THREE ESSAYS ON NONLINEAR TIME SERIES ECONOMETRICS

By

Zhengfeng Guo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2011

Nashville, Tennessee

Approved:

Professor Mototsugu Shintani

Professor Yanqin Fan

Professor Tong Li

Professor Ronald Masulis

To my husband, Bingyu

and

To my parents, with love and gratitude.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude to many people for their guidance, support and patience over the last few years. Without their personal and practical support, I would never have been able to finish my dissertation.

First of all, my sincere thanks go to my advisor, Professor Mototsugu Shintani, whose supervision and encouragement helped me overcome many difficult situations throughout the course of completing my dissertation. His knowledge and commitment to the highest standards inspired and motivated me. He is one of those rare advisors every student dreams of working with.

I am also very thankful to my committee members, Professor Yanqin Fan, Professor Tong Li and Professor Ronald Masulis, for their valuable comments and suggestions that substantially improved the quality of my dissertation. A special thank you goes to Professor Yanqin Fan for enabling my enrollment at Vanderbilt University and the Noel Dissertation Fellowship for financial support.

I would also like to thank my parents, who supported me and encouraged me in every way imaginable.

Finally, I would like to thank my husband who was my biggest cheerleader and who stood by me through the good and bad times.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The past decade has witnessed an impressive development of nonlinear modeling in both theory and practice. Many nonlinear models have been proposed in the literature, such as the bilinear models, the threshold autoregressive models and the Markov Switching models, to name only a few. These models can capture such features as nonnormality, asymmetry, leptokurtosis and volatility clustering, that are beyond the scope of their linear counterparts. In this context, the identification, testing and specification of nonlinear models is of great concern.

In time series analysis literature, one major task is to investigate the structural relationship between the present and past observations. Thus, it has been of considerable interest to correctly choose the exact number of lagged values to be included when explaining the variability of the present observation. Many techniques have been applied in an attempt to answer this model specification issue.

The model identification is generally realized by fitting autoregressive models of successive orders within a certain range, computing the estimates of the lag selection criteria, and adopting the one with the minimum value. To estimate autoregressive models, both parametric and nonparametric approaches have been widely applied. The classical procedures, including the Akaike information criterion (AIC), the final prediction error (FPE), the Bayesian information criterion (BIC) and the Hannan and Quinn criterion (HQ) are based on the parametric specification of autoregressive models. Despite their simplicity and intuitive appeal, classical procedures suffer from two drawbacks. The first is the consistency problem. Only the latter two procedures are consistent in the sense of picking the true order of the model with probability one asymptotically. The second drawback is that a complete parametric specification can be too restrictive in an applied context. As is now well documented, misspecification of the parametric models would eventually lead to the invalidity of the classical lag selection criteria.

Compared to their classical counterparts, lag selection criteria based on nonparametric techniques, such as the nonparametric version of FPE and cross-validation, are quite flexible in the sense that they can be applied to both linear and nonlinear models, where as some of the classical ones may fail to detect the right lags. Another advantage of the nonparametric FPE and cross-validation lies in their consistency, which is lacking in most classical procedures. However, like most nonparametric approaches, the flexibility comes at the cost of the curse of dimensionality, which refers to the problem that the finite and asymptotic properties of the nonparametric procedures deteriorate quickly as the dimension of the regressors increases. This problem makes most of the complete nonparametric procedures, including the nonparametric version of FPE and cross validation, impractical in empirical research.

In the first chapter, we address the model identification issue for nonlinear additive models. We develop a nonparametric lag selection criterion based on the final prediction error. The immediate advantage of this approach is that little prior information on the model structure is assumed. It offers an effective alternative to lag selection procedures based on classical criteria such as the AIC, the BIC and the FPE.

The first chapter proposes a new approach that is consistent and free of the curse of dimensionality. This approach uses a similar idea as the nonparametric FPE but differs by imposing an additive structure in model specification. The appeal of the additive model is that the fitted model is free from restrictive parametric assumptions, just as any other nonparametric method. However, unlike most of its counterparts, the effects of individual covariates can be easily interpreted, regardless of their number. Fan and Yao (2003) show how additive models can be used to improve the predictions of multiperiod volatility of aggregate returns. More importantly, most of the classical time series models such as AR(p) model, ARMA(p, q) model, ARCH(p) model and GARCH(p, q) model, belong to the family of additive models.

Following the idea presented in Tjostheim and Auestad (1994b) and using the marginal integration method by Linton and Nielsen (1995), we derive the nonparametric version of the FPE for additive models and show the new criterion is consistent. As a result of the additive structure, this new approach circumvents that serious drawback of nonparametric techniques : the curse of dimensionality.

We implement a Monte-Carlo study to investigate the performance of our nonparametric lag selection criterion. We compare the performance of this new approach with that of existing ones for a wide range of processes, including linear and nonlinear processes. Our findings show that this new approach generally outperforms the existing ones for general autoregressive models.

The first chapter proposes a nonparametric lag selection criterion that is applicable to additive processes. In contrast, the classical approaches may fail completely because of model specification issue and the nonparametric lag selection procedures may fail when too many lags are included. The simulation results show that this new method generally outperforms the nonparametric FPE for a wide range of additive models.

An individual time series, such as GDP, a stock price index, prices of commodities can wander extensively and yet some pair of series may move together due to underlying long-run equilibrium relations. Examples might be expenditures and household income, short and long term interest rates and prices of the same commodity in different markets. The concept of cointegration, defined first by Granger (1981, 1983) has been widely used to capture this type of long run linear relationship among two or more unit root processes. Cointegration allows for the estimation of structural parameters without the need to impose exogeneity assumption. Additionally, it plays an important role in evaluating the veracity of propositions in economics theories, for example, the theory of purchasing power parity.

Much of the empirical and theoretical work on cointegration has been conducted in the context of parametric models, including the well-known likelihood ratio test by Johansen (1991). More recently, Shintani (2001) and Cheng and Phillips (2008) have considered the use of model-free cointegration test and selection procedure for identifying the cointegrating rank of a process. The latter approaches are more flexible as compared to the likelihood ratio test in the sense that they do not require the specification of the structure of the data-generating process.

Cointegration is used to measure the linear relation among nonstationary variables only. Quite a few macroeconomic time series that are not unit root processes may behave like cointegrated processes in that the series move together over time. Furthermore, when it comes to variables with deterministic trends, the current econometric practices generally assume simple linear functions of time in all variables. However, there is empirical evidence that some long macroeconomic time series are more in accord with a nonlinear trend-stationary process. The concept of cotrending was introduced into the literature to fulfill a role similar to that of cointegration in a trend-stationary system. Cotrending is the phenomenon that one or more linear combinations of the time series would eliminate the deterministic trend. When we deal with variables with nonlinear trends or with structural breaks, many approaches fail to detect the correct cointegrating and cotrending rank.

The second chapter suggests a simple cotrending rank selection criterion. This approach utilizes the information contained in the data sample covariance matrix to investigate the cotrending relations. Specifically, when a system contains stationary elements, nonstationary elements and time trends, the sample covariance matrix would diverge at different rates. By exploring these different rates of divergence, we propose a novel cotrending rank selection criterion. The selection procedure is shown to be consistent in the sense of picking the true cointegrating and cotrending rank of the model with probability one when the sample size is large.

This chapter contributes to the literature in many aspects. First of all, it gives rise to a new tool that is model-free and simple to implement. Therefore, it is not necessary to build a complete model and is often desirable in case of nonlinear trends and structural breaks. Second, determining the cointegrating and cotrending ranks via a purely data-driven selection criterion has certain attractions over hypothesis testing procedures. One advantage is that it is not necessary to obtain the asymptotic distributions and the critical values. Lastly, this novel selection procedure has not been previously considered in the presence of both stochastic and deterministic trends..

Modeling and forecasting volatility (the covariance structure of asset returns) is important in the sense that volatility is considered as a measure of risk, and investors demand a premium for investing in risky assets. First introduced by Engle(1982), models of Autoregressive Conditional Heteroskedasticity (ARCH) and their extensions Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models form the most popular way to model the dependency of the conditional second moment and yield relatively accurate forecasts.

Recently, however, growing evidence suggests an asymmetric response of the conditional variance to positive and negative news. Several extensions of GARCH models aim at accommodating this asymmetry in the response. These include the GJR-GARCH model, the asymmetric GARCH models and Threshold GARCH models. Most of these models, however, are characterized by the existence of two regimes of volatilities: low volatility and high volatility, which are triggered by positive and negative shocks, respectively. Gonzalez-Rivera (1998) introduced the smooth transition GARCH model, where the regime transition function is continuous and exhibit a continuum of regimes. The smooth transition GARCH model generalizes the modeling of asymmetry in variance and nests a threshold specification.

The third chapter develops a smooth transition GARCH model with an asymmetric transition function, which allows an asymmetric response of volatility to the size and sign of shocks, and an asymmetric transition dynamic for positive and negative shocks. This specification encompasses a wide array of GARCH specifications and can yield much better fits to actual financial time series.

To test for asymmetry, we propose two testing procedures. One is based on the linearization of the transition function. The other is a supremum LM test with unidentified parameters under the null. From our simulation experiments, we find that the LM test is preferred, because it only requires the estimation of the model under the null. We apply our model to the empirical financial data: the NASDAQ index and the individual daily stock returns of IBM. The empirical evidence shows that our model outperforms many existing GARCH specifications.

Research into the time series properties of conditional second moments of returns has been an active area of empirical research. The chapter contributes to the rich literature by proposing a new more general model to capture the asymmetric effect of bad news and good news on the conditional second moments. The new specification can be more fitted to the financial time series than existing GARCH models.

CHAPTER II

NONPARAMETRIC LAG SELECTION FOR NONLINEAR ADDITIVE AUTOREGRESSIVE MODELS

Introduction

The final prediction error (FPE) criterion, as an alternative to the cross-validation criterion, provides a consistent lag selection procedure for the kernel-based nonparametric estimation of nonlinear autoregressive (AR) models. Under very general assumptions on the autoregressive function and on the function of conditional heteroskedasticity, Tschernig and Yang (2000) prove the consistency of the combinations of the lagged variables obtained by minimizing the nonparametric version of the FPE originally proposed by Tjøstheim and Auestad (1994).[1] In particular, using an optimal bandwidth that minimizes the asymptotic FPE, both probabilities of including too many lags (overfit) and missing some lags (underfit) approach zero as the sample size increases. Unfortunately, despite the desirable asymptotic property of the FPE procedure, Tschernig and Yang (2000) also point out its poor finite sample performance, namely, the fact that overfitting models are selected too often when the sample size is small. For this reason, they recommend making a multiplicative correction to the FPE in order to avoid overfitting. The possibility of developing a more effective lag selection procedure based on the FPE designed for special multidimensional models, such as additive models, is mentioned in section 3 of Tjøstheim and Auestad (1994) and in the conclusion of Tschernig and Yang (2000). However, the formal investigation for such an additive FPE procedure has not yet been conducted.

[1] In contrast, the lag selection using the original FPE of Akaike (1969) is not consistent for parametric time series models.

In this chapter, we introduce additivity in the autoregressive function and investigate the effect of placing such a simplifying structure on the properties of the FPE-like lag selection. We provide the conditions required for the consistency of the lag selection procedure using a variant of the FPE designed for the additive nonparametric regression. In contrast to the unrestricted FPE procedure without the additivity assumption, our additive nonparametric FPE-like procedure turns out to perform reasonably well in small samples. Indeed, the probability of overfitting becomes much smaller than in the unrestricted case so that there is no need for the finite sample correction used by Tschernig and Yang (2000).

The advantage of an additivity assumption in the nonparametric lag selection found in this chapter also suggests the potential for a similar procedure designed for more complex additive models, such as generalised additive models and generalised structured models (Mammen and Nielsen, 2003).

The remaining of the chapter is organized as follows. In section 2, we introduce the model and discuss the asymptotic properties of the procedure. Its finite sample performance is evaluated using Monte-Carlo simulation in section 3. All the proofs are provided in the Appendix A.

The nonparametric FPE for additive models

We consider the problem of selecting the combination of lags $S = \{i_1, i_2, ..., i_m\}$, where $i_j > i_k$ for $j > k$, in an additive AR model of the form,

$$Y_t = c + \sum_{i \in S} f_i(Y_{t-i}) + \sigma(\mathbf{X}_t)\xi_t$$

for $t = 1, ..., n$, where $\mathbf{X}_t = (Y_{t-i_1}, Y_{t-i_2}, ..., Y_{t-i_m})'$, and $\xi_t \sim$i.i.d.$(0, 1)$ with a finite fourth moment. In a typical nonparametric lag selection problem without an additive structure, the largest lag i_m in the model can be very large but the total number of lagged Y_t's, denoted by m, is required to be a small number due to the curse of dimensionality. Here, we do not need such a restriction on m ($\leq i_m$) since the convergence rate of additive regression estimators we employ does not depend on the dimension of the model. Below, we have a set of assumptions that are similar to the ones used in Tschernig and Yang (2000) except for the last assumption on the additive nonparametric regression estimator.

Assumptions A.

(A1) For some integer $M \geq i_m$, the vector process $X_{M,t} = (Y_{t-1}, Y_{t-2}, ..., Y_{t-M})'$ is strictly stationary and β-mixing with $\beta(n) \leq c_0 n^{-(2+\delta)/\delta}$ for some $\delta > 0$ and $c_0 > 0$.

(A2) The stationary distribution of the process $X_{M,t}$ has a continuous differentiable density $\mu(x_M)$.

(A3) The autoregression function $f_i(\cdot)$ for $i \in S$ is twice continuously differentiable while $\sigma(\cdot)$ is continuous and positive on the support of $\mu(\cdot)$.

(A4) The support of the weight function $w(\cdot)$ is compact with nonempty interior. The function $w(\cdot)$ is continuous, nonnegative and $\mu(x_M) > 0$ for $x_M \in$supp(w).

(A5) The kernel-based nonparametric additive regression estimator $\widehat{f}_i(x_i)$ for $i \in S$ converges to $f_i(x_i)$ at the one-dimensional rate of $\sqrt{nh}$ with its bias given by $r_i(x_i)\sigma_K^2 h^2/2$ where h is the bandwidth satisfying $h \to 0$, $nh \to \infty$ as $n \to \infty$, $\sigma_K^2 = \int K(u)u^2 du$, $K(\cdot)$ is a symmetric second order kernel function and $r_i(x_i)$ is positive and finite.

In estimating the additive AR model, we employ a kernel regression approach combined with the marginal integration proposed by Linton and Nielsen (1995): $\widehat{f}_i(x_i) = \int \widehat{f}(x) dQ(x_{-i}) - \widehat{c}$ where $\widehat{f}(x)$ is a nonparametric estimator of the nonlinear AR function without an additivity assumption, $\widehat{c}$ is an estimator of c such as $n^{-1}\sum_{t=1}^{n} Y_t$, x_{-i} represents all the elements in $x = (x_i, x_{-i})$ excluding x_i, and Q is a weighting function satisfying

$\int dQ(u) = \int q(u)du = 1$. Under some conditions, Assumption (A5) is satisfied with

$$r_i(x_i) = \int Tr\{\bigtriangledown^2 \sum f_i(x_i)\} q(x_{-i}) dx_{-i}$$

when the local linear estimator is used for $\widehat{f}(x)$, or with

$$r_i(x_i) = \int [Tr\{\bigtriangledown^2 \sum f_i(x_i)\} + 2 \bigtriangledown^T \mu(x) \bigtriangledown \sum f_i(x_i)/\mu(x)] q(x_{-i}) dx_{-i}$$

when the local constant (Nadaraya-Watson) estimator is used for $\widehat{f}(x)$. We focus on the marginal integration estimator instead of using the backfitting estimator for additive models because the former is computationally simple and its statistical properties are well-established.

By using an analogy to the asymptotic FPE of Tschernig and Yang (2000), the second term in formula (7) of Tjøstheim and Auestad (1994b) is decomposed as follows

$$\begin{aligned} & E\left[\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S} \widehat{f}_i(Y_{t-i})\right]^2 w(\mathbf{X}_{M,.t}) \\ = \quad & E\left[\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S} E\widehat{f}_i(Y_{t-i}) + \sum_{i\in S} E\widehat{f}_i(Y_{t-i}) - \sum_{i\in S'} \widehat{f}_i(Y_{t-i})\right]^2 w(\mathbf{X}_{M,.t}) \\ = \quad & E\left[(I' + II')^2 w(X_{M,t})\right] \end{aligned}$$

Using the results from Linton and Nielsen (1995) and setting the bandwidths in all the dimensions to h, we have

$$E\left[(II')^2 w(\mathbf{X}_{M,.t})\right] = h^4 \frac{\sigma_K^4}{4} \int \left[\sum_{i\in S} r_i(x_i)\right]^2 w(x_M)\mu(x_M) dx_M$$

By using the same argument as in Tschernig and Yang (2000), the cross term $E\left[I' II' w(\mathbf{X}_{M,.t})\right]$

is negligible. Now we derive $E\left[(I^{'})^2 w(\mathbf{X}_{M,.t})\right]$. Since

$$\begin{aligned}
\widehat{f}_i(Y_{t-i}) - E\widehat{f}_i(Y_{t-i}) &= \int \left[n^{-1} \sum_s \frac{K_h(Y_{t-i}-y_{i,s}) K_h(\mathbf{X}_{-i,t}-x_{-i,s})\sigma(X_t)\epsilon_t}{\widehat{\mu}(\mathbf{X}_t)} \mu(x_{-i}) \right] dx_{-i} \\
&= n^{-1} \sum_s \frac{K_h(Y_{t-i}-y_{i,s}) q(x_{-i})\sigma(X_t)\epsilon_t}{\widehat{\mu}(\mathbf{X}_t)},
\end{aligned}$$

Then

$$\begin{aligned}
& E\left[(I^{'})^2 w(\mathbf{X}_{M,.t})\right] \\
\approx & \int \left[n^{-1} \sum_s \frac{K_h(y-x) q(x_{-i})\sigma(x)\epsilon_s}{\mu(x)} \right]^2 \mu(y)\mu(x_M) w(x_M) dx_M.
\end{aligned}$$

which becomes

$$\frac{1}{n} \int \left\{ \frac{K_h(y-x) q(x_{-i})\sigma(x)}{\mu(x)} \right\}^2 \mu(y)\mu(x_M) w(x_M) dx_M \tag{II.1}$$

where the cross terms are left out by a U-statistic argument as in Tjøstheim and Auestad (1994b). The precedent equation can be re-written as

$$\frac{1}{n}\|K\|_2^2 \int \frac{\sigma^2(x)}{\mu(x)} q^2(x_{-i})\mu(x_M) w(x_M) dx_M$$

From our additivity assumption, $E\left[(I^{'})^2 w(\mathbf{X}_{M,.t})\right]$ converges to $\frac{1}{nh}\|K\|_2^2 B$.

Next, we introduce

$$AFPE = A + \frac{1}{nh}\|K\|_2^2 B + h^4 \frac{\sigma_K^4}{4} C \tag{II.2}$$

with

$$\begin{aligned}
A &= \int \sigma^2(x_M) w(x_M)\mu(x_M) dx_M, \\
B &= \int \frac{\sigma^2(x)}{\mu(x)} \left\{ \sum_{i \in S} q^2(x_{-i})\mu(x_i) \right\} w(x_M)\mu(x_M) dx_M, \text{ and} \\
C &= \int \left[\sum_{i \in S} r_i(x_i) \right]^2 w(x_M)\mu(x_M) dx_M
\end{aligned}$$

where $\|K\|_2^2 = \int K^2(u)du$, $\sigma_K^2 = \int K(u)u^2 du$ and $r_i(x_i)$ is the term appears in the asymptotic bias $r_i(x_i)\sigma_K^2 h^2/2$ of the estimator $\widehat{f}_i(x_i)$. The optimal bandwidth, h_{opt}, minimizes (II.2) is given by

$$h_{opt} = \left\{\|K\|_2^2 \sigma_K^{-4} B C^{-1}\right\}^{1/5} n^{-1/5}.$$

In principle, we can replace employ other estimators, e.g., the smooth backfitting estimator which is a useful practical variant of the classical backfitting estimator (see Mammen, Linton and Nielsen, 1999, and Nielsen and Sperlich, 2005). Following the same argument, we obtain

$$E\left[(II')^2 w(\mathbf{X}_{M,.t})\right] = h^4 c_h^4 \sigma_K^4 \int \left[\sum_{i\in S} \beta_i(x_i)\right]^2 w(x_M)\mu(x_M)dx_M$$

with c_h being the limit of $n^{1/5}h$ and $\beta_i(x_i) = \sum_{i\in S}\left[\frac{\nabla f_i(x_i)}{\mu(x)}\frac{\partial \mu(x)}{\partial x_i} + \frac{1}{2}\nabla^2 f_i(x_i)\right]$ when the local linear estimator is used for $\widehat{f}(x)$, or with $\beta_i(x_i) = \nabla^2 f_i(x_i) - \int \nabla^2 f_i(x_i)\mu_j(x_j)dx_j$ when the local constant (Nadaraya-Watson) estimator is used for $\widehat{f}(x)$.

Similarly, we can show that $E\left[(I')^2 w(\mathbf{X}_{M,.t})\right]$ converges to

$$\frac{1}{nh}\|K\|_2^2 \int \left\{\sum_{i\in S} \frac{\sigma_i^2(x_i)}{c_h \mu(x_i)}\right\} w(x_M)\mu(x_M)dx_M,$$

where $\sigma_i^2(x_i) = var(Y - f(x)|X_i = x_i)$. Therefore, we can define $AFPE$ with modified B and C if we employ the smooth backfitting estimator.

Our criterion for additive AR models motivated by the unrestricted FPE takes the form

$$\widehat{FPE}(S) = \widehat{A} + \frac{1}{nh_{opt}^{(m-1)\eta+1}} 2K(0)\widehat{B},$$

where $\eta \in [0,1]$,

$$\widehat{A} = n^{-1}\sum_{t=1}^{n}\left(Y_t - \sum_{i\in S}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t})$$

and

$$\widehat{B} = n^{-1}\sum_{t=1}^{n}\frac{(Y_t - \sum_{i\in S}\widehat{f}_i(Y_{t-i}))^2}{\widehat{\mu}(\mathbf{X}_t)}\left\{\sum_{i\in S} q(\mathbf{X}_{-i,t})\right\} w(\mathbf{X}_{M,.t}).$$

The first term in $\widehat{FPE}(S)$ is analogous to the measure of regression fit in traditional information criteria for the model selection, while the second term can be a penalty for an increased dimension m, depending on a tuning parameter η.[2] We follow Tschernig and Yang (2000) and focus on the case when the optimal bandwidth h_{opt} is used for $\widehat{f}_i(x_i)$ in $\widehat{A}$, but any bandwidth of order $n^{-1/5}$ can be used for $\widehat{f}_i(x_i)$ in $\widehat{B}$. We select the subset $\widehat{S} = \{\widehat{i}_1, \widehat{i}_2, ..., \widehat{i}_{\widehat{m}}\} \subseteq \{1,2,...,M\}$ which minimizes $\widehat{FPE}(S)$ among all possible combinations of $\{1,2,...,M\}$. The selected $\widehat{S} = S'$ overfits if $S' \supseteq S$ and $S' \neq S$ and underfits if it does not overfit and $S' \neq S$. The lag selection procedure is consistent if the probability of $\widehat{S} = S$ approaches unity as $n \to \infty$.

Theorem 1 *Under Assumption (A1)-(A5) and $\eta \in (0,1]$, as $n \to \infty$,*

$$\frac{\widehat{FPE}(S') - A}{\widehat{FPE}(S) - A} \to +\infty,$$

for any overfitting combination $S' = \{i'_1, i'_2, ..., i'_{m'}\}$.

The overfitting $\widehat{FPE}(S')$ asymptotically becomes larger than the correctly specified $\widehat{FPE}(S)$ because the penalty term of the former converges at a rate slower than the latter as long as $\eta > 0$. Note that h'_{opt} used for $\widehat{FPE}(S')$ differs from h_{opt} because B and C are replaced by $B' = \int \frac{\sigma^2(x)}{\mu(x')}\left\{\sum_{i\in S'} q^2(x_{-i})\mu(x_i)\right\} w(x_M)\mu(x_M)dx_M$ and

[2] When $\eta = 1$, the rate of the penalty term becomes same as that of the unrestricted nonparametric FPE. When $\eta = 0$, the rate becomes the one discussed in Tjøstheim and Auestad (1994) for the additive case.

$C' = \int \left[\sum_{i \in S'} r_i'(x_i)\right]^2 w(x_M)\mu(x_M)dx_M$, respectively, where x' represents a larger lag vector nesting x. Unlike the unrestricted FPE, however, the convergence rates of two bandwidths are the same even if the dimensions of the regressors are different. This is the reason why $\eta = 0$ is not desirable in excluding overfitting models.

To investigate the underfitting case, we focus on the case of a proper subvector x' of the true lag vector $x = (x', x'')$ for notational simplicity. The following assumption corresponds to the assumption A8 of Tschernig and Yang (2000).

(A6) The weighted squared projection error, defined as

$$c^2 = \int \left[\sum_{i \in S} f_i(x_i)\right]^2 w(x_M)\mu(x_M)dx_M - \int E^2\{\sum_{i \in S} f_i(x_i)|x'\} w(x_M)\mu(x_M)dx_M,$$

is positive.

Theorem 2 *Under Assumption (A1)-(A6) and $\eta \in [0,1]$, as $n \to \infty$,*

$$\widehat{FPE}(S') - \widehat{FPE}(S) \xrightarrow{p} c^2 > 0,$$

for any underfitting combination $S' = \{i_1', i_2', ..., i_{m'}'\}$.

A combination of Theorems 2.1 and 2.2 yields the following consistency result.

Theorem 3 *Under Assumption (A1)-(A6) and $\eta \in (0,1]$, as $n \to \infty$,*

$$P\left[\widehat{S} = S\right] \to 1.$$

Remarks

1. If $\eta > 0$, the probability of the procedure failing to completely identify the correct model diminishes as the sample size increases. Similar to the unrestricted FPE procedure, the consistency of the additive FPE-like procedure holds for both local linear and local constant estimators.

2. If $\eta = 0$, our FPE-like criterion is asymptotically equivalent to the asymptotic FPE. Therefore, the optimal bandwidth h_{opt} can be consistently estimated by searching for the bandwidth which minimizes $\widehat{FPE}(S)$ using $\eta = 0$. Once h_{opt} is estimated, the same bandwidth can be used for other FPEs with any $\eta \in (0, 1]$ for the purpose of consistent lag selection.

3. While both the unrestricted FPE procedure and our procedure for additive models are consistent, the latter procedure can be expected to perform better in the finite sample. The unrestricted nonparametric FPE procedure of Tjøstheim and Auestad (1994) performs poorly, mainly because overfitted models are selected too often. A comparison of the rates of divergence of the ratios in our Theorem 2.1 and Corollary 3.1 of Tschernig and Yang (2000) shows that the former is faster than the latter as long as $(m+4)(m'+4)\eta > 5$. This suggests that the probability of selecting overfitting models based on our procedure approaches zero faster for many combinations of lags.

Monte-Carlo simulation

We conduct a Monte-Carlo simulation to investigate the finite sample properties of our FPE-like procedure designed for the additive models. The performance of the proposed procedure is evaluated using both local linear and local constant estimators for various values of $\eta(= 0, 0.1, 0.5, 1.0)$ and is compared to the performance of the unrestricted FPE procedure. The artificial series are generated from 11 additive AR models given in Table 1, where ξ_t's are independent and identically distributed $N(0, 1)$ random variables. The processes are a collection of linear and nonlinear additive models previously used in similar simulation studies. The first three linear models (AR1-AR3) and the following three nonlinear models (NLAR1-NLAR3) are used in Tschernig and Yang (2000). The next two models (NLAR4 and NLAR5) are taken from Chen, Liu and Tsay (1995), and the three

Table 1. Data generating processes used in simulation

Model	Function
AR1	$Y_t = 0.5Y_{t-1} + 0.4Y_{t-2} + \xi_t$
AR2	$Y_t = -0.5Y_{t-1} + 0.4Y_{t-2} + \xi_t$
AR3	$Y_t = -0.5Y_{t-6} + 0.5Y_{t-10} + \xi_t$
NLAR1	$Y_t = -0.4(3 - Y_{t-1}^2)/(1 + Y_{t-1}^2) + 0.6\left\{3 - (Y_{t-2} - 0.5)^3\right\} / \left\{1 + (Y_{t-1} - 0.5)^4\right\} + \xi_t$
NLAR2	$Y_t = \left\{0.4 - 2\exp(-50Y_{t-6}^2)\right\} Y_{t-6} + \left\{0.5 - 0.5\exp(-50Y_{t-10}^2)\right\} Y_{t-10} + \xi_t$
NLAR3	$Y_t = \left\{0.4 - 2\cos(40Y_{t-6})\exp(-30Y_{t-6}^2)\right\} Y_{t-6} + \left\{0.5 - 0.5\exp(-50Y_{t-10}^2)\right\} Y_{t-10} + \xi_t$
NLAR4	$Y_t = 2\exp(-0.1Y_{t-1}^2)Y_{t-1} - \exp(-0.1Y_{t-2}^2)Y_{t-2} + \xi_t$
NLAR5	$Y_t = -2Y_{t-1}I(Y_{t-1} \leq 0) + 0.4Y_{t-1}I(Y_{t-1} > 0) + \xi_t$
NLAR6	$Y_t = 0.8\log(1 + 3Y_{t-1}^2) - 0.6\log(1 + 3Y_{t-3}^2) + \xi_t$
NLAR7	$Y_t = 1.5\sin((\pi/2)Y_{t-2}) - 1.0\sin((\pi/2)Y_{t-3}) + \xi_t$
NLAR8	$Y_t = (0.5 - 1.1\exp(-50Y_{t-1}^2))Y_{t-1} + (0.3 - 0.5\exp(-50Y_{t-3}^2))Y_{t-3} + \xi_t$

Notes: $I(x)$ is an indicator which takes a value 1 if x holds and 0 otherwise. $\xi_t \sim N(0,1).iid$

other models (NLAR6-NLAR8) are taken from Chen and Tsay (1993).

For each process, the first 120 observations of 220 realizations are discarded to generate a series of size 100 used in the nonparametric regression. Additive AR models are estimated using the marginal integration method applied to both the local constant and local linear estimators with a Gaussian kernel. In particular, we follow Sperlich, Linton and Härdle (1999) and use the empirical distribution function $Q_n(x_{-i})$ as the weighting function to obtain $\widehat{f}_i(x_i)$.[3] To find a combination S which minimizes $\widehat{FPE}(S)$, we employ the algorithm explained in Tjøstheim and Auestad (1994) with a maximum possible total number of lags set to $M = 13$. For the choice of bandwidth we employ the procedure used by Tschernig and Yang (2000). In particular, $h_m = \sqrt{\widehat{var}(Y_t)}\,\{4/(m+2)\}^{1/(m+4)}\, n^{-1/(m+4)}$ is used for $\widehat{\mu}(\mathbf{X}_t)$ in $\widehat{B}$ and h_1 is used for $\widehat{f}_i(Y_{t-i})$ in $\widehat{B}$. For the estimation of the optimal bandwidth h_{opt} used in $f_i(Y_{t-i})$ in $\widehat{A}$, we find a value which minimizes $\widehat{FPE}(S)$ by searching over the interval $[0.2h_1, 2h_1]$. The same optimal bandwidth, for each of the corresponding combination of lags, is also used in the FPEs with $\eta \neq 0$. We replicate each experiment 100 times and report the empirical frequencies of selecting the correct model along with

[3] We substitute the smoothed empirical density based on a Gaussian kernel, for the density of weight function, $q(\mathbf{X}_{-i,t})$, required in $\widehat{B}$.

overfitting frequencies. Tables (3) and (3) show the results for the local linear and local constant estimators, respectively.

The results of the simulation are summarized as follows. First, for almost all cases, frequencies of selecting correct combinations of lags based on the additive FPE-like procedure are higher than those based on the unrestricted FPE procedure. The only exception is the NLAR1 process for which both FPEs are performing very poorly. Thus, in general, gains from knowing the additive structure somewhat depend on the data generating process. Second, the FPE procedures work much better when the local constant estimator is used. This is true for both the unrestricted FPE and the FPE for additive models. Third, there is a significant reduction in the frequencies of overfits when the FPE for additive models is used, compared to the unrestricted FPE. This is consistent with our theoretical prediction. Finally, the performance of the additive FPE is not very sensitive to the choice of η for a wide range of values. However, setting $\eta = 1$ is not recommended for most of the cases. It is interesting to note that $\eta = 0$ often shows the best finite sample properties despite the fact that such a choice does not provide a consistent selection procedure. When the local constant estimator is employed, $\eta = 0.1$ or 0.5 provides the best result in many cases.

Conclusion

The possibility of using the FPE criterion in the lag selection of additive AR models has been previously discussed in the literature, but no formal proof on its asymptotic properties was available. We have shown that the FPE criterion designed for the additive model provides the consistent lag selection procedure under very general conditions. In addition, simulation results suggest the effectiveness of the additive FPE procedure in finite samples. Unlike the unrestricted FPE, the finite sample correction to reduce overfits might not be necessary. Our results are also in line with good finite sample properties of traditional information criteria for the nonlinear additive spline estimation recently reported by Huang and Yang (2004).

Finally, in this chapter, we focus on marginal integration because it does not involve the iterative computation. However, we note that better finite sample properties of the backfitting method over marginal integration have been often reported in simulation studies (e.g., Sperlich, Linton and Härdle, 1999, and Martins-Filho and Yang, 2007). The performance of our procedure based on other estimators, such as the smooth backfitting estimator, remains to be investigated in future work.

Table 2. Frequencies of selecting correct lags using a local linear estimator

	Unrestricted		FPE for additive models							
Model	FPE		$\eta = 0$		0.1		0.5		1.0	
LAR1	0	(89)	16	(29)	15	(32)	10	(48)	5	(56)
LAR2	0	(87)	18	(30)	17	(34)	8	(60)	5	(70)
LAR3	1	(98)	77	(19)	65	(31)	19	(79)	17	(82)
NLAR1	13	(73)	2	(38)	3	(43)	3	(53)	2	(44)
NLAR2	4	(84)	43	(42)	41	(46)	32	(58)	24	(66)
NLAR3	1	(84)	45	(34)	41	(40)	30	(56)	22	(65)
NLAR4	0	(99)	24	(50)	20	(55)	12	(65)	11	(72)
NLAR5	6	(94)	38	(47)	34	(53)	16	(78)	9	(84)
NLAR6	1	(53)	29	(28)	27	(34)	17	(47)	18	(46)
NLAR7	59	(20)	82	(13)	81	(14)	76	(12)	58	(12)
NLAR8	14	(84)	56	(41)	54	(42)	34	(62)	20	(73)

Notes: Frequencies of selecting the correct specification are computed from 100 simulation runs. Numbers in parentheses are frequencies of overfitting.

Table 3. Frequencies of selecting correct lags using a local constant estimator

	Unrestricted		FPE for additive models							
Model	FPE		$\eta = 0$		0.1		0.5		1.0	
LAR1	3	(91)	42	(2)	46	(2)	49	(3)	36	(19)
LAR2	10	(83)	49	(2)	52	(2)	48	(10)	29	(34)
LAR3	24	(75)	93	(4)	93	(5)	68	(30)	43	(57)
NLAR1	48	(7)	60	(4)	33	(13)	35	(2)	21	(3)
NLAR2	17	(77)	71	(13)	78	(8)	62	(22)	39	(41)
NLAR3	12	(81)	63	(6)	63	(7)	58	(15)	31	(40)
NLAR4	26	(71)	69	(3)	52	(7)	46	(6)	33	(19)
NLAR5	53	(9)	79	(0)	79	(0)	73	(0)	50	(1)
NLAR6	4	(59)	40	(6)	42	(5)	44	(10)	32	(18)
NLAR7	37	(0)	98	(0)	99	(0)	99	(0)	69	(0)
NLAR8	70	(11)	80	(9)	85	(8)	88	(0)	67	(1)

Notes: Frequencies of selecting the correct specification are computed from 100 simulation runs. Numbers in parentheses are frequencies of overfitting.

CHAPTER III

CONSISTENT COTRENDING RANK SELECTION WHEN BOTH STOCHASTIC AND NONLINEAR DETERMINISTIC TRENDS ARE PRESENT

Introduction

For decades, one of the most important issues in the analysis of macroeconomic time series has been how to incorporate a trend. Two popular approaches that have often been employed in the literature are (i) to consider a stochastic trend, with or without a linear deterministic trend, such as the one suggested in Nelson and Plosser (1982), and (ii) to consider a nonlinear deterministic trend such as the one with trend breaks considered in Perron (1989, 1997). Cointegration, introduced by Engle and Granger (1987), is a useful concept in understanding the nature of comovement among variables based on the first approach. In cointegration analysis, the cointegrating rank, defined as the number of linearly independent cointegrating vectors, provides valuable information regarding the trending structure of a multivariate system with stochastic trends. Several model-free consistent cointegrating rank selection procedures have been developed in the literature. Analogous to cointegration analysis is the analysis of comovement based on the second approach, namely, the nonlinear deterministic trend. The cotrend analyses of Bierens (2000), Hatanaka (2000) and Hatanaka and Yamada (2003) lie along this line of research. However, a consistent selection procedure of the cotrending rank, defined similarly as the cointegrating rank with a stochastic trend replaced by a nonlinear deterministic trend, has not yet been developed.

This chapter proposes a model-free consistent cotrending rank selection procedure when both stochastic and nonlinear deterministic trends are present in a multivariate system. Consistency here refers to the property that the probability of selecting the wrong cotrending rank approaches zero as sample size tends to infinity. Our procedure selects the cotrending rank by minimizing the von Neumann criterion, similar to the one used by Shintani (2001) and Harris and Poskitt (2004) in their analyses of cointegration. This approach exploits the fact that identification of cotrending rank can be interpreted as identification among three groups of eigenvalues of the generalized von Neumann ratio. Using this property of the von Neumann criterion, we propose two types of cotrending rank selection procedures that are (i) invariant to linear transformations of the data; (ii) robust to model misspecification; and (iii) valid not only with a break in the trend, but also with a broader class of nonlinear trend functions. The simulation results also suggest that our cotrending rank selection procedures perform well in finite samples.

Our analysis is closely related to that of Harris and Poskitt (2004) and Cheng and Phillips (2009), who propose consistent cointegrating rank procedures that do not require a parametric vector autoregressive model of cointegration such as the one in Johansen (1991). While we provide some examples of nonlinear trend functions, including trend breaks and smooth transition trend models, our cotrending rank selection procedure does not require the parametric specification of the trend function, or the parametric specification of serial dependence structure. Thus, our approach generalizes the results of Harris and Poskitt (2004) and Cheng and Phillips (2009) in the sense that it allows both common stochastic trends and common deterministic trends. Consequently, we can also use our procedure to determine the cointegrating rank in the absence of nonlinear deterministic trends. To illustrate this feature, we include both cointegrated and cotrended cases in our simulation analysis.

As emphasized in Stock and Watson (1988), the cointegrated system can be interpreted as a factor model with a stochastic trend being a common factor. Thus, determining the cointegrating rank is identical to determining the number of common stochastic trends because the latter is the difference between the dimension of the system (number of variables) and the cointegrating rank.[1] In the presence of both stochastic and nonlinear deterministic trends, however, the number of common nonlinear deterministic trends does not correspond to the difference between the dimension and the cotrending rank. Because the number of common deterministic trends also contains valuable information about the trending structure, we introduce the notion of weak cotrending rank, so that the difference between the dimension and the weak cotrending rank becomes the number of common deterministic trends.

Our two alternative definitions of cotrend are natural consequence of the notion of a common feature introduced in Engle and Kozicki (1993). They define the common feature as "a feature that is present in each of a group of series but there exists a non-zero linear combination of the series that does not have the feature". When such a feature is a *broad class* of trends, namely, a mixture of both stochastic and deterministic trends, the definition of cotrend requires a linear combination which eliminates both types of trends at the same time. In contrast, when such a feature is the *dominant* trend, namely the deterministic trend only, a linear combination should eliminate the deterministic trend but not necessary the stochastic trend. Since the latter type of cotrend nests the former type, we distinguish the two by referring the latter type as a weaker version of the cotrending relationship. Our procedure can select both the cotrending rank and weak cotrending rank.

The remainder of this chapter is organized as follows. Section 2 introduces some key concepts in the system of common stochastic and deterministic trends. The main theoretical results are provided in section 3. Section 4 reports Monte Carlo simulation

[1] PANIC method proposed by Bai and Ng (2004) utilizes the consistent selection of the number of common stochastic trends in a very large dynamic factor system based on information criteria. See also Bai and Ng (2002) for the case of consistent selection of the number of stationary common factors.

results to show the finite sample performance of our procedures. In section 5, we apply our procedures to the Japanese money demand function. Section 6 concludes, and the technical proofs are presented in the Appendix B.

Motivation

Cotrending ranks

Our cotrend analysis begins with an assumption that all the variables contain deterministic trends. This presumption is similar to the case of traditional cointegration analysis which requires all the variables to follow I(1) processes so that at least one stochastic trend is present in each variable of interest. The following simple bivariate examples illustrate the motivation of our cotrend analysis. In the presence of deterministic trends, a pair of variables, $y_t = (y_{1t}, y_{2t})'$, can be decomposed as

$$\begin{aligned} y_{1t} &= d_{1t} + s_{1t}, \\ y_{2t} &= d_{2t} + s_{2t}, \end{aligned} \tag{III.1}$$

where $d_t = (d_{1t}, d_{2t})'$ represents a deterministic trend component and $s_t = (s_{1t}, s_{2t})'$ represents a stochastic component that can be either I(0) or I(1) process. Suppose a simple bivariate linear trend model given by

$$\begin{aligned} y_{1t} &= c_1 + \mu_1 t + \varepsilon_{1t}, \\ y_{2t} &= c_2 + \mu_2 t + \varepsilon_{2t}, \end{aligned}$$

where ε_{1t} and ε_{2t} are zero mean I(0) error terms, $\mu_1 \neq 0$ and $\mu_2 \neq 0$. Then, this model has a representation (III.1) with

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= \varepsilon_{1t}, \quad \text{and} \quad s_{2t} = \varepsilon_{2t}. \end{aligned} \tag{III.2}$$

According to the definition of Engle and Kozicki (1993), a feature is said to be common if a linear combination of the series fails to have the feature. Since the deterministic trend is the main feature of interest, two variables are cotrended if the trend is eliminated by taking a particular linear combination (see also Bierens, 2000, Hatanaka, 2000, and Hatanaka and Yamada, 2003). In the case of a linear deterministic trend in (III.2), there is a trivial cotrending relationship since the vector $(1, -\mu_1/\mu_2)$ can eliminate the trend. Likewise, if m variables are generated from a multivariate linear trend model, there are $m-1$ trivial cotrending relationships since there are $m-1$ linearly independent non-zero cotrending vectors.

In our analysis, stochastic trends can be either included or excluded. When stochastic trends are present, there will be two layers of potential cotrending relationships. For example, suppose a pair of variables are generated from two independent random-walk-with-drift processes:

$$\begin{aligned} y_{1t} &= \mu_1 + y_{1t-1} + \varepsilon_{1t}, \\ y_{2t} &= \mu_2 + y_{2t-1} + \varepsilon_{2t}, \end{aligned}$$

where ε_{1t} and ε_{2t} are zero mean iid error terms, $\mu_1 \neq 0$ and $\mu_2 \neq 0$. Then, the model has a representation (III.1) with

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= s_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{III.3}$$

so that s_{1t} and s_{2t} are I(1) processes, or stochastic trends. In this case, the vector $(1, -\mu_1/\mu_2)$ eliminates the linear deterministic trend, but no linear combination can eliminate the stochastic trend. However, since the dominant trend, namely the deterministic trend can still be eliminated, we refer to the vector $(1, -\mu_1/\mu_2)$ as a weak cotrending vector. In contrast, if (III.3) is replaced by

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= (\mu_2/\mu_1) s_{1t} + \varepsilon_{2t}, \end{aligned} \tag{III.4}$$

the weak cotrending vector $(1, -\mu_1/\mu_2)$ eliminates not only the linear deterministic trend, but also the stochastic trend. Since both type of trends are eliminated by a single vector $(1, -\mu_1/\mu_2)$, we view such a case as the stronger version of the cotrending relationship.

In a system of m variables with both stochastic and deterministic trends, one of our goals is to identify the total number of linearly independent vectors that can eliminate both stochastic and deterministic trends at the same time. In this chapter, we refer to the number of such cotrending vectors as the cotrending rank and denote it by r_1. The cotrending rank can be any integer value in the range of $0 \leq r_1 < m$. In addition to r_1, we also introduce the weak cotrending rank (denoted by r_2) as the total number of linearly independent vectors that can eliminate the deterministic trend, regardless of whether such vectors can eliminate the stochastic trend at the same time. Since all the cotrending vectors are also weak cotrending vectors, r_2 should satisfy $r_1 \leq r_2 < m$. While it is not the stronger version of cotrending rank based on a broader notion of trends, the identification of r_2 is also important in the presence of both stochastic and deterministic trends, since $m - r_2$ in the m-variable-system corresponds to the total number of common deterministic trends. In the above example of $m = 2$, a vector $(1, -\mu_1/\mu_2)$ can eliminate the deterministic trend regardless of the values of μ_1 and μ_2. Thus, the weak cotrending rank r_2 of both models (III.3) and (III.4) is 1. However, the cotrending rank is 0 for model (III.3) and is 1 for model (III.4). In this chapter, we propose a simple procedure to identify both r_1 and r_2 in a system of m variables, in the presence of both stochastic and deterministic trends.

As discussed above, the elimination of the deterministic trend is of primary interest in our cotrend analysis. This differs from traditional cointegration analysis where elimination of the stochastic trend is its main interest even if a deterministic trend is included in the system. To see this point, consider another model with stochastic trends given by

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= s_{1t} + \varepsilon_{2t}. \end{aligned} \tag{III.5}$$

Here the cointegrating vector $(1, -1)$ can always eliminate the stochastic trend, but not the deterministic trend unless $\mu_1 = \mu_2$. For the purpose of distinguishing between (III.4) and (III.5) in cointegration analysis, Ogaki and Park (1997) introduced the notions of stochastic cointegration and deterministic cointegration. In their terminology, stochastic cointegration refers to the case in which only the stochastic trend is eliminated by the cointegrating vector. In contrast, deterministic cointegration refers to the case in which both stochastic and deterministic trends are eliminated by the same cointegrating vector. In our cotrend analysis, however, two models differ because the (strong) cotrending rank r_1 is 1 for (III.4) but 0 for (III.5).

Trend breaks and smooth transition trends

So far, we have only seen an obvious cotrending relationship with a linear trend for the purpose of introducing the notion of cotrending ranks. However, cotrend analysis becomes more meaningful when variables contain various forms of nonlinear deterministic trends so that the system can have more than one common deterministic trend. Here, we provide some examples of nonlinear trends to highlight the class of deterministic trends that are allowed in our consistent cotrending rank selection procedure.

As discussed in Mills (2003), many macroeconomic time series data, including GDP of the UK and Japan and stock prices in the U.S., violate the assumption of stable growth over the typical sample periods. A convenient approach to allow for multiple shifts in the growth rate, while maintaining the continuity of the trend function, is to consider a kinked trend, or a piece-wise linear trend structure in each segment of the whole sample period. When there are h time shifts in the (log) growth rate, the segmented linear trend can be written as

$$d_t^{KINK} = \mu_0 t + \sum_{i=1}^{h} \mu_i (t - T_i) 1[t > T_i],$$

where T_i is the trend break point and $1[x]$ is an indicator that takes the value of 1 if x is true and 0, otherwise. The segmented linear trend implies that the growth rate corresponds to μ_0, during the first subperiod $t < T_1$, and corresponds to $\mu_0 + \sum_{i=1}^{j} \mu_i$, in the remaining subperiods, $T_j \leq t < T_{j+1}$ for $j = 1, ..., h$.

Recall that in the preceding bivariate example with a linear trend, the deterministic trend terms d_{1t} and d_{2t} are by definition proportional to a common linear deterministic trend, say $d_t^{LIN} = t$, ignoring the constant. Therefore, we can always find at least one linear combination that eliminates the trend, and the cotrending relationship is trivial. However, if the linear trend functions in d_{1t} and d_{2t} are replaced by segmented trend functions, a linear combination can eliminate the deterministic trend if and only if (i) all the break points, T_i's, are the same and (ii) all the piece-wise trend slope coefficients, $\mu_i's$ are proportional between the two trend functions. If either of the two conditions fails to hold, the two nonlinear deterministic trends are linearly independent and no common deterministic trend exists. This fact also shows how our cotrend analysis differs from the cobreaking analysis of Hendry and Mizon (1998) and Clements and Hendry (1999). In the presence of a trend break, cobreaking is a necessary condition of cotrending, but not a sufficient condition.

Although the segmented trend function d_t^{KINK} imposes continuity, its first derivative is not continuous, suggesting an abrupt change of the growth rate at each break point. To allow for a gradual change in the growth rate, we may replace the indicator function in d_t^{KINK} with a smooth transition function. This substitution of the trend function leads to a smooth transition trend model. The smooth transition trend model was originally proposed by Bacon and Watt (1971) and has been discussed by Lin and Teräsvirta (1994) and Leybourne, Newbold and Vougas (1998). While there are many types of smooth transition trend functions, one most frequently used one is the logistic transition function given by

$$G(\gamma_i, T_i) = \frac{1}{1 + \exp(-\gamma_i(t - T_i))},$$

where γ_i (> 0) is the scaling parameter that controls the speed of transition, and T_i becomes the timing of the transition midpoint instead of the break point. The nonlinear deterministic trend component of a multiple-regime logistic smooth transition trend (LSTT) model takes the form of

$$d_t^{LST} = \mu_0 t + \sum_{i=1}^{h} \mu_i (t - T_i) G(\gamma_i, T_i).$$

It should be noted that, as γ_i approaches infinity, the logistic transition function $G(\gamma_i, T_i)$ approaches the indicator function $1[t > T_i]$. Thus, the deterministic trend d_t^{LST} nests both the kinked trend d_t^{KINK} and the linear trend d_t^{LIN} as special cases. Figure 1 shows the typical shape of kinked and smooth transition trends when $h = 1$. The former contains a one-time abrupt change in the first derivative, while the latter shows continuous change in the first derivative.

Both segmented and smooth transition type models of trend shift are allowed in our cotrending rank selection procedure. Furthermore, other types of nonlinear deterministic trend functions can be also included as long as they belong to a class of trend functions so that their order of magnitude is identical to that of a linear trend. Let $\{d_t^{KINK}\}_{t=1}^{T}$, and $\{d_t^{LST}\}_{t=1}^{T}$ be the deterministic sequences where $T_i = k_i T$, $0 < k_0 < k_1 < ... < k_h < 1$, and γ_i's are fixed. Then, both trend sequences have the same order of magnitude as the linear trend sequence $\{d_t^{LIN}\}_{t=1}^{T}$ in the sense that both $\sum_{t=1}^{T} d_t^{KINK} / \sum_{t=1}^{T} d_t^{LIN}$ and $\sum_{t=1}^{T} d_t^{LST} / \sum_{t=1}^{T} d_t^{LIN}$ approach a non-zero constant as T tends to infinity. Similarly, our analysis remains valid for any nonlinear deterministic trend sequence $\{d_t^{*}\}_{t=1}^{T}$ such that $\sum_{t=1}^{T} d_t^{*} / \sum_{t=1}^{T} d_t^{LIN}$ approaches some non-zero constant as T tends to infinity. In the following section, we propose a procedure to identify both r_1 and r_2 in a system of m variables, which is valid for any nonlinear deterministic trend functions that belong to this class of nonlinear trends.[2] An important feature of our procedure is that estimation of parametric nonlinear trend functions is not required. In this sense, our procedure can be

[2] We focus on this class of trends since the trend breaks are most frequently used forms of nonlinear trends in practice. However, we can easily extend our approach to incorportate other class of trend functions such as the one for quadratic trends or cubic trends.

viewed as a nonparametric approach to cotrending rank selection.

Theory

We assume that an m-variate time series, $y_t = [y_{1t}, \cdots, y_{mt}]'$, is generated by

$$y_t = d_t + s_t, \quad t = 1, \cdots T, \tag{III.6}$$

where $d_t = [d_{1t}, \cdots, d_{mt}]'$ is a nonstochastic trend component, $s_t = [s_{1t}, \cdots, s_{mt}]'$ is a stochastic process, respectively defined below, and neither d_t nor s_t is observable. We denote a random (scalar) sequence x_T by $O_p(T^\lambda)$ if $T^{-\lambda}x_T$ is bounded in probability and by $o_p(T^\lambda)$ if $T^{-\lambda}x_T$ converges to zero in probability. For a deterministic sequence, we use $O(T^\lambda)$ and $o(T^\lambda)$, if $T^{-\lambda}x_T$ is bounded and converges to zero, respectively. The first difference of x_t is denoted by Δx_t. Below, we employ a set of assumptions that are similar to those in Hatanaka and Yamada (2003).

Assumptions B.

(B1) $s_t = s_{t-1} + \xi_t$ and $\xi_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$, $C_0 = I_n$, $\sum_{j=0}^{\infty} j^2 \|C_j\| < \infty$, where ε_t is *iid* with zero mean and covariance matrix $\Sigma_{\varepsilon\varepsilon} > 0$.

(B2) Each element of $\sum_{t=1}^{T} d_t$ is $O(T^2)$ and is not $o(T^2)$.

(B3) There exists an $m \times m$ orthogonal full rank matrix $B = [B_\perp \ B_2 \ B_1 \ ]$, such that each element of $\sum_{t=1}^{T} B_1' y_t$ is $O_p(T^{1/2})$, each element of $\sum_{t=1}^{T} B_2' y_t$ is $O_p(T)$ and is not $o_p(T)$, and each element of $\sum_{t=1}^{T} B_\perp' y_t$ is $O_p(T^2)$ is not $o_p(T^2)$, where B_1, B_2, $B_\perp$ are $m \times r_1$, $m \times (r_2 - r_1)$ and $m \times (m - r_2)$, respectively.

Under Assumptions B , B_1 represents a set of cotrending vectors that eliminates both deterministic and stochastic trends. B_2 represents a set of vectors eliminating only deterministic trends, but not stochastic trends. $B_\perp$ consists of vectors orthogonal to B_1 and B_2.

In the scalar case, the von Neumann ratio is defined as the ratio of the sample second moment of the differences to that of the level of a time series. The multivariate generalization of the von Neumann ratio is defined as $S_{11}^{-1}S_{00}$ where

$$S_{11} = T^{-1}\sum_{t=1}^{T} y_t y_t^{'}, \quad \text{and} \quad S_{00} = T^{-1}\sum_{t=2}^{T} \Delta y_t \Delta y_t^{'}.$$

Shintani (2001) and Harris and Poskitt (2004) also use this multivariate version of the von Neumann ratio in cointegration analysis. Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_m \geq 0$ be the eigenvalues of $S_{11}^{-1}S_{00}$. We summarize the statistical properties of $\widehat{\lambda}_i^{'}$s in the presence of both stochastic and deterministic trends in the following lemma.

Lemma 1 *Under Assumptions B, we have: (i) a sequence of* $[\widehat{\lambda}_1, \cdots, \widehat{\lambda}_{r_1}]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*; (ii) a sequence of* $T[\widehat{\lambda}_{r_1+1}, \cdots, \widehat{\lambda}_{r_2}]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*, provided* $r_2 - r_1 > 0$*; and (iii) a sequence of* $T^2[\widehat{\lambda}_{r_2+1}, \cdots, \widehat{\lambda}_m]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*, provided* $m - r_2 > 0$*.*

From Lemma 1, the eigenvalues of $S_{11}^{-1}S_{00}$ can be classified into three groups depending on their rates of convergence, namely, $O_p(1)$, $O_p(T^{-1})$ and $O_p(T^{-2})$. The number of eigenvalues in each group corresponds to the number of cotrending relationships (r_1), the difference between weak cotrending and (strong) cotrending relationships ($r_2 - r_1$) and the number of common deterministic trends ($m - r_2$), respectively. We exploit this property to construct the following two types of consistent cotrending rank selection procedures based on the von Neumann criterion, which is defined as a sum of the partial sum of eigenvalues and a penalty term. The first is a 'paired' procedure which independently selects the cotrending rank r_1 and the weak cotrending rank r_2 by minimizing each of

$$\begin{aligned} VN_1(r_1) &= -\sum_{i=1}^{r_1} \widehat{\lambda}_i + f(r_1)\frac{C_T}{T}, \text{ and} \\ VN_2(r_2) &= -\sum_{i=1}^{r_2} \widehat{\lambda}_i + f(r_2)\frac{C_T^{'}}{T^2}, \end{aligned}$$

or

$$\begin{aligned} \widehat{r}_1 &= \arg \min_{0 \le r_1 \le m} VN_1(r_1), \text{ and} \\ \widehat{r}_2 &= \arg \min_{0 \le r_2 \le m} VN_2(r_2) \end{aligned}$$

where $f(r)$, C_T and $C_T^{'}$ are elements of penalty function defined in detail below.

The second procedure is a 'joint' procedure that simultaneously determines both r_1 and r_2 by minimizing

$$VN(r_1, r_2) = -\sqrt{T}\sum_{i=1}^{r_1} \widehat{\lambda}_i - \sum_{i=r_1+1}^{r_2} \widehat{\lambda}_i + f(r_1)\frac{C_T}{T} + f(r_2)\frac{C_T^{'}}{T^2},$$

or

$$(\widehat{r}_1, \widehat{r}_2) = \arg \min_{0 \le r_1, r_2 \le m} VN(r_1, r_2).$$

The main theoretical result is provided in the following proposition.

Proposition 1 *(i) Suppose Assumptions B holds, and $f(r)$ is an increasing function of r, $C_T, C_T^{'} \to \infty$, $C_T/T, C_T^{'}/T \to 0$, then the paired procedure using $VN_1(r_1)$ and $VN_2(r_2)$ yields,*

$$\lim_{T\to\infty} P(\widehat{r}_1 = r_1, \widehat{r}_2 = r_2) = 1.$$

(ii) Suppose Assumptions B holds, and $f(r)$ is an increasing function of r, $C_T/\sqrt{T}, C_T^{'}/\sqrt{T} \to \infty$, $C_T/T, C_T^{'}/T \to 0$, then the joint procedure using $VN(r_1, r_2)$ yields,

$$\lim_{T\to\infty} P(\widehat{r}_1 = r_1, \widehat{r}_2 = r_2) = 1.$$

Remarks:

(a) The proposition shows that both of the two cotrending rank selection procedures are consistent in selecting a cotrending rank without specifying a parametric model as long as the trend belongs to a certain class of nonlinear functions. The joint selection procedure requires slightly stronger assumptions on C_T and C_T' than the paired selection procedure.

(b) Commonly employed C_T in the literature of information criteria includes $C_T = \ln(T)$, $2\ln(\ln(T))$, and 2, which respectively leads to the Bayesian information criterion (BIC), Hannan-Quinn criterion (HQ), and Akaike information criterion (AIC). Part (i) of the proposition implies that the paired cotrending rank selection procedure is consistent when BIC and HQ type penalties are employed, but is inconsistent when an AIC type penalty is employed. In contrast, part (ii) of the proposition implies that C_T (and C_T') should diverge at the rate faster than $\sqrt{T}$ for the joint cotrending rank selection procedure, thus none of $C_T = \ln(T)$, $2\ln(\ln(T))$, and 2 yield consistency.

(c) By the definition of $VN(r_1, r_2)$, cotrending ranks selected by the joint procedure always satisfy $\widehat{r}_1 \leq \widehat{r}_2$. For the paired procedure, selected cotrending ranks will satisfy $\widehat{r}_1 \leq \widehat{r}_2$ if $C_T' = T^{\alpha} C_T$, where $0 \leq \alpha < 1$. This fact can be demonstrated by the following argument. The selected cotrending rank $\widehat{r}_1$ implies that $VN_1(r) > VN_1(\widehat{r}_1)$ for all $r < \widehat{r}_1$. The result is equivalent to the partial sum of eigenvalues $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i$ being greater than $\{f(\widehat{r}_1) - f(r)\} C_T T^{-1}$ (note that $\widehat{\lambda}_i \geq 0$ and $f(\widehat{r}_1) - f(r) > 0$). To see if $VN_2(r) > VN_2(\widehat{r}_1)$ for the corresponding r and $\widehat{r}_1$, it suffices to show that $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i$ is greater than $\{f(\widehat{r}_1) - f(r)\} C_T' T^{-2}$. By substituting $C_T' = T^{\alpha} C_T$ the latter becomes $\{f(\widehat{r}_1) - f(r)\} C_T T^{-1} \times T^{-(1-\alpha)}$. Since $T^{-(1-\alpha)} < 1$, $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i > \{f(\widehat{r}_1) - f(r)\} C_T T^{-1} > \{f(\widehat{r}_1) - f(r)\} C_T T^{-1} \times T^{-(1-\alpha)}$. Because we have shown that $VN_2(r) > VN_2(\widehat{r}_1)$ for all $r < \widehat{r}_1$., it implies $\widehat{r}_1 \leq \widehat{r}_2$.

(d) The criterion function $VN_1(r_1)$ in the paired procedure can solely be used to select cointegrating rank in a system of stochastic trends without nonlinear deterministic trends. It nests the criterion function considered in Harris and Poskitt (2004) as a special case. Their criterion $\Gamma_{C,T}$, in their notation, is identical to $VN_1(r_1)$ combined with $C_T = \ln(T)$ and $f(r) = 2r(2m - r + 1)$. Thus, part (i) of the proposition extends the result of Harris and Poskitt (2004) to the cointegrating rank selection for general choice of C_T and $f(r)$.

(e) For consistency of our procedures, $f(r)$ can be any increasing function of r. In this chapter, we follow Harris and Poskitt (2004) and employ $f(r) = 2r(2m - r + 1)$, the function used in their consistent cointegrating rank selection criterion. This choice satisfies the required condition of an increasing function since $df(r)/dr = 4(m - r) > 0$. Other choices of functions, such as $f(r) = 2mr - r^2$ and $f(r) = 2mr - r(r + 1)/2$, are also discussed in Cheng and Phillips (2009) based on the reduced rank regression structure of the cointegrated system.

Experimental evidence

Stochastic trends and cointegrating rank

The proposed cotrending rank selection procedures are justified based on the asymptotic theory. Thus, it is of interest to examine their finite sample properties by means of Monte Carlo analysis. This section reports the results under different settings of the true cotrending ranks, and of various penalty terms.

Before we present the main simulation results of cotrending rank selection in a system with stochastic and nonlinear deterministic trends, let us first consider the case of a cointegrated system without deterministic trends. Understanding the basic characteristics of the multivariate von Neumann ratio-based procedure in a simple system with stochastic trends only, will help us justify the use of the similar procedure in a more complicated system. Recall, that the von Neumann ratio criterion $VN_1(r_1)$ in the paired procedure can be used to determine the cointegrating rank in the cointegrated system, and that it nests the cointegrating rank selection procedure of Harris and Poskitt (2004) as a special case. Since estimation of the cointegrating vector and serial correlation structure is not required, our procedure and the procedure by Harris and Poskitt (2004) may be viewed as a nonparametric approach to cointegrating rank selection. In contrast, the information criteria for selecting cointegrating rank in Cheng and Phillips (2009) are based on the eigenstructure of a reduced rank regression model. While serial correlation structure is not estimated, cointegrating vectors are estimated. In this sense, their procedure may be viewed as a semiparametric approach to cointegrating rank selection. Here, we use the same simulation design as in Cheng and Phillips (2009) and compare the finite sample performance of two alternative approaches.

A bivariate time series $y_t = (y_{1t}, y_{2t})'$ is generated from

$$\Delta y_t = \alpha\beta' y_{t-1} + u_t, , \quad t = 1, \cdots T,$$

where u_t follows a $VAR(1)$ process with a VAR coefficient $0.4 \times I_2$ and a mutually independent standard normal error term. By setting $\alpha\beta' = 0$,

$$\alpha\beta' = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix},$$

and

$$\alpha\beta' = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.15 \end{pmatrix},$$

we generate a multivariate system with the true cointegrating rank $r_1 = 0, 1$ and 2, respectively. We evaluate the finite sample performance of both semiparametric and nonparametric approaches by the frequencies of selecting the true cointegrating rank in 20,000 replications for the sample sizes $T = 50$, 100 and 400[3]. For the reduced rank regression procedure of Cheng and Phillips (2009), we employ the AIC, BIC and HQ criteria and denote them by RRR-AIC, RRR-BIC and RRR-HQ, respectively. The von Neumann ratio criterion $\Gamma_{C,T}$ of Harris and Poskitt (2004) is equivalent to $VN_1(r_1)$ with $f(r) = 2r(2m - r + 1)$ and $C_T = \ln(T)$. Since it involves a BIC-type penalty, we refer to this procedure by VN-BIC. In addition, we also consider the AIC-type penalty $C_T = 2$, as well as an HQ type penalty $C_T = 2\ln(\ln(T))$, and denote corresponding criteria by VN-AIC and VN-HQ, respectively. It should be noted that theoretical analysis implies that both RRR-AIC and VN-AIC are inconsistent in selecting true cointegrating rank.

Table 4 reports the performance of the cointegrating rank selection procedures based on six criteria with frequencies of correctly selecting true rank shown in bold fonts. The results of the simulation can be summarized as follows.

First, the semiparametric approach by Cheng and Phillips (2009) and our nonparametric approach seem to complement to each other because their relative performance depends on the data generating processes. If true cointegrating rank is $r_1 = 0$, the nonparametric von Neumann ratio-based procedures uniformly outperform the semiparametric reduced rank regression-based procedures for all the sample sizes under consideration. In contrast, if the true cointegrating rank is $r_1 = 2$ and the sample size is small ($T = 50$ and 100), each of the reduced rank regression procedures, RRR-AIC, RRR-BIC and RRR-HQ, works better than each counterpart of the von Neumann ratio procedures, VN-AIC, VN-

[3]Here, we follow Cheng and Phillips (2009) and the first 50 observations are discarded to eliminate the effect of the initial values $y_0 = 0$ and $u_0 = 0$.

BIC and VN-HQ, respectively. If the true cointegrating rank is $r_1 = 1$, the semiparametric reduced rank regression procedure works better with a BIC type penalty (RRR-BIC) when the sample size is as small as $T = 50$, but the nonparametric von Neumann ratio procedures dominates for the other cases.

Second, for the von Neumann ratio-based procedures, the AIC type penalty often works well when the sample size is small, despite the fact that it provides theoretically inconsistent rank selection. In particular, it dominates other types of penalties if the true cointegrating rank is the largest ($r_1 = 2$), mainly because the penalty for higher rank is much smaller with $C_T = 2$ than with $C_T = \ln(T)$ or $C_T = 2\ln(\ln(T))$. However, even in the case of low frequencies of selecting the true rank when the sample size is small, they quickly approach one when the sample size increases to $T = 400$. On the whole, it seems fair to say that the von Neumann criterion is at least as useful as the information criterion based on the reduced rank regression in selecting cointegrating rank.

Deterministic trends and cotrending rank

In this subsection, we evaluate the finite sample performance of our proposed procedure using the three-dimensional vector series $y_t^* = (y_{1t}^*, y_{2t}^*, y_{3t}^*)'$ with different combinations of cotrending and weak cotrending ranks ($m = 3$).

To consider the case with only one common (nonlinear) deterministic trend, we first generate the data using

$$\begin{aligned} y_{1t}^* &= \rho_1 y_{1t-1}^* + \varepsilon_{1t}, \\ y_{2t}^* &= \rho_2 y_{2t-1}^* + \varepsilon_{2t}, \\ y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau T \\ c + (\mu_0 - \mu_1)\tau T + \mu_1 t & \text{if } t > \tau T \end{cases}, \end{aligned} \tag{III.7}$$

with $(\varepsilon_{1t}, \varepsilon_{2t})' = iidN(0, \Sigma_\varepsilon)$ where

$$\Sigma_\varepsilon = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Note that here y_{1t}^* and y_{2t}^* do not have a deterministic trend, but the transformed system becomes equivalent to y_t given in equation (III.6), where each element contains a deterministic trend and a stochastic component. We can use any nonsingular matrix A such that $y_t = Ay_t^* = d_t + s_t$. Because the eigenvalues for the von Neumann ratio are invariant to any nonsingular transformation of the data, we can directly use y_t^* in the computation of our rank selection criteria in place of $y_t = Ay_t^*$ in the simulation. For example, a transformation using a matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

yields

$$\begin{aligned} y_{1t} &= y_{1t}^* + y_{2t}^* + y_{3t}^* = d_{1t} + s_{1t} \\ y_{2t} &= -y_{1t}^* + y_{2t}^* + y_{3t}^* = d_{2t} + s_{2t} \\ y_{3t} &= y_{1t}^* + y_{3t}^* = d_{3t} + s_{3t}. \end{aligned}$$

For the case of $\rho_1 = 0.5$ and $\rho_2 = 1.0$, a vector $(1, -1, 0)$ becomes a cotrending vector since $y_{1t} - y_{2t} = 2y_{1t}^*$ is stationary, and a vector $(1, 0, -1)$ becomes a weak cotrending vector since $y_{1t} - y_{3t} = y_{2t}^*$ contains a stochastic trend but not a deterministic trend. Since other cotrending vectors can be also incorporated by a different choice of a nonsingular matrix

A, a very large class of cotrended system can be covered by our simple simulation design.

We consider three cases by using different combinations of $\rho_i \in \{0.5, 1.0\}$ for $i = 1, 2$, in (III.7) and generate the data with $(r_1, r_2) = (2, 2)$, $(1, 2)$, and $(0, 2)$. In particular, setting $\rho_1 = \rho_2 = 0.5$ implies $(r_1, r_2) = (2, 2)$, $\rho_1 = 0.5$ and $\rho_2 = 1.0$ implies $(r_1, r_2) = (1, 2)$, and $\rho_1 = \rho_2 = 1.0$ implies $(r_1, r_2) = (0, 2)$. The parameters for the kinked trend function are set to $c = 0.5$, $\mu_0 = 2$, $\tau = 0.5$, and $\mu_1 = 0.5$.

Second, we consider the cases of two deterministic trends using

$$
\begin{aligned}
y_{1t}^* &= \rho_1 y_{1t-1}^* + \varepsilon_{1t}, \\
y_{2t}^* &= c + \mu_0 t \\
y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau T \\ c + (\mu_0 - \mu_1)\tau T + \mu_1 t & \text{if } t > \tau T \end{cases},
\end{aligned}
\tag{III.8}
$$

with $\varepsilon_{1t} = iidN(0, 1)$, $\rho_1 \in \{0.5, 1.0\}, c = 0.5$, $\mu_0 = 2$, $\tau = 0.5$. This system generates the data with $(r_1, r_2) = (1, 1)$ when $\rho_1 = 0.5$, and $(r_1, r_2) = (0, 1)$ when $\rho_1 = 1.0$.

Finally, we consider the three-deterministic trend case using

$$
\begin{aligned}
y_{1t}^* &= c + \mu_0 t + \varepsilon_{1t}, \\
y_{2t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau_1 T \\ c + (\mu_0 - \mu_1)\tau_1 T + \mu_1 t & \text{if } t > \tau_1 T \end{cases} \\
y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau_2 T \\ c + (\mu_0 - \mu_1)\tau_2 T + \mu_1 t & \text{if } t > \tau_2 T \end{cases},
\end{aligned}
\tag{III.9}
$$

with $\varepsilon_{1t} = iidN(0, 1)$, $c = 0.5$, $\mu_0 = 2$, $\tau_1 = 0.5, \tau_2 = 1/3$ and $\mu_1 = 0.5$. This system generates the data with $(r_1, r_2) = (0, 0)$.

We employ two paired cotrending rank selection procedures and two joint cotrending rank selection procedures. For the paired procedures, we employ a BIC type penalty

$C_T = \ln(T)$ for $VN_1(r_1)$. Recall that selected cotrending ranks from the paired procedure always satisfy $\widehat{r}_1 \leq \widehat{r}_2$ as long as $C'_T = T^{\alpha} C_T$, where $0 \leq \alpha < 1$. Here, we employ $C'_T = \sqrt{T}\ln(T)$ for $VN_2(r_2)$ and denote corresponding paired procedure by 'paired BIC.' In addition, we also consider the case with a weaker penalty for $VN_2(r_2)$ by replacing the penalty with $C'_T = \sqrt{T}\ln(\ln(T))$. Since $VN_1(r_1)$ is the same as before but the penalty for $VN_2(r_2)$ somewhat resembles that of the HQ type penalty, we denote the procedure by 'paired BIC-HQ.'

For the joint selection procedures, $\ln(T)$ cannot be used for C_T(and C'_T), since consistency requires the penalty diverges at a rate faster than $\sqrt{T}$. Therefore, we consider $VN(r_1, r_2)$ with the penalty $C_T = C'_T = \sqrt{T}\ln(T)$, and denote the procedure by 'joint BIC.' We additionally consider the pair of slower rate $C_T = C'_T = \sqrt{T}\ln(\ln(T))$ and denote the corresponding procedure by 'joint HQ.' As in the case of cointegration analysis, we employ $f(r) = 2r(2m - r + 1)$.

Table 5, Table 6 and Table 7 report the frequencies of selecting cotrending rank r_1 and weak cotrending rank r_2 by four procedures for sample sizes $T = 50$, 100 and 400 in 20,000 replications.[4] For each data generating process, the pair $(\widehat{r}_1, \widehat{r}_2)$ is selected by minimizing the von Neumann criterion among $(r_1, r_2) = (2, 2)$, $(1, 2)$, $(0, 2)$, $(1, 1)$, $(0, 1)$ and $(0, 0)$.[5] Frequencies of selecting the true model are shown in a bold font in the table. The results of the simulation can be summarized as follows.

First, both the paired procedures and joint procedures work well even when the sample size is as small as $T = 50$. When there is only one common deterministic trend and $T = 50$, paired procedures, paired BIC and paired BIC-HQ, work better than the joint

[4]For the stationary AR(1) part of the equations, the initial values are generated from its stationary distribution. For the other equations, initial values are set at 0.

[5]We only report the results from raw series version of the von Neumann criterion in the simulation since the demeaned version yielded similar results. The full simulation results are available upon request.

procedures, joint BIC and joint BIC-HQ, for the cases $(r_1, r_2) = (2, 2)$ and $(1, 2)$, but the latter works better for the case of $(r_1, r_2) = (0, 2)$. However, as the sample size increases, the frequencies of selecting the true rank become close to one for both types of procedures and thus the performance of the two procedures become almost indistinguishable.

Second, when there are two common deterministic trends and $T = 50$, the paired procedures perform better for the case of $(r_1, r_2) = (1, 1)$ and the joint procedures perform better for the case of $(r_1, r_2) = (0, 1)$. When $T = 100$, both procedures yield sufficiently high frequencies of selecting the true rank.

Finally, when there are three deterministic trends in the system, or $(r_1, r_2) = (0, 0)$, the performance highly depends on the choice of penalty terms. In particular,paired BIC and joint BIC select true rank all the time even if the sample size is $T = 50$. In contrast, the frequencies are very low for the paired BIC-HQ and joint BIC-HQ when sample size is small ($T = 50$), and frequencies become close to unity only when the sample size is $T = 400$.

Smooth transition trends and cotrending rank

In this section, we study the effect of nonlinearity in the trend function on the performance of our cotrending rank selection procedure. To this end, we consider the logistic smooth transition trend model and control the shape of the deterministic function by controlling the scale parameters in the logistic transition function. We generate the artificial data with $(r_1, r_2) = (0, 1)$ using

$$\begin{aligned} y_{1t}^* &= y_{1t-1}^* + \varepsilon_{1t}; \\ y_{2t}^* &= c_0 + \mu_0 t; \qquad\qquad \text{(III.10)} \\ y_{3t}^* &= (c_0 + \mu_0 t) G(\gamma, \tau T) + (c_1 + \mu_1 t)(1 - G(\gamma, \tau T)) \end{aligned}$$

where $G(\gamma, \tau T)$ is a logistic transition function defined in the previous section and $\varepsilon_{1t} = iidN(0,1)$, $c = 0.5$, $\mu_0 = 2$, $\tau = 0.5$, and $\mu_1 = 0.5$. As noted above, the scale parameter γ controls the speed of transition. As γ approaches infinity, the logistic function collapses to an index function $I(t > \tau T)$ and (III.10) become (III.7) with $\rho_1 = 1.0$. On the other hand, as γ approaches zero, the smooth transition trend model approaches to a linear trend. In this scenario, we can always find the linear combination that eliminates the trend function. In other words, when γ is close to zero, the system of two common deterministic trends ($r_2 = 1$) becomes closer to the system of one common deterministic trend ($r_2 = 2$). Therefore, for a small value of γ, we expect that it will be difficult for our procedure to identify $r_2 = 1$ from $r_2 = 2$.

Table 8 presents the simulation results given different choices of the scale parameter $\gamma \in \{0.001, 0.005, 0.01\}$ when $T = 400$. Note that we can use the result of Table 7 for $(r_1, r_2) = (0, 1)$ and $T = 400$ as the benchmark limit case with a large γ. It turns out that the procedure works well in selecting the true rank even γ is as small as 0.01. Consistent with our prediction, two of the four procedures (joint BIC-HQ and joint BIC-HQ) select $r_2 = 2$ when $\gamma = 0.01$, and all the procedures select $r_2 = 2$ when $\gamma = 0.001$.

Application

The simulation results in the previous section show that our procedures perform well in various experimental set-ups. In this section, we apply our procedures to the Japanese money demand function to investigate the cotrending relations among money demand, income and interest rate ($m = 3$). A seasonally adjusted quarterly series of real GDP, two definitions of monetary aggregates, $M1$ and $M2$, and the call rate for the sample period from 1980:Q1 to 2010:Q4, are plotted in Figures 2 to 5. The figures show the possibility of kinked deterministic trends in these variables.

We follow Bae, Kakkar and Ogaki (2006) and consider following three different specifications of money demand functions,

$$
\begin{aligned}
\text{Model 1} &: \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 i_t + \varepsilon_t, \\
\text{Model 2} &: \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 \ln(i_t) + \varepsilon_t, \text{ and} \\
\text{Model 3} &: \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 \ln\left(\frac{i_t}{1+i_t}\right) + \varepsilon_t,
\end{aligned}
$$

where M_t is the money demand, P_t is the aggregate price level, y_t is real GDP and i_t is the nominal interest rate.

We apply both paired and joint cotrending rank selection procedures to the vectors $(\ln(M_t/P_t), \ln(y_t), i_t)$, $(\ln(M_t/P_t), \ln(y_t), \ln(i_t))$, and $(\ln(M_t/P_t), \ln(y_t), \ln(i_t/(1+i_t)))$. Table 9 reports the empirical results for all three different specifications of the functional form for interest elasticity of money demand. The results are somewhat mixed depending on the choice of the penalty of the criteria and the choice of the variables. However, it is important to note that none of the procedures select $(r_1, r_2) = (0, 0)$. This implies that there are, at least, either cotrending or weak cotrending relationships in Japanese money demand in the long-run. When $M2$ is used as the monetary aggregate and when demeaned version of the von Neumann ratio is used, $(r_1, r_2) = (0, 2)$ is selected for all cases, implying that the kinked trend is likely to be a single common deterministic trend among three variables.

Conclusion

This paper has proposed a model-free cotrending rank selection procedure to use when both stochastic and nonlinear deterministic trends are present in a multivariate system. The procedure selects two types of cotrending ranks by minimizing two new criteria based on the generalized von Neumann ratio. Our approach is invariant to the linear transformation of data, robust to misspecification of the model and consistent under very general conditions. Monte Carlo experiments have suggested good finite sample performance of the proposed procedure. An empirical application to the money demand function in Japan has also suggested the usefulness of our procedure in detecting cotrending relationships when nonlinear deterministic trends are present in data.

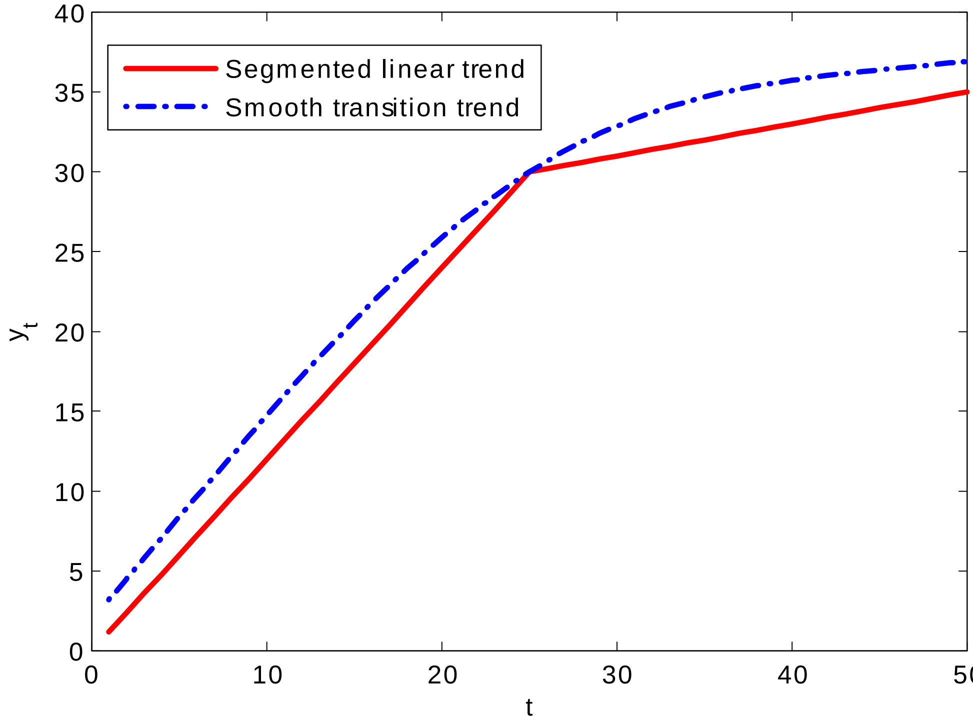

Figure 1. Segmented linear trend and smooth transition trend

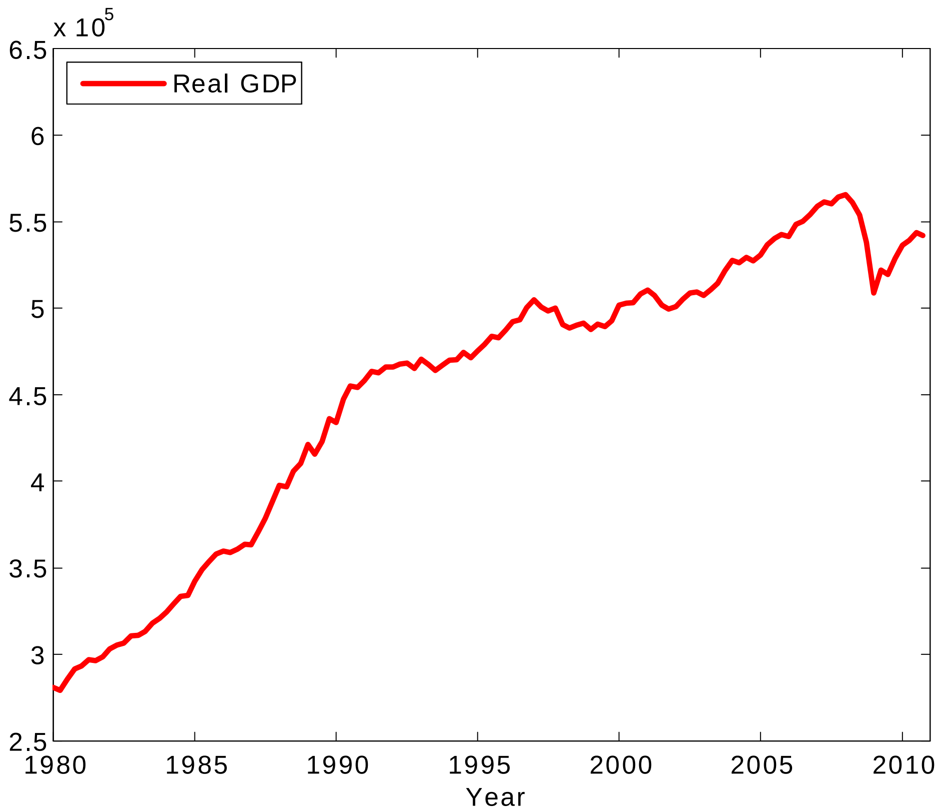

Figure 2. Real GDP

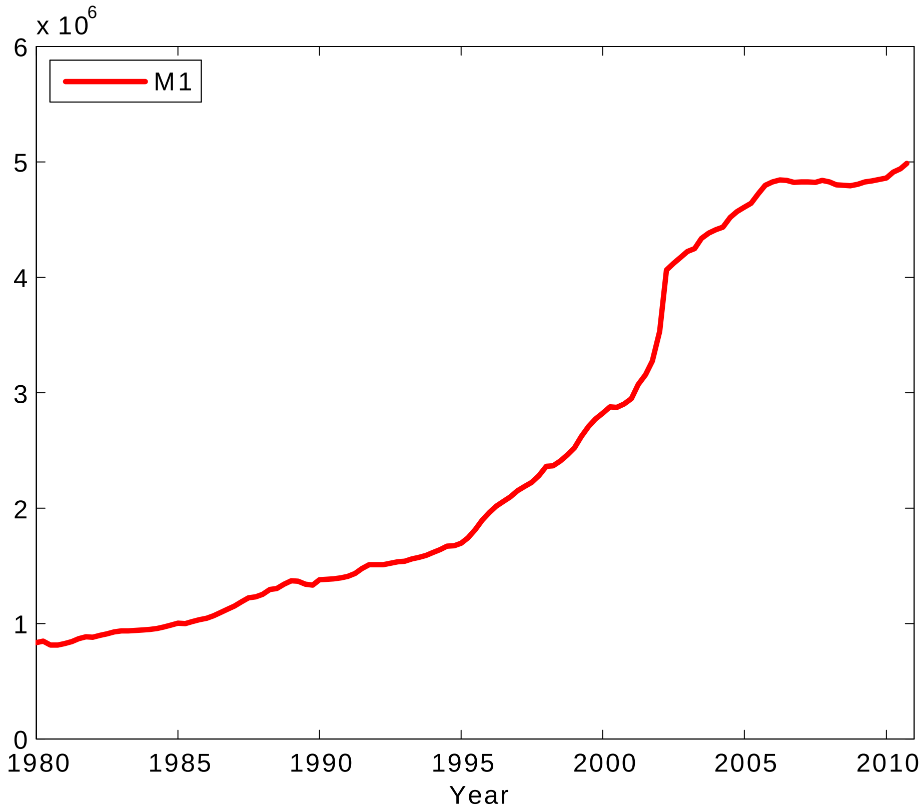

Figure 3. M1

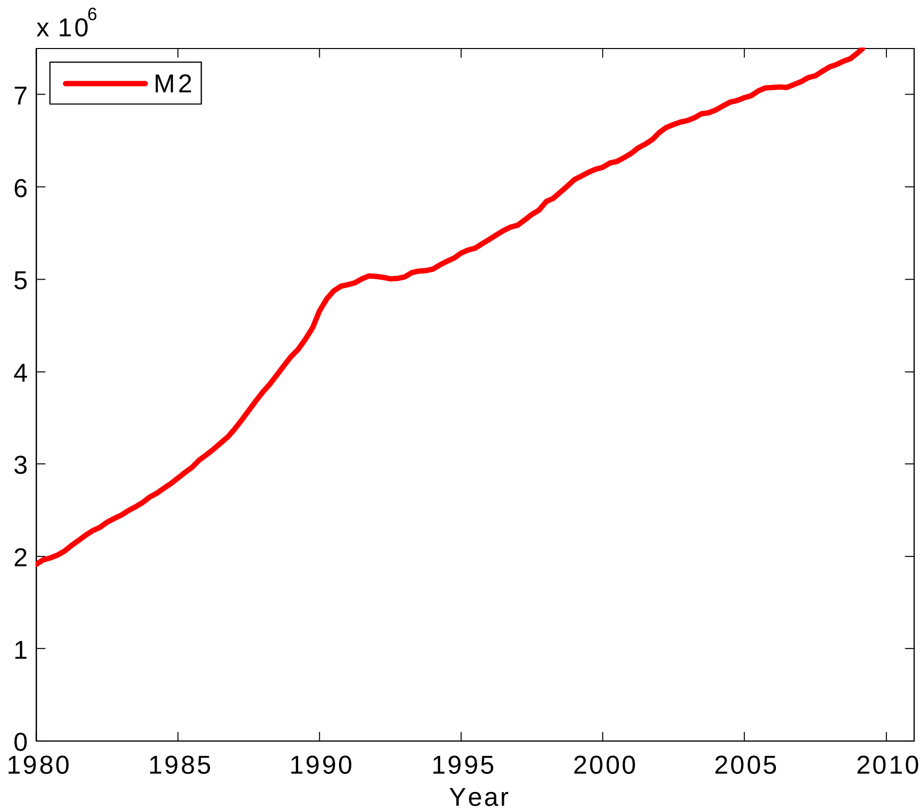

Figure 4. M2

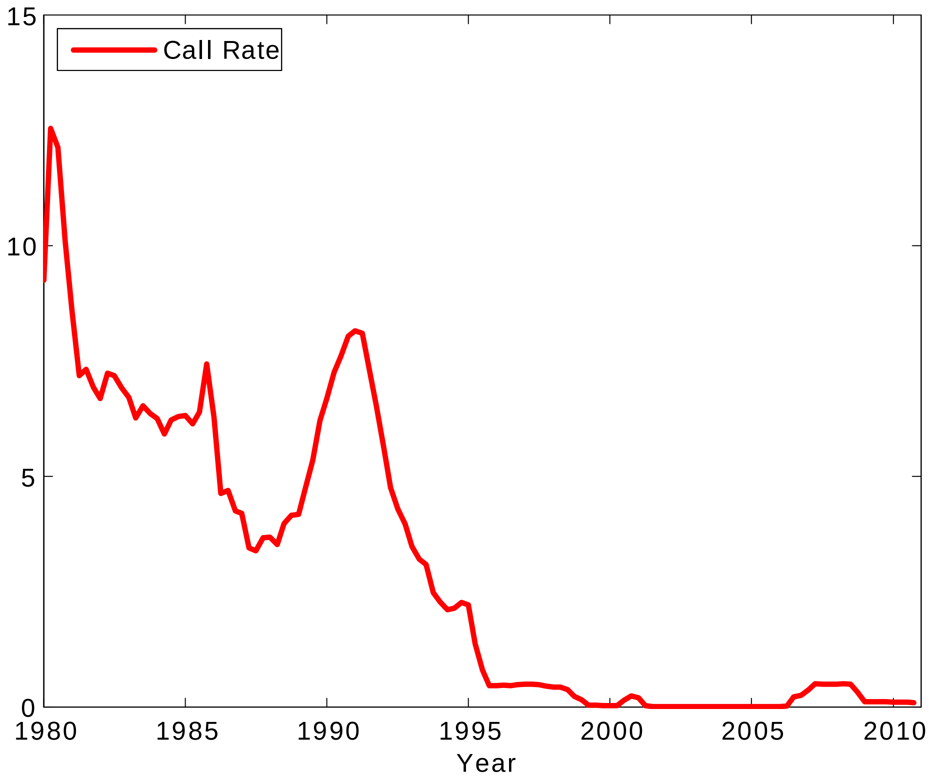

Figure 5. Call rate

Table 4. Two dimensional cointegrating rank selection

T=50	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$
RRR-AIC	**0.46**	0.41	0.13	0.00	**0.78**	0.22	0.02	0.55	**0.43**
RRR-BIC	**0.81**	0.17	0.03	0.00	**0.92**	0.08	0.45	0.45	**0.10**
RRR-HQ	**0.62**	0.30	0.07	0.00	**0.85**	0.15	0.13	0.61	**0.26**
VN-AIC	**0.97**	0.03	0.00	0.04	**0.96**	0.00	0.01	0.80	**0.19**
VN-BIC	**1.00**	0.00	0.00	0.65	**0.35**	0.00	0.47	0.52	**0.01**
VN-HQ	**0.99**	0.01	0.00	0.21	**0.79**	0.00	0.09	0.84	**0.07**
T=100	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$
RRR-AIC	**0.49**	0.39	0.12	0.00	**0.78**	0.22	0.00	0.25	**0.75**
RRR-BIC	**0.88**	0.11	0.01	0.00	**0.94**	0.06	0.05	0.73	**0.22**
RRR-HQ	**0.70**	0.25	0.05	0.00	**0.87**	0.13	0.00	0.51	**0.49**
VN-AIC	**0.98**	0.02	0.00	0.00	**1.00**	0.00	0.00	0.39	**0.61**
VN-BIC	**1.00**	0.00	0.00	0.05	**0.95**	0.00	0.01	0.95	**0.05**
VN-HQ	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.76	**0.24**
T=400	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$
RRR-AIC	**0.52**	0.37	0.11	0.00	**0.76**	0.24	0.00	0.00	**1.00**
RRR-BIC	**0.95**	0.05	0.00	0.00	**0.96**	0.04	0.00	0.02	**0.98**
RRR-HQ	**0.80**	0.18	0.03	0.00	**0.89**	0.11	0.00	0.00	**1.00**
VN-AIC	**0.97**	0.03	0.00	0.00	**1.00**	0.00	0.00	0.00	**1.00**
VN-BIC	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.06	**0.94**
VN-HQ	**1.00**	0.00	0.00	0.00	**1.00**	0.10	0.00	0.00	**1.00**

Note: Frequencies of selecting each cointegrating rank are reported.

Table 5. Three dimensional cotrending rank selection: T=50

T=50	**(2,2)**	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**0.93**	0.06	0.01	0.00	0.00	0.00
Paired BIC-HQ	**0.93**	0.06	0.01	0.00	0.00	0.00
Joint BIC	**0.81**	0.12	0.08	0.00	0.00	0.00
Join BIC-HQ	**0.79**	0.13	0.08	0.00	0.00	0.00
T=50	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.02	**0.71**	0.18	0.07	0.02	0.00
Paired BIC-HQ	0.02	**0.78**	0.20	0.00	0.00	0.00
Joint BIC	0.01	**0.46**	0.35	0.14	0.04	0.00
Join BIC-HQ	0.01	**0.57**	0.41	0.00	0.00	0.00
T=50	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.05	**0.79**	0.01	0.16	0.00
Paired BIC-HQ	0.00	0.05	**0.94**	0.00	0.00	0.00
Joint BIC	0.00	0.02	**0.82**	0.01	0.16	0.00
Joint BIC-HQ	0.00	0.02	**0.98**	0.00	0.00	0.00
T=50	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**0.81**	0.19	0.00
Paired BIC-HQ	0.00	0.00	0.00	**0.81**	0.19	0.00
Joint BIC	0.00	0.00	0.00	**0.62**	0.38	0.00
Joint BIC-HQ	0.00	0.00	0.00	**0.60**	0.40	0.00
T=50	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.03	**0.96**	0.01
Paired BIC-HQ	0.00	0.00	0.00	0.03	**0.97**	0.00
Joint BIC	0.00	0.00	0.00	0.01	**0.97**	0.01
Joint BIC-HQ	0.00	0.00	0.00	0.01	**0.99**	0.00
T=50	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.98	**0.02**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Join BIC-HQ	0.00	0.00	0.00	0.00	0.98	**0.02**

Note: The first and the second elements in the parenthesis denote cotrending and weak cotrending rank r_1 and r_2, respectively. Numbers are frequencies of selecting each pair of cotrending ranks.

Table 6. Three dimensional cotrending rank selection: T=100

T=100	**(2,2)**	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Paired BIC-HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint BIC-HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
T=100	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.02	**0.93**	0.00	0.06	0.00	0.00
Paired BIC-HQ	0.02	**0.98**	0.00	0.00	0.00	0.00
Joint BIC	0.01	**0.82**	0.00	0.16	0.00	0.00
Joint BIC-HQ	0.01	**0.99**	0.00	0.00	0.00	0.00
T=100	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.03	**0.87**	0.00	0.10	0.00
Paried BIC-HQ	0.00	0.03	**0.97**	0.00	0.00	0.00
Joint BIC	0.00	0.01	**0.89**	0.00	0.10	0.00
Joint BIC-HQ	0.00	0.02	**0.98**	0.00	0.00	0.00
T=100	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Paired BIC-HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint BIC-HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
T=100	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.02	**0.98**	0.01
Paired BIC-HQ	0.00	0.00	0.00	0.02	**0.98**	0.00
Joint BIC	0.00	0.00	0.00	0.01	**0.98**	0.01
Joint BIC-HQ	0.00	0.00	0.00	0.01	**0.99**	0.00
T=100	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint BIC-HQ	0.00	0.00	0.00	0.00	0.00	**1.00**

Note: See note for Table 5.

Table 7. Three dimensional cotrending rank selection: T=400

T=400	**(2,2)**	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Paired BIC-HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
T=400	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	**0.99**	0.00	0.01	0.00	0.00
Paired BIC-HQ	0.00	**1.00**	0.00	0.00	0.00	0.00
Joint BIC	0.00	**0.94**	0.00	0.06	0.00	0.00
Joint HQ	0.00	**1.00**	0.00	0.00	0.00	0.00
T=400	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.01	**0.98**	0.00	0.02	0.00
Paired BIC-HQ	0.00	0.01	**0.99**	0.00	0.00	0.00
Joint BIC	0.00	0.00	**0.98**	0.00	0.02	0.00
Joint HQ	0.00	0.00	**1.00**	0.00	0.00	0.00
T=400	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Paired BIC-HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
T=400	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
T=400	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint HQ	0.00	0.00	0.00	0.00	0.00	**1.00**

Note: See note for Table 5.

Table 8. Cotrending rank selection with smooth transition trend models: T=400

	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
			(i) $\gamma = 0.001$			
Paired BIC	0.00	0.00	0.91	0.00	**0.09**	0.00
Paired BIC-HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
Joint BIC	0.00	0.00	0.91	0.00	**0.09**	0.00
Joint HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
			(ii) $\gamma = 0.005$			
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
			(iii) $\gamma = 0.01$			
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	0.00	0.00	**1.00**	0.00

Note: Frequencies of selecting each cointegrating rank are reported.

Table 9. Cotrending relationship among money, income and interest rates

	Model 1		Model 2		Model 3	
	VN	VN-μ	VN	VN-μ	VN	VN-μ
			(i) $M1$			
Paired BIC	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)
Paired BIC-HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint BIC	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)
Joint HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
			(ii) $M2$			
Paired BIC	(0,1)	(0,2)	(0.1)	(0,2)	(0,1)	(0,2)
Paired BIC-HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint BIC	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)

Note: For each pair of numbers, the first element denotes the cotrending rank, r_1, and the second element denotes the weak cotrending rank, r_2. The first column represents the results of the von Neumann criteria from raw series (VN), and the second column is from demeaned series (VN-μ).

CHAPTER IV

AN ASYMMETRIC SMOOTH TRANSITION GARCH MODEL

Introduction

Modeling and forecasting returns volatility in financial market is one of the most important issue in financial econometrics. Over the years, a number of different features of returns volatility have emerged, such as positive dependence in the volatility process, volatility clustering, high persistence and nonlinearity. The most widely used class of models to estimate and forecast volatility is represented by the autoregreesive conditional heteroskedasticity (ARCH model by Engle, 1982) and the generalized autoregreesive conditional heteroskedasticity (GARCH model by Bollerslev et al., 1990).

One key criticism of the GARCH specifications comes from the modeling of conditional variance as a function of past squared residuals, which makes the sign of the residuals irrelevant in predicting volatility. The symmetric treatment of positive and negative residuals contradicts the stylized fact, first noted by Black(1976), that stock market returns become more volatile after a negative shock, than they do after a positive shock of the same magnitude. One possible explanation, known as "the leverage effect", is that negative excess return reduce the equity value, hence the leverage ratio, of a given firm increase, thus raising its riskiness and the future volatility of its assets. Nelson's (1991) Exponential GARCH model is one of the first of many specifications, for example, threshold GARCH (TGARCH) model proposed by Rabemananjara and Zakoian(1993), the asymmetric power ARCH model developed by Ding et al (1993), and so on, that involves asymmetric functions of the residuals. It is well known in the literature that the specifications which allow for "leverage effect" dominate the standard GARCH specifications.

More recently, several authors introduced smooth transition specifications (Hagerud, 1997, Gonzalez-Rivera, 1998, Anderson et al., 1999 and Medeiros and Veiga, 2009), in modeling the asymmetric response of conditional variance to positive versus negative news. The smooth transition models can be thought of as a regime switch model with a continuum of regimes. For certain parameter values, it nests with the threshold specifications that only allows for a finite number of regimes. The smooth transition specifications in some sense generalize the modeling of asymmetry in variance and the empirical evidence in favor of the smooth transition specification is also reported by these authors.

The main purpose of this chapter is to propose a new smooth transition GARCH model, which allows both sign asymmetry and transition asymmetry. The smooth transition specifications in the volatility literature assume a transition function that is symmetric around its midpoint, which implies that negative shocks and positive shocks will have the same transition phases. The symmetry in the transition phases may be too restricted for practical purposes. Following Nelder (1961) and Sollis et al.(1998), we introduce a generalized logistic function that allows for both sign asymmetry and transition asymmetry, to model conditional variance.

This chapter contributes to the literature in many aspects. First of all, our model is a generalization of the smooth transition GARCH models by Hagerud (1997), Gonzalez-Rivera (1998)) and Anderson et al. (1999), and can nest with a lot of existing specifications, such as the DGE model of Ding, Granger, and Engle(1993), and the GJR model of Glosten, Jagannathan, and Runkle (1993), for certain range of parameter values. Secondly, our model allows both the ARCH parameters and GARCH parameters to vary with shocks, which gives rise to a news impact curve that changes shape as volatility varies. Therefore, the shocks of the same size and same magnitude may have different impact on current volatility depending on past volatility levels. Similar to the asymmetric nonlinear smooth GARCH model by Anderson et al. (1999), our model is nonlinear in both past shocks and past volatilities. Thirdly, to test for asymmetry, we propose two testing procedures, one is based on the linearization of the transition function and the other is a supremum LM test with unidentified parameters under the null, following Davies (1977, 1987). We find that the LM test is preferred, because it only requires estimation of the model under the null.

The remainder of this chapter is organized as follows. In section 2, we introduce the asymmetric smooth transition GARCH (ASTGARCH for abbreviation) model and its statistical properties. In section 3, we address the problem of testing for the existence of a smooth transition mechanism. We propose two test statistics to test for asymmetry and conduct a Monte-Carlo experiment to examine their finite sample performance. In section 4, we offer an application to NASDAQ stock index daily returns and IBM daily stock returns, and in section 5, we conclude the chapter and summarize this work.

An asymmetric adjustment smooth transition model

The model

Let r_t denote the rate of returns of a financial asset from time $t-1$ to time t and let Ψ_{t-1} be the investors' information set which contains relevant information at time t. The unexpected shock is denoted by ε_t, which is given by $r_t - E(r_t|\Psi_{t-1})$. The conditional variance of returns, $h_t = Var(r_t|\Psi_{t-1})$, is a measure of volatility, first proposed by Engle(1982). In the literature, $\varepsilon_t|\Psi_{t-1}$ is generally assumed to follow a normal distribution with mean zero and variance h_t. This distribution, however, can be relaxed to more general ones, for example, the standardized distribution (Bollerslev 1987) and the generalized error distribution (Nelson 1991). We assume conditional normality of $\varepsilon_t|\Psi_{t-1}$ in this chapter. Furthermore, we assume

$$\varepsilon_t = u_t\sqrt{h_t}$$

where u_t is an $i.i.d.n$ sequence with zero mean and unit variance.

The first volatility model that incorporates the smooth transition specification is by Hagerud (1997) and Gonzalez-Rivera (1998), which is given

$$h_t = w_0 + \sum_{i=1}^{p} \alpha_{0i}\varepsilon_{t-i}^2 + \left(\sum_{i=1}^{p} \alpha_{1i}\varepsilon_{t-i}^2\right) F(s_{t-1},\gamma) + \sum_{i=1}^{q} \beta_{0i}h_{t-i}.$$

The smooth transition model generalizes the modeling of variance with the introduction of a smooth transition specification in the sense that it allows for intermediate transition states. It also encompasses a wide array of ARCH specifications, such as the DGE model of Ding, Granger, and Engle(1993), and the GJR model of Glosten, Jagannathan, and Runkle (1993), and the threshold ARCH model of Rabemananjara and Zakoian (1993).

As discussed by Fonari and Mele(1997), the main restriction of the smooth transition model is the effects of ε_{t-1} and h_{t-1} on the volatility are additively separable. In other words, the impact of ε_{t-1} on conditional variance does not depend on past volatility values and is always the same for a given value of ε_{t-1}. Anderson et al. (1998) introduce the asymmetric nonlinear smooth transition GARCH models (ANTSGARCH), a class of models that extends the smooth transition GARCH model and allows the nonlinearity in both GARCH and ARCH parameters, which is given by

$$h_t = w_0 + \sum_{i=1}^{p} \alpha_{0i}\varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_{0i}h_{t-i} + F(s_{t-1},\gamma)\left[w_1 + \left(\sum_{i=1}^{p} \alpha_{1i}\varepsilon_{t-i}^2\right) + \sum_{i=1}^{q} \beta_{1i}h_{t-i}\right].$$

In this chapter, we introduce an asymmetric transition function: the generalized logistic function introduced by Nedler (1991) and Sollis et al. (1998) and propose the following new specification.

Definition 1 *An asymmetric smooth transition GARCH model, is defined by the model*

$$h_t = w_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h_{t-1} + F(s_{t-1}, \lambda, \gamma)(w_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}) \tag{IV.1}$$

where

$$F(s_{t-1}, \lambda, \gamma) = [1 + \exp(\lambda s_{t-1}/\gamma)]^{-\gamma}$$

is the transition function. s_{t-1} is the transition variable and λ is the smooth parameter. Possible transition asymmetry is introduced through the parameter γ where $\gamma = 1$ implies no asymmetry.

It is too restrictive for practical purposes to assume that transition functions are symmetric around its mid-point, which implies that positive shocks and negative shocks will have the same transition phases. Our model departs from most of the existing GARCH specification by making use of this asymmetric transition function. The intuition behind this assumption resides in the two asymmetries found in the volatility literature, i.e, the leverage effect and reversion of asymmetric effect. Given these two properties, the positive shocks and negative shocks in generally have different impact dynamics on volatility. Our model may easily capture these two asymmetry properties. In this chapter, we focus on the asymmetric smooth transition GARCH (1,1) model (ASTGARCH(1,1) for abbreviation). Other variants can also be obtained following the same methodology. For convenenience, we denote the parameter vector $\theta \equiv (w_0, \alpha_0, \beta_0, w_1, \alpha_1, \beta_1, \lambda, \gamma,)$.

The equivalence between our specification and that of Anderson et al. (2005) can be readily established by setting $\gamma = 1$. In the case of $\gamma = 1$, $w_1 = 0$ and $\beta_1 = 0$, our model collapses to the smooth transition GARCH model of Hagerud (1997) and Gonzalez-Rivera (1998). By using the same reasoning as Gonzalez-Rivera (1998), we can easily show the equivalence of our model with the DGE model, the GJR model and so on.

An important tool, widely used in the literature to capture the impact of innovations on volatility, is the news impact curve(NIC) introduced by Engle and Ng (1991). The idea is to examine the implied relation between ϵ_{t-1} and h_t, holding constant the information set prior to $t-2$ and earlier. For a standard GARCH(1,1) model with w_1, α_1 and β_1 all being set equal to zero, the news impact curve is characterized by

$$NIC(\epsilon_{t-1}|h_{t-1} = h) = w_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h$$

The $NIC(\epsilon_{t-1}|h_{t-1} = h)$ is a quadratic function of ε_{t-1} and the sign of ε_{t-1} is irrelevant in this function. The past volatility h only changes the level of NIC, but not the shape of the curve.

A key feature of our model is the sign and size asymmetry. In other words, shocks of the same size and same sign may have different effects on volatility depending on past volatility levels. The news impact curve is

$$NIC(\epsilon_{t-1}|h_{t-1} = h) = w_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h + F(s_{t-1}, \lambda, \gamma)(w_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h)$$

which depends nonlinearly on ε_{t-1} and h. So we draw a set of news impact to show the relationship between ϵ_{t-1} and h_t, following Bollerslev, et al. (1994).

Figure 6 presents the news impact curve, conditional on different past volatility levels, of our new model where the GARCH(1,1) model is used as a benchmark. In this figure, the ASTGARCH(1,1) is generated by setting $w_0 = 0.01, \alpha_0 = 0.1, \beta_0 = 0.1, \lambda = 1$ and

$$h_t = 0.01 + 0.1\epsilon_{t-1}^2 + 0.1\overline{h} + F(\epsilon_{t-1}, 1, 0.2)(0.1 + 0.01\epsilon_{t-1}^2 + 0.5\overline{h}).$$

Meanwhile, the GARCH(1,1) model is generated by

$$h(t) = 0.01 + 0.1\epsilon_{t-1}^2 + 0.1\overline{h}.$$

Here, we set $\overline{h}$ =0.5 and 5, respectively.

As we can see from Figure 6, for the GARCH(1,1) model, the news impact curve does not change the shape as we change the initial values of $\overline{h}$. In contrast, the shape of news impact curve for ASTGARCH(1,1) varies with the initial values of $\overline{h}$, which implies a nonlinear relationship between ϵ_{t-1}^2 and $h(t)$. We can also observe that the shock of the same size and magnitude may have different effects on volatility, depending on past volatility levels.

Estimation

In this section, we consider maximum likelihood estimation of the asymmetric smooth transition model and alternatively quasi maximum likelihood estimation in case of a nonnorrmal distribution of the error terms. We limit our results to ANSTGARCH(1,1) model, while observing that

the results are quite similar for more complicated models.

Let ε_t be the unexpected return, which is given by $\epsilon_t = r_t - E(r_t|\Psi_{t-1})$, the log-likelihood function for a sample of T observations is, apart from a constant:

$$l_T(Y;\theta) = -\frac{1}{2}\ln h_t - \frac{1}{2}\epsilon_t^2 h_t^{-1}.$$

Differencing with respect to the variance parameters yields

$$\frac{\partial l_t}{\partial \theta} = \frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \theta}(\epsilon_t^2 h_t^{-1} - 1),$$

$$\frac{\partial^2 l_t}{\partial \theta \partial \theta'} = (\epsilon_t^2 h_t^{-1} - 1)\frac{\partial}{\partial \theta'}\left[\frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \theta}\right] - \frac{1}{2}h_t^{-2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta'}\epsilon_t^2 h_t^{-1}.$$

where

$$\frac{\partial h_t}{\partial \theta} = z_t + \beta_0\frac{\partial h_{t-1}}{\partial \theta} + F(s_{t-1}, \lambda, \gamma)\beta_1\frac{\partial h_{t-1}}{\partial \theta}.$$

where z_t is defined as a 8×1 row vector

$$\left[1, \varepsilon_{t-1}^2, h_{t-1}, F(s_{t-1}, \lambda, \gamma), F(s_{t-1}, \lambda, \gamma)\varepsilon_{t-1}^2, F(s_{t-1}, \lambda, \gamma)h_{t-1}, \frac{\partial F(s_{t-1}, \lambda, \gamma)}{\partial \lambda}, \frac{\partial F(s_{t-1}, \lambda, \gamma)}{\partial \gamma}\right]$$

where

$$\frac{\partial F(s_{t-1}, \lambda, \gamma)}{\partial \lambda} = -s_{t-1}(w_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}) \left[1 + \exp(\lambda s_{t-1}/\gamma)\right]^{-\gamma-1} \exp(\lambda s_{t-1}/\gamma)$$

and

$$\begin{aligned} \frac{\partial F(s_{t-1}, \lambda, \gamma)}{\partial \gamma} &= (w_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}) \left[1 + \exp(\lambda s_{t-1}/\gamma)\right]^{-\gamma} \\ &\quad \left[\frac{\exp(\lambda s_{t-1}/\gamma)}{1 + \exp(\lambda s_{t-1}/\gamma)} \lambda s_{t-1}/\gamma - \ln\left(1 + \exp(\lambda s_{t-1}/\gamma)\right)\right] \end{aligned}$$

By the law of total expectations, the information matrix involves only the first derivative

$$J = E\left[\frac{\partial^2 l_t}{\partial \theta \partial \theta'} | \Psi_{t-1}\right] = -\frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial \theta} \frac{\partial h_t}{\partial \theta'} \epsilon_t^2 h_t^{-1},$$

which can be estimated by its sample analogue. However, the first derivative can only be estimated recursively.

The MLE, denoted by $\widehat{\theta}_{T,ML}$, which maximizes $l_T(Y;\theta)$, is consistent and its asymptotic distribution is normal when the true parameter vector θ_0 is not on the boundary of its parameter space, the conditional density is correctly specified and regularity conditions apply.

Covariance stationarity

In the ARCH literature, a key issue is to know whether shocks to variance is persistent or not. Moreover, the estimation of parameters generally imposes covariance stationarity. In linear models, it is quite straightforward to obtain sufficient and necessary conditions for statioanarity. However, the same problem is more complicated in a nonlinear framework. It is customary to analyze dynamics by examining the stationarity properties of the limiting processes.

Following Bollerslev(1986) and Gonzalez-Rivera(1996), the values of $F(s_{t-1}, \lambda, \gamma)$ lies between 0 and 1. In the upper regime, $F(s_{t-1}, \lambda, \gamma) = 1$, so the process is covariance stationary if and only if

$$\alpha_0 + \beta_0 + \alpha_1 + \beta_1 < 1.$$

In the lower regime, $F(s_{t-1}, \lambda, \gamma) = 0$, so the process is covariance stationary if and only if

$$\alpha_0 + \beta_0 < 1.$$

Similar conditions can be found for any other regimes. However, it is noteworthy that covariance stationarity of the upper regime implies covariance stationarity in any other regimes, but not vice versa.

Positivity of the variance is achieved by imposing the restrictions that $w_0 > 0$, $\alpha_0 > 0$, $\beta_0 > 0$, $w_0 + w_1 > 0$, $\alpha_0 + \alpha_1 > 0$ and $\beta_0 + \beta_1 > 0$.

Specification test for asymmetry

In this section, we introduce several ways to test for asymmetry in volatility in our asymmetric smooth transition GARCH model. We are interested in two types of asymmetries: the asymmetry of volatility in response of negative news and positive news, and the transition asymmetry. Therefore, we are concerned with two null hypotheses tests: $H_{10} : \lambda = 0$ and $H_{20} : \gamma = 1$. If H_{10} is true, there is no leverage effect. In contrast, if H_{20} is true, there is no transition asymmetry and our model collapses to that of Anderson et al.

(1998). To test for $H_{20} : \gamma = 1$, we can make use of the usual Wald test or t test statistic.

A testing procedure with a Taylor expansion

The complication arises from the fact that when λ is equal to zero, the parameters w_1, α_1, β_1 and θ are unidentified. Following Luukkonen et al (1988), this problem is solved by replacing $F(.)$ with a suitable linear approximation, i.e, a second order Taylor expansion of the transition function around $\lambda = 0$. The approximation of $F(.)$ is then inserted into equation (IV.1), and testing statistic such as LM test and F test in a linear framework can be implemented.

In our context, the transition function $F(s_{t-1}, \lambda, \gamma)$ can be approximated by its Taylor expansion at $\lambda = 0$, which is

$$T_1 = F(0) + F^{'}(0)\lambda$$

Thus the asymmetric smooth transition model can be approximated by

$$h_{lt} = w_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h_{lt-1} + \left[F(0) + F^{'}(0)\lambda\right](w_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{lt-1})$$

Reparameterize the above model and insert $F^{'}(0) = -2^{-\gamma-1}s_{t-1}$, we obtain the model

$$h_{lt} = w + \alpha\varepsilon_{t-1}^2 + \beta h_{lt-1} + \beta_{00} s_{t-1}\varepsilon_{t-1}^2 + \beta_{11} s_{t-1} h_{lt-1} \tag{IV.2}$$

where $\beta_{00} = -2^{-\gamma-1}\alpha_1\lambda$ and $\beta_{11} = -2^{-\gamma-1}\beta_1\lambda$.

Equation (IV.2) is a purely auxiliary model to obtain the test statistics, which may have some undesirable property of being explosive (see Granger and Andersen, 1978, p.28). The null hypothesis of $H_{10} : \lambda = 0$ can be rewritten as

$$H_{10} : \beta_{00} = \beta_{11} = 0.$$

Given the residual, ε_t is conditionally normal, a Lagrange multiplier test statistic for this hypothesis is

$$\frac{1}{2}\left\{\sum_{t=1}^{T}\frac{1}{2}\widetilde{h}_{lt}^{-1}(\epsilon_t^2\widetilde{h}_{lt}^{-1}-1)\frac{\partial h_{lt}}{\partial\theta_0}\right\}\left\{\widetilde{h}_{lt}^{-1}\frac{\partial h_{lt}}{\partial\theta_0}\left[\widetilde{h}_t^{-1}\frac{\partial h_{lt}}{\partial\theta_0}\right]'\right\}^{-1}\left\{\sum_{t=1}^{T}\frac{1}{2}\widetilde{h}_{lt}^{-1}(\epsilon_t^2\widetilde{h}_{lt}^{-1}-1)\frac{\partial h_{lt}}{\partial\theta_0}\right\} \tag{IV.3}$$

where $\theta_0 \equiv (\alpha, \beta, \beta_{00}, \beta_{11})$ is the vector of parameters in equation (IV.2), $\widetilde{h}_{lt}$ is the conditional variance under the null of GARCH(1,1) model, and $\frac{\partial h_{lt}}{\partial\theta_0}$ is the partial derivative of h_{lt} with respect to θ_0 in equation (IV.2) under the null, see Hagerud (1997) for detail.

As is pointed out by Luukkonen et al (1988), a potential problem of the testing is that λ can not be separated from β_{00} and β_{11}. When the values of γ, α_1 and β_1 are relatively small, the test may not have satisfactory power against the alternative.

A supremum LM-test with unidentified parameters

As we discussed above, the key issue in our testing procedure comes from the unidentificability problem. In other words, when the null hypothesis $H_{10} : \lambda = 0$ is true, then the parameter vector $\theta_1 \equiv (\gamma, w_1, \alpha_1, \beta_1)$ can take any values. Following Davies (1997,1998) and Gonzalez-Rivera(1998), we keep the unidentified parameters θ_1 fixed, the parameter for which the score is calculated is $\theta_0 \equiv (w_0, \alpha_0, \beta_0, \lambda)$. Under conditional normality assumption, a general form of the LM test statistic for $H_{10} : \lambda = 0$ is the same as in equation (IV.3). However, because the parameter vector θ_1 is unknown, the LM test statistic is a function of θ_1 and therefore is not feasible. We denote this test statistic $LM(\theta_0)$. Davies(1977) suggested the following test

$$T(\theta_1) = \sup_{\theta_1} LM(\theta_1)$$

for which the probability distribution is unknown. Hansen (1996) proposed a simulation based method to find the null distribution of the forgoing supremum test statistic. Let $\theta_1^* = \arg\max_{\theta_1} LM(\theta_1)$, to simulate the null distribution of $T(\theta_1)$, we need to draw $T \times J$ *iid* random variables u_{tj} from $N(0,1)$, $t = 1 \cdots T$, $j = 1 \cdots J$, and generate a sample of J-scores and $J-$test statistics:

$$
\begin{aligned}
\widehat{s}_{n,j}(\theta_1^*, 0) &= \frac{1}{2}\sum_t \widetilde{h}_t^{-1} \frac{\partial \widetilde{h}_t}{\partial \theta_0} (\epsilon_t^2 \widetilde{h}_t^{-1} - 1) u_{tj} \\
T_{n,j}(\theta_1^*, 0) &= n \widehat{s}_{n,j}(\theta_1^*, 0)' \widehat{V}^{-1}(\theta_0^*) \widehat{s}_{n,j}(\theta_1^*, 0)
\end{aligned}
$$

The approximate p-value of the supremum test statistic $T(\theta_1)$ is simply the frequency with which $T_{n,j}(\theta_1^*, 0) > T(\theta_1)$ occurs. In our simulation section, we set $J = 200$.

Monte-Carlo simulation

In this section, we conduct a Monte-Carlo simulation to investigate the finite sample performance of the two test statistics. Table 10, Table 11 and Table 12 display the size and power of the test statistics under three data generating processes. The experiment consists of 100 replications.

In Table 10, the artificial data is generated by using $\theta = [0.2, 0.2, 0.2, 0.1, 0.3, 0.3, 100, 0.8]$ in equation (IV.1). Table 10 reports the actual rejection frequencies of the two testing procedures. In addition, we can only estimate $\beta_{00} = -2^{\gamma-1}\alpha_1\lambda$ and $\beta_{11} = -2^{\gamma-1}\beta_1\lambda$, λ can not be separated from β_{00} and β_{11}. This nonidentifiability indicates that the test may not have a satisfactory power in case of small α_1, β_1 and γ. Because of all the above reasons, the linearization test does not have a good finite sample performance. In Table 11, the artificial data is generated by setting $\theta = [0.2, 0.2, 0.2, 0.1, 0.1, 0.1, 5, 0.1]$ in equation (IV.1). In Table 12, the artificial data is generated by GARCH(1,1) model by setting

$\theta = [0.2, 0.2, 0.2, 0, 0, 0, 0, 0]$ in equation (IV.1).

As we can see from the above tables, the supremum LM test has a good size for both sample sizes, but has a relatively small power when the sample size is 500. However, in case of large sample size(1500 in our simulation,) the test becomes more powerful.

Application

In this section, an asymmetric smooth transition GARCH model is estimated to financial data, and the smooth transition GARCH model, GARCH(1,1) model and the asymmetric nonlinear smooth transition GARCH model are estimated as benchmarks. The first data set is daily returns of the valued weighted NASDAQ index from January 2,1990 to December 31, 2007, consisting of 4540 observations. The second data is comprised of 4792 daily observations for the individual stock IBM, from January 2,1990 to December 31, 2008. These data have been extracted from Center for Research on Stock Prices(CRSP) database.

Table 13 reports the summary statistics for the NASDAQ index daily returns and the IBM daily returns. We find that the distribution of the daily returns depart from normality distribution by their skewness and kurtosis, and has a fat tail, which are the two key stylized facts of financial asset returns.

Table 14 presents the estimation coefficient and likelihoods to NASDAQ index for our new model, the asymmetric nonlinear smooth transition model, smooth transition model and the GARCH(1,1) model. It is apparent that there is a smooth transition between volatility regimes. We also test for the significance of the coefficients λ and θ and find that the null $\lambda = 0$ and the null $\theta = 1$ are both rejected at the 5% significance level. One noteworthy point is the the t statistic for λ does not have the standard t distribution for

$\lambda = 0$, because of the unidentificability issue we have addressed. Therefore, we need to rely on the two test statistics we propose in the previous section to test $\lambda = 0$.

Table 15 presents the estimation of conditional variance to IBM daily returns. Also, the null hypothesis $\lambda = 0$ and the null hypothesis $\theta = 1$ are rejected at the 5% significance level.

Conclusion

The asymmetric response of volatility to positive shocks and negative shocks, best known as the leverage effect has been well addresses in the financial econometrics literature. A lot of empirical models have been proposed to capture this effect with applications to stock returns and exchange rates and so on. In this chapter, we have introduced an asymmetric smooth transition model, which permits both the asymmetric responses and asymmetric transition dynamics of the shocks on the volatility. This model is a generalization of the asymmetric nonlinear smooth transition models by Anderson, Nam and Vahid(1999) and the smooth transition model by Hagerud (1997) and Gonzalez-Rivera (1998). Under certain conditions, this model nests with a lot of existing specification, such as the threshold model by Zokanian and the widely used asymmetric power model of DGE and the GJR model. Two test statistics are suggested to test whether there exists the leverage effect. A Monte-Carlo experiment has shown that the supremum LM test is preferred when the small sample is small, due to its estimation simplicity. The empirical result also shows the advantage of our new model, which is more flexible in capturing the features of financial asset return volatility.

Table 10. Simulated power of two test statistics

Test	Sample Size	Actual Rejection Frequencies(%)	
		Nominal size: 5%	Nominal size: 10%
LM test	T=500	76%	82%
	T=1500	90%	92%
Supremum LM test	T=500	85%	93%
	T=1500	100%	100%

Table 11. Simulated power of two test statistics

Test	Sample Size	Actual Rejection Frequencies(%)	
		Nominal size: 5%	Nominal size: 10%
LM test	T=500	25%	33%
	T=1500	36%	44%
Supremum LM test	T=500	41%	51%
	T=1500	91%	95%

Table 12. Simulated size of two test statistics

Test	Sample Size	Actual Rejection Frequencies(%)	
		Nominal size: 5%	Nominal size: 10%
LM test	T=500	20%	32%
	T=1500	25%	32%
Supremum LM test	T=500	17%	27%
	T=1500	11%	21%

Table 13. Summary statistics

	Mean	Median	Skewness	Kurtosis	St. Dev	Max	Min
NASDAQ	0.00	0.00	0.26	6.64	0.02	0.13	-0.16
IBM	0.00	0.00	0.16	6.38	0.01	0.14	-0.10

Table 14. Estimation of conditional variance : NASDAQ index

	ASTGARCH	ASNGARCH	STGARCH	GARCH
ω_0	0.00	0.00	0.00	0.00**
α_0	0.06**	0.06**	0.05**	0.08**
β_0	0.90**	0.90**	0.77**	0.92**
ω_1	0.00**	0.00**		
α_1	0.04**	0.04**	0.04**	
β_1	0.00	0.00**		
λ	495.69	561.13**	133.00**	
γ	0.53**			
Log Likelihood	13785.09	13786.52	13459.10	13761.42

Table 15. Estimation of conditional variance : IBM daily returns

	ASTGARCH	ASNGARCH	STGARCH	GARCH
ω_0	0.00**	0.00	0.00**	0.00**
α_0	0.05**	0.04**	0.05**	0.05**
β_0	0.92**	0.93**	0.77**	0.94**
ω_1	0.00**	0.00**		
α_1	0.02**	0.03**	0.08**	
β_1	0.01*	0.00**		
λ	56.21*	194.47	133.00**	
γ	0.09**			
Log Likelihood	12667.60	12678.32	12397.97	12650.30

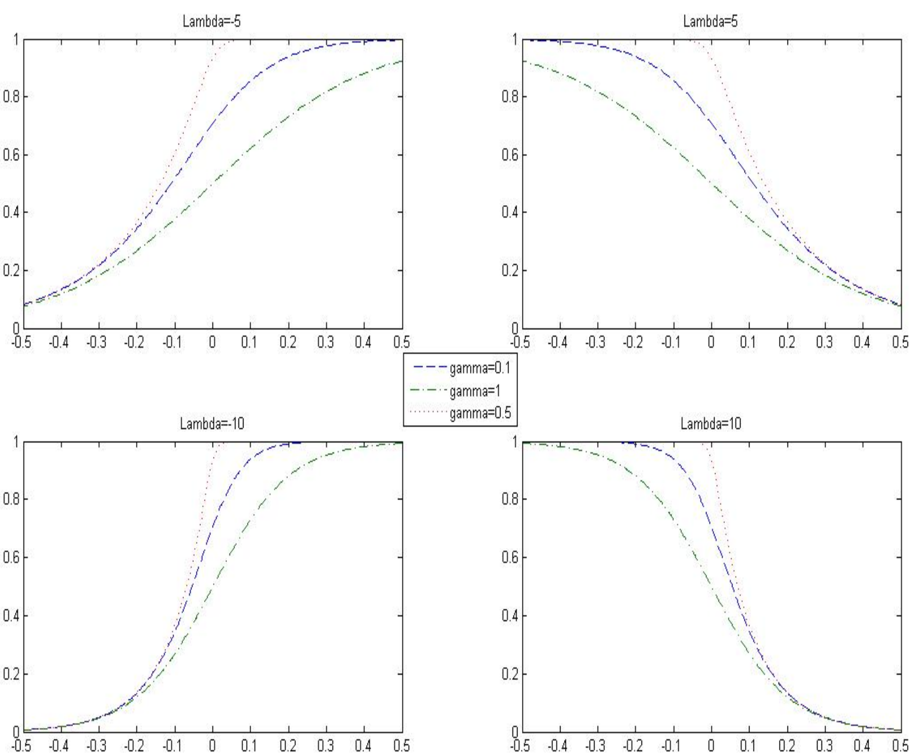

Figure 6. Transition function $F(\epsilon_{t-1}, \lambda, \gamma)$ with different λ and γ

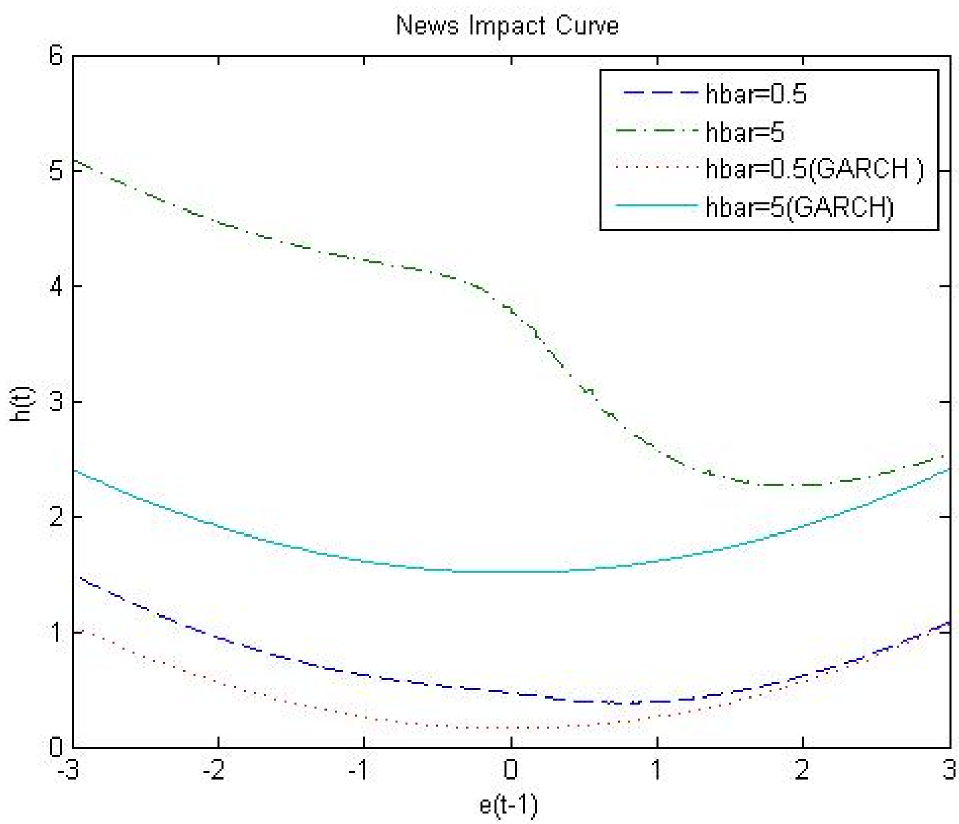

Figure 7. News impact curve for GARCH(1,1) model and the asymmetric smooth transition GARCH model for $h_{t-1} = 0.5$ and $h_{t-1} = 5$

APPENDIX A

PROOFS OF NONPARAMETRIC LAG SELECTION FOR NONLINEAR ADDITIVE AUTOREGRESSIVE MODELS

Proof of Theorem 1

Proof. We first show the asymptotic equivalence of $AFPE$ and $\widehat{FPE}(S)$ for $\eta = 0$ and $\eta > 0$. We then show that the rates of convergence for the overfitting case become slower only if $\eta > 0$.

Under Assumptions A, using the argument in the proof of Theorem 3.1 of Tschernig and Yang (2000), we have

$$
\begin{aligned}
\widehat{A} &= n^{-1}\sum_{t=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) + \sigma(\mathbf{X}_t)\xi_t - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t}) \\
&= n^{-1}\sum_{t=1}^{n}\sigma(\mathbf{X}_t)^2\xi_t^2 w(\mathbf{X}_{M,.t}) + n^{-1}\sum_{t=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t}) \\
&\quad + n^{-2}\sum_{t=1}^{n}\sum_{j=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)\sigma(\mathbf{X}_j)\xi_j w(\mathbf{X}_{M,.t})
\end{aligned}
$$

Taking expectation of $\widehat{A}$ and follow the same argument as in TA paper, we have the first terms contributes to A and the second term contributes to$\frac{1}{nh}\|K\|_2^2 B + h^4\frac{\sigma_K^4}{4}C$. Taking expectation of the third term, we have

$$
\begin{aligned}
& E\left[n^{-1}\sum_{t=1}^{n} 2\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)\sigma(\mathbf{X}_t)\xi_t w(\mathbf{X}_{M,.t})\right] \\
= \;& 2E\left[n^{-1}\sum_{t=1}^{n}\sum_{j=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) - E\left(\sum_{i\in S} f_i(Y_{t-i})\right) + E\left(\sum_{i\in S} f_i(Y_{t-i})\right) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)\sigma(\mathbf{X}_t)\xi_t w(\mathbf{X}_{M
\end{aligned}
$$

If $t = j$, the above equation can be written as

$$
\begin{aligned}
& -2E\left[\frac{K_h(Y_{t-i} - y_{i,s})}{\widehat{\mu}(\mathbf{X}_t)}\left\{\sum_{i\in S} q(x_{-i})\right\}\sigma^2(X_t)^2\xi_t^2\right] \\
= & \quad -2\frac{1}{nh}K(0)\int\frac{\sigma^2(x)}{\mu(x)}\left\{\sum_{i\in S} q(x_{-i})\right\}w(x_M)\mu(x_M)dx_M
\end{aligned}
$$

If $s \neq t$, the contributions are of order $O(T^{-1})$, following the argument from Tjøstheim and Auestad (1994) paper.

If $\eta = 0$,

$$
\begin{aligned}
\widehat{FPE}(S) &= \widehat{A} + \frac{1}{nh_{opt}}2K(0)\widehat{B} \\
&= AFPE + o_p\{(nh_{opt})^{-1}\} \\
&= A + \frac{1}{nh_{opt}}\left\{\|K\|_2^2 B + \frac{\sigma_K^4}{4}C\right\} + o_p\{(nh_{opt})^{-1}\}.
\end{aligned}
$$

If $\eta > 0$,

$$
\begin{aligned}
\widehat{FPE}(S) &= \widehat{A} + \frac{1}{nh_{opt}^{(m-1)\eta+1}}2K(0)\widehat{B}, \\
&= AFPE + \frac{1}{nh_{opt}^{(m-1)\eta+1}}2K(0)B + o_p\{(nh_{opt})^{-[(m-1)\eta+1]}\} \\
&= A + \frac{1}{nh_{opt}^{(m-1)\eta+1}}2K(0)B + o_p\{(nh_{opt})^{-[(m-1)\eta+1]}\}.
\end{aligned}
$$

Asymptotic properties of $\widehat{FPE}(S')$ with $m' > m$ are similarly obtained by replacing B and C of $AFPE$ by B' and C' and consider the new limit $AFPE'$. The result follows from the fact that $(m'-1)\eta + 1 > (m-1)\eta + 1$ as long as $\eta > 0$. ■

Proof of Theorem 2

Proof. Under Assumptions A and $\eta \in [0, 1]$, we have

$$\widehat{FPE}(S) = A + O_p\{(nh_{opt})^{-[(m-1)\eta+1]}\}$$

for correct specification. For underfitting combinations S', we have

$$\widehat{FPE}(S') = \widehat{A}' + \frac{1}{nh_{opt}^{(m-1)\eta+1}} 2K(0)\widehat{B}'$$

where

$$\widehat{A}' = n^{-1}\sum_{t=1}^{n}\left(Y_t - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t})$$

and

$$\widehat{B}' = n^{-1}\sum_{t=1}^{n}\frac{(Y_t - \sum_{i\in S'}\widehat{f}_i(Y_{t-i}))^2}{\widehat{\mu}(\mathbf{X}_t)}\left\{\sum_{i\in S'} q(\mathbf{X}_{-i,t})\right\} w(\mathbf{X}_{M,.t}).$$

The decomposition of $\widehat{A}'$ yields

$$\begin{aligned}
\widehat{A}' &= n^{-1}\sum_{t=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) + \sigma(\mathbf{X}_t)\xi_t - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t}) \\
&= n^{-1}\sum_{t=1}^{n}\sigma(\mathbf{X}_t)^2\xi_t^2 w(\mathbf{X}_{M,.t}) + n^{-1}\sum_{t=1}^{n}\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t}) \\
&\quad + n^{-1}\sum_{t=1}^{n} 2\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)\sigma(\mathbf{X}_t)\xi_t w(\mathbf{X}_{M,.t})
\end{aligned}$$

Using the argument in the proof of Theorem A.2 of Tschernig and Yang (2000), the first term converges to A, the second term converges to

$$E\left[\left(\sum_{i\in S} f_i(Y_{t-i}) - \sum_{i\in S'}\widehat{f}_i(Y_{t-i})\right)^2 w(\mathbf{X}_{M,.t})\right] = c^2 + O(h_{opt}'^2),$$

and the third term converges to zero. Thus,

$$\widehat{FPE}(S') = A + c^2 + O_p(h_{opt}'^2) + O_p\{(nh_{opt})^{-[(m-1)\eta+1]}\}$$

and the result follows by subtracting $\widehat{FPE}(S)$ from $\widehat{FPE}(S')$. ■

Proofs of Theorem 3

Proof. The result directly follows from Theorems 1 and 2. ■

APPENDIX B

PROOFS OF CONSISTENT COTRENDING RANK SELECTION WHEN BOTH STOCHASTIC AND DETERMINISTIC TRENDS ARE PRESENT

Proof of Lemma 1

Proof. We want to show that $\widehat{\lambda}_1,\cdots,\widehat{\lambda}_{r_1}$ is $O_p(1)$ but is not $o_p(1)$, $\widehat{\lambda}_{r_1+1},\cdots,\widehat{\lambda}_{r_2}$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$, and $\widehat{\lambda}_{r_2+1}$, $\cdots$, $\widehat{\lambda}_m$ is $O_p(T^{-2})$ but is not $o_p(T^{-2})$ if all the eigenvalues of $S_{11}^{-1}S_{00}$ are arranged in a descending order. We employ the data matrix notation, $Y^{'} = [y_1, \cdots, y_T]$, $D^{'} = [d_1, \cdots, d_T]$ and $S^{'} = [s_1, \cdots, s_T]$.

We have constructed an orthogonal full rank matrix $[B_\perp\ B_2\ B_1]$ in Assumption 1 and further define

$$M_{11} = B^{'}S_{11}B, \text{ and } M_{00} = B^{'}S_{00}B$$

Due to the orthogonality of the matrix $[B_\perp\ B_2\ B_1]$, the eigenvalues of $S_{11}^{-1}S_{00}$ arise as the same solutions to

$$\det(\lambda M_{11} - M_{00}) = 0.$$

Our proof can be established in the following two steps.

Step 1:

We assume $G = \lim_{T\to\infty} T^{-3}\sum_{t=1}^{T} d_t d_t'$ exists and $T^{-3}\sum_{t=1}^{T} d_t d_t' - G$ is $O(T^{-1/2})$. The eigenvalues of $T^2 M_{11}^{-1}M_{00}$ are equivalent to the eigenvalues $\lambda^{'}s$ that solve

$$\det(\lambda T^{-2}M_{11} - M_{00}) = 0$$

For the matrix $T^{-2}M_{11}$, the only block matrix that is not equal to zero is $B_\perp^{'}Y^{'}YB_\perp$, which converge to $B_{\perp}^{'} G B_{\perp}$ under Assumptions B. Because the eigenvalues are continuous functions of the matrix,

$$p \lim_{T \longrightarrow \infty} \lambda_i (T^2 M_{11}^{-1} M_{00}) = \lambda_i (p \lim_{T \longrightarrow \infty} T^2 M_{11}^{-1} M_{00}).$$

It can be easily shown that M_{00} is $O_p(1)$ but is not $o_p(1)$. Therefore, for $i = r_2 + 1, \cdots, m$, we are led to

$$\lambda_i (T^2 M_{11}^{-1} M_{00}) = O_p(1) \text{ but is not } o_p(1).$$

This leads to the result that $T^2 \widehat{\lambda}_i$ is $O_p(1)$ but not $o_p(1)$ for $i = r_2 + 1, \cdots, m$.

Step 2:

Let $D_T = diag[I_{m-r_2}, T^{1/2} I_{r_2}]$, the roots of

$$\det(\lambda T^{-2} M_{11} - M_{00}) = 0$$

are equivalent to

$$\det(D_T \left[\lambda T^{-2} M_{11} - M_{00}\right] D_T) = 0 \tag{B.1}$$

The matrix $\lambda T^{-2} M_{11}$ can be rewritten as

$$\begin{pmatrix} \lambda T^{-3} B_{\perp}^{'} Y^{'} Y B_{\perp} & \lambda T^{-3} B_{\perp}^{'} Y^{'} Y [\begin{array}{cc} B_2 & B_1 \end{array}] \\ \lambda T^{-3} [\begin{array}{c} B_2^{'} \\ B_1^{'} \end{array}] Y^{'} Y B_{\perp} & \lambda T^{-3} [\begin{array}{c} B_2^{'} \\ B_1^{'} \end{array}] Y^{'} Y [\begin{array}{cc} B_2 & B_1 \end{array}] \end{pmatrix},$$

and we denote

$$Y_a = \lambda T^{-3} B_{\perp}^{'} Y^{'} Y B_{\perp} - B_{\perp}^{'} \Delta Y^{'} \Delta Y B_{\perp}^{'},$$

$$Y_b = \lambda T^{-2} \begin{pmatrix} B_2^{'} Y^{'} Y B_2 & B_2^{'} Y^{'} Y B_1 \\ B_1^{'} Y^{'} Y B_2 & B_1^{'} Y^{'} Y B_1 \end{pmatrix} - \begin{pmatrix} T B_2^{'} \Delta Y^{'} \Delta Y B_2 & T^{1/2} B_2^{'} \Delta Y^{'} \Delta Y B_1 \\ T^{1/2} B_1^{'} \Delta Y^{'} \Delta Y B_2 & T B_1^{'} \Delta Y^{'} \Delta Y B_1 \end{pmatrix},$$

and

$$Y_c = \lambda T^{-\frac{5}{2}} \begin{pmatrix} B_2' Y' Y B_\perp \\ \\ B_1' Y' Y B_\perp \end{pmatrix} - T^{1/2} \begin{pmatrix} B_2' \Delta Y' \Delta Y B_\perp \\ \\ B_1' \Delta Y' \Delta Y B_\perp \end{pmatrix}.$$

Then equation (B.1) is rewritten as

$$\det(Y_a) \det[Y_b - Y_c' Y_a^{-1} Y_c] = 0 \tag{B.2}$$

The first determinant can on the LHS of (B.2) cannot be equal to zero, implying the second determinant must be zero. Concerning the first part of Y_b,only its first $r_2 \times r_2$ diagonal block is nonzero, and the second part of Y_b and $Y_c' Y_a^{-1} Y_c$ is $O_p(T)$ but is not $o_p(T)$. Hence, we are led to

$$\det(\lambda_i T^{-2} B_1' Y' Y B_1 - O_p(T)) = 0$$

for $i = r_1 + 1, \cdots, r_2$. While we let T goes to infinity and the solutions λ_i solves the above equation satisfies

$$\lambda_i(T^2 M_{11}^{-1} M_{00}) = O_p(T) \text{ but is not } o_p(T) \text{ for } i = r_1 + 1, \cdots, r_2.$$

Therefore, one can conclude that $\widehat{\lambda}_i$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$ for $i = r_1 + 1, \cdots, r_2$. Analogously, one can show that $\widehat{\lambda}_i$ is $O_p(1)$ but is not $o_p(1)$ for $i = 1, \cdots, r_1$. ■

Proof of Proposition 2

Proof. (i) Let r_1 be the true cotrending rank, which is estimated by minimization of $VN_1(r_1)$ for $0 \leq r_1 \leq m$. To check the consistency of this estimator, we need to show $VN(r_1') > VN(r_1)$ if r_1' is not equal to the true cotrending rank r_1.

When $r_1^{'} < r_1$,

$$VN_1(r_1^{'}) - VN_1(r_1) = \sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1}$$

In order to consistently select r_1 with probability 1 as $T \to \infty$, we need

$$\sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1} > 0, \text{ as } T \to \infty.$$

From Proposition 1, we know the first term is a positive number that is bounded away from zero and the second term is a negative number of order $O(C_T T^{-1})$. As long as $C_T T^{-1} \to 0$ as $T \to \infty$, the above inequality holds and we are led to the conclusion that $VN_1(r_1^{'}) > VN_1(r_1)$ when $r_1^{'} < r_1$.

When $r_1^{'} > r_1$,

$$VN_1(r_1^{'}) - VN_1(r_1) = -\sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1}$$

From Proposition 1, we know that $\widehat{\lambda}_i$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$ for $i = r_1 + 1, \cdots r_2$, By multiplying both sides by T, we have

$$T\left(VN_1(r_1^{'}) - VN_1(r_1)\right) = -T\sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T.$$

As long as $C_T \to \infty$ as $T \to \infty$, the second term on the right hand side dominates, which leads to $VN_1(r_1^{'}) > VN_1(r_1)$ when $r_1^{'} > r_1$. Thus the consistency of $VN_1(r_1)$ in selecting true cotrending rank is established. Analogously, one can establish the consistency of the estimator of the true weak cotrending rank by $VN_2(r_2)$.

(ii) To show the consistency of the joint selection procedure, consider all the possible cases as follows.

Case 1: $r_1^{'} < r_1$,

We have

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = \sqrt{T} \sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + O_p(\frac{C_T}{T}),$$

where $\widehat{\lambda}_i$ for $i = r_1^{'} + 1, \cdots, r_1$ is $O_p(1)$ but is not $o_p(1)$.

From Proposition 1 and Lemma 1, the first term dominates, which leads to $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ when $r_1^{'} < r_1$.

Case 2: $r_1^{'} > r_1$.

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = -\sqrt{T} \sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))\frac{C_T}{T} + O_p(\frac{C_T}{T^2}),$$

where $\widehat{\lambda}_i$ is $O_p(T^{-1})$ for $i = r_1 + 1, \cdots, m$.

The dominant term in the above equation is $(f(r_1^{'}) - f(r_1))\frac{C_T}{T}$ provided that $\frac{C_T}{\sqrt{T}} \to \infty$, the inequality $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ holds in this case.

Case 3: $r_1^{'} = r_1$.

When $r_2^{'} > r_2$,

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = -\sqrt{T} \sum_{i=r_2+1}^{r_2^{'}} \widehat{\lambda}_i + (f(r_2^{'}) - f(r_1))\frac{C_T^{'}}{T^2},$$

where $\widehat{\lambda}_i$ is $O_p(T^{-2})$ for $r_2^{'} + 1, \cdots, m$.

Then, we have

$$T^2 \Big(VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2)\Big) = -\sqrt{T} \sum_{i=r_2+1}^{r_2^{'}} T^2\widehat{\lambda}_i + (f(r_2^{'}) - f(r_2))C_T^{'}.$$

Provided that $\frac{C_T}{\sqrt{T}} \to \infty$, the dominant term is $(f(r_2^{'}) - f(r_2))C_T^{'}$, which is greater than zero. Hence $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ in this case.When $r_2^{'} < r_2$,

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = \sqrt{T} \sum_{i=r_2^{'}+1}^{r_2} \widehat{\lambda}_i + (f(r_2^{'}) - f(r_2))\frac{C_T^{'}}{T^2}.$$

The first term on the right hand side is $O_p(T^{-3/2})$ but is not $o_p(T^{-3/2})$, dominate the second term, provided that $\frac{C_T^{'}}{T} \to 0$. Hence $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ in this case.Combining the conditions on C_T and $C_T^{'}$, for all the preceding cases, it follows that the joint selection procedure will lead to consistent estimation of the cotrending and weak cotrending rank ■

BIBLIOGRAPHY

[1] Akaike, H., 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical. Mathematics*, 21, 243-247.

[2] Akaike, H., 1970. Statistical predictor identification. *Annals of the Institute of Statistical. Mathematic 2, 2*03-217s*, 2.*

[3] Anderson, H. M., Nam, K., Vahid., F., 1999. Asymmetric nonlinear smooth transition GARCH models. *Nonlinear Time Series Analysis of Economic and Financial Data,* 191-207*.*

[4] Bae, Y., Kakkar, V., Ogaki, M., 2006. Money demand in Japan and nonlinear cointegration. *Journal of Money, Credit, and Banking* 38, 1659-1667.

[5] Bacon, D. W., Watts, D. G., 1971. Estimating the transition between two intersecting straight lines. *Biometrika* 58, 525–534.

[6] Bai, J., Ng, S., 2002. Determine the number of factors in approximate factor models. *Econometrica* 70, 191-221*.*

[7] Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127-1177.

[8] Bierens, H. J., 2000. Nonparametric nonlinear co-trending analysis, with an application to inflation and interest in the U.S. *Journal of Business and Economic Statistics* 18, 323–337.

[9] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[10] Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics* Vol 69, No.3, 542-547.

[11] Bollerslev, T., Engle, R.F., Nelson, D., 1994. ARCH models. Handbook of Econometrics, Vol.4, Chp. 49, 2959-3038.

[12] Chen, R., Liu, J. S., Tsay, R. S., 1995. Additivity test for nonlinear autoregression. *Biometrika* 80, 369-383.

[13] Chen, C.W.S., Chiang, T.C., So, M.K.P., 2003. Asymmetric reaction to US stock return news: Evidence from major stock markets based on a double-threshold model. *Journal of Economics and Business* 55, 487-502.

[14] Chen, R., Tsay, R.S., 1993. Nonlinear additive ARX model. *Journal of the American Statistical Association* 88, 955-967.

[15] Cheng, X., Phillips, P. C. B., 2009. Semiparametric cointegrating rank selection. *Econometrics Journal* 12, S83–S104.

[16] Clements, M. P., Hendry, D. F., 1999. On winning forecasting competitions in economics. *Spanish Economic Review* 1, 123–160.

[17] Engle, R. F., Granger, C. W. J., 1987. Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251-276.

[18] Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254.

[19] Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33-43.

[20] Engle, R.F., 1983. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50,987-1008.

[21] Engle, R. F., Kozicki, S., 1993. Testing for common features. *Journal of Business & Economic Statistics* 11, 369-38.

[22] Fan, J , Yao, Q., 2005. Nonlinear Time Series: Nonparametric and Parametric. Springer Series in Statistics.

[23] Fornari, F., Mele, M., 1997. Sign and volatility switching ARCH models: theory and applications to international stock markets. *Journal of Applies Econometrics*, 12,49-65.

[24] Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749-1778.

[25] Glosten, L., Jagannanthan, R., Runkle, D.E., 1993. On the relationship between the expected value and volatility of the nominal excess returns on stocks. *Journal of Finance* 48, 1779-1801.

[26] Gonzalez-Rivera, G.,1996. Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3, 61-78.

[27] Hagerud, G.E., 1997. Specification tests for asymmetric GARCH. working paper, Stockolm School of Economics.

[28] Hagerud, G.E., 1997. Specification tests for asymmetric GARCH. working paper, Stockolm School of Economics.

[29] Hardle, W., Lutkepohl, H., Chen, R. (1997) A review of nonparametric time series analysis. *International Statistical Review* 65, 49-72.

[30] Harris, D., Poskitt, D. S., 2004. Determination of cointegrating rank in partially non-stationary process via a generalised von-Neumann criterion. *Econometrics Journal* 7, 191-217.

[31] Hatanaka, M., 2000. How to determine the number of relations among deterministic trends., *The Japanese Economic Review* 51, 349-373.

[32] Hatanaka, M., Yamada, H., 2003. Co-trending: A Statistical System Analysis of Economic Trends, Springer.

[33] Hendry, D. F., Mizon, G. E., 1998. Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics* 23, 267-294.

[34] Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64,413-430.

[35] Huang, J. Z., Yang, L., 2004. Identification of non-linear additive autoregressive models. *Journal of the Royal Statistical Society* B 66, 463-477.

[36] Laakkonen, H., Lanne, M., 2008. Asymmetric News Effects on Volatility: Good vs. Bad News in Good vs. Bad Times. MPRA Paper 8296, University Library of Munich, Germany

[37] Leybourne, S., Newbold, P., Vougas, D., 1998. Unit roots and smooth transitions. *Journal of Time Series Analysis* 19, 83–97.

[38] Lin, C.-F. J., Teräsvirta, T., 1994. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* 62, 211-228.

[39] Linton, O. B., Nielsen, J. P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93-100.

[40] Johansen, S., 1991. Estimation and hypothesis test of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

[41] Mammen, E., Nielsen, J. P., 2003. Generalised structured models. *Biometrika* 90, 551-566.

[42] Mammen, E., Linton, O., Nielsen, J., 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Annals of Statistics 27, 1443-1490.

[43] Hardle, W., Lutkepohl, H. and Chen, R., 1997. A review of nonparametric time series analysis. *International Statistical Review* 65, 49-72.

[44] Martins-Filho, C., Yang, K., 2007. Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion.. *Journal of Nonparametric Statistics* 19, 23-62.

[45] Medeiros, M.C., Veifa. A., 2009. Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model. *Econometric Theory*, 25, 117-161.

[46] Mills, T., 2003. Modelling Trends and Cycles in Economics Time Series, Palgrave Macmillan.

[47] Nelder,J..A., 1961. The fitting of a generalisation of the logic curve. *Biometrics*, 17, 89-1\10.

[48] Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, Vol 59, 347-370.

[49] Nelson, C. R., Plosser C. R., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* 10, 139-162.

[50] Ogaki, M., Park, J. Y., 1998. A cointegration approach to estimating preference parameters. *Journal of Econometrics* 82, 107–134.

[51] Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361-1401.

[52] Perron, P., 1997. Further evidence from breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355-385.

[53] Rabemananjara, R., Zakoian, J.M., 1993. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics*, Vol. 8, 31-49.

[54] Rappoport, P., Reichlin, L., 1989. Segmented trends and non-stationary time series. *Economic Journal* 99, 168-77.

[55] Sperlich, S., Linton, O. B., Härdle, W. 1999. Integration and backfitting methods in additive models - finite sample properties and comparison, *TEST* 8, 419-458.

[56] Shintani, M., 2001. A simple cointegrating rank test without vector autoregression. *Journal of Econometrics* 105, 337-362.

[57] Stock, J. H., Watson, M. W., 1988. Testing for common trends. *Journal of the American Statistical Association* 83, 1097-1107.

[58] Sollis, R., Leybourne, S. and Newbold, P., 1998. Unit root and asymmetric smooth transition. *Journal of Time Series Analysis* Vol 20, No.6, 671-688.

[59] Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208-218.

[60] Tjøstheim, D., Auestad, B., 1994. Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association* 89, 1410-1419.

[61] Tschernig, R., Yang, L., 2000. Nonparametric lag selection for time series. *Journal of Time Series Analysis* 21, 457-487.

[62] Yoshihara, K., 1976. Limiting Behavior of U-statisticsfor stationary absolutely regular processes. Zeitschrift fur Wahrscheinlichkeitstheories undverwandte Gebiete, 35, 237-252.