) - g(x)$. We have the following theorem

Theorem 1. Under assumptions A.1–A.4,

$$\sqrt{n(T)h^d}(\widehat{g}(x) - g(x)) \rightarrow_d N(0, p_s^{-1}(x)\sigma^2(x) \int K^2(u)du).$$

Remark 1. In this theorem, stochastic normalization is used. As suggested by Equation (A.3) in Appendix A, we also have

$$\sqrt{u(T)h^d}(\widehat{g}(x) - g(x)) \rightarrow_d MN(0, M_\beta^{-1}(1)p_s^{-1}(x)\sigma^2(x) \int K^2(u)du),$$

where $MN(\cdot)$ denotes a mixed-normal variable. See also the discussion in [1].

Remark 2. Combining this theorem and Lemma 1, we have

$$\sqrt{\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)}(\widehat{g}(x) - g(x)) \rightarrow_d N(0, \sigma^2(x) \int K^2(u)du).$$

This self-normalization quantity is the same as that used in [2]. $\int K^2(u)du$ is known when a specific kernel function is used². So that for statistical inference purpose or construction of confidence band, we need to estimate $\sigma^2(x)$.

In [15,16], different methods for the variance estimation are proposed. Similar to [4] and [17], we estimate this quantity by a localized version of the usual residual-based method, *i.e.*,

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^T (y_t - \hat{g}(x))^2 K(\frac{x_t-x}{h_\sigma})}{\sum_{t=1}^T K(\frac{x_t-x}{h_\sigma})}, \tag{5}$$

where h_σ is another bandwidth. This is a two-step estimator which corresponds to local constant regression of the square of the residuals on the regressors. To investigate Equation (5), we impose the following assumption on h_σ .

A.5. h_σ satisfies same restrictions as h in Assumption A.4.

Then we have following theorem.

Theorem 2. Under assumptions A.1–A.5,

$$\hat{\sigma}^2(x) \rightarrow_p \sigma^2(x).$$

Remark 1. Theorem 2 shows that the estimator is consistent. In this paper, we focus on the mean function estimation. The investigation of more efficient estimation of $\sigma^2(x)$ (*cf.* [16]) is left for future research.

Remark 2. Combining Theorem 1 and Theorem 2, and using the Slutsky theorem, we have

$$\sqrt{\frac{\sum_{t=1}^T K(\frac{x_t-x}{h})}{\hat{\sigma}^2(x) \int K^2(u)du}} (\hat{g}(x) - g(x)) \rightarrow_d N(0, 1). \tag{6}$$

Thus we can construct the 95% confidence interval for the mean function estimator as

$$\hat{g}(x) \pm 1.96 \times \sqrt{\frac{\hat{\sigma}^2(x) \int K^2(u)du}{\sum_{t=1}^T K(\frac{x_t-x}{h})}}. \tag{7}$$

4. Monte Carlo Simulation

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of the nonparametric estimator. We focus on the case where $d = 2$. The performance of nonparametric estimation is evaluated by the root mean squared error (RMSE), which is defined by:

$$RMSE = \sqrt{\frac{1}{N} \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N (\hat{g}(x_{1,t_n}, x_{2,t_n}) - g(x_{1,t_n}, x_{2,t_n}))^2}, \tag{8}$$

² For the Quartic kernel used in the Monte Carlo simulation, it is equal to $(\frac{5}{7})^d$.

where x_{i,t_n} with $i = 1$ or 2 denotes the observation at time- t in the n -th replication with total replication number $N = 1000$. For both variables, the Quartic kernel (i.e., $k(u) = \frac{15}{16}(1 - u^2)^2 1(|u| \leq 1)$) is used. It is well known that the kernel selection plays little role in performance of nonparametric estimation. Bandwidth selection plays an important role in the performance of nonparametric estimates such that a large bandwidth may lead to large bias but small variance and *vice versa*. From the theoretical analysis, we know that the variance of the estimator is of order $(T^\beta L_s(T)h^d)^{-1}$ and the square of bias is of order h^4 . So that the optimal bandwidth minimizing the mean square error (MSE) for the estimator should be of order $(T^{-\beta} L_s^{-1}(T))^{\frac{1}{4+d}}$.

In the simulation, we concentrate on the case $\beta = 1/2$ and $d = 2$, so that we report the simulation results with bandwidth being $cT^{-1/12}$ with some pre-specified values of c . We also report the results when $h = h^*$, with h^* chosen by the leave-one-out cross validation method, which is widely used when the data is positive recurrent (cf. p. 50, [22]). From the simulation results below, we can find that the method performs quite well even when the regressors are null recurrent.

Specifically, we consider following models:

Model 1:

$$\{x_{1,t}\} \sim i.i.d.N(0, 1),$$

$$x_{2,t} = x_{2,t-1} + u_t,$$

with $\{u_t\} \sim i.i.d.N(0, 1)$

$$y_t = g(x_{1,t}, x_{2,t}) + \varepsilon_t,$$

where $\{\varepsilon_t\} \sim i.i.d.N(0, 1)$ and is independent of $\{x_{1,t}\}$ and $\{x_{2,t}\}$. In this model, $\{x_{1,t}\}$ is *i.i.d.*, $\{x_{2,t}\}$ is a random walk and $\{x_t\}$ is a 1/2-null recurrent process because it is a special case of Example 1 in Section 2 (it is also a special case of [10]). We let $g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$ or $g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$. The results for Model 1 are summarized in Table 1.

Table 1. RMSEs for Model 1.

Functional Form	c	T = 200	T = 400	T = 800
$g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$	1	0.3707	0.2996	0.2427
	2	0.3046	0.2678	0.2357
	3	0.4104	0.3777	0.3454
	4	0.5490	0.5131	0.4770
	h^*	0.3183	0.2629	0.2192
$g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$	1	0.3636	0.2915	0.2360
	2	0.2456	0.2071	0.1765
	3	0.2295	0.2010	0.1750
	4	0.2425	0.2111	0.1818
	h^*	0.2471	0.2054	0.1744

From Table 1, we can see that because of the trade-off between bias and variance, either a too large or a too small bandwidth will make the estimator less precise by the RMSE criterion. The bandwidth selected

by the cross validation method balances the bias and variance and performs reasonable especially when the sample size is large.

To further assess the finite sample approximation, we compare the normalized quantity in Equation (6) with the standard normal density. We compute the quantity at $[x_1 \ x_2] = [0 \ 0]$, the sample sizes are respectively 200, 400 and 800, the bandwidth used is $2T^{-1/12}$, the functional form is $g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$, the variance is 1 (we use the true function of the variance rather than the estimated quantity), and the replication number is 1000.

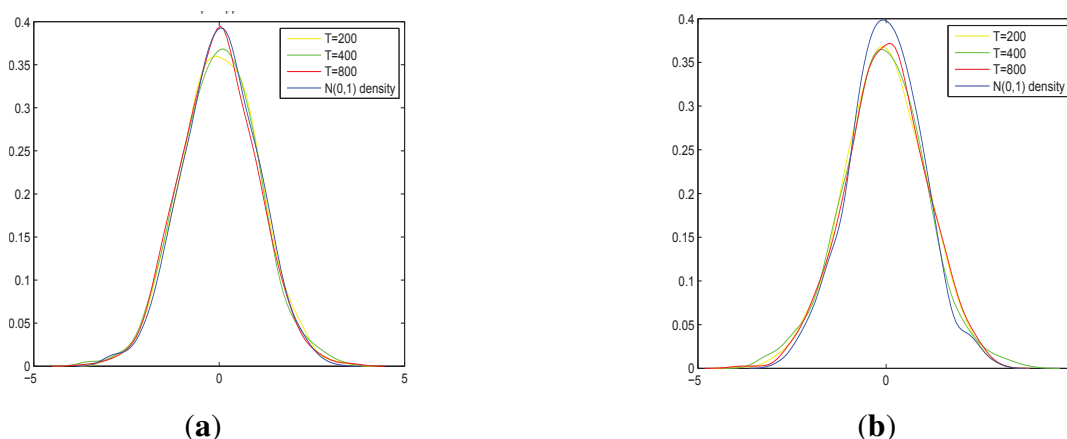


Figure 1. In (a) we compute $\sqrt{\frac{\sum_{t=1}^T K(\frac{x_t-x}{h})}{\int K(u)^2 du}} (\hat{g}(x) - g(x))$ at point $[0 \ 0]$ in each replication; in (b) we compute the normalized quantity in $\sqrt{\frac{\sum_{t=1}^T K(\frac{x_t-x}{h})}{\int K(u)^2 du}} (\hat{g}(x) - g(x))$ at the median of each replication. The number of replications is 1000.

From Figure 1a above, we can see that the approximation to normality is quite good. However, this is not always the case. For other choices of points or functional forms for evaluation, the performance may be much worse. For example, we find that when the true functional form is $g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$, there is a systematic bias in the estimation when the evaluation point is $[0 \ 0]$. This phenomenon looks strange at first glance, however, it is typical in the situation when the regressors are not positive recurrent. When the regressors are null recurrent, the simulated realizations may cover very different regions of the x -axis. Hence, for a fixed evaluation point, for some replications, there may be many observations in the neighbourhood while for other replications, there may be few observations (see more discussion in [1]). Figure 1b provides the finite sample approximation of the quantity using different evaluation points for different replications, *i.e.*, for each replication, the evaluation point is the median of the observations.

In the second model, we assume $\{x_t\}$ is a cointegrated process. Specifically, the model is:

Model 2:

$$x_t = Ax_{t-1} + u_t,$$

with $\{u_t\} \sim i.i.d.N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$.

$$y_t = g(x_{1,t}, x_{2,t}) + \varepsilon_t,$$

where $\{\varepsilon_t\} \sim i.i.d.N(0, 1)$ and is independent of $\{x_{1,t}\}$ and $\{x_{2,t}\}$. In this model, $\{x_t\}$ is 1/2-null recurrent process. We let $g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$ or $g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$. This model falls into the category of Example 2 in Section 2. The results for Model 2 are summarized in Table 2.

Table 2. RMSEs for Model 2.

Functional Form	c	T = 200	T = 400	T = 800
$g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$	0.1	0.5098	0.3891	0.2922
	0.2	0.3853	0.2914	0.2234
	0.4	0.3038	0.2518	0.2437
	0.8	0.3481	0.3892	0.4826
	h^*	0.2941	0.2445	0.2009
$g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$	0.5	0.2513	0.1869	0.1412
	1	0.1971	0.1560	0.1323
	2	0.1890	0.1683	0.1467
	3	0.2044	0.1773	0.1583
	4	0.2111	0.1864	0.1654
	h^*	0.1991	0.1637	0.1334

From Table 2, we can see that for different data generating mechanism of $\{x_t\}$ or function form of g , the choice of h should be quite different. In addition, the cross validation method serves as a good choice.

Next, we consider the model where the regressors are threshold cointegrated.

Model 3:

$$x_t = Ax_{t-1}1(x_{t-1} \in D) + Bx_{t-1}1(x_{t-1} \in D^c) + u_t,$$

with $\{u_t\} \sim i.i.d.N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 3/4 & 1/8 \\ 1 & 1/2 \end{pmatrix}$, $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$, $D = [\tau_1 \ \tau_2] \times [\tau_3 \ \tau_4] = [-3 \ 3] \times [-2 \ 2]$ and

$$y_t = g(x_{1,t}, x_{2,t}) + \varepsilon_t,$$

where $\{\varepsilon_t\} \sim i.i.d.N(0, 1)$ and is independent of $\{x_{1,t}\}$ and $\{x_{2,t}\}$. We know in this model, $\{x_{1,t}\}$ and $\{x_{2,t}\}$ are threshold cointegrated processes with different cointegrated coefficients in different regimes. According to [21], they form a bivariate 1/2-null recurrent system. We let $g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$ or $g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$. The results for Model 3 are summarized in Table 3.

Table 3. RMSEs for Model 3.

Functional Form	c	T = 200	T = 400	T = 800
$g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$	0.1	0.8941	0.7909	0.6446
	0.5	0.4422	0.3715	0.3806
	1	0.4502	0.5221	0.6176
	2	0.7344	0.8827	1.1070
	h^*	0.4104	0.3410	0.3031
$g(x_1, x_2) = x_1 \times \frac{1}{1+x_2^2}$	0.5	0.4230	0.3037	0.2298
	1	0.3042	0.2625	0.2417
	2	0.3411	0.3173	0.2874
	3	0.3732	0.3412	0.3042
	h^*	0.3175	0.2564	0.2155

Our conclusion for Table 3 is same as that for Table 2.

The above three models assume the error term to be homoscedastic. In the final example, heteroscedasticity will be taken into account. The model is:

Model 4:

$$\{x_{1,t}\} \sim i.i.d.N(0, 1),$$

$$x_{2,t} = x_{2,t-1} + u_t,$$

with $\{u_t\} \sim i.i.d.N(0, 1)$

$$y_t = g(x_{1,t}, x_{2,t}) + \sigma(x_{1,t}, x_{2,t})e_t,$$

where $\{e_t\} \sim i.i.d.N(0, 1)$ and is independent of $\{x_{1,t}\}$ and $\{x_{2,t}\}$. We let $g(x_1, x_2) = x_1 + \frac{1}{1+x_2^2}$ with $\sigma^2(x_1, x_2) = \frac{x_1^2}{1+x_2^2}$. We estimate the conditional variance function using Equation (5) and then calculate the RMSE. In our simulation, the same bandwidth for the mean and variance functions is used.

The results for Model 4 are summarized in Table 4.

Table 4. RMSEs for Model 4.

Functions	c	T = 200	T = 400	T = 800
$g(\cdot)$	1	0.2224	0.1796	0.1542
	2	0.2615	0.2352	0.2140
	3	0.3955	0.3674	0.3395
	4	0.5415	0.5082	0.4741
	h^*	0.2837	0.2565	0.2167
$\sigma^2(\cdot)$	1	0.3890	0.3196	0.2689
	2	0.3507	0.2996	0.2651
	3	0.3788	0.3369	0.2971
	4	0.4321	0.3897	0.3410
	h^*	0.7900	0.3686	0.3093

From Table 4, we can see that in general the cross validation method is not as good as in the homoscedasticity case. For some choice of fixed bandwidth, the RMSEs for mean and variance functions can be smaller. However, the cross validation is still a reasonable choice because in practice we do not know the true DGP which makes it difficult to use some pre-specified bandwidth.

In summary, we can see that the choice of bandwidth has large impact on the performance of the nonparametric estimates. The leave-one-out cross validation method is still a reasonable option for the multivariate nonparametric estimates with β -null recurrent regressor. In the empirical analysis in the next section, this method will be used.

5. Empirical Application to the Relationship of Interest Rates

In this section, we apply the proposed method to study the relationship of three interest rates: the effective Federal funds rate (FF), 3-month Treasury bill rate (TB3m) and 5-year Treasury bill rate (TB5y). The Federal Reserve (Fed) implements monetary policy by targeting the effective FF (see, e.g., [23]); the TB3m is a preeminent risk-free rate in the U.S. money market and is often used by researchers as a proxy for the risk-free asset (see, e.g., [24]), and TB5y is often used to represent the long term interest rate (see, e.g., [25]).

In the literature (*cf.* [23,26]), it is often argued that the interest rates move together according to the expectation hypothesis (EH) such that the Treasury bill rates are equal to market's expectation for the FF over the term of TB rates plus a risk premium. So that according to conventional EH/monetary policy views, FF "anchors" the U.S. money market. However, according to [27], the short run T-bill rate adjusts before the FF, rather than *vice versa*. This is possibly due to the fact that if the market anticipates changes in the FF, the T-bill rates will move in advance of the Federal funds rate. Thus, the market should have anticipated the information of the FF before its announcement. Under the same reasoning, if the short term T-bill rate contains short term information of the market, we expect that the long term T-bill rate contains long term information of the market. So that both short term and long term T-bill rates will influence the FF.

We use monthly data of FF, TB3m and TB5y from the website of the Federal Reserve Bank of St Louis. The sample period is from January 1962 to May 2014 and the sample size is 629. The ADF test suggests that all the three series contain unit roots with p -values for TB3m, TB5y and FF being respectively 0.2807, 0.3513 and 0.2648. We also perform ADF test for the interest rate differential series $\{TB3m_t - TB5y_t\}$ and the result suggests that the series is stationary with p -value being less than 0.01. This means that the short term and long term T-bill rates are cointegrated with cointegrated vector $[1 \ -1]$ ³. Thus TB3m and TB5y form a 1/2-recurrent system so that we can apply the proposed method to estimate the relationship of FF with TB3m and TB5y⁴.

The time series plot of the three series is shown in Figure 2.

³ Recently, there are many papers considering the nonlinear dynamics of the term structure of interest rate where the nonlinear error correction models are used, see, e.g., [28,29]. This kind of process is also covered by our theory.

⁴ The bandwidth is selected by the leave-one-out cross validation method and is 0.1130.

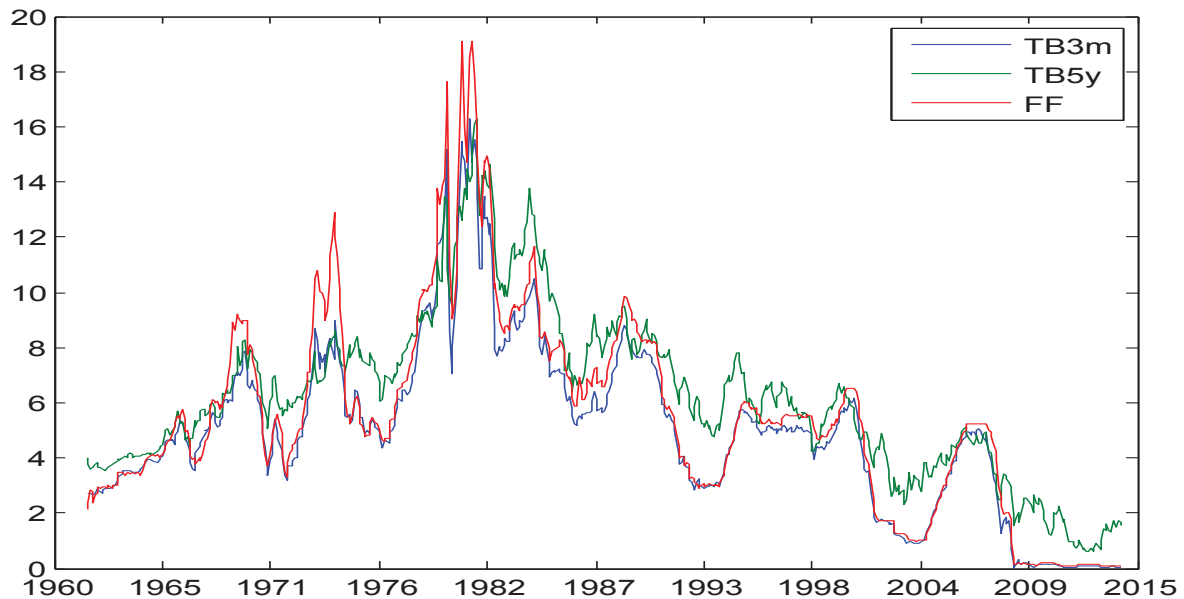


Figure 2. Time series plot of FF, TB3m and TB5y: 1962–2014.

From Figure 2, we can see that the three series move together. In particular, FF and TB3m seem to have a close relationship.

To study the relationship of the variables, in an exploratory phase we first examine the relationship of FF with TB5y when TB3m is fixed. Specifically, we plot the graph of fitted values against TB5y when TB3m is 4 or 8. Notice that because TB5y and TB3m are cointegrated, when TB3m is fixed, we can only study the relationship of FF with TB5y when TB5y is within some small regions. Otherwise, there will be insufficient observations for the nonparametric estimate⁵. So that when TB3m is 4, we study the relationship with TB5y within $[3.8 \ 4.4]$ and when TB3m is 8, we study the relationship with TB5y within $[7.8 \ 8.4]$. In addition, we can plot the 95% point-wise confidence intervals according to Equation (7). The results are shown in Figure 3.

⁵ That's also the reason why we do not try to display the joint estimation results, but instead, we display the relationship of FF and TB5y with TB3m fixed.

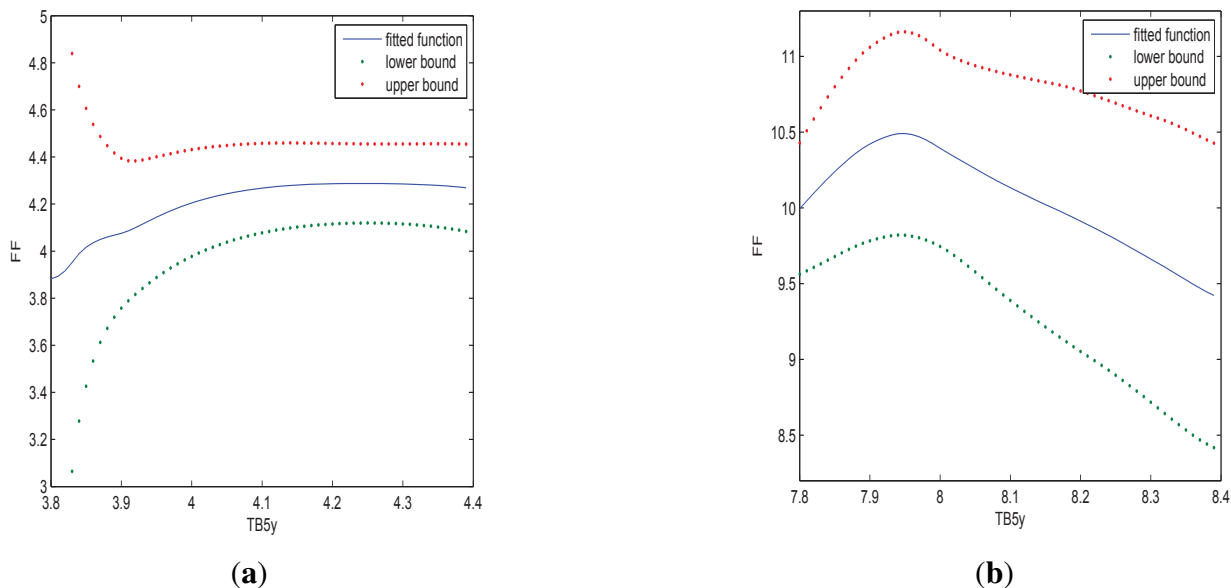


Figure 3. In (a) we estimate the relationship of FF and TB5y when TB3m is 4; in (b) we estimate the relationship of FF and TB5y when TB3m is 8.

Figure 3 indicates that the relationship of FF with TB5y is nonlinear and the relationship is largely affected by TB3m.

Similarly, we can examine the relationship of FF with TB3m when TB5y is fixed. When TB5y is 4, we study the relationship with TB3m within [3.7 4.3] and when TB5y is 8, we study the relationship with TB3m within [7.7 8.3]. The results are shown in Figure 4.

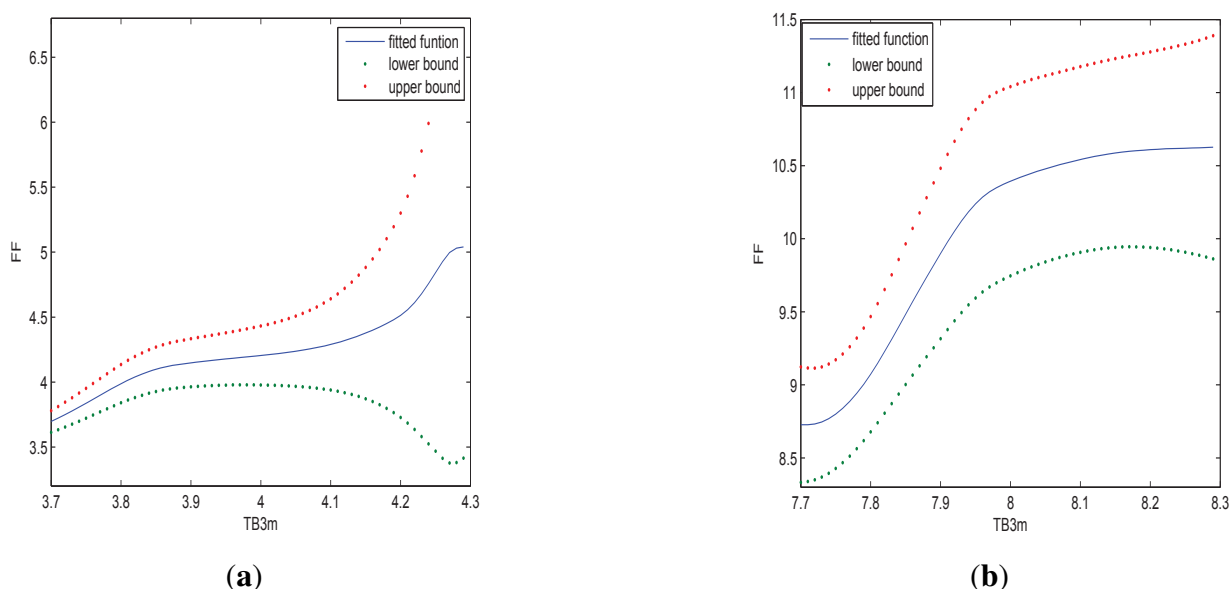


Figure 4. In (a) we estimate the relationship of FF and TB3m when TB5y is 4; in (b) we estimate the relationship of FF and TB3m when TB5y is 8.

From Figure 4, we can see that the relationship of FF with TB3m is nonlinear and different TB5y does not make large difference in the relationship.

In summary, the pairwise exploratory phase suggests that the relationship of FF and TB3m and TB5y may not be linear. The nonlinearity may be due to transaction cost as suggested by [28] or the policy interventions as suggested by [30].

Next, we estimate our model (1) by estimating g nonparametrically. We compare the in-sample mean square error of the nonparametric model with the linear model. Define the following MSE

$$MSE = \frac{1}{T} \sum_{t=1}^T (FF_t - \widehat{FF}_t)^2,$$

where \widehat{FF} is the fitted value from the nonparametric model or the linear model (regress FF on $TB3m$, $TB5y$ and a constant). The MSE for the nonparametric model is 0.0995 while the MSE for the linear model is 0.2802. It can be seen that the nonparametric model outperforms the linear model.

However, it is well known that comparison based on the in-sample performance as done above is sensitive to outliers and data mining, see, e.g., [31]. Empirical evidence of out-of-sample forecast performance is generally regarded as more trustworthy. In addition, out-of-sample forecasts can better reflect the information available to forecasters in “real time” (cf. [32]). As emphasized by [33], out-of-sample forecast is the “ultimate test of forecasting model”.

We study the out-of-sample forecasts of the nonparametric model and linear model by comparing their *one-step-ahead* forecasting performance. Specifically, we define the out-of-sample MSE (OMSE) as

$$OMSE = \frac{1}{m} \sum_{t=T-m+1}^T (FF_t - \widetilde{FF}_t)^2,$$

where \widetilde{FF}_t is the fitted value at $(TB3m_t, TB5y_t)$ with the model estimated using the observations up to time $t - 1$ ⁶. Furthermore, out-of-sample size m is taken to be $[1/10T]$, $[1/15T]$, $[1/20T]$, where T is the full sample size. The reason for choosing relatively small proportion of out-of-sample evaluation period is mainly due to the fact that for the nonparametric forecast, because of the nature of local estimate, a too small proportion of in-sample size may make the forecast impossible if the observation of the period we are going to forecast is an “outlier” such that we do not have enough observations in the neighborhood.

The results for the out-of-sample forecasts are reported Table 5.

Table 5. OMSE Comparison.

Model	$m = [1/10 T]$	$m = [1/15 T]$	$m = [1/20 T]$
Linear model	0.1018	0.0671	0.0519
Nonparametric model	0.0021	0.0022	0.0015

From Table 5, we can see that the nonparametric model consistently outperforms the linear model by the out-of-sample evaluation. It is also interesting that the OMSE is smaller than the in-sample

⁶ *I.e.*, we use recursive (expanding) windows in estimation. Compared with the rolling-window forecasting which uses fixed in-sample size, the recursive window utilizes all the data when one forecasts for the next period, thus may increase the precision of the in-sample estimation, which in turn, leads to better out-of-sample forecasting.

MSE. It is possible because the volatility increases with regressors (see Figure 5 below) and our out-of-sample forecasting periods are the periods where the interest rates are very low as is seen from Figure 2. The comparison of OMSE provides additional strong evidence that there exists nonlinearity in the relationship.

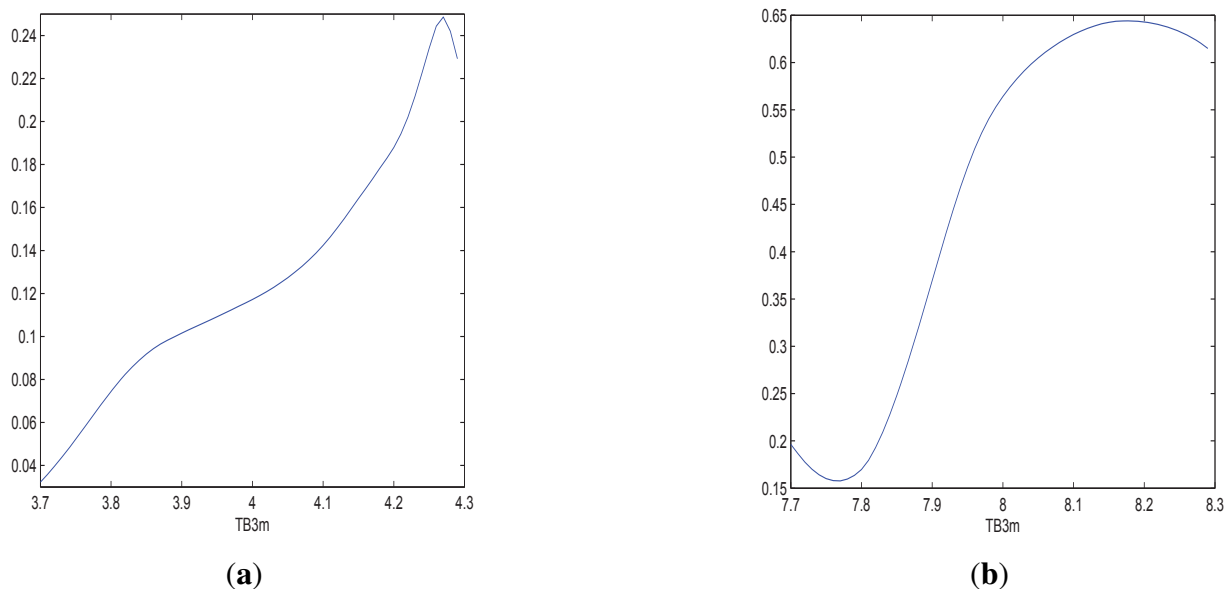


Figure 5. In (a) we estimate the shape of the variance with TB3m when TB5y = 4; in (b) we estimate the shape of the variance with TB3m when TB5y = 8.

Similarly, we can estimate the conditional variance function using Equation (5) (with the same bandwidth as that used for the mean function estimation). Figure 5 studies the variance conditional on TB3m when TB5y is fixed. When TB5y is 4, we study shape of the variance with TB3m within [3.7 4.3] and when TB5y is 8, we study the relationship with TB3m within [7.7 8.3].

Figure 5 indicates that there exists conditional heteroscedasticity⁷. The conditional heteroscedasticity of FF is also found in [34], where a univariate model with ARCH-type (*cf.* [12]) volatility function is used. Chan *et al.* [35] study different models for the short term interest rate encompassed in the following stochastic differential equation (SDE):

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^\gamma dW_t, \tag{9}$$

where r_t is the (spot) interest rate, W_t is the standard Brownian motion, α , β , σ and γ are some constants. When $\gamma = 0$, Equation (9) is the Vasicek [36] model and when $\gamma = 1/2$, Equation (9) is the famous Cox-Ingersoll-Ross (CIR, [37]) model. We can see that in [36], the process is conditional homoscedastic and when $\gamma \neq 0$, there exists conditional heteroscedasticity. The empirical findings of [35] suggest that γ is significantly larger than 0, so that the volatility increases with the interest rate. Our model setup is regression rather than autoregression. However, the interest rates are cointegrated and move in the same direction, so that the finding of [35] implies the volatility in our model also increases with the regressors. This is consistent with the empirical results in our paper.

⁷ We can also find that the volatility with TB3m fixed is not constant. To save space, we don't report the results here.

6. Conclusions

In this paper, we establish the asymptotic theory of nonparametric estimation for multivariate β -null recurrent processes in presence of heteroscedasticity. The Monte Carlo simulation results suggest that the nonparametric estimator performs well using bandwidth selected by the cross validation method. The application to the relationship of Federal funds rate with short term and long term T-bill rates indicates the existence of nonlinearity.

In the current paper, the variance function estimation is not discussed in details as for the mean function estimation. This is left for future research. In empirical applications, because of different ranges of different regressors especially in this multivariate nonstationary case, different bandwidths may be used for different regressors to make the estimate more efficient (*cf.* [18]).

Acknowledgements

We thank the editor and three anonymous referees for invaluable feedback which greatly improved the structure of the paper.

Author Contributions

Biqing Cai is the main author of this paper and both authors contribute to this paper.

Appendix

A. Some Markov Theory

To make this paper self-contained, in this Appendix, we introduce necessary notions of Markov chain properties. They are fundamental for deriving the asymptotic theory in this paper. We refer to [1,2] and [38] for more comprehensive treatments.

Let $\{X_t, t \geq 0\}$ be a ϕ -irreducible Markov chain on the state space (E, \mathcal{E}) with transition probability P . This means that for any set $A \in \mathcal{E}$ with $\phi(A) > 0$, we have $\sum_{t=1}^{\infty} P^t(x, A) > 0$ for all $x \in E$. In this paper, we take $E \subseteq \mathbb{R}^d$. We further assume that the ϕ -irreducible Markov chain $\{X_t\}$ is Harris recurrent.

Definition 1. A Markov chain $\{X_t\}$ is Harris recurrent if, given a neighborhood N_x of x with $\phi(N_x) > 0$, $\{X_t\}$ returns to N_x with probability one, for any $x \in E$.

The Harris recurrent chain is positive recurrent if there exists an invariant *probability* measure such that $\{X_t\}$ is strictly stationary and is null recurrent otherwise. The Harris recurrence allows one to construct a split chain, which decomposes the partial sum of functions of $\{X_t\}$ into blocks of i.i.d. parts and two asymptotically negligible remaining parts. Let τ_k be the regeneration times, T the number of observations and $n(T)$ the number of regenerations as in [1]⁸.

⁸ They use the notation $T(n)$ instead of $n(T)$.

For the process $\{G(X_t) : t \geq 0\}$, defining

$$U_k = \begin{cases} \sum_{t=0}^{\tau_0} G(X_t), & k = 0 \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} G(X_t), & 1 \leq k \leq n(T), \\ \sum_{t=\tau_{n(T)}+1}^T G(X_t), & k = n(T) + 1 \end{cases}$$

where $G(\cdot)$ is a real function defined on \mathbb{R}^d , then we have

$$S_n(G) = \sum_{t=0}^T G(X_t) = U_0 + \sum_{k=1}^{n(T)} U_k + U_{n(T)+1}. \tag{A.1}$$

From [38], we know that $\{U_k, k \geq 1\}$ is a sequence of i.i.d. random variables, and U_0 and $U_{n(T)+1}$ converge to zero almost surely (a.s.) when they are divided by the number of regenerations $n(T)$ (using Lemma 3.2 in [1]).

The general Harris recurrence only yields stochastic rates of convergence in asymptotic theory of the nonparametric estimators, where distribution and size of the number of regenerations $n(T)$ have no a priori known structure but fully depend on the underlying process $\{X_t\}$. To obtain a specific rate of $n(T)$ in our asymptotic theory for the null recurrent process, we next impose some restrictions on the tail behavior of the distribution of the recurrence times of the Markov chain.

Definition 2. A Markov chain $\{X_t\}$ is β -null recurrent if there exist a small nonnegative function f , an initial measure λ , a constant $\beta \in (0, 1)$, and a slowly varying function $L_f(\cdot)$ such that

$$E_\lambda\left(\sum_{t=1}^T f(X_t)\right) \sim \frac{1}{\Gamma(1 + \beta)} T^\beta L_f(T), \tag{A.2}$$

as $T \rightarrow \infty$, where E_λ stands for the expectation with initial distribution λ and $\Gamma(1 + \beta)$ is the Gamma function with parameter $1 + \beta$.

The definition of a small function f in the above definition can be found in some existing literature (cf. p. 15, [38]). Assuming β -null recurrence restricts the tail behavior of the recurrence time of the process to be a regularly varying function. In fact, for all small functions f , by Lemma 3.1 in [1], we can find an $L_s(\cdot)$ such that Equation (A.2) holds for the β -null recurrent Markov chain with $L_f(\cdot) = \pi_s(f)L_s(\cdot)$, where π_s is an invariant measure of the Markov chain $\{X_t\}$, $\pi_s(f) = \int f(x)\pi_s(dx)$ and s is the small function in the minorization inequality (3.4) of [1]. Letting $L_s(T) = L_f(T)/(\pi_s(f))$ and following the argument in [1], we may show that the regeneration number $n(T)$ of the β -null recurrent Markov chain $\{X_t\}$ has the following asymptotic distribution

$$\frac{n(T)}{T^\beta L_s(T)} \rightarrow_d M_\beta(1), \tag{A.3}$$

where $M_\beta(1)$ is the Mittag-Leffler distribution with parameter β (cf. [39]). Since $n(T) < T$ a.s. for the null recurrent case by Equation (A.3), the rates of convergence for the nonparametric kernel estimators are slower than those for the stationary time series case (cf. [2]). We also denote $u(T) = T^\beta L_s(T)$, which is used in the main text of this paper.

In Section 2 of this paper, some typical examples of multivariate β -null recurrent processes are provided.

B. Mathematical Proofs

Proof of Lemma 1. According to the split chain technique (cf. (A.1)), we have

$$\frac{1}{n(T)h^d} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) = \frac{1}{n(T)h^d} \{U_0(K_h) + \sum_{k=1}^{n(T)} U_k(K_h) + U_{n(T)+1}(K_h)\},$$

where $K_h = \prod_{i=1}^d k_i(\frac{x_{i,t} - x_i}{h})$. We have for $k = 1, \dots, n(T)$

$$E(U_k(K_h)) = \pi_s(K_h) = \int k_1\left(\frac{y_1 - x_1}{h}\right) \dots k_d\left(\frac{y_d - x_d}{h}\right) p_s(y_1, \dots, y_d) dy.$$

Denoted $\frac{y_1 - x_1}{h} = z_1, \dots, \frac{y_d - x_d}{h} = z_d$, then

$$\begin{aligned} \int k_1\left(\frac{y_1 - x_1}{h}\right) \dots k_d\left(\frac{y_d - x_d}{h}\right) p_s(y_1, \dots, y_d) dy &= h^d \int k_1(z_1) \dots k_d(z_d) p_s(x_1 + z_1 h, \dots, x_d + z_d h) dz \\ &= h^d p_s(x) + o(h^d). \end{aligned}$$

And similar to the proof of Lemma 3.2 of [1], we have

$$P_\lambda(|U_0(K_h)| < \infty) = 1$$

and

$$P_\lambda(|U_{n(T)+1}(K_h)| < \infty) = 1,$$

where λ is an arbitrary initial measure.

Then Lemma 1 follows from the weak law of large numbers because $n(T)h^d \rightarrow \infty$ as $T \rightarrow \infty$. \square

Proof of Lemma 2. We have

$$\begin{aligned} \frac{1}{\sqrt{n(T)h^d}} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) \varepsilon_t &= \frac{1}{\sqrt{n(T)h^d}} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) \sigma(x_t) e_t \\ &= \frac{1}{\sqrt{n(T)h^d}} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) \sigma(x) e_t + \frac{1}{\sqrt{n(T)h^d}} \sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) [\sigma(x_t) - \sigma(x)] e_t \\ &\equiv I_1 + I_2. \end{aligned} \tag{B.1}$$

It follows from Theorem 3.5 of [2] (it is easy to show that the conditions of the theorem hold under our assumptions A.1–A.4) that

$$I_1 \rightarrow_d N(0, p_s(x) \sigma^2(x) \int K^2(u) du). \tag{B.2}$$

For the term I_2 , by the i.i.d. assumption on $\{e_t\}$

$$\begin{aligned} E\left[\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right) [\sigma(x_t) - \sigma(x)] e_t\right]^2 &= E \sum_{t=1}^T K^2\left(\frac{x_t - x}{h}\right) [\sigma(x_t) - \sigma(x)]^2 \\ &\leq C^2 h^2 E \sum_{t=1}^T K^2\left(\frac{x_t - x}{h}\right) = o(u(T)h^d) \end{aligned}$$

by Assumption A.1 on $\sigma(\cdot)$. So that by the Markov inequality

$$I_2 = o_P(1). \tag{B.3}$$

Combining the results (B.1) (B.2) and (B.3), we have proved this lemma.

□

Proof of Lemma 3. Using the same split chain technique as in the proof of Lemma 1,

$$\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) = U_0(L_h) + \sum_{k=1}^{n(T)} U_k(L_h) + U_{n(T)+1}(L_h),$$

with $L_h = K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x))$. We have

$$E(U_k(L_h)) = \pi_s(L_h) = \int K\left(\frac{y - x}{h}\right)(g(y) - g(x))p_s(y)dy.$$

Using the change of variables

$$= h^d \int k_1(z_1) \cdots k_d(z_d)(g(x_1 + z_1h, \dots, x_d + z_dh) - g(x_1, \dots, x_d))p_s(x_1 + z_1h, \dots, x_d + z_dh)dz$$

and by Taylor expansion

$$= h^d \left\{ \int k_1(z_1) \cdots k_d(z_d)(g'_1(x_1, \dots, x_d)z_1h + \cdots + g'_d(x_1, \dots, x_d)z_dh + 1/2 \sum_{i=1}^d \sum_{j=1}^d g''_{ij}(\tilde{x}_1, \dots, \tilde{x}_d)z_i z_j h^2)p_s(x_1 + z_1h, \dots, x_d + z_dh)dz \right\},$$

where \tilde{x}_i is a real number between x_i and $x_{i,t}$. So that

$$E(U_k(L_h)) = O(h^{2+d})$$

and

$$E\left(\frac{1}{n(T)h^d} \sum_{k=1}^{n(T)} U_k(L_h)\right) = O(h^2). \tag{B.4}$$

Moreover, we have

$$E(U_k^2(L_h)) = \int K^2\left(\frac{y - x}{h}\right)(g(y) - g(x))^2 p_s(y)dy = O(h^{2d+2})$$

and

$$E\left(\frac{1}{n(T)h^d} \sum_{k=1}^{n(T)} U_k(L_h)\right)^2 = O\left(\frac{1}{u(T)h^{2d}} \times h^{2d+2}\right). \tag{B.5}$$

By Equations (B.4) and (B.5) and the Markov inequality

$$\frac{1}{n(T)h^d} \sum_{k=1}^{n(T)} U_k(L_h) = O_P(h^2) + O_P\left(\frac{h}{\sqrt{n(T)}}\right) = O_P(h^2) \tag{B.6}$$

by Assumption A.4.

Using the fact that

$$|U_0(L_h)| = \left| \sum_{t=1}^{\tau_0} K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) \right| \leq \sum_{t=1}^{\tau_0} \left| K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) \right|$$

and by the definition of τ_0 ,

$$\begin{aligned} E\left(\sum_{t=1}^{\tau_0} \left| K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) \right|\right) &\leq E\left(\sum_{t=\tau_0}^{\tau_1} \left| K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x)) \right|\right) \\ &= \int K\left(\frac{y - x}{h}\right) |g(y) - g(x)| p_s(y) dy = O(h^{d+1}). \end{aligned}$$

So that by the Markov inequality

$$\frac{1}{n(T)h^d} U_0(L_h) = O_P\left(\frac{h}{n(T)}\right) = o_P(h^2) \tag{B.7}$$

by Assumption A.4. Similarly, we have

$$\frac{1}{n(T)h^d} U_{n(T)+1}(L_h) = o_P(h^2). \tag{B.8}$$

Lemma 3 holds because of Equations (B.6)–(B.8).

□

Proof of Theorem 1. We have

$$\begin{aligned} \widehat{g}(x) - g(x) &= \frac{\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)(g(x_t) - g(x))}{\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)} + \frac{\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)\varepsilon_t}{\sum_{t=1}^T K\left(\frac{x_t - x}{h}\right)} \\ &\equiv I_1 + I_2. \end{aligned}$$

So that Theorem 1 follows from Lemmas 1–3 and the Slutsky theorem.

□

Proof of Theorem 2. We have

$$\widehat{\sigma}^2(x) = \frac{\sum_{t=1}^T (y_t - \widehat{g}(x))^2 K\left(\frac{x_t - x}{h_\sigma}\right)}{\sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)} \equiv \frac{B_1 + B_2 + B_3}{B_0},$$

where $B_0 = \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)$, $B_1 = \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)\varepsilon_t^2$, $B_2 = \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x_t) - \widehat{g}(x))^2$ and $B_3 = 2 \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x_t) - \widehat{g}(x))\varepsilon_t$. We have

$$\frac{B_1}{B_0} = \frac{\sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)\sigma^2(x) + \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(\varepsilon_t^2 - \sigma^2(x_t)) + \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(\sigma(x_t)^2 - \sigma^2(x))}{\sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)}$$

$$= \sigma^2(x) + O_P((u(T)h_\sigma^d)^{-1/2}) + O_P(h_\sigma) = \sigma^2(x) + o_P(1)$$

by Lemma 2 and Assumption A.1 on $\sigma(\cdot)$.

Thus Theorem 2 holds if we can prove $\frac{B_2}{B_0} = o_P(1)$ and $\frac{B_3}{B_0} = o_P(1)$.

Using Taylor expansion, we have

$$\begin{aligned} B_2 &= \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x_t) - g(x) + g(x) - \widehat{g}(x))^2 \\ &\equiv B_{21} + B_{22} + B_{23}, \end{aligned}$$

where $B_{21} = \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x_t) - g(x))^2$, $B_{22} = \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x) - \widehat{g}(x))^2$, and $B_{23} = 2 \sum_{t=1}^T K\left(\frac{x_t - x}{h_\sigma}\right)(g(x_t) - g(x))(g(x) - \widehat{g}(x))$.

By Lemma 3, we have

$$\frac{B_{21}}{B_0} = o_P(1).$$

And by Theorem 1,

$$\frac{B_{22}}{B_0} = (g(x) - \widehat{g}(x))^2 = o_P(1).$$

By the Cauchy-Schwarz inequality,

$$B_{23}^2 \leq B_{21}^2 B_{22}^2,$$

which implies

$$\frac{B_{23}}{B_0} = o_P(1).$$

Again by the Cauchy-Schwarz inequality,

$$B_3^2 \leq B_1^2 B_2^2.$$

Thus

$$\frac{B_3}{B_0} = o_P(1).$$

So that

$$\widehat{\sigma}^2(x) \rightarrow_p \sigma^2(x).$$

□

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Karlsen, H.A.; Tjøstheim, D. Nonparametric estimation in null recurrent time series. *Ann. Statist.* **2001**, *29*, 372–416.
2. Karlsen, H.A.; Myklebust, T.; Tjøstheim, D. Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.* **2007**, *35*, 252–299.

3. Wang, Q.; Phillips, P.C.B. Asymptotic theory for local time density and nonparametric cointegrating regression. *Econometr. Theory* **2009a**, *25*, 710–738.
4. Wang, Q.; Phillips, P.C.B. Structural nonparametric cointegrating regression. *Econometrica* **2009b**, *77*, 1901–1948.
5. Myklebust, T.; Karlsen, H.A.; Tjøstheim, D. Null recurrent unit root processes. *Econometr. Theory* **2012**, *28*, 1–41.
6. Park, J.; Phillips, P.C.B. Nonlinear regression with integrated time series. *Econometrica* **2001**, *69*, 117–161.
7. Dong, C.; Gao, J.; Tjøstheim, D.; Yin, J. *Specification Testing for Nonlinear Multivariate Cointegrating Regressions*; Monash University working paper; Monash University: Melbourne, Australia, 2014.
8. Schienle, M. *Nonparametric Nonstationary Regression*; University of Mannheim working paper; University of Mannheim: Mannheim, Germany, 2008.
9. Chen, J.; Gao, J.; Li, D. Estimation in semi-parametric regression with non-stationary regressors. *Bernoulli* **2012**, *18*, 678–702.
10. Gao, J.; Phillips, P.C.B. *Functional Coefficient Nonstationary Regression with Non- and Semi-parametric Cointegration*; Monash University working paper; Monash University: Melbourne, Australia, 2013.
11. Bingham, N.H.; Goldie, C.M.; Teugels, J.L. *Regular Variation*; Cambridge University Press: Cambridge, UK, 1989.
12. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **1982**, *50*, 987–1008.
13. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econometr.* **1986**, *31*, 307–327.
14. Bollerslev, T.; Chou, R.Y.; Kroner, K.F. ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econometr.* **1992**, *52*, 5–59.
15. Härdle, W.; Tsybakov, A. Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometr.* **1997**, *81*, 223–242.
16. Fan, J.; Yao, Q. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **1998**, *85*, 645–660.
17. Wang, Q.; Wang, Y. Nonparametric cointegrating regression with NNH errors. *Econometr. Theory* **2013**, *29*, 1–27.
18. Yang, L.; Tschernig, R. Multivariate bandwidth selection for local linear regression. *J. R. Statist. Soc.: Ser. B* **1999**, *61*, 793–815.
19. Kallianpur, G.; Robbins, H. The sequence of sums of random variables. *Duke Math. J.* **1954**, *21*, 285–307.
20. Gao, J.; Tjøstheim, D.; Yin, J. Estimation in threshold autoregressive models with a stationary and a unit root regime. *J. Econometr.* **2013**, *172*, 1–13.
21. Cai, B.; Gao, J.; Tjøstheim, D. *A New Class of Bivariate Threshold Cointegration Models*; University of Bergen working paper; University of Bergen: Bergen, Norway, 2014.

22. Pagan, A.; Ullah, A. *Nonparametric Econometrics*; Cambridge University Press: Cambridge, UK, 1999.
23. Rudebusch, G.D. Federal reserve interest rate, targeting rational expectations, and the term structure. *J. Monet. Econ.* **1995**, *35*, 245–274.
24. Anderson, T.G.; Lund, J. Estimating continuous-time stochastic volatility models of the short-term interest rate. *J. Econometr.* **1997**, *77*, 343–377.
25. Shiller, R.J.; Campbell, J.Y.; Schoenholtz, K.L.; Weiss, L. Forward rates and future policy: Interpreting the term structure of interest rates. *Brook. Papers Econ. Activ.* **1983**, *1*, 173–223.
26. Rudebusch, G.D. Term structure evidence on interest rate smoothing and monetary policy inertia. *J. Monet. Econ.* **2002**, *49*, 1161–1187.
27. Sarno, L.; Thornton, D.L. The dynamic relationship between the federal funds rate and the Treasury bill rate: An empirical investigation. *J. Bank. Finan.* **2003**, *27*, 1079–1110.
28. Anderson, H.M. Transaction costs and non-linear adjustment towards equilibrium in the US treasury bill market. *Oxford Bull. Econ. Statist.* **1997**, *59*, 465–484.
29. Clarida, R.H.; Sarno, L.; Taylor, M.P.; Valenter, G. The role of asymmetries and regime shifts in the term structure of interest rates. *J. Bus.* **2006**, *79*, 1193–1224.
30. Balke, N.S.; Fomby, T.B. Threshold cointegration. *Int. Econ. Rev.* **1997**, *38*, 627–645.
31. White, H. A reality check for data snooping. *Econometrica* **2000**, *68*, 1097–1126.
32. Diebold, F.X.; Rudubusch, G. Forecasting output with the composite leading index: A real-time analysis. *J. Am. Statist. Assoc.* **1991**, *86*, 603–610.
33. Stock, J.; Watson, M. Why has U.S. inflation become harder to forecast? *J. Money Credit Bank.* **2007**, *39*, 3–33.
34. Hamilton, J. The daily market for Federal funds. *J. Polit. Econ.* **1996**, *104*, 26–56.
35. Chan, K.C.; Karolyi, G.A.; Longstaff, F.A.; Sanders, A.B. An empirical comparison of alternative models of the short-term interest rate. *J. Finan.* **1992**, *47*, 1209–1227.
36. Vasicek, O. An equilibrium characterization of the term structure. *J. Finan. Econ.* **1977**, *5*, 177–188.
37. Cox, J.C.; Ingersoll, J.E.; Ross, S.A. A theory of term structure of interest rates. *Econometrica* **1985**, *53*, 385–407.
38. Nummelin, E. *General Irreducible Markov Chains and Non-negative Operators*; Cambridge University Press: Cambridge, UK, 1984.
39. Kasahara, Y. Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *J. Math. Kyoto Univ.* **1984**, *24*, 521–538.