FORECASTING FINANCIAL RETURNS: A COPULA-BASED METHOD AND A ROBUST TEST

A Dissertation

by

GUANNAN LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ke-Li Xu
Co-Chair of Committee,	Qi Li
Committee Members,	Dennis W. Jansen
	Ximing Wu
Head of Department,	Timothy J. Gronberg

May 2016

Major Subject: Economics

ABSTRACT

My dissertation includes two essays studying the forecasting of financial returns. In the first essay, I study the temporal dependence structures of financial returns by using a mixture copula model. A mixture copula is a linear combination of several single copulas. It is more flexible than a single copula and can capture various dependence structures in financial data. Therefore, instead of choosing a single copula based on certain statistical criteria, I propose to use a model average approach to estimate the temporal dependence structure of a stationary Markov process in a mixture copula framework. The asymptotic properties of the model average estimator are established under some regularity conditions. Simulations show that the model average approach gives the most accurate estimation and predicting results compared to some competing methods, when the working mixture model is misspecified. Using a real data example, we demonstrate the usefulness of our proposed method.

In the second essay, I suggest a robust test that is a data-dependent weighted average of the regression-based test and the covariance-based test. This new test allows for multivariate cases and yields chi-squared inference regardless of whether predictors are stationary, local-to-unity or $I(1)$. No prior knowledge of the orders of integration or bias corrections are required. Furthermore, the new test does not force the dependent variable and predictors to share the same order of integration under the alternative hypothesis. It is very important because in practice the dependent variable usually appears to be stationary while predictors may be (near) nonstation-ary. This test shows good simulation results. In the empirical application section, we test for the predictability of excess stock returns using a large set of predictors.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation includes two essays, the name of the first essay is "Investigating Temporal Dependence in Financial Data via Mixture Copulas", and the name of the second essay is "A Robust Test for Predictability with Unknown Persistence".

1.1 Introduction to the First Essay

A substantial body of literature is available in economics and finance, which studies the temporal dependence of financial returns. A good knowledge of the temporal dependence structures is essential for many important financial applications, such as risk management and financial variables forecasting. However, much literature focuses only on linear temporal dependence through linear autocorrelation analysis, although more general dependence patterns could exist. To model nonlinear temporal dependence in a time series, a copula-based method can be a good choice because it is flexible enough to separate the temporal dependence from the marginals. Therefore, by choosing different forms of margins and copula functions, one can model a wide variety of temporal dependence properties (such as asymmetry or clusters) and marginal behaviors (such as fat tails).

By the theorem shown in Sklar (1959), any multivariate joint distribution can be written in terms of its marginal distributions and a copula function. Because of its flexibility, the copula model has been widely used in finance and economics. For example, Li (2000) and Frey and McNeil (2001) propose the use of copula models to estimate default correlations. Chollete et al. (2005), Hu (2006), Chollete et al. (2009), and Long et al. (2015) apply the copula method to study the contemporaneous dependence structure among international stock markets. Cherubini et al. (2004) uses a copula-based approach to measure the portfolio Value-at-Risk.

To describe the temporal dependence of a univariate time series, Joe (1997) studies a class of stationary Markov models by using parametric margins and copula functions. He applies the parametric copula method to daily environmental data. Bouyé et al. (2002) measure the temporal dependence in a time series by using different types of Archimedean copula functions for nonlinear dependence. Chen and Fan (2006) and Chen et al. (2009) propose a semiparametric copula approach characterized by nonparametric margins and parametric copula functions.

For empirical studies, a crucial question is how to choose an appropriate copula to satisfactorily capture the dependence patterns. In the context of modeling the contemporaneous dependence among several variables, there have been some efforts in the literature to choose an individual copula from a candidate set, usually based on certain statistical criteria[1]. One may consider extending these methods to univariate time series. However, two disadvantages of using an individual copula should be considered. First of all, there are many different varieties of copulas. In fact, one can always create a new copula by making some transformations on an existing copula[2]. However, taking a large candidate set is usually very inefficient in practice. Most empirical users only consider several commonly used copulas, such as Gaussian, Clayton, and Gumbel, to build their candidate sets. Considering the fact that the true copula model is never known to econometricians, it is highly probable that one's candidate set fails to include the true copula or all the candidates are not very close to the true dependence function. Under these circumstances, the selected "most appropriate" copula might fail to capture the true dependence structure. Second,

[1]For example, Chen, Fan, and Patton (2003), Fermanian (2005), and Scaillet (2007) propose the use of Goodness-of-Fit (GoF) tests to select a copula. Manner and Reznikova (2012), Patton (2012), and Fan and Patton (2014) compare the log-likelihood function values of each copula and choose the one having the largest value.

[2]For example, Patton (2006) proposes a Joe-Clayton copula by taking a particular Laplace transformation on the BB7 copula of Joe (1997).

there is usually no single copula that applies to all cases in practice. For example, when studying the dependence structure across international stock markets, different pairs of markets may exhibit different dependence structures; therefore, they can not be captured by one individual copula (see Hu, 2006).

To overcome the shortcomings of using an individual copula function, Chollete, Peña, and Lu (2005) and Hu (2006) suggest using mixture copulas in the context of multivariate models. A mixture copula is a linear combination of several individual copulas. The weight of each individual copula is non-negative and the weights add up to one. Compared to an individual copula, a mixture copula is more flexible and can generate dependent structures that do not belong to any single copula. Therefore, by arranging different weights on components of copulas in a mixture model, one can capture various dependence structures in the financial data, such as tail dependence or asymmetric dependence. For example, Chollete et al. (2005) and Hu (2006) consider mixture models, including Gaussian, Gumbel, and rotated Gumbel copulas, to estimate the dependence structures among international stock markets. They find strong left tail dependence as the weight parameter associated with the Gumbel copula is almost zero, while the rotated Gumbel tends to be significant due to its positive weight.

When modeling the temporal dependence structure of a univariate time series, most of the previous literature rely only on a single copula (e.g., Bouyé et al., 2002, Chen and Fan, 2006[3], and Chen et al., 2009). Because of the disadvantages associated with the use of a single copula, we use a mixture copula model to study the temporal dependence in time series (copula-based stationary Markov models) in this paper. Although a mixture copula model is much more flexible than a single copula,

[3]Although the two-step Quasi-MLE method proposed in Chen and Fan (2006) can be easily extended to estimate a mixture copula model.

it is still highly probable that one's working model is misspecified. That is, the true copula may not be included in the working mixture copula candidate set because the true data dependence structure is unknown to econometricians. To handle copula misspecification cases, we propose to use a model average approach to estimate a mixture copula[4]. Specifically, we first fit observations to each component copula in the mixture model, and then we estimate their associated weights by minimizing a cross-validation criterion in a manner similar to the one proposed by Hansen (2007). Similar to Chen and Fan (2006), we estimate the unknown margins by nonparametric methods, such as the rescaled empirical distribution function, while assuming that copulas belong to some parametric families. We show that our model average approach can generate an asymptotically optimal estimator in the sense of achieving the infeasible lowest possible squared estimation losses. That is, the model average approach asymptotically minimizes the distance between the estimated mixture copula and the unknown true model when the working model is misspecified. This is important in practice, considering the fact that empirical researchers usually take a small copula candidate set, and the misspecification problem should be common.

The simulation results indicate the superiority of the model average approach in capturing the temporal dependence in the time series. Compared with a mixture copula estimated by Chen and Fan's (2006) two-step Quasi-MLE method (MML) and a single copula selected by the standard BIC method (BIC), our model average approach generates estimation results with the smallest mean square errors, especially when the copula model is misspecified. Estimating conditional quantiles (VaR) of financial returns is important in risk management; therefore, we also compare the

[4]In a different context, Long et al., 2015 use the model average approach to estimate the contemporaneous dependence patterns among several variables. To the best of our knowledge, there is no other paper that considers using a mixture copula and/or a model average approach to estimate the temporal dependence of copula-based Markov models.

plug-in estimators of conditional quantiles estimated via different methods. Again, our model average approach surpasses the other two methods by registering the smallest errors. In the empirical application, we apply the model average approach to estimate the temporal dependence of the daily returns of several equity indexes (NASDAQ, SHASHR, KOSPI, and TAIEX). The empirical results show our model average approach can be a useful tool in describing and predicting the temporal dependence structures of financial returns and in risk management.

Finally, I would like to clarify that the model average method and the results shown in this paper can be easily extended to copula-based Markov processes of any finite order, although we only present results for first-order Markov models for the sake of clarity in this paper.

1.2 Introduction to the Second Essay

Many papers in economics and finance study the predictability of financial returns. The traditional framework assumes a linear regression

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t, \tag{1.1}$$

where y_t is financial returns such as stock or bond returns during period t, x_{t-1} is a lagged variable which could be, for example, dividend yields or interest rates at the end of period $t-1$, and ε_t is usually assumed to be serially uncorrelated and $E(\varepsilon_t \mid x_{t-1}, x_{t-2}, ..., x_1) = 0$, $E(\varepsilon_t \mid x_t) \neq 0$. When x_t is stationary and the sample size is large, this kind of predictive regression works well and so does the standard t-ratio test, since the finite sample bias disappears asymptotically. However, when

x_t is local-to-unity, for example,

$$x_t = \rho x_{t-1} + \upsilon_t,$$
$$\rho = (1 + c/T) \text{ and } c < 0,$$

and

$$E(\varepsilon_t \upsilon_t) \neq 0,$$

it is well-known that the bias is still present even for a large sample and is uncorrectable as it depends on the local-to-unity parameter c, which can not be consistently estimated (e.g., Cavanagh, Elliott, and Stock, 1995). Thus, predictive regression tests suffer from substantial size distortion.

Moreover, the predictive regression forces the dependent variable and the independent variable to have the same orders of integration under a fixed alternative. In other words, if x_t is (near) nonstationary, y_t is implicitly assumed to be (near) nonstationary and cointegrated with x_t. With y_t (such as stock returns) empirically showing only weak serial correlation, it is reasonable to think the dependent variable is always stationary. This predictive regression does not allow the case that x_t is (near) nonstationary while y_t is stationary unless β_1 is zero. In consequence, the power of predictive regression tests is very weak and the test statistics do not diverge to infinity under fixed and unbalanced alternatives.

The size distortion problem mentioned above has recently generated a large volume of literature aimed at correcting inference. However, most of the existing literature is based on the predictive regression and studies test performance and asymptotics only under the null hypothesis (and the local alternatives), forgoing problems of imbalance under the fixed alternative (e.g., Cavanagh, Elliott, and Stock, 1995;

Torous, Valkanov, and Yan, 2004; Valkanov, 2003; Campbell and Yogo, 2006; Magdalinos and Phillips, 2009b; Kostakis et al., 2010; Phillips and Lee, 2013; Cai and Wang, 2014). Therefore, although these procedures can effectively control the size, they are usually hard to maintain a strong power against fixed alternatives. Moreover, a common assumption in the predictive model literature is that there is no drift in the AR(1) model for the regressors. That is because for the predictive regression models, fitting a drift to the regressiors often affects asymtotics in nonstationary cases. However, this assumption may not hold for all applications.

In contrast to regression-based tests, Maynard and Shimotsu (2009) develop a test based on the covariance between y_t and x_{t-1}. This kind of covariance-based test allows unbalanced fixed alternatives that are not covered by the regression model, i.e., when x_t is $I(1)$ or local-to-unity, y_t can still maintain stationary behavior. The covariance-based test allows either a nonzero intercept or a linear trend in (y_t, x_t) and has a same limiting distribution using demeaned or detrended residuals.

Another favorable property of the covariance-based method is that it generates a single t-type test that has the same asymptotic distribution under either unit root or local-to-unity assumption. That is, when x_t is local-to-unity, the t-type statistic has a standard normal limit distribution which does not depend on the local-to-unity parameter c. Therefore, there is no need to worry about the uncorrectable bias problem as in the predictive regression model. This covariance-based test is also shown to be consistent against the local $o(n^{-1/2})$ versions of the traditional linear regression alternative although it can not provide consistency against local regression alternatives of order n^{-1}.

However, the covariance-based t test performs poorly when both of y_t and x_t are modeled as stationary processes. The test is undersized and the test statistic converges to zero in probability under the null. Another drawback is that Maynard and Shimotsu (2009) retrict the covariance-based test to the case of a single predictive variable, which is not general in practice.

From what has been discussed above, if we know the orders of integration of the dependent variable and the predictors before testing, we may choose either the standard test from the predictive regression model which is good enough when both y_t and x_t are stationary, or, the covariance-based test when y_t and x_t have different orders of integration. However, in practice, we are not equipped with much prior information about whether a predictor is stationary, local-to-unity or $I(1)$, although the dependent variable can be always treated as a stationary process. Therefore, we construct a new test statistic that is a data-dependent weighted average of the regression-based test and the covariance-based test. Our test can automatically select the regression-based test statistic when predictors are stationary, and select the covariance-based test statistic when predictors are (near) nonstationary.

This test has many desirable properties. First of all, it allows the existence of unbalanced alternatives. Second, no bias corrections are needed. Third, we extend Maynard and Shimotsu (2009)'s covariance-based test to a multivariate system. Our test allows more than one predictor in the model. Fourth, neither a nonzero intercept nor a linear trend in (y_t, x_t) would affect the limiting distribution. Finally, our test yields a test statistic that has a standard χ^2 limiting distribution regardless of whether the regressors are stationary, local-to-unity or $I(1)$, i.e., no prior knowledge of the orders of integration is required, which is very important empirically. This test could successfully control the size while maintaining a strong power against both local and fixed alternatives in either balanced or unbalanced case.

2. INVESTIGATING TEMPORAL DEPENDENCE IN FINANCIAL DATA VIA MIXTURE COPULAS

2.1 Copula-Based Markov Models

Several previous papers, including Darsow et al. (1992), Victor et al. (2006), and Ibragimov (2009) have presented characterizations of a copula-based time series to be a Markov process. A brief review can be found in Nelsen (2006, section 6.4). In this paper, we denote $\{Y_t\}_{t=1}^{T}$ as a first-order stationary Markov process with continuous state space and a marginal distribution which is denoted as G. Then, the probabilistic properties of $\{Y_t\}_{t=1}^{T}$ are captured by the joint distribution function of Y_t and Y_{t-1}, which is denoted as $F(y_1, y_2)$. According to Sklar (1959), one can always model $F(y_1, y_2)$ by modeling the marginal distribution G and the copula function of Y_{t-1} and Y_t separately, and the copula model does not constrain the choice of margins

$$F(y_1, y_2) \equiv C(G(y_1), G(y_2)).$$

Thus, the copula approach is a useful tool in modeling the temporal dependence of a stationary Markov model. Because of the flexibility of the copula approach, we can model a wide variety of marginal behaviors (such as fat tails and/or skewness) and temporal dependence properties (such as asymmetry and/or clusters) by coupling different forms of margins and copula functions together.

Following Chen and Fan (2006) and Chen et al. (2009), we have the assumption about the true data generating process (DGP):

Assumption 1. (i) $\{Y_t\}_{t=1}^{T}$ is a sample of a strictly stationary first-order Markov process generated from $(G_0(\cdot), C_0(\cdot, \cdot; \theta_0))$, where $G_0(\cdot)$ is the true invariant distribution that is absolutely continuous with respect to the Lebesgue measure on the real line; $C_0(\cdot,\cdot;\theta_0)$ is the true parametric copula for Y_{t-1} and Y_t up to the unknown value θ_0 and is absolutely continuous with respect to the Lebesgue measure on $[0,1]^2$. (ii) The true marginal density $g_0(\cdot)$ of $G_0(\cdot)$ is positive on its support and the true copula density $c_0(\cdot,\cdot;\theta_0)$ of $C_0(\cdot,\cdot;\theta_0)$ is positive on $(0,1)^2$.

Remark 1. The absolute continuity assumption of the bivariate copula $C_0(\cdot,\cdot;\theta_0)$ stated in Assumption 1(i) rules out the Fréchet–Hoeffding upper and lower bounds, as well as their linear combinations.

Remark 2. As indicated by Chen and Fan (2006), any process satisfying Assumption 1 is a β-mixing process.

Under Assumption 1(i), one can easily derive the true conditional density function of Y_t given Y_{t-1} (say $f_0(\cdot \mid Y_{t-1})$):

$$f_0(\cdot \mid Y_{t-1}) = g_0(\cdot)c_0(G_0(Y_{t-1}), G_0(\cdot);\theta_0),$$

where $g_0(\cdot)$ is the marginal density function and $c_0(\cdot,\cdot;\theta_0)$ is the copula density function.

We can easily see that under Assumption 1(i), the transformed process $U_t \equiv G_0(Y_t)$ is a stationary first-order Markov process with uniform margins. The true copula for U_{t-1} and U_t is $C_0(\cdot,\cdot;\theta_0)$. Then, $C_{2|1}[\cdot \mid u;\theta_0] \equiv \frac{\partial}{\partial u}C_0(u,\cdot;\theta_0) \equiv C_1(u,\cdot;\theta_0)$ is the conditional distribution of $U_t \equiv G_0(Y_t)$ given $U_{t-1} = u$; and $C_{2|1}^{-1}[q \mid u;\theta_0]$ is the qth, $q \in (0,1)$, conditional quantile of U_t given $U_{t-1} = u$. Finally, the qth conditional quantile of Y_t given Y_{t-1} is

$$Q_q^Y(y) = G_0^{-1}(C_{2|1}^{-1}[q \mid G_0(y);\theta_0]).$$

By definition, $C_{2|1}^{-1}[q \mid u; \theta_0]$ is monotonic across different q values. Therefore, $Q_q^Y(y)$, the qth conditional quantile of Y_t given Y_{t-1}, also increases in q. As indicated by Chen and Fan (2006) and Chen et al. (2009), this is another attractive feature of the copula-based Markov model [1].

2.2 An Introduction to Mixture Copula

Since the true joint distribution and the true copula function are always unknown, one may use a mixture copula to estimate the joint distribution. A mixture copula is a weighted average of several individual copulas. Specifically, a mixture copula model is formulated as

$$C(\mathbf{u}; \boldsymbol{\theta}, \omega) = \sum_{k=1}^{L} \omega_k C_k(\mathbf{u}; \theta_k) = \sum_{k=1}^{L} \omega_k C_k \left\{G_0(y_1), G_0(y_2); \theta_k\right\}, \tag{2.1}$$

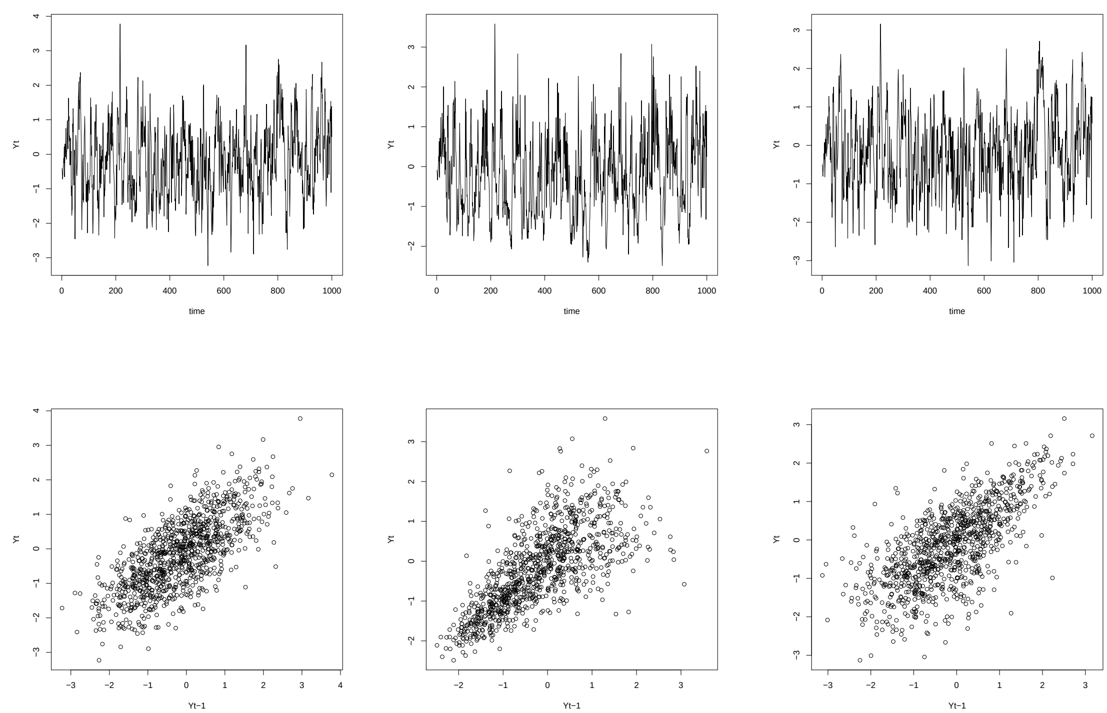
where $\{C_1(\cdot), ..., C_L(\cdot)\}$ is a set of candidate copulas with unknown associated parameters $\theta = (\theta_1, \ldots, \theta_L)^{\mathrm{T}}$ and marginal distribution $\mathbf{u} = (G_0(y_1), G_0(y_2))$.

The unknown parameters that needed to be estimated in equation (1) could be separated into three categories: $\theta = (\theta_1, \ldots, \theta_L)^{\mathrm{T}}$ is a vector of dependence parameters which control the degree of dependence. $\omega = \{\omega_1, \ldots, \omega_L\}$ denote the weight parameters. The weight parameters are also called shape parameters as they reflect the shape or the structure of dependence for the mixture copula and indicate how much credence we should place in the estimated dependence parameters for the associated candidate copula. As ω represents the weight, constraints should be imposed to guarantee that $0 \leq \omega_l \leq 1$ and $\sum_{l=1}^{L} \omega_l = 1$. The unknown marginal distributions can be estimated using either parametric or nonparametric method.

It is straightforward to show that a mixture copula is also a copula and thus have all copula properties. Comparing with individual copula, mixture copula is more

[1] See also Bouyé and Salmon (2009)

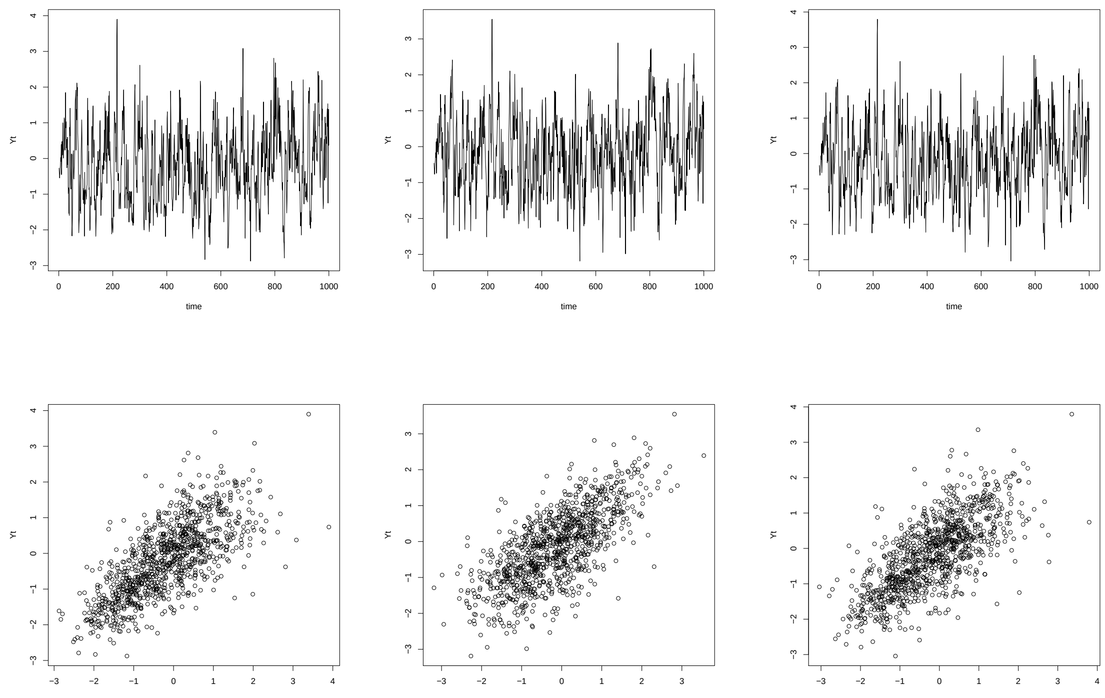
flexible and can generate dependent structures that do not belong to any individual copula. By arranging weights on different component copulas in the mixture model, we can capture various dependence structures in data. Although ideally one should include as many as possible component copulas into the mixture model to cover every possible dependent pattern, this would make the mixture model too complicated and the estimation burden too heavy.



(a) Gaussian Copula (b) Clayton Copula (c) Gumbel Copula

Figure 2.1: Time series plots and scatter plots for three single copulas

In practice, we usually consider only a few candidate copulas, which is simple but flexible enough to capture the data dependence structure. Figure 2.1 display time



(a) Gaussian-Clayton Mixture (b) Gaussian-Gumbel Mixture (c) Clayton-Gumbel Mixture

Figure 2.2: Time series plots and scatter plots for three mixture copulas

series plots and the corresponding scatter plots of realizations generated from three individual copula functions: Gaussian, Clayton, and Gumbel. By selecting the copula parameter for each copula function, we generate these three different dependence structures with the same degree of dependence (Kendall's $\tau = 0.5$). Moreover, all margins are generated from the standard normal distributions. We could observe from Figure 2.1 that Gaussian copula exhibits symmetric dependence structure, while Clayton copula and Gumbel copula show asymmetric dependence structures. In particular, Clayton copula shows strong left tail dependence and Gumbel copula displays strong right tail dependence. We then generate three different mixture copulas with equal weights (1/2) on each component, shown in Figure 2.2. It could be observed that after mixing individual copulas, a mixture copula could exhibit a totally new dependence structure which is not existing in component copulas. For example, the mixture copula of Gaussian and Clayton (or Gumbel) shows only moderate asymmetric tail dependence, and the mixture of Clayton and Gumbel exhibit an almost symmetric dependence structure.

In the next section, we discuss our procedure of estimating a mixture copula model and prove that our estimation method lead to asymptotically minimum estimation loss.

2.3 Theoretic Model

We use a mixture copula model to estimate the true copula function $C_0(\cdot, \cdot; \theta_0)$. The estimation procedure includes three stages. In the first stage, we estimate the marginal distribution using a nonparametric method. Then, the copula parameters in each component copula can be estimated by the Quasi-MLE (QMLE) after replacing the unknown margins with the estimators obtained from the first stage. In the last stage, we select the weight of each component copula by the Cross-Validation (CV) method.

Specifically, assuming our mixture copula candidate set includes $K-1$ individual copulas, we denote each candidate copula as

$$C_k(\mathbf{u};\boldsymbol{\theta}_k) = C_k\left\{G_0(y_1), G_0(y_2);\boldsymbol{\theta}_k\right\}, \qquad k = 1,\ldots,K-1, \tag{2.2}$$

where $G_0(\cdot)$ is the true (but unknown) marginal density of Y_t, $\mathbf{u} = (G_0(y_1), G_0(y_2))$ is an arbitrary point in $[0,1]^2$, and $\boldsymbol{\theta}_k$ is a finite dimensional parameter associated with the k^{th} copula.

Write $C_0(\mathbf{u};\boldsymbol{\theta}_0) = C_0\left\{G_0(y_1), G_0(y_2);\boldsymbol{\theta}_0\right\}$ as the true copula. Note that $C_0(\mathbf{u};\boldsymbol{\theta}_0)$ can be neither an individual copula inside the candidate set nor a mixture copula that is a linear combination of $\{C_1(\mathbf{u};\boldsymbol{\theta}_1),\ldots,C_{K-1}(\mathbf{u};\boldsymbol{\theta}_{K-1})\}$. In other words, the working mixture model may be misspecified.

In the first stage, we use the rescaled empirical distribution function to estimate the marginal distribution of Y_t:

$$\widetilde{F}(y) = \frac{1}{T+1}\sum_{t=1}^{T} I\left\{Y_t \le y\right\}.$$

Then we plug the estimated marginal distribution $\widetilde{F}$ into each component copula function and estimate the parameters of the k^{th} copula $\boldsymbol{\theta}_k$ by QMLE, as stated in Chen and Fan (2006). In other words, when estimating $\boldsymbol{\theta}_k$, we assume $\{Y_t\}_{t=1}^{\mathrm{T}}$ is generated solely from the k^{th} copula C_k. Let $\widetilde{\mathbf{u}} = (\widetilde{F}(y_1), \widetilde{F}(y_2))$, and we denote the resulting estimator as

$$C_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k) = C_k\left\{\widetilde{F}(y_1), \widetilde{F}(y_2);\widehat{\boldsymbol{\theta}}_k\right\}, \qquad k = 1,\ldots,K-1. \tag{2.3}$$

Because if the candidate set includes only single copulas, a mixture of them can lead to poor fit, especially when the true copula is not close to any of these single copulas. To make our estimation method more general, we also include a mixture of the $K-1$ copulas into the candidate set, denoted as the K^{th} copula. We use the rescaled empirical distribution to replace the unknown margins, and implement the QMLE method to estimate the finite dimensional parameter of the mixture.

Let

$$C_K(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_K) = C_K\left\{\widetilde{F}(y_1),\widetilde{F}(y_2);\widehat{\boldsymbol{\theta}}_K\right\} \equiv \sum_{k=1}^{K-1} \breve{\omega}_k C_k\left\{\widetilde{F}(y_1),\widetilde{F}(y_2);\breve{\boldsymbol{\theta}}_k\right\},$$

where $\breve{\omega}_1,\ldots,\breve{\omega}_{K-1}$ and $\breve{\boldsymbol{\theta}}_1,\ldots,\breve{\boldsymbol{\theta}}_{K-1}$ are the QML estimators. $\breve{\omega}_k$ is constrained to be between 0 and 1 and add up to 1. For each copula k, $\breve{\boldsymbol{\theta}}_k$ also has its own constraint. For example, the parameter (the correlation coefficient) for Gaussian copula need to be between -1 and 1.

Denote $\mathbf{w} = (w_1,\ldots,w_K)^{\mathrm{T}}$ as weight vector, taking values in the following set

$$\mathcal{W} = \left\{\mathbf{w}\in[0,1]^K : \sum\nolimits_{k=1}^{K} w_k = 1\right\}.$$

Then, the third stage of our estimation procedure is to use the model average method to select the weights of all candidate copulas in the following model

$$C(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}},\mathbf{w}) = \sum_{k=1}^{K} w_k C_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k), \tag{2.4}$$

where $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\theta}}_1^{\top},\ldots,\widehat{\boldsymbol{\theta}}_K^{\top}\right)^{\top}$. Note that the model specified in (2.4) is very general and could cover many popular copula estimation methods mentioned in previous literature. For example, if w_1 equals one and all other weights are zeros, then (2.4) is equivalent to estimating an individual copula through QMLE.

In this paper we select the weights by using the Cross-Validation (CV) method, which is similar to the method presented in Hansen (2007). Specifically, denote an empirical estimator of $C_0(\mathbf{u};\boldsymbol{\theta}_0)$ as

$$\widetilde{C}(y_1, y_2) = \frac{1}{T}\sum_{t=1}^{T} I(Y_t \le y_1 \text{ and } Y_{t-1} \le y_2), \tag{2.5}$$

where $\mathbf{y} = (y_1, y_2)$ is an arbitrary point in $\mathcal{R}^2$, $I(\cdot)$ is an indicator function.

Define $U_{0t} \equiv (G_0(Y_t), G_0(Y_{t-1}))$ and $\widetilde{U}_t \equiv (\widetilde{F}(Y_t), \widetilde{F}(Y_{t-1}))$, and we use the following CV criterion to select the weight $\mathbf{w}$

$$CV(\mathbf{w}) = \sum_{t=1}^{T}\left\{C(\widetilde{U}_t;\widehat{\boldsymbol{\theta}},\mathbf{w}) - \widetilde{C}(Y_t, Y_{t-1})\right\}^2. \tag{2.6}$$

and

$$\widehat{\mathbf{w}} = \text{argmin}_{\mathbf{w}\in\mathcal{W}} CV(\mathbf{w}).$$

Replacing $\mathbf{w}$ in (2.4) with $\widehat{\mathbf{w}}$, we estimate the true copula $C_0(\mathbf{u};\boldsymbol{\theta}_0)$ by the model average estimator $C(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}},\widehat{\mathbf{w}})$.

The notations used in this paper are summarized as follows.

The $T \times 1$ vector of the true copula evaluated at $(Y_0, Y_1), ..., (Y_{T-1}, Y_T)$:

$$\mathbf{C}_0 = \{C_0(U_{01};\boldsymbol{\theta}_0), \ldots, C_0(U_{0T};\boldsymbol{\theta}_0)\}^\top, \tag{2.7}$$

the $T \times 1$ vector of the estimated k^{th} candidate copula using all observations (when $k = K$, the candidate copula is a mixture of the candidate copulas) evaluated at

$(Y_0, Y_1), ..., (Y_{T-1}, Y_T)$:

$$\widehat{\mathbf{C}}_k = \left\{ C_k(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}_k), \ldots, C_k(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}_k) \right\}^\top, \tag{2.8}$$

the vector of weighted average of estimated candidate copulas $\left\{ \widehat{\mathbf{C}}_1, \ldots, \widehat{\mathbf{C}}_K \right\}$ evaluated at $(Y_0, Y_1), ..., (Y_{T-1}, Y_T)$:

$$\widehat{\mathbf{C}}(\mathbf{w}) = \sum_{k=1}^{K} w_k \widehat{\mathbf{C}}_k = \left\{ C(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}, \mathbf{w}), \ldots, C(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}, \mathbf{w}) \right\}^\top, \tag{2.9}$$

the vector of empirical estimator of $\mathbf{C}_0$ evaluated at $(Y_0, Y_1), ..., (Y_{T-1}, Y_T)$:

$$\widetilde{\mathbf{C}} = \left\{ \widetilde{C}(Y_1, Y_0), ..., \widetilde{C}(Y_T, Y_{T-1}) \right\}^\top, \tag{2.10}$$

and the vector of model average estimator of $\mathbf{C}_0$ evaluated at $(Y_0, Y_1), ..., (Y_{T-1}, Y_T)$:

$$\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) = \left\{ C(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}), \ldots, C(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}) \right\}^\top. \tag{2.11}$$

Please note that in general we use the ‘hat’ notation to define estimators based on parametric model estimation methods, and use the ‘tilde’ notation to define estimators based on nonparametric estimation methods (such as empirical function).

Now, the CV criterion can be rewritten as

$$CV(\mathbf{w}) = \left\| \widehat{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 = \mathbf{w}^\top \breve{\mathbf{H}}^\top \breve{\mathbf{H}} \mathbf{w},$$

where $\breve{\mathbf{h}}_k = \widehat{\mathbf{C}}_k - \widetilde{\mathbf{C}}$ and $\breve{\mathbf{H}} = (\breve{\mathbf{h}}_1, \ldots, \breve{\mathbf{h}}_K)$. Therefore, the CV criterion $CV(\mathbf{w})$ is a quadratic function of $\mathbf{w}$ and can be minimized with respect to $\mathbf{w}$ easily.

We define a quadratic loss function of the model average estimator as

$$L_T(\mathbf{w}) = \|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\|^2. \tag{2.12}$$

Like previous papers on model selection and model averaging such as Shao (1997) and Hansen (2007), our goal is to use model averaging to reduce the quadratic loss. The following theorem demonstrates that our method can asymptotically minimize the quadratic loss function.

Theorem 1 *Under Assumption 1 and Conditions C.1 – C.3 presented in Appendix A.1,*

$$\frac{L_T(\widehat{\mathbf{w}})}{\inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})} \to 1 \quad \text{in probability (as } T \to \infty). \tag{2.13}$$

Theorem 1 shows that the model average estimator $\widehat{\mathbf{C}}(\widehat{\mathbf{w}})$ is asymptotically optimal by registering the infeasible lowest possible squared estimation losses: the squared loss of $\widehat{\mathbf{C}}(\widehat{\mathbf{w}})$ is asymptotically identical to that of the infeasible best possible model average estimator. The proof for Theorem 1 can be found in the Appendix.

2.4 Numerical Studies

We compare estimation results of the proposed model average approach on a mixture copula (MA) with two other methods: Chen and Fan (2006)'s QMLE method on a mixture copula (MML), and a standard BIC method which selects only a single copula from the candidate set (BIC). For Monte Carlo setup, we consider two cases of simulation. In Case I simulation, a strictly stationary first-order Markov process $\{Y_t\}_{t=1}^T$ is generated from copulas which are included in the mixture copula model. So the working mixture copula is correctly specified in that case. On the contrary, in Case II simulation, the working mixture copula model is misspecified. That is, $\{Y_t\}_{t=1}^T$ is generated from copulas which are not components of the working mixture model. We want to compare which method can better describe the temporal dependence of $\{Y_t\}_{t=1}^T$ under the two types of settings.

Following Chen et al. (2009), a strictly stationary first-order Markov process can be generated by following steps:

1. Generate an i.i.d. sequence of uniform random variables $\{V_t\}_{t=1}^T$.
2. Set $U_1 = V_1$, and $U_t = C_{2|1}^{-1}[V_t \mid U_{t-1}; \theta_0]$.
3. Set $Y_t = G_0^{-1}(U_t)$ for $t = 1, ..., T$.

The true marginal distribution of the Markov process is set to be a standard normal distribution. The simulation considers the sample size $T = 500$ and 1000. For $T = 500$ cases, we generate a time series with 2000 observations, but delete the first 1000. For the remaining 1000 observations, we use the first 500 data points to estimate and fit the mixture model, and the last 500 to evaluate the out-of-sample predicting performance. For $T = 1000$ cases, we generate a time series with 3000 observations. After deleting the first 1000, we use the first half of the remaining 2000 to estimate, and the second half of the remaining 2000 to predict. Therefore, the number of out-of-sample observations is equal to the number of in-sample observations. All simulations are repeated 500 times.

2.4.1 Simulation Case I

Case I simulation considers the scenario that data are generated from copulas which are included in our working model. the working mixture model includes three individual copulas: Gaussian, Clayton, and Gumbel. These three copulas are widely used in empirical studies. Gaussian copula shows symmetric dependence structure, while Clayton copula exhibits strong left tail dependence and Gumbel copula displays strong right tail dependence. The presumed mixture can be constructed as

$$C(u, v; \boldsymbol{\theta}, \boldsymbol{\omega}) = \omega_{Ga} C_{Ga}(u, v; \theta_1) + \omega_{Cl} C_{Cl}(u, v; \theta_2) + \omega_{Gu} C_{Gu}(u, v; \theta_3),$$

where C_{Ga}, C_{Cl}, and C_{Gu} stand for Gaussian, Clayton, and Gumbel copula, respectively, and u, v denote the two margins.

Observations are simulated from different copulas by the following DGPs. We simulate three single copulas (the weights on the other two copulas are zero) and three mixture copulas with two components. Specifically, we have the following six cases for the setup of weights:

Case 1: $\omega_{Ga} = 1$, $\omega_{Cl} = 0$, $\omega_{Gu} = 0$;

Case 2: $\omega_{Ga} = 0$, $\omega_{Cl} = 1$, $\omega_{Gu} = 0$;

Case 3: $\omega_{Ga} = 0$, $\omega_{Cl} = 0$, $\omega_{Gu} = 1$;

Case 4: $\omega_{Ga} = 1/2$, $\omega_{Cl} = 1/2$, $\omega_{Gu} = 0$;

Case 5: $\omega_{Ga} = 1/2$, $\omega_{Cl} = 0$, $\omega_{Gu} = 1/2$;

Case 6: $\omega_{Ga} = 0$, $\omega_{Cl} = 1/2$, $\omega_{Gu} = 1/2$.

For every case of weight above, we consider two sets of copula parameters:

Parameter setting 1: $\theta_{Ga} = 0.3$, $\theta_{Cl} = 2.0$, $\theta_{Gu} = 2.0$;

Parameter setting 2: $\theta_{Ga} = 0.6$, $\theta_{Cl} = 5.0$, $\theta_{Gu} = 3.5$.

Thus we will totally have $6 \times 2 = 12$ groups of DGPs.

Table 2.1 presents how close the estimated copula is to the true copula in terms of mean squared errors (MSEs) using the three methods we mentioned at the beginning of this section (MA, MML and BIC). To save space, we only present the out-sample predicting errors of each competing method. For expositional ease, the MSEs in each case is multiplied by 1000.

Table 2.1: Mean of squared out-of-sample prediction losses for Type I simulation

Sample Size=500; Out=500						
	$\theta_{Ga}=0.3, \theta_{Cl}=2.0, \theta_{Gu}=2.0$			$\theta_{Ga}=0.6, \theta_{Cl}=5.0, \theta_{Gu}=3.5$		
	MA	MML	BIC	MA	MML	BIC
$\omega_{Ga}=1, \omega_{Cl}=0, \omega_{Gu}=0$	0.5276	0.5450	0.5330	0.9596	0.9903	0.9630
$\omega_{Ga}=0, \omega_{Cl}=1, \omega_{Gu}=0$	1.6650	1.8634	1.6901	6.8191	7.0379	6.7958
$\omega_{Ga}=0, \omega_{Cl}=0, \omega_{Gu}=1$	1.5113	1.5536	1.5264	4.3578	4.3900	4.3452
$\omega_{Ga}=0.5, \omega_{Cl}=0.5, \omega_{Gu}=0$	0.7327	0.7296	0.7858	1.5419	1.5361	1.6589
$\omega_{Ga}=0.5, \omega_{Cl}=0, \omega_{Gu}=0.5$	0.7993	0.8046	0.8220	1.5712	1.5693	1.6066
$\omega_{Ga}=0, \omega_{Cl}=0.5, \omega_{Gu}=0.5$	1.2887	1.2820	1.3521	3.7591	3.7580	3.8774
Sample Size=1000; Out=1000						
	$\theta_{Ga}=0.3, \theta_{Cl}=2.0, \theta_{Gu}=2.0$			$\theta_{Ga}=0.6, \theta_{Cl}=5.0, \theta_{Gu}=3.5$		
	MA	MML	BIC	MA	MML	BIC
$\omega_{Ga}=1, \omega_{Cl}=0, \omega_{Gu}=0$	0.2412	0.2547	0.2424	0.4438	0.4666	0.4435
$\omega_{Ga}=0, \omega_{Cl}=1, \omega_{Gu}=0$	0.8059	0.9554	0.8222	3.3414	3.4751	3.3762
$\omega_{Ga}=0, \omega_{Cl}=0, \omega_{Gu}=1$	0.7142	0.7396	0.7211	2.3615	2.3905	2.3594
$\omega_{Ga}=0.5, \omega_{Cl}=0.5, \omega_{Gu}=0$	0.3592	0.3570	0.4188	0.7213	0.7231	0.8476
$\omega_{Ga}=0.5, \omega_{Cl}=0, \omega_{Gu}=0.5$	0.3603	0.3685	0.3823	0.7886	0.7877	0.8337
$\omega_{Ga}=0, \omega_{Cl}=0.5, \omega_{Gu}=0.5$	0.6003	0.5998	0.6413	1.8218	1.8260	1.9008

We make a few observations from Table 2.1. First, when the data are generated from a single copula (the first three rows of each subtable in Table 2.1), MML which uses Quasi-MLE to estimate a mixture copula gives the largest predicting errors in all settings. Our model average approach and the BIC method show similar performances. Second, when the true copula is a mixture copula (the last three rows of each subtable in Table 2.1), The BIC method which selects a single copula performs worst. The performances of the two mixture copula estimation methods: MA and MML are quite similar to each other. Hence, Table 2.1 shows in terms of out-of-sample predicting errors, our model average method gives satisfactory estimation

Table 2.2: Mean of squared estimation losses of 0.01 conditional quantile for Type I simulation

Sample Size=500; Out=500						
	$\theta_{Ga}=0.3,\theta_{Cl}=2.0,\theta_{Gu}=2.0$			$\theta_{Ga}=0.6,\theta_{Cl}=5.0,\theta_{Gu}=3.5$		
	MA	MML	BIC	MA	MML	BIC
$\omega_{Ga}=1,\omega_{Cl}=0,\omega_{Gu}=0$	29.33	34.00	29.98	30.32	34.06	28.68
$\omega_{Ga}=0,\omega_{Cl}=1,\omega_{Gu}=0$	20.03	71.20	21.51	20.77	51.54	35.84
$\omega_{Ga}=0,\omega_{Cl}=0,\omega_{Gu}=1$	34.90	51.54	37.20	36.35	44.73	42.07
$\omega_{Ga}=0.5,\omega_{Cl}=0.5,\omega_{Gu}=0$	31.16	31.38	63.82	35.70	36.28	75.83
$\omega_{Ga}=0.5,\omega_{Cl}=0,\omega_{Gu}=0.5$	29.72	32.13	36.75	30.00	30.21	52.26
$\omega_{Ga}=0,\omega_{Cl}=0.5,\omega_{Gu}=0.5$	43.12	42.90	62.85	30.92	30.63	59.60
Sample Size=1000; Out=1000						
	$\theta_{Ga}=0.3,\theta_{Cl}=2.0,\theta_{Gu}=2.0$			$\theta_{Ga}=0.6,\theta_{Cl}=5.0,\theta_{Gu}=3.5$		
	MA	MML	BIC	MA	MML	BIC
$\omega_{Ga}=1,\omega_{Cl}=0,\omega_{Gu}=0$	14.81	18.08	14.44	14.94	17.62	14.53
$\omega_{Ga}=0,\omega_{Cl}=1,\omega_{Gu}=0$	11.66	41.10	12.02	12.62	33.64	17.13
$\omega_{Ga}=0,\omega_{Cl}=0,\omega_{Gu}=1$	19.05	34.22	19.35	20.63	30.08	23.96
$\omega_{Ga}=0.5,\omega_{Cl}=0.5,\omega_{Gu}=0$	17.17	16.41	54.32	19.92	20.21	63.63
$\omega_{Ga}=0.5,\omega_{Cl}=0,\omega_{Gu}=0.5$	15.52	16.02	22.20	15.83	15.93	41.79
$\omega_{Ga}=0,\omega_{Cl}=0.5,\omega_{Gu}=0.5$	15.84	15.33	33.93	16.69	16.49	46.78

results no matter whether the true copula is a single copula or a mixture.

Since estimating conditional quantiles (VaR) of financial returns is important in risk management, we also compare the plug-in estimators of the 1% conditional quantiles via the three competing methods. The results are presented in Table 2.2. Again, our model average approach shows relatively small estimation errors no matter whether the true copula is a single copula or a mixture.

2.4.2 Simulation Case II

In Case II simulations, we assume the working mixture model is misspecified. That is, observations are generated from copulas which are out of our candidate set. Considering the fact that the true dependence structure is always unknown to econometricians, the misspecification cases should be common in empirical studies. Therefore, we want to see how our model average approach performs in these cases.

In the simulation setup, our working mixture model still include Gaussian, Clayton, and Gumbel copulas but the true observations are generated from a mixture copula which is a weighted average of Frank, Survival Gumbel (SG) and Survival Clayton (SC) copulas, another three widely used copulas. Frank copula is similar to Gaussian copula as both exhibit symmetric dependence structure, while Frank copula has relatively stronger dependence in the center of the distribution. Survival Gumbel (Clayton) copula is a 180° rotation of Gumbel (Clayton), so it exhibits left (right) tail dependence as the Clayton (Gumbel) copula does. Similarly to what we did in Simulation Case I, we consider six cases of weighting setups with three single copulas and three mixture copulas:

$$\text{Case 1:} \quad \omega_{Frank} = 1,\, \omega_{SG} = 0,\, \omega_{SC} = 0;$$

$$\text{Case 2:} \quad \omega_{Frank} = 0,\, \omega_{SG} = 1,\, \omega_{SC} = 0;$$

$$\text{Case 3:} \quad \omega_{Frank} = 0,\, \omega_{SG} = 0,\, \omega_{SC} = 1;$$

$$\text{Case 4:} \quad \omega_{Frank} = 1/2,\, \omega_{SG} = 1/2,\, \omega_{SC} = 0;$$

$$\text{Case 5:} \quad \omega_{Frank} = 1/2,\, \omega_{SG} = 0,\, \omega_{SC} = 1/2;$$

$$\text{Case 6:} \quad \omega_{Frank} = 0,\, \omega_{SG} = 1/2,\, \omega_{SC} = 1/2;,$$

and two different copula parameter settings:

$$\text{Parameter setting 1:} \quad \theta_{Frank} = 2.0,\, \theta_{SG} = 2.0,\, \theta_{SC} = 2.0;$$

$$\text{Parameter setting 2:} \quad \theta_{Frank} = 5.7,\, \theta_{SG} = 3.5,\, \theta_{SC} = 5.0.$$

In Table 2.3 we present the out-of-sample predicting performance using the three methods when the mixture model is misspecified. We could observe from Table 2.3 that our model average approach performs the best among three methods and gives

Table 2.3: Mean of squared out-of-sample prediction losses for Type II simulation

	Sample Size=500; Out=500					
	$\theta_{Frank}=2.0, \theta_{SG}=2.0, \theta_{SC}=2.0$			$\theta_{Frank}=5.7, \theta_{SG}=3.5, \theta_{SC}=5.0$		
	MA	MML	BIC	MA	MML	BIC
$\omega_F=1, \omega_{SG}=0, \omega_{SC}=0$	0.5852	0.6440	0.6100	1.4253	1.4843	1.5100
$\omega_F=0, \omega_{SG}=1, \omega_{SC}=0$	1.5028	1.5764	1.6149	4.7470	4.8817	4.9128
$\omega_F=0, \omega_{SG}=0, \omega_{SC}=1$	2.2280	2.3376	2.2316	8.2011	8.2833	8.4504
$\omega_F=0.5, \omega_{SG}=0.5, \omega_{SC}=0$	0.8567	0.8983	0.9254	2.0935	2.2006	2.2375
$\omega_F=0.5, \omega_{SG}=0, \omega_{SC}=0.5$	0.9299	1.0138	0.9459	2.2517	2.4923	2.3500
$\omega_F=0, \omega_{SG}=0.5, \omega_{SC}=0.5$	1.4682	1.5086	1.5087	4.0592	4.2301	4.1841
	Sample Size=1000; Out=1000					
	$\theta_{Frank}=2.0, \theta_{SG}=2.0, \theta_{SC}=2.0$			$\theta_{Frank}=5.7, \theta_{SG}=3.5, \theta_{SC}=5.0$		
	MA	MML	BIC	MA	MML	BIC
$\omega_F=1, \omega_{SG}=0, \omega_{SC}=0$	0.2704	0.3122	0.2865	0.6648	0.7020	0.7420
$\omega_F=0, \omega_{SG}=1, \omega_{SC}=0$	0.6217	0.6730	0.7328	2.0133	2.1365	2.1615
$\omega_F=0, \omega_{SG}=0, \omega_{SC}=1$	1.0832	1.1831	1.0862	3.9171	4.1216	3.9406
$\omega_F=0.5, \omega_{SG}=0.5, \omega_{SC}=0$	0.3487	0.3803	0.4138	0.9183	0.9862	1.0387
$\omega_F=0.5, \omega_{SG}=0, \omega_{SC}=0.5$	0.4025	0.4357	0.4118	1.0786	1.2043	1.1427
$\omega_F=0, \omega_{SG}=0.5, \omega_{SC}=0.5$	0.6349	0.6619	0.6738	1.9116	2.0132	2.0003

the most accurate results in terms of the predicting errors. These results support the point that the model average method outperforms other methods especially in misspecification cases.

We also check the performances of the three methods in estimating the conditional quantiles in misspecification cases. The results are displayed in Table 2.4. As expected, conditional quantile estimated by our model average method is much more accurate than those estimated by the MML and the BIC methods.

2.5 An Empirical Study

In this section we consider an empirical example to evaluate the performance of our method. Specifically, we consider the daily log-returns for four indices: Nasdaq 100 (NASDAQ), Shanghai Stock Exchange A Share Index (SHASHR), Korea Composite Stock Price Index (KOSPI), and Taiwan Capitalization Weighted Stock Index (TAIEX), from January 1991 (or from January 1996 for SHASHR) to March 2015. The sample size is $T = 6320$ for NASDAQ, KOSPI and TAIEX, or $T = 5016$

Table 2.4: Mean of squared estimation losses of 0.01 conditional quantile for Type II simulation

Sample Size=500; Out=500						
	$\theta_{Frank}=2.0, \theta_{SG}=2.0, \theta_{SC}=2.0$			$\theta_{Frank}=5.7, \theta_{SG}=3.5, \theta_{SC}=5.0$		
	MA	MML	BIC	MA	MML	BIC
$\omega_F=1, \omega_{SG}=0, \omega_{SC}=0$	34.07	44.57	39.47	54.06	64.49	62.44
$\omega_F=0, \omega_{SG}=1, \omega_{SC}=0$	27.76	63.07	35.72	24.86	40.79	41.29
$\omega_F=0, \omega_{SG}=0, \omega_{SC}=1$	115.49	185.46	115.71	159.61	293.93	159.96
$\omega_F=0.5, \omega_{SG}=0.5, \omega_{SC}=0$	39.02	44.59	71.17	47.95	63.76	74.13
$\omega_F=0.5, \omega_{SG}=0, \omega_{SC}=0.5$	42.15	52.90	43.32	71.61	130.37	92.21
$\omega_F=0, \omega_{SG}=0.5, \omega_{SC}=0.5$	38.31	57.47	53.39	64.11	90.66	84.97
Sample Size=1000; Out=1000						
	$\theta_{Frank}=2.0, \theta_{SG}=2.0, \theta_{SC}=2.0$			$\theta_{Frank}=5.7, \theta_{SG}=3.5, \theta_{SC}=5.0$		
	MA	MML	BIC	MA	MML	BIC
$\omega_F=1, \omega_{SG}=0, \omega_{SC}=0$	18.68	26.03	22.90	39.98	47.83	46.08
$\omega_F=0, \omega_{SG}=1, \omega_{SC}=0$	20.69	48.21	27.93	21.19	28.41	38.09
$\omega_F=0, \omega_{SG}=0, \omega_{SC}=1$	108.67	178.65	110.66	159.88	229.80	159.92
$\omega_F=0.5, \omega_{SG}=0.5, \omega_{SC}=0$	24.70	27.87	60.27	32.68	44.94	59.91
$\omega_F=0.5, \omega_{SG}=0, \omega_{SC}=0.5$	28.43	35.53	29.95	57.99	123.72	78.91
$\omega_F=0, \omega_{SG}=0.5, \omega_{SC}=0.5$	31.03	38.95	42.34	50.22	74.27	71.09

Table 2.5: The summary statistics for daily log-returns

	NASDAQ	SHASHR	KOSPI	TAIEX
Mean	0.0495	0.0393	0.0174	0.0130
Median	0.0663	0.0000	0.0000	0.0000
min	-11.110	-10.450	-12.800	-9.9360
max	17.200	9.4810	11.280	9.0590
S.D.	1.7465	1.6379	1.7120	1.5169
Skewness	0.0798	-0.2458	-0.1532	-0.1118
Kurtosis	8.9987	8.5920	8.3973	6.4268

for SHASHR. The daily log-returns of the four markets are calculated from their respective price indices in their own currencies.[2] The main purpose of the empirical study is to model the temporal dependence of returns for the four indices.

We divide the data equally into two groups. The first group contains the first half sample (3160 observations for NASDAQ, KOSPI and TAIEX, and 2508 observations

[2]We also consider using price indices whose currencies are converted into US dollars. The results are similar and do not change the conclusion.

for SHASHR), and the second group includes the remained data points. We use the first group (training set) to estimate and fit the mixture model, and the second group (testing set) to evaluate the out-sample predicting performance. The summary statistics for daily log-returns for four indices are displayed in Table 2.5. During the sample period, NSDAQ showed the highest average and median daily return. The skewness is negative for SHASHR, KOSPI and TAIEX, indicating higher probability in having extreme daily losses. Kurtosis for all four markets is greater than 3, implying a deviation from Normality. These statistics show that it can be difficult to correctly specify marginal distributions in practice and nonparametric methods should be used to estimate the margins.

We fit the data into a mixture copula model which is a weighted average of Gaussian, Clayton, and Gumbel. These three component copulas are widely used and are applied in Long et al. (2015) for constructing their mixture models. We implement the model average method, Chen and Fan (2006)'s QMLE method on a mixture copula, and the BIC method to estimate the mixture copula model respectively. Since the true copula is unknown, we compare the performances across different methods following the procedure described in Genest and Rivest (1993). The purpose of this exercise is to examine whether our model average estimation results have relatively smaller estimation losses and satisfactorily capture the temporal dependence structures of financial returns.

Specifically, for each of the four returns, we build four 7×7 cross-classifications. These cross-classifications are available upon request. For each time series, the observed frequencies are displayed by the cross-classifications in the first column, and the estimated results through the three methods (MA, MML and BIC)are presented in the second, third and fourth columns, respectively. Let H denote a cross-classification table with 7 rows and 7 columns, and $H(i,j)$ represent the cell in the

ith row and the jth column of H, where $i, j = 1, ..., 7$. The number recorded in $H(i, j)$ is the number of times that Y_t is between the $(i-1)/7$ and the $i/7$ percentile of its range, while Y_{t-1} is within the $(j-1)/7$ and $j/7$ percentile of its range. Specifically, define u_i and u_j as the $i/7$ and $j/7$ percentiles for $\{Y_t\}$. Then $H(i, j)$ shows the number of paris of observations (Y_t, Y_{t-1}) such that $u_{i-1} < Y_t \leq u_i$ and $u_{j-1} < Y_{t-1} \leq u_j$. For example, the cell $(5, 6)$ records the number of times that Y_t is within the 57th (4/7) and the 71st (5/7) percentile of its range, while Y_{t-1} is between 71st (5/7) and 86th (6/7) percentile of its range. Therefore, if Y_t and Y_{t-1} are perfectly positively correlated, it could be observed that the numbers on the principal diagonal are much bigger than those on other places. If Y_t and Y_{t-1} are independent, then the number in each cell of the cross-classification should be similar to each other. If the process has a symmetric temporal dependence structure, then the numbers displayed in the cells at the top-left (top-right) and the bottom-right (bottom-left) of the cross-classification should be close.

We first focus on the cross-classifications of the observed frequencies. Taking NASDAQ as an example. The cell at the top-left displays the frequency that Y_t and Y_{t-1} are both below the 14th (1/7) percentile of the range; Correspondingly, the cell at the bottom-right represents the number of times that Y_t and Y_{t-1} are between the 86th (6/7) and 100th (7/7) percentile of the range. It could be observed that during the period between January 1991 and March 2015, there are 176 times that Y_t and Y_{t-1} are both lower than their 14th percentile, which is much higher than 126, the number of times that Y_t and Y_{t-1} are both higher than their 86th percentile. Therefore, the daily returns of NASDAQ shows a left tail dependence structure. SHASHR and KOSPI display a similar left tail dependence structure. Such a pattern is not significant in TAIEX.

The estimated frequencies are obtained as follows: we first estimate the true

Table 2.6: Mean of in-sample estimation errors based on MA, MML and BIC

	Q_{MA}	Q_{MML}	Q_{BIC}
NASDAQ	178.72	179.76	203.79
SHASHR	260.20	294.06	445.63
KOSPI	74.76	88.20	117.51
TAIEX	124.50	137.34	142.14

copula of the four markets using one of the three methods and denote the estimated copula as $\widehat{C}$. Then we calculate the estimated probability $\widehat{p}_{ij}$ for each cell $H(i,j)$ through the following equation

$$\widehat{p}_{ij} = \begin{cases} \widehat{C}(u_i, u_j) & \text{for } i = 1 \text{ and } j = 1, \\ \widehat{C}(u_i, u_j) - \widehat{C}(u_i, u_{j-1}) & \text{for } i = 1 \text{ and } j = 2, ..., 7, \\ \widehat{C}(u_i, u_j) - \widehat{C}(u_{i-1}, u_j) & \text{for } i = 2, ..., 7 \text{ and } j = 1, \\ \widehat{C}(u_i, u_j) - \widehat{C}(u_i, u_{j-1}) - \widehat{C}(u_{i-1}, u_j) + \widehat{C}(u_{i-1}, u_{j-1}) & \text{for } i, j = 2, ..., 7, \end{cases}$$

and the estimated frequency in each cell $H(i,j)$ can be obtained by multiplying the estimated probability $\widehat{p}_{ij}$ to the number of observations.

We examine the performance of the three methods by comparing their estimated frequencies with the observed frequencies of all the cells. Denote $H_{i,j}$, $H_{i,j}^{MA}$, $H_{i,j}^{MML}$ and $H_{i,j}^{BIC}$ as the observed frequency and the estimated frequency by MA, MML and BIC in cell (i,j) respectively. The estimation errors are defined as:

$$\begin{aligned} Q_{MA} &= \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (H_{i,j} - H_{i,j}^{MA})^2, \\ Q_{MML} &= \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (H_{i,j} - H_{i,j}^{MML})^2, \\ Q_{BIC} &= \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (H_{i,j} - H_{i,j}^{BIC})^2, \end{aligned}$$

Table 2.7: Mean of out-of-sample predicting errors based on MA, MML and BIC

	Q_{MA}	Q_{MML}	Q_{BIC}
NASDAQ	125.65	144.69	128.85
SHASHR	162.80	207.20	239.36
KOSPI	120.14	126.97	128.99
TAIEX	183.81	196.05	207.56

where $k = 7$. We present the estimation errors in Table 2.6. Among all cases, the MA method generates the smallest estimation errors, while the BIC method which selects an individual copula based on the comparison of BIC among Gaussian, Clayton, and Gumbel copulas exhibits the largest estimation errors. The real data example again demonstrates that the model average estimators could satisfactorily capture the temporal dependence structures compared with results estimated by other competing methods.

We next evaluate the out-sample predicting accuracy among different methods based on the second half of the data set. We calculate the out-sample predicting losses in the same way as the in-sample estimation errors are calculated. Table 2.7 shows the results. It can be seen from Table 2.7 that the model average approach gives satisfactory estimates that achieve the smallest predicting losses. Thus, the model average method also provides relatively satisfactory predicting performance.

3. A ROBUST TEST FOR PREDICTABILITY WITH UNKNOWN PERSISTENCE

3.1 Literature Review

Section 3.1 provides literature review.

3.1.1 The Bonferroni Method

Cavanagh, Elliott, and Stock (1995) (and then Campbell and Yogo, 2006) provide a Bonferroni method that first finds a confidence interval for c as in Stock (1991), and then for each possible value of c, calculates a confidence interval for β given c. Finally, the confidence interval for β without c could be obtained as

$$CI_{\beta}(\alpha) = \bigcup_{CI_c(\alpha_c)} CI_{\beta|c}(\alpha_\beta)$$

where α_c, α_β and α are significant levels in each step respectively. By Bonferroni's inequality, α is no more than $\alpha_c + \alpha_\beta$.

When the predictor is local-to-unity, this method could successfully control the size while maintain the local power. However, Phillips (2014) points out that Stock's confidence intervals are seriously biased asymptotically in the stationary case and vicinities of unity that are wider than $O(n^{-1/3})$. Therefore, predictive regression tests based on the Bonferroni method do not work well when x_t is stationary. Moreover, it is hard to extend this approach to multivatiate models. Finally, since this method still assumes a linear regression model like (1), y_t has to share the same order of integration as x_t under a fixed alternative, as previously mentioned.

3.1.2 A Quasi Restricted Likelihood Ratio Test (QRLRT)

Chen, Deo and Yi (2013) suggested a quasi restricted likelihood ratio test (QRLRT). Their idea, which bases on predictive regression models, is as follows. Although it has been shown in Chen and Deo (2009) that the restricted likelihood (RL) of AR processes have good properties such as very small deviation of the restricted likelihood ratio test (RLRT) distribution from the χ^2 when ρ approaches unity, it is hard to obtain the limiting distribution of the RLRT when ρ is close to one. Therefore, they construct a weighted least squares approximation to the RL (WLSRL) which is easy to estimate and share the good properties of the RL estimators. Then they have the asymptotic distribution of the QRLRT based on the WLSRL. The QRLRT statistic is defined as

$$\Lambda_T = Q_T(0, \hat{\rho}^0_{WLS}) - Q_T(\hat{\beta}_{WLS}, \hat{\rho}_{WLS}),$$

and (1), If $\rho \in (-1,1)$ and ρ is fixed, $\Lambda_T \to_d \chi^2_1$

(2), If $\rho = 1 - c/k_T$, where $k_T = T^\lambda$,$\lambda \in (0,1)$, $\Lambda_T \to_d \chi^2_1$

(3), If $\rho = 1 - c/T$, $\Lambda_T \to_d \Lambda_{c,\sigma}$, where

$\Lambda_{c,\sigma} = \left(\sigma g^{1/2}_{c,\sigma}\tau_c + \sqrt{1-\sigma^2 g_{c,\sigma}}Z\right)^2$,$\tau_c = \frac{\int J_c(\lambda)dW(\lambda)}{\sqrt{\int J^2_c(\lambda)d(\lambda)}}$,$Z \sim N(0,1)$. And it has been shown that the right tail of $\Lambda_{c,\sigma}$ is very close to that of a χ^2_1. Their simulations show that the resulting sup bound test maintain size without significant power loss. The undesirable properties include weak test power (only about 10%) when c is large (20) and β_1 is not far from zero.

3.1.3 A Robust Bootstrap and Subsampling Approach

Camponovo, Scaillet and Trojani (2013) provide a robust bootstrap and subsampling approach. They claim that conventional hypothesis testing methods including bias correction methods and bootstrap and subsampling tests may suffer from non-resistance to even small fractions of anomalous observations in the data. Thus they develop a class of robust resampling tests. This method bases on the predictive regression model

$$y_t = \theta' w_{t-1} + \varepsilon_t,$$

with $\theta = (\beta_0, \beta_1)'$ and $w_{t-1} = (1, x_{t-1})'$ and denote $z_t = (y_t, w'_{t-1})'$.

Given a positive constant a, $\hat{\theta}_T^R$ is the M-estimatior solve the equation

$$\psi_{T,a}(z_{(T)}, \hat{\theta}_T^R) = 1/T \sum_{t=1}^{T} (y_t - w'_{t-1}\hat{\theta}_T^R) w_t h_a(z_t, \hat{\theta}_T^R) = 0$$

where $h_a(z_t, \theta) = \min(1, \frac{a}{\|(y_t - w'_{t-1}\theta) w_{t-1}\|})$, and

$$\hat{\theta}_T^R = (\sum_{t=1}^{T} (w_{t-1} w'_{t-1}) h_{at})^{-1} \sum_{t=1}^{T} (y_t w_{t-1}) h_{at},$$

Thus, the Huber weight $0 \leq h_{at} \leq 1$ reduces the influence of potential anomalous observations. The sampling distribution of the nonstudentized statistic $t_T^{NS} = \sqrt{T}(\hat{\theta}_T^R - \theta_0)$ can be estimated using the robust fast resampling distribution

$$L_{T,m}^{NS,K*}(x) = 1/N \sum_{t=1}^{N} II(\sqrt{k}(-\left[\nabla_\theta \psi_{T,a}(z_{(T)}, \hat{\theta}_T)\right]^{-1} \psi_{k,a}(z_{(T,m),s}^{K*}, \hat{\theta}_T)) \leq x), K = B, S$$

where s indexes the N possible random samples generated by the bootstrap and subsampling procedures, $k = T$ for the block bootstrap and $k = m$ for the subsampling.

3.1.4 Differencing Transformations

Camponovo (2012) develops a class of estimators and test statistics based on differencing transformations. Specifically, he proves that the instruments $w_t =$

$\Delta x_{t-l}^{t-1} + (1-\rho^{l-1})\Delta x_{t-l-1}^{t-1}$ satisfy the moment conditions that

$$E\left[(\Delta y_{t-1}^{t} - \beta \Delta x_{t-l-1}^{t-1})w_t\right] = 0.$$

where $\Delta x_{t-l}^{t} = x_t - x_{t-l}$, $\Delta y_{t-l}^{t} = y_t - y_{t-l}$, $t = l+1, ..., T$, and $l \geq 2$. Therefore, he defines the new class of estimators as

$$\hat{\beta}_{T,l,\rho} = \frac{\sum_{t=l+1}^{T} \Delta y_{t-l}^{t} w_t}{\sum_{t=l+1}^{T} \Delta x_{t-l-1}^{t-1} w_t}.$$

and shows the test statistic

$$t_{T,l}(\beta_0) = \sqrt{T}\left(\frac{\sqrt{V_{l,\hat{\rho}_T}}}{J_{l,\hat{\rho}_T}}\right)^{-1}(\hat{\beta}_{T,l,\hat{\rho}_T} - \beta_0) \rightarrow_d N(0,1),$$

under $H_0 : \beta = \beta_0$, for $\rho \in (-1, 1]$, and $l/T \rightarrow \infty$, as $l \rightarrow \infty$ and $T \rightarrow \infty$. $\hat{\rho}_T$ is a consistent estimator of ρ. $V_{l,\hat{\rho}_T}$ and $J_{l,\hat{\rho}_T}$ are functions of $\hat{\rho}_T$, l, σ_ε^2, σ_υ^2 and $\sigma_\varepsilon \sigma_\upsilon$.Therefore, the test statistic has a normal limit distribution no matter whether the predictor is stationary, nonstationary or local-to-unity. The author also provides a data-driven method to select l for the finite sample.

The undesirable points include that people have to choose the optimal l each time when applying this approach in empirical studies.

3.1.5 The Linear Projection Method

Amihud and Hurvich (2004) and Amihud et al. (2009) propose the two-stage least squares (linear projection method) estimator to correct the finite sample bias of the OLS estimates assuming the regressors are stationary. Cai and Wang (2014) apply this method to (near) integrated cases and derive the asymptotic distribution of the two-step estimator. As we know, for predictive regression models, fitting an drift to the regressior often affects asymtotic properties when the regessor is (near) nonstationary. Therefore, they consider two cases: without and with drift. The asymptotic distribution is the mixed normal for the former, and normal for the latter. Since the limiting distribution in the zero-drift case is nonstandard, when doing tests, one needs to use Monte Carlo simulation method to get critical values. According to their simulation results, the size and the local power in both cases are proper. The limitation of this method includes the nonstandard asymptotic distribution and requiring the prior knowledge of the order of integration and the drift. Also, based on a linear regression model, the tests lack power under fixed alternatives.

3.2 The IVX Approach

Magdalinos and Phillips (2009b), Kostakis et al. (2010), and Phillips and Lee (2013) suggest another solution to predictive regressions, which is called the IVX method. This approach has many good properties comparing with other tests based on predictive regressions. First, it is robust to a very general class of degree of persistence in the regressors, ranging from mildly integrated to mildly explosive processes. Second, this framework is easy to extend to multivariate systems while the Bonferroni method is restricted to the case of a scalar predictor. Third, the resulting test statistic has standard chi-squared inference under the null hypothesis and is convenient to implement. The IVX approach assume the following predictive regression framework

$$
\begin{aligned}
y_t &= \beta \mathbf{x}_{t-1} + \varepsilon_t, \\
\mathbf{x}_t &= R\mathbf{x}_{t-1} + \upsilon_t, \\
R &= I_p + \frac{C}{T^\alpha},\ C \leq 0 \text{ and } \alpha > \frac{1}{2}
\end{aligned}
$$

The idea of this method is to generate an instrument that is mildly integrated. More specifically, the instrument takes the form

$$\begin{aligned}\widetilde{\mathbf{z}}_t &= \sum_{j=1}^{t} R_z^{t-j}\Delta\mathbf{x}_j,\\ R_z &= I_p + \frac{C_z}{T^{\delta}},\ C_z < 0 \text{ and } \delta \in (\frac{2}{3}, 1)\end{aligned}$$

Then,

$$\widehat{\beta}_{IVX} = (Y'\widetilde{Z} - T\ \widehat{\Lambda}_{\varepsilon\upsilon})(X'\widetilde{Z})^{-1}$$

where $Y = [y_1, ..., y_T]'$, $X = [\mathbf{x}_1', ..., \mathbf{x}_T']'$ and $\widetilde{Z} = [\widetilde{\mathbf{z}}_1', ..., \widetilde{\mathbf{z}}_T']'$. $\widehat{\Lambda}_{\varepsilon\upsilon}$ is the estimator of the one-sided long run covariance between ε_t and υ_t, $\Lambda_{\varepsilon\upsilon}$. And

$$\widehat{\Lambda}_{\varepsilon\upsilon} = \frac{1}{T}\sum_{h=1}^{M}(1 - \frac{h}{M+1})\sum_{t=h+1}^{T}\widehat{\varepsilon}_t\widehat{\upsilon}_{t-h}'$$

where

$$\begin{aligned}\widehat{\varepsilon}_t &= y_t - \widehat{\beta}_{OLS}\mathbf{x}_{t-1}\\ \widehat{\upsilon}_{t-h} &= \mathbf{x}_{t-h} - \widehat{R}_{OLS}\mathbf{x}_{t-h-1}.\end{aligned}$$

Let $\widehat{\Omega}_{\varepsilon\varepsilon}$ be any consistent estimator of $\Omega_{\varepsilon\varepsilon}$, the two-sided long run covariance between ε_t and υ_t. The Wald statistic,

$$\begin{aligned}W_{IVX} &= vec(\widehat{\beta}_{IVX} - \beta)'[(X'P_{\tilde{Z}}X)^{-1} \otimes \widehat{\Omega}_{\varepsilon\varepsilon}]^{-1}vec(\widehat{\beta}_{IVX} - \beta)\\ &\to {}_d\chi^2(p),\end{aligned}$$

no matter whether the regressors are mildly integrated, local-to-unity or $I(1)$. However, this approach maintains a predictive regression setting that forces y_t and $\mathbf{x}_t$

to have the same orders of integration under a fixed alternative. In fact, as shown below, as long as we allow y_t and $\mathbf{x}_t$ to have different orders of integration, the IVX test statistic never converges to infinity, even if $cov(y_t, \mathbf{x}_{t-1}) \neq 0$.

Theorem 1. Let

$$\begin{aligned} \mathbf{x}_t &= R\mathbf{x}_{t-1} + \upsilon_t, \\ R &= I_p + \frac{C}{T^\alpha},\ C \leq 0 \text{ and } \alpha > \frac{1}{2} \end{aligned}$$

where $\mathbf{x}_t$ is an $p \times 1$ vector and $C = diag(c_1, c_2, \cdots, c_p)$ is a diagonal matrix of localizing coefficients. And

$$\mathbf{u}_t = \begin{pmatrix} y_t \\ \upsilon_t \end{pmatrix} = F(L)\varsigma_t = \sum_{j=0}^{\infty} F_j \varsigma_{t-j}, \quad \sum_{j=0}^{\infty} j \parallel F_j \parallel < \infty, \quad \delta > 1;$$

where $F(z) = \sum_{j=0}^{\infty} F_j z^j$, $F_0 = I_{p+1}$, $F(1)$ has *full rank* and $\varsigma_t \sim IID(0, \Sigma)$ satisfying $\Sigma > 0$ and the moment condition $E \parallel \varsigma_1 \parallel^4 < \infty$. Therefore, $\mathbf{x}_t$ could be integrated, local-to-unity or mildly integrated regressors while y_t is always stationary. Then we have

$$\begin{aligned} W_{IVX} &= vec(\widehat{\beta}_{IVX})'[(X'P_{\tilde{Z}}X)^{-1} \otimes \widehat{\Omega}_{\varepsilon\varepsilon}]^{-1} vec(\widehat{\beta}_{IVX}) \\ &\rightarrow {}_d\chi^2(p). \end{aligned}$$

Remark 1. In the previous theorem and proof, we do not impose any restrictions on the covariance between y_t and $\mathbf{x}_{t-1}$. Therefore, no matter whether the true covariance is zero or not, the IVX test statistic does not diverge to infinity.

3.3 Model and Estimation

Instead of testing coefficients of predictors in predictive regressions, we are interested in testing

$$H_0 : cov(y_t, \mathbf{x}_{t-1}) = 0 \tag{3.1}$$

This restriction allows for more general cases that y_t and $\mathbf{x}_t$ are of different orders of integration. If both y_t and $\mathbf{x}_t$ are stationary, then testing (3.1) are equivalent to testing the coefficient of $\mathbf{x}_{t-1}$ equals zero.

In practice, the predictor $\mathbf{x}_t$, such as interest rates and/or dividend-price ratios, can be either (near) nonstationary or stationary. On the other hand, y_t, such as a stock return, is usually stationary. Therefore, we consider two situations: (1) $\mathbf{x}_t$ is $I(1)$ or local-to-unity, y_t is stationary; (2) both of $\mathbf{x}_t$ and y_t are stationary.

3.3.1 Predictors Are I(1) or Local-To-Unity

When $\mathbf{x}_t$ is $I(1)$ or local-to-unity and y_t is $I(0)$ and generated by

Assumption 1.

$$\mathbf{x}_t = R\mathbf{x}_{t-1} + \upsilon_t, \ t = 2, ..., T, \tag{3.2}$$

$$R = I_p + \frac{C}{T} \tag{3.3}$$

where $\mathbf{x}_t$ is an $p \times 1$ vector and $C = diag(c_1, c_2, \cdots, c_p)$ is a diagonal matrix of localizing coefficients with $C = 0$ or $C \leq 0$, and we allow general linear dependence in υ_t

$$\mathbf{z}_t = \begin{pmatrix} y_t \\ \upsilon_t \end{pmatrix} = F(L)\varsigma_t = \sum_{j=0}^{\infty} F_j \varsigma_{t-j}, \quad \sum_{j=0}^{\infty} j^{\delta} \parallel F_j \parallel < \infty, \quad \delta > 1;$$

where $F(z) = \sum_{j=0}^{\infty} F_j z^j$, $F_0 = I_{p+1}$, $F(1)$ has *full rank* and $\varsigma_t \sim IID(0, \Sigma)$ satisfying

$\Sigma > 0$ and the moment condition $E \parallel \varsigma_1 \parallel^4 < \infty$,

$$\sum_{h=-\infty}^{\infty} | h |^{\delta} \| \Gamma(h) \| < \infty, \quad \Gamma(h) = E\mathbf{z}_t\mathbf{z}'_{t+h},$$

they are of different orders of integration. Tests based on predictive regressions such as IVX tests can not maintain a strong power against the fixed alternative. A covariance-based test should be used. However, as $\mathbf{x}_t$ is $I(1)$, $cov(y_t, \mathbf{x}_{t-1})$ is no longer a constant and depends on t. Therefore, We follow Maynard and Shimotsu (2009, MS henceforth) that define $\lambda_{y,\Delta\mathbf{x}} = \lim_{t\to\infty} cov(y_t, \mathbf{x}_{t-1})$ and base the test on $\lambda_{y,\Delta\mathbf{x}}$

$$H_0 : \lambda_{y,\Delta\mathbf{x}} = 0. \tag{3.4}$$

And

$$\lambda_{y,\Delta\mathbf{x}} = (\lambda_{y,\Delta x_1}, \lambda_{y,\Delta x_2}, \cdots \lambda_{y,\Delta x_p})',$$

where $\lambda_{y,\Delta x_i} = \lim_{t\to\infty} cov(y_t, x_{it-1})$ for $i = 1, \cdots, p$.

If $\mathbf{x}_t$ is $I(1)$ ($C = 0$) and we assume $\lim_{t\to\infty} cov(y_t, \mathbf{x}_0) \to 0$, then

$$\lambda_{y,\Delta\mathbf{x}} = \lim_{t\to\infty} cov(y_t, \mathbf{x}_{t-1}) = \sum_{h=1}^{\infty} cov(y_t, \Delta\mathbf{x}_{t-h}), \tag{3.5}$$

which is well defined when $\sum_{h=1}^{\infty} | cov(y_t, \Delta x_{it-h}) | < \infty$ for $i = 1, \cdots, p$.

If $\mathbf{x}_t$ is local-to-unity ($C \leq 0$) with $\mathbf{x}_t \equiv 0$ for $t \leq 0$, then directly following equation (6) in MS, we obtain

$$\lambda_{y,\Delta\mathbf{x}} = \sum_{h=1}^{\infty} cov(y_t, \upsilon_{t-h}). + O(T^{-1}) = \lambda_{y,\upsilon} + O(T^{-1}),$$

and is well defined as long as $\sum_{h=1}^{\infty} | cov(y_t, \upsilon_{it-h}) | < \infty$ for $i = 1, \cdots, p$.

The one-sided kernel estimator of $\lambda_{y,\Delta\mathbf{x}}$ is

$$\widehat{\lambda}_{y,\Delta\mathbf{x}} = \sum_{h=1}^{T-1} k(\frac{h-1}{m})\widehat{\Gamma}_{\Delta\mathbf{x}y}(h),\ \widehat{\Gamma}_{\Delta\mathbf{x}y}(h) = \frac{1}{T}\sum_{t=h+1}^{T} y_t \Delta\mathbf{x}_{t-h},$$

where m is the bandwidth and $k(x)$ is the kernel. And

$$\widehat{\lambda}_{y,\Delta\mathbf{x}} = (\widehat{\lambda}_{y,\Delta x_1}, \widehat{\lambda}_{y,\Delta x_2}, \cdots \widehat{\lambda}_{y,\Delta x_p})',$$

where $\widehat{\lambda}_{y,\Delta x_i} = \sum\limits_{h=1}^{T-1} k(\frac{h-1}{m})\widehat{\Gamma}_{\Delta x_i y}(h) = \sum\limits_{h=1}^{T-1} k(\frac{h-1}{m})(\frac{1}{T}\sum\limits_{t=h+1}^{T} y_t \Delta x_{it-h})$ for $i = 1, \cdots, p$.

Assumption 2. The kernel $k(x)$ is continuous at $x = 0$ and uniformly bounded with $k(0) = 1$, $\int_0^\infty \bar{k}(x)dx < \infty$, and $\lim_{x\to 0+}(1-k(x))/(\mid x \mid^q) = k_q < \infty$ with $\delta \geq q$, where $\bar{k}(x) = \sup_{y \geq x} \mid k(y) \mid$.

Assumption 3. $\frac{1}{m} + \frac{m^{\max\{1,q\}}}{T} \longrightarrow 0$ as $T \longrightarrow \infty$.

First of all, if $C = 0$ and $\mathbf{x}_t$ is integrated, we have the following properties.

Lemma 1. If Assumption 1, 2 and 3 hold, then

(i) $\lim_{T\to\infty} m^q E(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) = -k_q \sum\limits_{h=1}^{\infty} \Gamma_{\Delta\mathbf{x}y}(h)h^q$,

(ii) $\widehat{\lambda}_{y,\Delta\mathbf{x}} \underset{p}{\longrightarrow} \lambda_{y,\Delta\mathbf{x}}$ as $T \to \infty$,

(iii) $\lim_{T\to\infty} Tm^{-1}var(\widehat{\lambda}_{y,\Delta\mathbf{x}}) = \mathbf{V} \equiv 4\pi^2 f_{yy}(0)[f_{\Delta x_i \Delta x_j}(0)]_{p\times p} \int_0^\infty k^2(x)dx$, where $f_{yy}(\tau)$ denote the spectral density of y_t,and $f_{\Delta x_i \Delta x_j}(\tau)$ denote the cross spectral density between Δx_{it} and Δx_{jt}.

Lemma 2. If Assumption 1, 2 and 3 hold, $\mathbf{V} > 0$ and $m^2/T + T/m^{2q+1} \to 0$, then

$$\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) \to_d \mathbf{N}(\mathbf{0}, \mathbf{V}), \quad \text{as } T \to \infty,$$

As stated in MS, neither a nonzero intercept in $(y_t, \mathbf{x}_t)$ nor a linear trend in $\mathbf{x}_t$ affects the limiting distribution.

When $C \leq 0$, $\mathbf{x}_t$ is local-to-unity.

Define $\widehat{\lambda}_{y,\mathbf{v}}$, $\widehat{\lambda}_{y,v_i}$, $\widehat{\Gamma}_{vy}(h)$ and $\widehat{\Gamma}_{v_iy}(h)$ analogously to $\widehat{\lambda}_{y,\Delta\mathbf{x}}$, $\widehat{\lambda}_{y,\Delta x_i}$, $\widehat{\Gamma}_{\Delta\mathbf{x}y}(h)$ and $\widehat{\Gamma}_{\Delta x_iy}(h)$ respectively, with $\mathbf{v}_t$ or v_{it} replacing $\Delta\mathbf{x}$ or Δx_{it}.

By Lemma 3 of MS,

$$\widehat{\lambda}_{y,\Delta x_i} = \sum_{h=1}^{T-1} k(\frac{h-1}{m})\widehat{\Gamma}_{v_iy}(h) + O_p(m/T) \underset{p}{\longrightarrow} \widehat{\lambda}_{y,v_i}$$

for each i. Therefore, we have the following properties.

Lemma 3. Suppose Assumption 1, 2 and 3 hold. Then $\widehat{\lambda}_{y,\Delta\mathbf{x}} \underset{p}{\longrightarrow} \widehat{\lambda}_{y,\mathbf{v}} \underset{p}{\longrightarrow} \lambda_{y,\mathbf{v}} \underset{p}{\longrightarrow} \lambda_{y,\Delta\mathbf{x}}$. If, in addition, $\mathbf{V_v} > 0$ and $m^2/T + T/m^{2q+1} \to 0$, then

$$\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) \underset{p}{\longrightarrow} \sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\mathbf{v}} - \lambda_{y,\mathbf{v}}) \to_d \mathbf{N}(\mathbf{0}, \mathbf{V_v}),$$

where $\mathbf{V_v} = 4\pi^2 f_{yy}(0)[f_{v_iv_j}(0)]_{p\times p}\int_0^\infty k^2(x)dx$, and $f_{v_iv_j}(\tau)$ denote the cross spectral density between v_{it} and v_{jt}.

Thus, when $\mathbf{x}_t$ are local-to-unity, the test has the same limiting distribution as $I(1)$ cases and does not depend on the local-to-unity parameter C.

Assumption 4. The kernel $\widetilde{k}(x)$ satisfies Assumption 2 with $\widetilde{q}$ replacing q, $\widetilde{k}(x) = 0$ if $| x |> 1$, and $1/\widetilde{m} + \widetilde{m}^{\max\{1,\widetilde{q}\}}/T \to 0$ as $T \to \infty$.

Lemma 4. If Assumption 1, 2, 3 and 4 hold, then $\widetilde{\mathbf{V}} \underset{p}{\longrightarrow} \mathbf{V}$ or $\mathbf{V_v}$ as $T \to \infty$.

The definition of $\widetilde{\mathbf{V}}$ can be found in Appendix B. $\widetilde{k}(x)$ and $\widetilde{m}$ are a kernel and a bandwidth that are not necessarily the same as $k(x)$ and m. In the simulation part, we take Bartlett kernel for both $k(x)$ and $\widetilde{k}(x)$, and set $\widetilde{m} = m^{0.9}$.

The proof is omitted because it is the same as that of Lemma 5 in MS.

Finally, W_1 can be defined as

$$W_1 \equiv \frac{T}{m}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}})'\widetilde{\mathbf{V}}^{-1}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}),$$

Corollary 5. If the assumptions of Lemma 2 or Lemma 3 and Lemma 4 hold, then

$$W_1 \to_d \chi^2(p) \text{ as } T \to \infty.$$

3.3.2 Predictors Are Stationary

When $\mathbf{x}_t$ is staionary, the covariance-based test does not have a χ^2 limiting distribution. Therefore, If we know both $\mathbf{x}_t$ and y_t are stationary before testing, a regression-based test should be implemented..

Assumption 5. $(y_t, \mathbf{x}_t)$ is generated by

$$\mathbf{w} = \begin{pmatrix} y_t \\ \mathbf{x}_t \end{pmatrix} = B(L)\varrho_t = \sum_{j=0}^{\infty} B_j \varrho_{t-j}, \quad \sum_{j=0}^{\infty} j \parallel B_j \parallel < \infty,$$

where $\varrho_t \sim IID(0, \Sigma)$ satisfying $\Sigma > 0$ and the moment condition $E \parallel \varrho_1 \parallel^4 < \infty$.

$$\sum_{h=-\infty}^{\infty} | h |^{\delta} \parallel \gamma(h) \parallel < \infty, \ \delta > 1; \ \gamma(h) = \begin{bmatrix} \gamma_{yy}(h) & \gamma_{y\Delta\mathbf{x}}(h) \\ \gamma_{\Delta\mathbf{x}y}(h) & \gamma_{\Delta\mathbf{x}\Delta\mathbf{x}}(h) \end{bmatrix} = E\mathbf{w}_t\mathbf{w}'_{t+h}.$$

Since $var(\mathbf{x}_{t-1})$ are finite in stationary cases, testing (3.1) is then equivalent to testing $\beta = 0$ in the following linear regression

$$y_t = \beta\mathbf{x}_{t-1} + \varepsilon_t, \ E(\varepsilon_t \mid I_{\mathbf{x},t-1}) = 0,$$

where $I_{\mathbf{x},t-1}$ denotes the information set which contains the past values of $\mathbf{x}_t$.We impose $E(\varepsilon_t \mid I_{\mathbf{x},t-1}) = 0$ to ensure that there is non-predictability of y_t under the null hypothesis.

Under this framework, the usual OLS estimators and tests are good enough since the finite sample bias disappears asymptotically. Denote $\widehat{\beta}$ and $\widehat{\varepsilon}_t$ as OLS estimators and residuals, respectively. And

$$\widehat{Avar(\widehat{\beta})} = (\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}')^{-1}\widehat{lrv(\mathbf{x}_{t-1}\varepsilon_t)}(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}')^{-1}.$$

where $\widehat{lrv(\mathbf{x}_{t-1}\varepsilon_t)}$ is Newey and West's (1987) estimator of the long run variance of $\mathbf{x}_{t-1}\varepsilon_t$.

Corollary 6. If Assumption 5 holds, then

$$W_0 \equiv T(\widehat{\beta}-\beta)'[\widehat{Avar(\widehat{\beta})}]^{-1}(\widehat{\beta}-\beta) \to_d \chi^2(p) \text{ as } T \to \infty. \tag{3.6}$$

3.4 The Robust Process

Since the integration order of $\mathbf{x}_t$ is unknown in practice, we need to construct a process such that W_0 is automatically selected when $\mathbf{x}_t$ is stationary and W_1 is selected when $\mathbf{x}_t$ is (near) nonstationary.

We follow the approach of Harvey, Leybourne and Taylor (2007, HLT henceforth) that constructs a new statistic which is a data-dependent weighted average of W_0 and W_1. We define the statistic W as

$$W = \{1-\lambda(\mathbf{U},\mathbf{S})\}W_0 + \lambda(\mathbf{U},\mathbf{S})W_1,$$

where $\lambda(\mathbf{U},\mathbf{S})$ is the weight that need to satisfy the following condition.

Assumption 6. $\lambda(\mathbf{U},\mathbf{S})$ is a function on [0,1] and for suitable choice of $\mathbf{U}$ and $\mathbf{S}$,

$$\lambda(\mathbf{U},\mathbf{S}) = \begin{cases} o_p(1), \text{ if } \mathbf{x}_t \text{ is } I(0) \\ 1+o_p(T^{-1/2}), \text{ if } \mathbf{x}_t \text{ is } I(1) \text{ or local-to-unity.} \end{cases}$$

Under this condition, $\lambda(\mathbf{U},\mathbf{S}) \xrightarrow{p} 0$ when predictors are $I(0)$, and $\lambda(\mathbf{U},\mathbf{S}) \xrightarrow{p} 1$

when predictors are local to unity or $I(1)$.

For the choice of the weight, we take

$$\lambda(\mathbf{U},\mathbf{S}) = \exp\left(-\frac{\sum_{i=1}^{p} U_i^2}{\sum_{i=1}^{p} S_i^2}\right),$$

where $U_i = DF - GLS_i^{\mu}$ and $S_i = \widehat{\eta}_{\mu i}$ are the local GLS-demeaned Dickey-Fuller statistic and the level-stationary KPSS statistic respectively for x_i.

In the stationary case, U_i diverges at at rate $O_e(T^{\delta_U})$[1], $\delta_U > 0$ and $S_i = O_p(1)$. We have

$$\lambda(\mathbf{U},\mathbf{S}) = \exp\left(-\frac{\sum_{i=1}^{p} \mid O_e(T^{2\delta_U}) \mid}{\sum_{i=1}^{p} \mid O_p(1) \mid}\right) = o_p(1).$$

In the (near) nonstationary case, $U_i = O_p(1)$ and S_i diverges at at rate $O_e(T^{\delta_S})$, $\delta_S > 1/4$. We have

$$\lambda(\mathbf{U},\mathbf{S}) = \exp\left(-\frac{\sum_{i=1}^{p} \mid O_p(1) \mid}{\sum_{i=1}^{p} \mid O_e(T^{2\delta_S}) \mid}\right) = 1 + O_e(T^{-2\delta_S}) = 1 + o_p(T^{-1/2}).$$

Next, we look at the orders of W_0 and W_1 under different assumptions.

Lemma 7. If Assumption 1 holds, $W_0 = T\widehat{\beta}'[\widehat{Avar(\widehat{\beta})}]^{-1}\widehat{\beta}$ is of $O_p(1)$ under either the null or the alternative hypothesis.

When both $\mathbf{x}_t$ and y_t are stationary, we have the following properties for W_1.

Lemma 8. If Assumption 5, 2, 3 and 4 hold, and $k(x)$ is *Lipschitz(1)*, then $W_1 = \frac{T}{m}(\widehat{\lambda}_{y,\Delta\mathbf{x}})'\widetilde{\mathbf{V}}^{-1}(\widehat{\lambda}_{y,\Delta\mathbf{x}})$ is of $O_p([m(\widetilde{m}^{-\widetilde{q}}+\widetilde{m}^{1/2}T^{-1/2})]^{-1})$ under the null hypothesis that $\lambda_{y,\Delta\mathbf{x}} = 0$, and of $O_p(T[m(\widetilde{m}^{-\widetilde{q}} + \widetilde{m}^{1/2}T^{-1/2})]^{-1})$ under the alternative hypothesis.

Theorem 2. If Assumptions 2, 3, 4 and 6 hold, and in addition, $m^2/T + T/m^{2q+1} \to 0$, $k(x)$ is *Lipschitz(1).* And any of the following conditions is satisfied: (1), $q = 1$ and $\widetilde{q} = 1$; (2), $q = 2$ and $\widetilde{q} = 1$, $\widetilde{m} \leq \min\{T^{1/3}, m\}$ or $\widetilde{m} > \max\{T^{1/3}, \frac{T}{m^2}\}$;

[1]For here, $O_e(T^k)$ denotes exact order in probability.

(3), $q=1$ and $\widetilde{q}=2$, $\widetilde{m} \leq \min\{T^{1/5}, m^{1/2}\}$ or $\widetilde{m} > \max\{T^{1/5}, \frac{T}{m^2}\}$ or (4), $q=2$ and $\widetilde{q}=2$, $\widetilde{m} \leq \min\{T^{1/5}, m^{1/2}\}$ or $\widetilde{m} > \max\{T^{1/5}, \frac{T}{m^2}\}$. Then we have under the null hypothesis,

$$W = \begin{cases} \{1+o_p(1)\} \cdot W_0 + o_p(1) \cdot O_p([m(\widetilde{m}^{-\widetilde{q}} + \widetilde{m}^{1/2}T^{-1/2})]^{-1})) \\ \qquad = W_0 + o_p(1), \text{ if } \mathbf{x}_t \text{ is } I(0) \\ o_p(T^{-1/2}) \cdot O_p(1) + \{1 + o_p(T^{-1/2})\} \cdot W_1 = W_1 + o_p(1), \\ \qquad \text{if } \mathbf{x}_t \text{ is } I(1) \text{ or local-to-unity,} \end{cases}$$

and under the althernatives,

$$W = \begin{cases} O_p(T), \text{ if } \mathbf{x}_t \text{ is } I(0) \\ O_p(T/m), \text{ if } \mathbf{x}_t \text{ is } I(1) \text{ or local-to-unity.} \end{cases}$$

Remark 3. MS (2009) impose that $\widetilde{k}(x)$ have to be the Bartlett kernel and $\widetilde{m}$ must be smaller than m to ensure the size of W_1 is conservative in stationary cases. While for here, our test allows more flexible choices for $\widetilde{k}(x)$ and $\widetilde{m}$. For the choice of m, from Lemma 1, we know the optimal bandwidth is $m^* = cT^{1/2q+1}$. However, to obain Lemma 2, m needs to grow faster than m^*. Therefore, in practice, we can select m as small as possible and slightly larger than the m^*.

Remark 4. Under the null hypothesis, the test statistic W is asymptotically equivalent to W_0 in stationary cases, and to W_1 in (near) nonstationary cases. Our test can sucessfully control the size. On the other hand, under the alternatives, the test has a very strong power no matter whether the predictors are $I(0)$, local-to-unity or $I(1)$.

Remark 5. When there is an intercept or a linear trend in $(y_t, \mathbf{x}_t)$, demeaned or detrended residuals should be employed and our test still works.

Remark 6. When there is only one predictor, one may also implement a t version of the test. That is,

$$t = \{1 - \lambda(U,S)\}t_0 + \lambda(U,S)t_1,$$

where

$$t_1 = \frac{\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x})}{\sqrt{\widetilde{V}}} \to_d N(0,1) \text{ as } T \to \infty,$$

$$t_0 = \frac{\sqrt{T}(\widehat{\beta} - \beta)}{\sqrt{\widehat{Avar(\widehat{\beta})}}} \to_d N(0,1) \text{ as } T \to \infty,$$

and

$$\lambda(U,S) = \exp\left(-\frac{U^2}{S^2}\right) = \exp(-(\frac{DF - GLS^{\mu}}{\widehat{\eta}_{\mu}})^2).$$

And under the assumptions of Theorem 2, we have

$$t = \begin{cases} \{1 + o_p(1)\} \cdot t_0 + o_p(1) \cdot O_p([m(\widetilde{m}^{-\widetilde{q}} + \widetilde{m}^{1/2}T^{-1/2})]^{-1/2}) = t_0 + o_p(1), \\ \text{if } x_t \text{ is } I(0) \\ o_p(T^{-1/2}) \cdot O_p(1) + \{1 + o_p(T^{-1/2})\} \cdot t_1 = t_1 + o_p(1), \\ \text{if } x_t \text{ is } I(1) \text{ or local-to-unity} \end{cases}$$

under the null hypothesis, and

$$t = \begin{cases} O_p(T^{1/2}), \text{ if } x_t \text{ is } I(0) \\ O_p(\sqrt{T/m}), \text{ if } x_t \text{ is } I(1) \text{ or local-to-unity} \end{cases}$$

under the alternatives. This t test version is more convenient when comparing with previous literature. In simulation part, we take this version[2].

[2]Similar results were found using the Wald test version.

3.5 Simulation Results

3.5.1 DGP

This section provides simulation results. For simplicity, we consider only single predictor cases. We use the following data generating process:

$$x_t = \rho x_{t-1} + u_{2t}, \tag{3.7}$$

where $\rho \in \{1, 0.99, 0.95, 0.90, 0.80, 0.50, 0.10, 0\}$.

For the null hypothesis, we specify

$$y_t = a + u_{1t} \tag{3.8}$$

where $a = 0$ and

$$\begin{aligned} (u_{1t}, u_{2t})' &= \Sigma^{1/2} v_t, \quad v_t \sim NIID(0, I_2) \\ \Sigma &= \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \end{aligned} \tag{3.9}$$

where $\sigma_{12} \in \{0, 0.25, 0.50, 0.75, 0.95\}$.

We choose two alternative hypotheses

$$y_t = a + \beta x_{t-1} + u_{1t}, \tag{3.10}$$

and

$$y_t = a + \gamma(x_{t-1} - (1 + c/T)x_{t-2}) + u_{1t} = a + \gamma u_{2,t-1} + u_{1t}, \tag{3.11}$$

The second one allows y_t to keep stationary when x_t is (near) $I(1)$.

Since our new test statistic t_W equals the weighted average of t_0 and t_1, we first need to estimate t_0 and t_1 separately. t_0 is estimated following equations (3.6). For t_1 part, to calculate $\widehat{\lambda}_{y,\Delta x}$ and $\widetilde{V}$, we need to choose appropriate kernel and bandwitch. Bartlett kernel is used for both $k(x)$ and $\widetilde{k}(x)$[3]. Following Andrews (1991), the optimal bandwidth m^* is estimated by

$$\widehat{m}^* = 1.1447(\widehat{\alpha}(1)T)^{1/3}, \tag{3.12}$$

and $\widehat{\alpha}(1)$ is got through equation (6.4) in Andrews (1991)

$$\widehat{\alpha}(1) = \sum_{a=1}^{p} w_a \frac{4\widehat{\rho}_a^2\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^6(1+\widehat{\rho}_a)^2} / \sum_{a=1}^{p} w_a \frac{\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^4}.$$

In the above equation, $(\widehat{\rho}_a, \widehat{\sigma}_a^2)$ are AR parameters and innovation variance from AR(1) models for $a = 1, ..., p$. In our simulation, we choose two AR(1) models $y_t = \widehat{\rho}_1 y_{t-1} + \widehat{\varepsilon}_t$ and $\triangle x_t = \widehat{\rho}_2 \triangle x_{t-1} + \widehat{\epsilon}_t$. We take $w_a = 1$ for $a = 1, 2$.

The bandwidth $\widetilde{m}$ in estimating V is set to $\widetilde{m} = (\widehat{m}^*)^{0.9}$.

For better comparison with tests developed under the commonly used assumption that y_t (for example, one-period returns) is a martingale difference sequence (MDS) under the null, we follow MS and simplify $\widetilde{V}$ as

$$\widetilde{V}_{MDS} = \frac{1}{m}\sum_{h'=1}^{T-1}\sum_{h=1}^{T-1} k(\frac{h'-1}{m})k(\frac{h-1}{m}) \times \widetilde{k}(\frac{h'-h}{\widetilde{m}})\widehat{\Gamma}_{\Delta x \Delta x}(h'-h)\widehat{\Gamma}_{yy}(0),$$

although our test allows for serial correlation in y_t.

[3]Other kernels were also considered, including the Parzen, Quadratic Spectral and Tukey–Hannning kernels. This do not change our results much.

For the weight $\lambda(U,S)$, based on simulation results, we set

$$\begin{aligned}\lambda(U,S) &= \exp[-c(\frac{DF-GLS^{\mu}}{\widehat{\eta}_{\mu}})^2],\\ c &= 0.006,\end{aligned}$$

to acquire numerically the best overall finite sample performance[4]. $DF-GLS^{\mu}$ is the local GLS-demeaned Dickey-Fuller statistics and the number of lagged different terms is determined by the MAIC method of Ng and Perron (2001), with the maximum at the integer part of $12(T/100)^{1/4}$. $\widehat{\eta}_{\mu}$ is the level-stationary KPSS statistic. Following Kwiatkowski, Phillips, Schmidt and Shin (1992), we have

$$\widehat{\eta}_{\mu} = \frac{\sum_{t=1}^{T}(\sum_{i=1}^{t} e_i)^2}{T^2\widehat{\omega}^2}, e_i = x_i - \bar{x},$$

where $\widehat{\omega}^2$ is the long run variance estimator and takes the form: $\widehat{\omega}^2 = T^{-1}\sum\limits_{t=1}^{T} e_t^2 + 2T^{-1}\sum\limits_{j=1}^{l} w(j,l)\sum\limits_{t=j+1}^{T} e_t e_{t-j}$; $w(j,l) = 1 - \frac{j}{l+1}$ and l is the integer part of $12(T/100)^{1/4}$.

To differentiate our test from those based on pre-test methods, we also construct two test statistics using the DF-GLS and the KPSS as pre test respectively. More specifically, in the first step, we test x_t by the DF-GLS (or the KPSS) at 5% significance level. And then, if we can reject the null hypothesis (x_t is nonstationary when the DF-GLS test is used, and stationary when the KPSS test is used), we set $t - DFGLS = t_0$ (or $t - KPSS = t_1$). If we can not reject the null hypothesis in the first step, we set $t - DFGLS = t_1$ (or $t - KPSS = t_0$).

All simulation results are based on 2000 replications and sample sizes of $T = 100, 200$ and 500.

[4] We discuss the choice of the parameter c in section 3.5.4.

3.5.2 Size

Table 3.1 shows rejection rates for the weighted average test t with a nominal significant level of 5%. These results are reliable in either stationary cases or non-stationary cases, even when the sample size T is not very large (for example, when $T = 100$).

By contrast, although rejection rates for the covariance-based test t_1, provided in Table 3.2, have good performance when x_t is $I(1)$ or local-to-unity, this test becomes quite conservative for $I(0)$ cases and the results do not improve as the sample size grows larger. For example, in the worst case for $\rho = 0.1$ and $\sigma_{12} = 0$, the size is only 0.007 even for a large sample size of 500. This corroborates that t_1 converges to zero in probability in stationary cases.

Table 3.3 shows that the rejection rates for the regression-based test statistic t_0 are reasonable when x_t is stationary or σ_{12} is small but very unreliable when x_t is nonstationary and σ_{12} is large. The rejection rates can even exceed 30%.

Table 3.1: The new t-statistic: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0405	0.0425	0.0505	0.0575	0.0675
0.990	0.0440	0.0395	0.0365	0.0595	0.0650
0.950	0.0360	0.0350	0.0420	0.0615	0.0640
0.900	0.0380	0.0380	0.0455	0.0525	0.0680
0.800	0.0405	0.0400	0.0615	0.0620	0.0530
0.500	0.0485	0.0450	0.0450	0.0505	0.0555
0.100	0.0450	0.0460	0.0445	0.0375	0.0450
0.000	0.0355	0.0380	0.0390	0.0440	0.0435
			$T = 200$		
1.000	0.0495	0.0465	0.0470	0.0455	0.0520
0.990	0.0370	0.0345	0.0435	0.0540	0.0575
0.950	0.0325	0.0420	0.0445	0.0510	0.0525
0.900	0.0370	0.0495	0.0440	0.0510	0.0570
0.800	0.0510	0.0465	0.0415	0.0510	0.0535
0.500	0.0420	0.0465	0.0470	0.0400	0.0515
0.100	0.0320	0.0375	0.0330	0.0325	0.0440
0.000	0.0360	0.0410	0.0285	0.0370	0.0385
			$T = 500$		
1.000	0.0425	0.0510	0.0360	0.0405	0.0410
0.990	0.0380	0.0355	0.0405	0.0380	0.0435
0.950	0.0300	0.0275	0.0355	0.0390	0.0450
0.900	0.0310	0.0385	0.0395	0.0460	0.0545
0.800	0.0505	0.0480	0.0475	0.0535	0.0435
0.500	0.0415	0.0395	0.0400	0.0350	0.0615
0.100	0.0350	0.0395	0.0390	0.0415	0.0385
0.000	0.0365	0.0345	0.0370	0.0325	0.0390

Table 3.2: Maynard and Shimotsu (2009) t-statistic: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0515	0.0470	0.0490	0.0470	0.0455
0.990	0.0535	0.0460	0.0315	0.0460	0.0300
0.950	0.0430	0.0420	0.0420	0.0410	0.0335
0.900	0.0380	0.0430	0.0465	0.0370	0.0315
0.800	0.0430	0.0415	0.0360	0.0350	0.0360
0.500	0.0290	0.0265	0.0300	0.0270	0.0275
0.100	0.0165	0.0145	0.0160	0.0155	0.0155
0.000	0.0100	0.0125	0.0105	0.0130	0.0170
			$T = 200$		
1.000	0.0530	0.0525	0.0475	0.0405	0.0415
0.990	0.0495	0.0435	0.0370	0.0425	0.0410
0.950	0.0435	0.0525	0.0450	0.0460	0.0360
0.900	0.0485	0.0485	0.0425	0.0445	0.0380
0.800	0.0435	0.0505	0.0390	0.0430	0.0355
0.500	0.0260	0.0270	0.0265	0.0215	0.0310
0.100	0.0120	0.0170	0.0100	0.0150	0.0160
0.000	0.0070	0.0110	0.0130	0.0120	0.0090
			$T = 500$		
1.000	0.0450	0.0530	0.0375	0.0410	0.0370
0.990	0.0480	0.0435	0.0420	0.0380	0.0380
0.950	0.0450	0.0355	0.0470	0.0455	0.0345
0.900	0.0445	0.0420	0.0505	0.0395	0.0440
0.800	0.0415	0.0425	0.0355	0.0335	0.0410
0.500	0.0220	0.0260	0.0190	0.0235	0.0315
0.100	0.0070	0.0140	0.0135	0.0125	0.0125
0.000	0.0095	0.0100	0.0115	0.0115	0.0090

Table 3.3: Regression t-statistic: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0860	0.1000	0.1640	0.2385	0.3485
0.990	0.0855	0.0915	0.1515	0.2040	0.2705
0.950	0.0930	0.0860	0.1040	0.1430	0.1570
0.900	0.0955	0.0840	0.0995	0.1170	0.1330
0.800	0.0740	0.0815	0.1000	0.0995	0.0985
0.500	0.0835	0.0780	0.0760	0.0880	0.1005
0.100	0.0800	0.0800	0.0820	0.0790	0.0825
0.000	0.0790	0.0775	0.0850	0.0850	0.0905
			$T = 200$		
1.000	0.0745	0.0815	0.1395	0.2065	0.3150
0.990	0.0685	0.0795	0.1135	0.1585	0.1990
0.950	0.0575	0.0710	0.0825	0.0890	0.1120
0.900	0.0710	0.0825	0.0750	0.0870	0.0895
0.800	0.0725	0.0815	0.0640	0.0710	0.0780
0.500	0.0655	0.0685	0.0710	0.0625	0.0680
0.100	0.0620	0.0675	0.0630	0.0635	0.0755
0.000	0.0600	0.0685	0.0515	0.0595	0.0685
			$T = 500$		
1.000	0.0720	0.0785	0.1375	0.1890	0.2950
0.990	0.0690	0.0660	0.0775	0.1030	0.1270
0.950	0.0650	0.0555	0.0555	0.0705	0.0760
0.900	0.0450	0.0635	0.0615	0.0680	0.0780
0.800	0.0655	0.0645	0.0620	0.0640	0.0575
0.500	0.0545	0.0570	0.0500	0.0475	0.0720
0.100	0.0530	0.0520	0.0545	0.0575	0.0510
0.000	0.0640	0.0555	0.0590	0.0575	0.0560

Table 3.4: DFGLS pre-test t-statistic: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0510	0.0495	0.0550	0.0650	0.0870
0.990	0.0525	0.0490	0.0555	0.0730	0.1045
0.950	0.0665	0.0685	0.0740	0.0975	0.1170
0.900	0.0650	0.0800	0.0710	0.0850	0.1005
0.800	0.0730	0.0710	0.0840	0.0740	0.0735
0.500	0.0535	0.0575	0.0555	0.0500	0.0420
0.100	0.0280	0.0295	0.0330	0.0260	0.0300
0.000	0.0250	0.0260	0.0300	0.0355	0.0235
			$T = 200$		
1.000	0.0525	0.0595	0.0625	0.0670	0.0775
0.990	0.0435	0.0610	0.0770	0.0795	0.1000
0.950	0.0570	0.0655	0.0725	0.0820	0.0985
0.900	0.0595	0.0685	0.0665	0.0750	0.0840
0.800	0.0630	0.0625	0.0710	0.0700	0.0535
0.500	0.0565	0.0545	0.0450	0.0525	0.0505
0.100	0.0335	0.0320	0.0360	0.0240	0.0375
0.000	0.0305	0.0315	0.0240	0.0230	0.0340
			$T = 500$		
1.000	0.0515	0.0495	0.0520	0.0680	0.0800
0.990	0.0500	0.0580	0.0620	0.0895	0.1105
0.950	0.0580	0.0565	0.0580	0.0640	0.0820
0.900	0.0635	0.0560	0.0675	0.0715	0.0715
0.800	0.0535	0.0565	0.0580	0.0665	0.0665
0.500	0.0450	0.0430	0.0380	0.0515	0.0465
0.100	0.0355	0.0245	0.0315	0.0345	0.0340
0.000	0.0380	0.0315	0.0375	0.0330	0.0300

Table 3.5: KPSS pre-test: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0595	0.0780	0.1080	0.1710	0.2290
0.990	0.0745	0.0740	0.1065	0.1570	0.2090
0.950	0.0875	0.0790	0.0920	0.1065	0.1405
0.900	0.0770	0.0875	0.0935	0.0940	0.1155
0.800	0.0885	0.0830	0.1015	0.1010	0.1065
0.500	0.0740	0.0895	0.0845	0.0795	0.0850
0.100	0.0735	0.0800	0.0775	0.0740	0.0855
0.000	0.0730	0.0690	0.0740	0.0910	0.0825
			$T = 200$		
1.000	0.0555	0.0645	0.0845	0.1225	0.1720
0.990	0.0520	0.0615	0.0975	0.1120	0.1485
0.950	0.0520	0.0720	0.0735	0.0880	0.0940
0.900	0.0550	0.0680	0.0705	0.0745	0.0885
0.800	0.0635	0.0640	0.0725	0.0730	0.0580
0.500	0.0730	0.0720	0.0590	0.0675	0.0690
0.100	0.0600	0.0595	0.0670	0.0550	0.0695
0.000	0.0560	0.0695	0.0560	0.0660	0.0665
			$T = 500$		
1.000	0.0470	0.0495	0.0655	0.0870	0.0910
0.990	0.0525	0.0575	0.0590	0.0820	0.0965
0.950	0.0620	0.0585	0.0545	0.0640	0.0810
0.900	0.0600	0.0520	0.0660	0.0690	0.0665
0.800	0.0510	0.0535	0.0615	0.0680	0.0670
0.500	0.0515	0.0445	0.0435	0.0600	0.0535
0.100	0.0625	0.0440	0.0460	0.0520	0.0505
0.000	0.0650	0.0545	0.0565	0.0550	0.0515

We next consider the two pretest-based statistics: $t - DFGLS$ and $t - KPSS$. The rejection rates are shown in Table 3.4 and 3.5. From these two tables, we find the pretest-based test can not completely solve the size problem and their performances are much worse than our test's, especially for a small sample size. For example, when $T = 100$, $\rho = 0.99$ and $\sigma_{12} = 0.95$, the size is over 20% for $t - KPSS$ and over 10% for $t - DFGLS$. On the other hand, when both ρ and σ_{12} approach zero, $t - DFGLS$ becomes quite conservative.

3.5.3 Power

We first consider the power of the weighted average test t against equation (3.11) with $\beta \neq 0$, which is the standard regression alternative. This alternative implies that x_t and y_t have the same order of integration. When ρ is equal to one, x_t and y_t are cointegrated, and when $\rho \ll 1$, both of x_t and y_t are stationary. Since y_t is usually stationary in practice, we should focus more on those results when ρ is not that close to one, for example, $\rho = 0.9, 0.8, 0.5, 0.1$ and 0. Table 3.6 gives the results. As expected, the power of the test is strong and increases in both sample size and β.

Table 3.6: The t-statistic: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.4605	0.7080	0.8595	0.9915	1.0000
	0.95	0.4310	0.6590	0.8350	0.9890	0.9995
$\rho = 0.99$	0.50	0.3755	0.6335	0.8490	0.9955	1.0000
	0.95	0.3445	0.5620	0.7810	0.9885	1.0000
$\rho = 0.95$	0.50	0.3925	0.6385	0.8085	0.9890	1.0000
	0.95	0.3835	0.6210	0.8005	0.9885	0.9995
$\rho = 0.90$	0.50	0.4285	0.6630	0.8310	0.9845	0.9995
	0.95	0.4205	0.6690	0.8385	0.9920	1.0000
$\rho = 0.80$	0.50	0.3690	0.6005	0.7625	0.9770	0.9985
	0.95	0.3985	0.6215	0.7860	0.9755	0.9995
$\rho = 0.50$	0.50	0.2255	0.4270	0.5875	0.9295	0.9965
	0.95	0.3065	0.5000	0.6715	0.9530	0.9975
$\rho = 0.10$	0.50	0.2480	0.4055	0.5775	0.9320	0.9980
	0.95	0.2115	0.3565	0.5455	0.9305	0.9955
$\rho = 0$	0.50	0.2310	0.3740	0.5550	0.9315	0.9955
	0.95	0.2125	0.3500	0.5405	0.9345	0.9975
			$T = 200$			
$\rho = 1$	0.50	0.6460	0.8920	0.9760	1.0000	1.0000
	0.95	0.6895	0.8960	0.9820	0.9995	1.0000
$\rho = 0.99$	0.50	0.5665	0.8635	0.9720	1.0000	1.0000
	0.95	0.6075	0.8800	0.9685	0.9995	1.0000
$\rho = 0.95$	0.50	0.6745	0.8890	0.9740	1.0000	1.0000
	0.95	0.6760	0.8875	0.9725	1.0000	1.0000
$\rho = 0.90$	0.50	0.7125	0.8955	0.9775	1.0000	1.0000
	0.95	0.7055	0.8805	0.9675	1.0000	1.0000
$\rho = 0.80$	0.50	0.5840	0.8360	0.9450	0.9990	1.0000
	0.95	0.6410	0.8750	0.9575	0.9985	1.0000
$\rho = 0.50$	0.50	0.4495	0.7115	0.8780	0.9995	1.0000
	0.95	0.4275	0.7185	0.8835	0.9990	1.0000
$\rho = 0.10$	0.50	0.3635	0.6075	0.8470	0.9985	1.0000
	0.95	0.3780	0.6620	0.8410	0.9985	1.0000
$\rho = 0$	0.50	0.3835	0.6585	0.8565	1.0000	1.0000
	0.95	0.3980	0.6565	0.8525	0.9995	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.9510	0.9980	0.9995	1.0000	1.0000
	0.95	0.9540	0.9990	1.0000	1.0000	1.0000
$\rho = 0.99$	0.50	0.9045	0.9985	1.0000	1.0000	1.0000
	0.95	0.9025	0.9960	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	0.9470	0.9960	1.0000	1.0000	1.0000
	0.95	0.9595	0.9970	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.9580	0.9960	0.9995	1.0000	1.0000
	0.95	0.9620	0.9960	0.9995	1.0000	1.0000
$\rho = 0.80$	0.50	0.9095	0.9910	0.9995	1.0000	1.0000
	0.95	0.9170	0.9925	1.0000	1.0000	1.0000
$\rho = 0.50$	0.50	0.7860	0.9795	0.9975	1.0000	1.0000
	0.95	0.7510	0.9640	0.9945	1.0000	1.0000
$\rho = 0.10$	0.50	0.7120	0.9430	0.9970	1.0000	1.0000
	0.95	0.7200	0.9495	0.9960	1.0000	1.0000
$\rho = 0$	0.50	0.6710	0.9430	0.9945	1.0000	1.0000
	0.95	0.6735	0.9335	0.9950	1.0000	1.0000

Table 3.7: Maynard and Shimotsu (2009) t-statistic: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.4070	0.6175	0.7615	0.9735	0.9990
	0.95	0.4035	0.6065	0.7585	0.9625	0.9965
$\rho = 0.99$	0.50	0.2835	0.4870	0.6670	0.9665	0.9985
	0.95	0.3095	0.4835	0.6705	0.9530	0.9975
$\rho = 0.95$	0.50	0.2015	0.3200	0.5110	0.9240	0.9970
	0.95	0.2340	0.3825	0.5510	0.9365	0.9965
$\rho = 0.90$	0.50	0.1845	0.3150	0.4640	0.8795	0.9925
	0.95	0.1800	0.3185	0.5075	0.9085	0.9950
$\rho = 0.80$	0.50	0.1640	0.2915	0.4100	0.8410	0.9795
	0.95	0.1860	0.3250	0.4715	0.8775	0.9900
$\rho = 0.50$	0.50	0.1785	0.3100	0.4660	0.8450	0.9865
	0.95	0.2145	0.3340	0.5020	0.8740	0.9885
$\rho = 0.10$	0.50	0.2420	0.3810	0.5400	0.9135	0.9970
	0.95	0.1865	0.3150	0.4935	0.9065	0.9940
$\rho = 0$	0.50	0.2120	0.3405	0.5245	0.9135	0.9935
	0.95	0.1955	0.3215	0.5010	0.9145	0.9970
			$T = 200$			
$\rho = 1$	0.50	0.5815	0.8205	0.9450	1.0000	1.0000
	0.95	0.6445	0.8300	0.9545	0.9990	1.0000
$\rho = 0.99$	0.50	0.4085	0.7275	0.8995	0.9995	1.0000
	0.95	0.4570	0.7560	0.9145	0.9980	1.0000
$\rho = 0.95$	0.50	0.2990	0.5480	0.8085	0.9970	1.0000
	0.95	0.3050	0.5670	0.8115	0.9985	1.0000
$\rho = 0.90$	0.50	0.2860	0.5025	0.7425	0.9955	1.0000
	0.95	0.2895	0.5110	0.7590	0.9950	1.0000
$\rho = 0.80$	0.50	0.2705	0.4825	0.6880	0.9885	1.0000
	0.95	0.2725	0.5380	0.7275	0.9930	1.0000
$\rho = 0.50$	0.50	0.3510	0.5555	0.7495	0.9940	1.0000
	0.95	0.3050	0.5480	0.7350	0.9910	1.0000
$\rho = 0.10$	0.50	0.3145	0.5570	0.8090	0.9950	1.0000
	0.95	0.3360	0.5865	0.8000	0.9960	1.0000
$\rho = 0$	0.50	0.3270	0.6065	0.8235	0.9985	1.0000
	0.95	0.3535	0.6020	0.8220	0.9995	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.9195	0.9960	0.9995	1.0000	1.0000
	0.95	0.9255	0.9950	0.9995	1.0000	1.0000
$\rho = 0.99$	0.50	0.7635	0.9900	0.9985	1.0000	1.0000
	0.95	0.7580	0.9810	0.9975	1.0000	1.0000
$\rho = 0.95$	0.50	0.6090	0.9240	0.9965	1.0000	1.0000
	0.95	0.6485	0.9480	0.9965	1.0000	1.0000
$\rho = 0.90$	0.50	0.5090	0.8595	0.9850	1.0000	1.0000
	0.95	0.5950	0.8975	0.9880	1.0000	1.0000
$\rho = 0.80$	0.50	0.5060	0.8265	0.9755	1.0000	1.0000
	0.95	0.5550	0.8780	0.9835	1.0000	1.0000
$\rho = 0.50$	0.50	0.6165	0.9155	0.9875	1.0000	1.0000
	0.95	0.5995	0.8970	0.9795	1.0000	1.0000
$\rho = 0.10$	0.50	0.6700	0.9275	0.9940	1.0000	1.0000
	0.95	0.6890	0.9390	0.9925	1.0000	1.0000
$\rho = 0$	0.50	0.6470	0.9340	0.9945	1.0000	1.0000
	0.95	0.6460	0.9210	0.9945	1.0000	1.0000

Table 3.8: Regression t-statistic: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.8795	0.9730	0.9955	1.0000	1.0000
	0.95	0.8975	0.9730	0.9890	1.0000	1.0000
$\rho = 0.99$	0.50	0.8340	0.9565	0.9880	1.0000	1.0000
	0.95	0.8445	0.9600	0.9855	1.0000	1.0000
$\rho = 0.95$	0.50	0.7075	0.8970	0.9650	0.9985	1.0000
	0.95	0.7470	0.9075	0.9655	0.9995	1.0000
$\rho = 0.90$	0.50	0.6445	0.8350	0.9455	0.9985	1.0000
	0.95	0.6045	0.8335	0.9315	0.9985	1.0000
$\rho = 0.80$	0.50	0.4375	0.6950	0.8625	0.9965	1.0000
	0.95	0.4600	0.6875	0.8460	0.9930	1.0000
$\rho = 0.50$	0.50	0.2595	0.4690	0.6725	0.9695	0.9990
	0.95	0.3500	0.5555	0.7290	0.9760	0.9995
$\rho = 0.10$	0.50	0.2590	0.4405	0.6145	0.9550	0.9985
	0.95	0.2440	0.4020	0.5965	0.9595	0.9995
$\rho = 0$	0.50	0.2565	0.4140	0.6230	0.9540	0.9985
	0.95	0.2605	0.4265	0.6225	0.9645	1.0000
			$T = 200$			
$\rho = 1$	0.50	0.9900	0.9995	1.0000	1.0000	1.0000
	0.95	0.9920	0.9995	1.0000	1.0000	1.0000
$\rho = 0.99$	0.50	0.9895	0.9990	1.0000	1.0000	1.0000
	0.95	0.9700	0.9970	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	0.9315	0.9930	1.0000	1.0000	1.0000
	0.95	0.9260	0.9855	0.9985	1.0000	1.0000
$\rho = 0.90$	0.50	0.8605	0.9820	0.9985	1.0000	1.0000
	0.95	0.8490	0.9710	0.9960	1.0000	1.0000
$\rho = 0.80$	0.50	0.6910	0.9160	0.9880	1.0000	1.0000
	0.95	0.6910	0.9210	0.9795	1.0000	1.0000
$\rho = 0.50$	0.50	0.4865	0.7730	0.9200	1.0000	1.0000
	0.95	0.4610	0.7620	0.9225	0.9990	1.0000
$\rho = 0.10$	0.50	0.3915	0.6475	0.8815	0.9990	1.0000
	0.95	0.3995	0.6990	0.8775	0.9995	1.0000
$\rho = 0$	0.50	0.4005	0.6880	0.8870	1.0000	1.0000
	0.95	0.4175	0.6885	0.8825	1.0000	1.0000
			$T = 500$			
$\rho = 1$	0.50	1.0000	1.0000	1.0000	1.0000	1.0000
	0.95	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.99$	0.50	1.0000	1.0000	1.0000	1.0000	1.0000
	0.95	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	0.9995	1.0000	1.0000	1.0000	1.0000
	0.95	0.9995	1.0000	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.9960	1.0000	1.0000	1.0000	1.0000
	0.95	0.9890	1.0000	1.0000	1.0000	1.0000
$\rho = 0.80$	0.50	0.9595	0.9990	1.0000	1.0000	1.0000
	0.95	0.9550	0.9995	1.0000	1.0000	1.0000
$\rho = 0.50$	0.50	0.8120	0.9860	0.9985	1.0000	1.0000
	0.95	0.7935	0.9795	0.9975	1.0000	1.0000
$\rho = 0.10$	0.50	0.7405	0.9570	0.9985	1.0000	1.0000
	0.95	0.7610	0.9645	0.9975	1.0000	1.0000
$\rho = 0$	0.50	0.6920	0.9500	0.9970	1.0000	1.0000
	0.95	0.7030	0.9490	0.9975	1.0000	1.0000

Table 3.9: DFGLS pre-test: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.4150	0.6410	0.7895	0.9825	0.9995
	0.95	0.4000	0.6090	0.7715	0.9700	0.9975
$\rho = 0.99$	0.50	0.2910	0.5225	0.7135	0.9720	0.9985
	0.95	0.3065	0.4980	0.7040	0.9670	0.9980
$\rho = 0.95$	0.50	0.2975	0.5015	0.6770	0.9610	0.9985
	0.95	0.2620	0.4930	0.6975	0.9635	0.9990
$\rho = 0.90$	0.50	0.4690	0.6600	0.7950	0.9645	0.9985
	0.95	0.4240	0.6545	0.8100	0.9785	1.0000
$\rho = 0.80$	0.50	0.3570	0.5825	0.7155	0.9350	0.9915
	0.95	0.4035	0.5865	0.7455	0.9545	0.9965
$\rho = 0.50$	0.50	0.2205	0.3935	0.5280	0.8755	0.9850
	0.95	0.2775	0.4480	0.6155	0.9150	0.9940
$\rho = 0.10$	0.50	0.2440	0.3925	0.5610	0.9155	0.9965
	0.95	0.1930	0.3250	0.5090	0.8975	0.9930
$\rho = 0$	0.50	0.2035	0.3355	0.5165	0.9035	0.9930
	0.95	0.2095	0.3380	0.5200	0.9150	0.9965
			$T = 200$			
$\rho = 1$	0.50	0.6180	0.8545	0.9590	1.0000	1.0000
	0.95	0.6650	0.8575	0.9665	0.9995	1.0000
$\rho = 0.99$	0.50	0.4995	0.7935	0.9415	0.9995	1.0000
	0.95	0.5240	0.8130	0.9490	0.9985	1.0000
$\rho = 0.95$	0.50	0.7560	0.8870	0.9590	1.0000	1.0000
	0.95	0.7865	0.9080	0.9670	0.9995	1.0000
$\rho = 0.90$	0.50	0.7945	0.9110	0.9615	0.9995	1.0000
	0.95	0.7900	0.9135	0.9615	0.9985	1.0000
$\rho = 0.80$	0.50	0.6045	0.8270	0.9275	0.9935	1.0000
	0.95	0.6315	0.8520	0.9405	0.9990	1.0000
$\rho = 0.50$	0.50	0.4025	0.6490	0.8210	0.9935	1.0000
	0.95	0.4025	0.6635	0.8390	0.9940	1.0000
$\rho = 0.10$	0.50	0.3435	0.5890	0.8160	0.9970	1.0000
	0.95	0.3585	0.6200	0.8175	0.9975	1.0000
$\rho = 0$	0.50	0.3650	0.6340	0.8355	0.9990	1.0000
	0.95	0.3655	0.6015	0.8105	0.9985	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.9470	0.9985	0.9995	1.0000	1.0000
	0.95	0.9470	0.9965	0.9995	1.0000	1.0000
$\rho = 0.99$	0.50	0.9010	0.9990	1.0000	1.0000	1.0000
	0.95	0.9040	0.9945	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	0.9930	0.9985	1.0000	1.0000	1.0000
	0.95	0.9925	0.9995	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.9910	0.9980	1.0000	1.0000	1.0000
	0.95	0.9845	0.9990	0.9990	1.0000	1.0000
$\rho = 0.80$	0.50	0.9380	0.9930	1.0000	1.0000	1.0000
	0.95	0.9415	0.9955	0.9995	1.0000	1.0000
$\rho = 0.50$	0.50	0.7665	0.9630	0.9945	1.0000	1.0000
	0.95	0.7235	0.9440	0.9870	1.0000	1.0000
$\rho = 0.10$	0.50	0.7005	0.9360	0.9950	1.0000	1.0000
	0.95	0.6860	0.9300	0.9925	1.0000	1.0000
$\rho = 0$	0.50	0.6375	0.9220	0.9925	1.0000	1.0000
	0.95	0.6345	0.9110	0.9920	1.0000	1.0000

Table 3.10: KPSS pre-test: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.6390	0.8515	0.9335	0.9945	1.0000
	0.95	0.5415	0.7770	0.9035	0.9960	0.9995
$\rho = 0.99$	0.50	0.5850	0.8125	0.9270	0.9970	1.0000
	0.95	0.5350	0.7685	0.8950	0.9955	1.0000
$\rho = 0.95$	0.50	0.5740	0.7705	0.8760	0.9870	0.9995
	0.95	0.6225	0.8110	0.9070	0.9955	1.0000
$\rho = 0.90$	0.50	0.5590	0.7545	0.8885	0.9850	1.0000
	0.95	0.5435	0.7785	0.8890	0.9930	1.0000
$\rho = 0.80$	0.50	0.4150	0.6575	0.8210	0.9810	0.9985
	0.95	0.4330	0.6480	0.8130	0.9825	1.0000
$\rho = 0.50$	0.50	0.2470	0.4555	0.6565	0.9535	0.9975
	0.95	0.3335	0.5420	0.7135	0.9735	0.9990
$\rho = 0.10$	0.50	0.2610	0.4370	0.6090	0.9465	0.9975
	0.95	0.2395	0.3950	0.5890	0.9535	0.9990
$\rho = 0$	0.50	0.2620	0.4135	0.6195	0.9535	0.9970
	0.95	0.2575	0.4195	0.6195	0.9595	0.9995
			$T = 200$			
$\rho = 1$	0.50	0.7630	0.9165	0.9775	1.0000	1.0000
	0.95	0.7710	0.9250	0.9860	0.9995	1.0000
$\rho = 0.99$	0.50	0.7265	0.9050	0.9695	1.0000	1.0000
	0.95	0.7650	0.9275	0.9790	0.9995	1.0000
$\rho = 0.95$	0.50	0.7565	0.8810	0.9570	1.0000	1.0000
	0.95	0.7740	0.9000	0.9675	1.0000	1.0000
$\rho = 0.90$	0.50	0.7660	0.9015	0.9615	0.9995	1.0000
	0.95	0.7560	0.9065	0.9655	0.9990	1.0000
$\rho = 0.80$	0.50	0.6445	0.8745	0.9570	0.9985	1.0000
	0.95	0.6575	0.8850	0.9530	0.9985	1.0000
$\rho = 0.50$	0.50	0.4820	0.7625	0.9120	1.0000	1.0000
	0.95	0.4590	0.7590	0.9125	0.9985	1.0000
$\rho = 0.10$	0.50	0.3885	0.6430	0.8690	0.9985	1.0000
	0.95	0.3990	0.6905	0.8650	0.9990	1.0000
$\rho = 0$	0.50	0.4030	0.6805	0.8835	0.9995	1.0000
	0.95	0.4145	0.6840	0.8720	0.9985	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.9475	0.9975	0.9995	1.0000	1.0000
	0.95	0.9485	0.9965	0.9995	1.0000	1.0000
$\rho = 0.99$	0.50	0.8710	0.9960	1.0000	1.0000	1.0000
	0.95	0.8865	0.9905	0.9985	1.0000	1.0000
$\rho = 0.95$	0.50	0.8755	0.9785	0.9990	1.0000	1.0000
	0.95	0.9105	0.9905	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.9260	0.9790	0.9980	1.0000	1.0000
	0.95	0.9300	0.9840	0.9990	1.0000	1.0000
$\rho = 0.80$	0.50	0.9190	0.9850	0.9990	1.0000	1.0000
	0.95	0.9160	0.9875	0.9995	1.0000	1.0000
$\rho = 0.50$	0.50	0.8030	0.9835	0.9985	1.0000	1.0000
	0.95	0.7850	0.9720	0.9965	1.0000	1.0000
$\rho = 0.10$	0.50	0.7380	0.9535	0.9980	1.0000	1.0000
	0.95	0.7545	0.9600	0.9970	1.0000	1.0000
$\rho = 0$	0.50	0.6980	0.9500	0.9970	1.0000	1.0000
	0.95	0.7010	0.9480	0.9980	1.0000	1.0000

For comparison, we also provide simulation results for the power of t_1 against equation (3.11) (Table 3.7). Comparing Table 3.7 with Table 3.6, it can be found that t_1 has a much weaker power than t in stationary cases ($\rho = 0.9$, 0.8 or 0.5). For instance, when $\rho = 0.9$, $\sigma_{12} = 0.50$, $\beta = 0.1$ and $T = 200$, the power for the covariance-based test t_1 is only about 29%, while the weighted average test t has a power of almost 70%. Moreover, the difference does not disappear when the sample size becomes larger. Even for a sample size of $T = 500$, the power for t_1 is only about 50% when $\rho = 0.8$, $\sigma_{12} = 0.50$ and $\beta = 0.1$, while the weighted average test t has a very strong power of more than 90%. These results are due to the fact that t diverges to infinity faster than t_1. Figures 3.1, 3.2 and 3.3 which present finite sample power curves for the tests also show our test ("——") dominates Maynard and Shimotsu (2009)'s test ("$\cdot\cdot\cdot$") in stationary cases.

Table 3.8, 3.9 and 3.10 present finite sample power for t_0, $t - DFGLS$ and $t - KPSS$. As expected, t_0 has the strongest power in stationary cases. But the difference between the power of t_0 and our weighted average test t is very small for a large sample size. As shown in Figures 3.1–3.3, the pretest-based tests $t - DFGLS$ ("$- \cdot -$") and $t - KPSS$ ("$--$") have similar power to t when both x_t and y_t behave in a stationary manner.

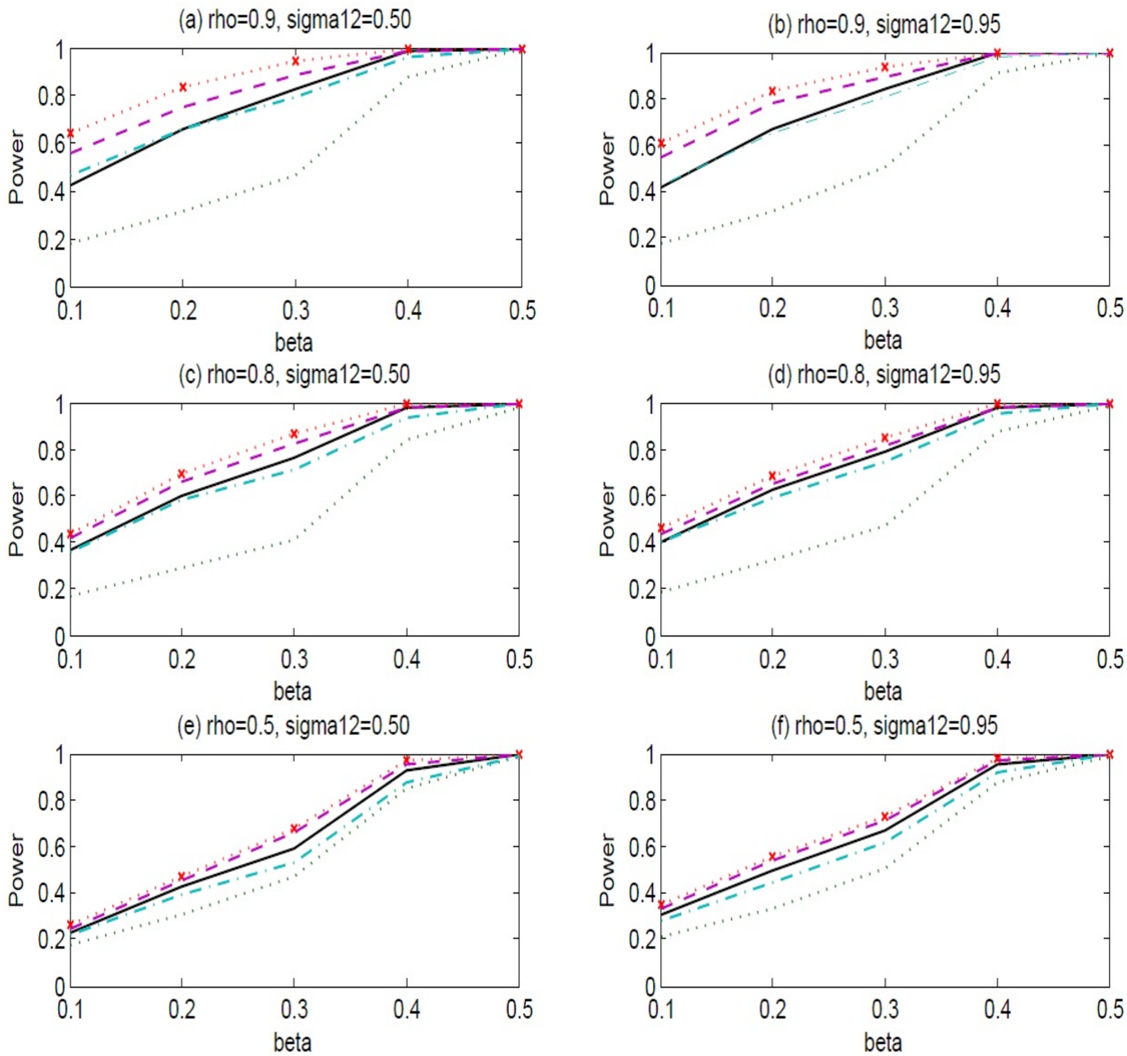


Figure 3.1: Finite sample power against equation (3.11), T = 100

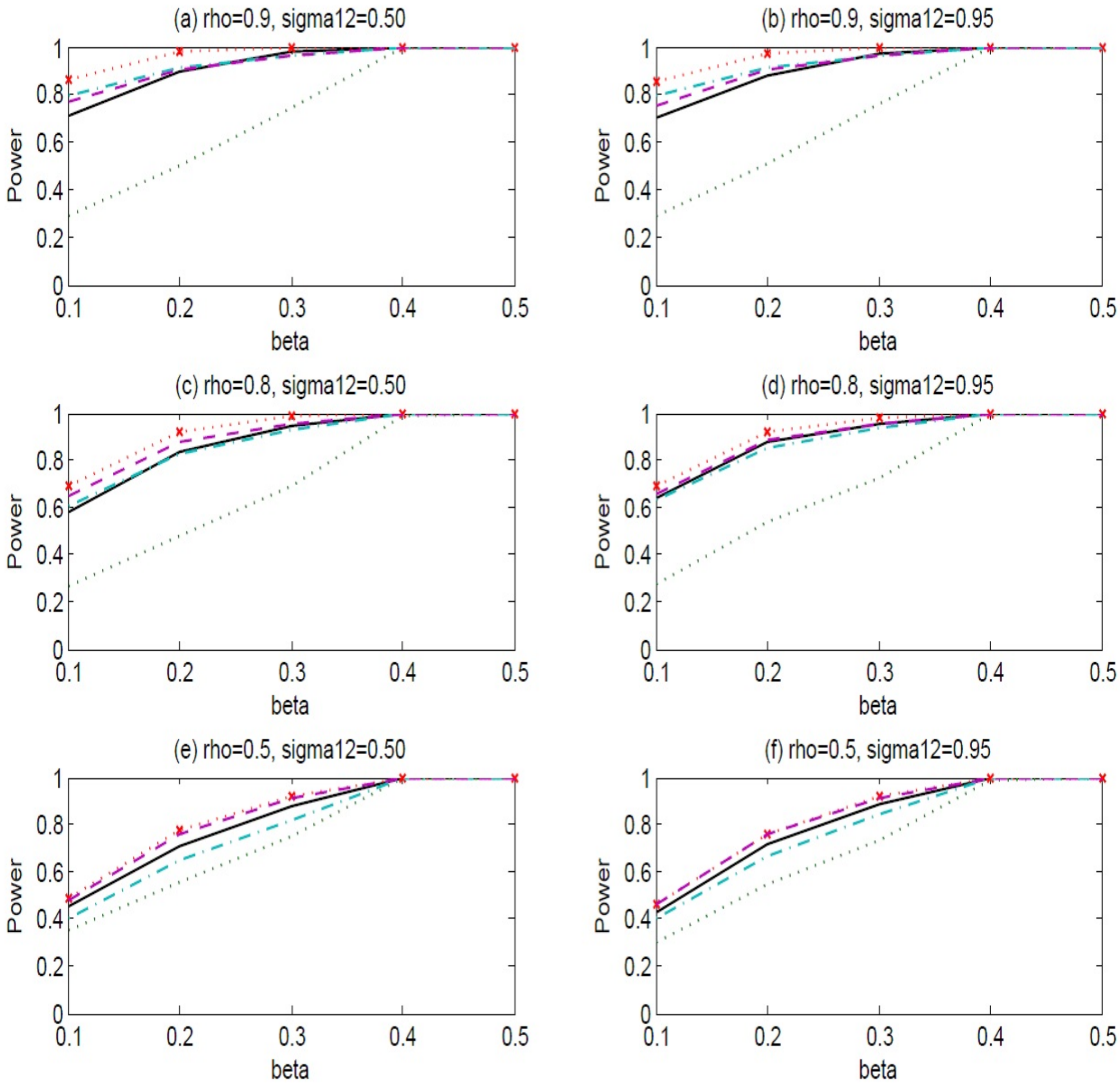


Figure 3.2: Finite sample power against equation (3.11), T = 200

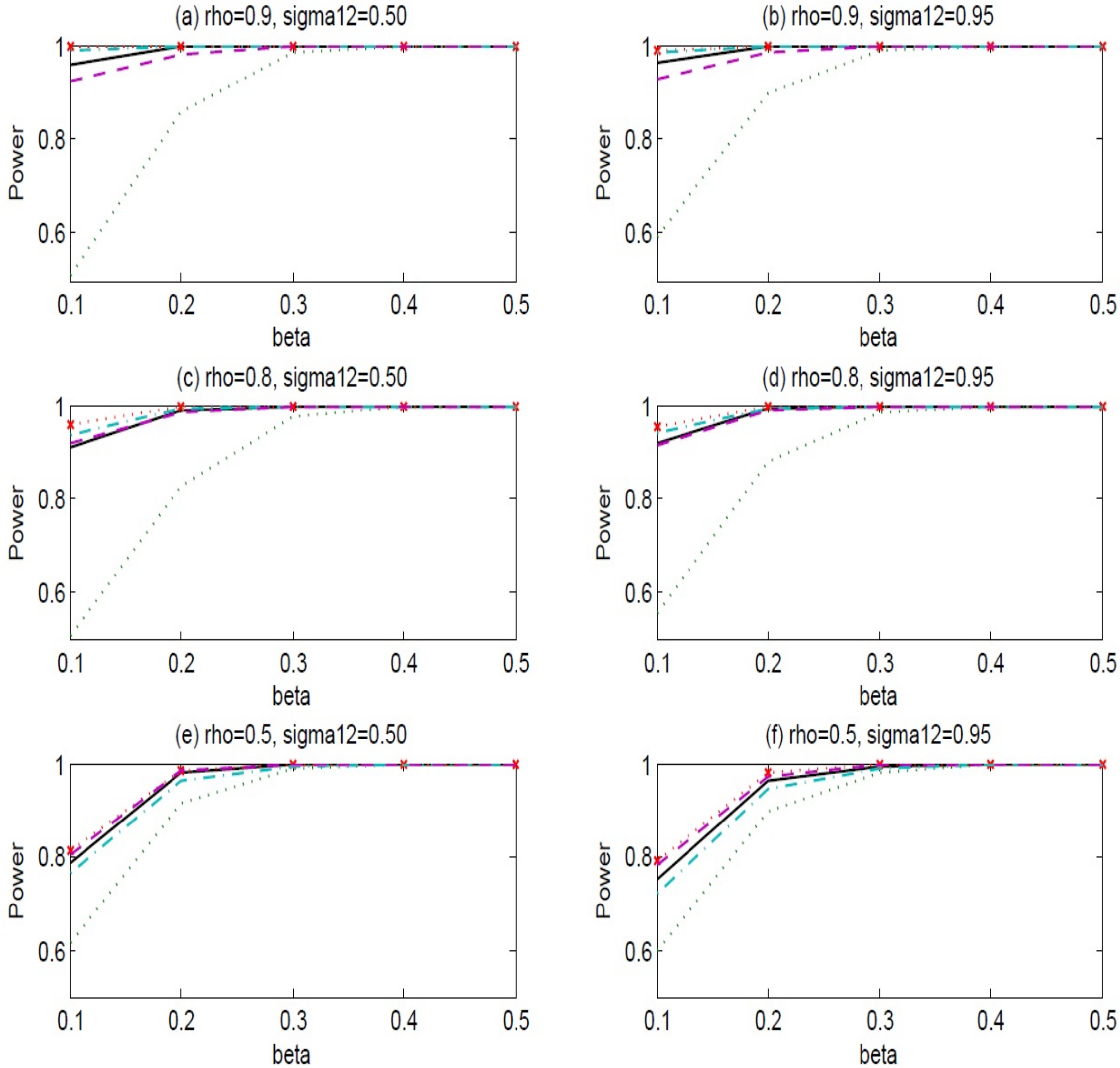


Figure 3.3: Finite sample power against equation (3.11), T = 500

We next look at the power of the weighted average test t and the covariance-based test t_1 against equation (3.12). This alternative allows that when x_t is $I(1)$ or local-to-unity, y_t can still keep stationary. Therefore, we should focus more on those results when ρ is equal or close to one, for example, $\rho = 1$, 0.99 and 0.95. Table 3.11 and 3.12 give the test results. As expected, t and t_1 have very similar power when x_t is $I(1)$ or local-to-unity since the weight for t_1, $\lambda(U, S)$, is almost one under nonstationary cases. This can be seen more clearly from Figures 3.4–3.6 in which the lines of t and t_1 are very close.

On the other hand, the regression-based test t_0 shows very weak power when x_t becomes nonstationary and the alternative is unbalanced (Table 3.13). Furthermore, the power does not improve as the sample size goes larger. For instance, the power is only about 5% for $\rho = 1$, $\sigma_{12} = 0.50$, $\gamma = 0.1$ and a large sample size of 500. These simulation results corroborate that t_0 is only $O_p(1)$ and does not diverge.

The pretest-based tests have generally much higher power than the regression-based test t_0 and the power improves in both sample size and γ in nonstationary and unbalanced cases (Table 3.14 and Table 3.15). However, from Figures 3.4–3.6, it is easy to see that the power of $t - KPSS$ ("$--$") is always much weaker than the power of t, although it dominates the power of t_0 ("$\cdot \times \cdot$"). On the other hand, the power of $t - DFGLS$ ("$- \cdot -$") is very close to the power of t when $\rho = 1$ and 0.99 (for $T = 500$, only when $\rho = 1$), and when $\rho = 0.95$, the differences are obvious. This is easy to understand. Since when $\rho = 1$ or 0.99, x_t is nonstationary for sure and the $DFGLS$ pretest can easily make right choices ($t - DFGLS = t_1$), and when $\rho = 0.95$, it is somewhat harder for the $DFGLS$ pretest to choose the right t statistics and the power is weaker.

Table 3.11: The t-statistic: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.2280	0.3435	0.5185	0.8535	0.9385
	0.95	0.1900	0.3400	0.4410	0.7475	0.8890
$\rho = 0.99$	0.50	0.2150	0.3490	0.5195	0.8315	0.9300
	0.95	0.2050	0.3200	0.4540	0.7410	0.8755
$\rho = 0.95$	0.50	0.2600	0.3880	0.5210	0.8265	0.9150
	0.95	0.2260	0.3225	0.4345	0.7210	0.8500
$\rho = 0.90$	0.50	0.2125	0.3175	0.4560	0.8005	0.9255
	0.95	0.2000	0.3130	0.4355	0.7250	0.8950
$\rho = 0.80$	0.50	0.2160	0.3335	0.4545	0.8215	0.9640
	0.95	0.1980	0.3460	0.4795	0.8395	0.9755
$\rho = 0.50$	0.50	0.2215	0.3445	0.5365	0.9090	0.9940
	0.95	0.1950	0.3660	0.5080	0.9200	0.9970
$\rho = 0.10$	0.50	0.2230	0.3595	0.5410	0.9235	0.9970
	0.95	0.2230	0.3880	0.5920	0.9470	0.9990
$\rho = 0$	0.50	0.2500	0.4365	0.6065	0.9400	0.9985
	0.95	0.2025	0.3560	0.5250	0.9160	0.9960
			$T = 200$			
$\rho = 1$	0.50	0.3195	0.5470	0.7460	0.9620	0.9840
	0.95	0.3410	0.5230	0.7055	0.9380	0.9730
$\rho = 0.99$	0.50	0.3430	0.5745	0.7470	0.9525	0.9730
	0.95	0.3040	0.4865	0.6765	0.8915	0.9505
$\rho = 0.95$	0.50	0.3095	0.4920	0.6745	0.8830	0.9575
	0.95	0.2825	0.4345	0.5830	0.8120	0.8970
$\rho = 0.90$	0.50	0.2800	0.4570	0.6145	0.8875	0.9750
	0.95	0.2635	0.3960	0.5680	0.8625	0.9765
$\rho = 0.80$	0.50	0.2480	0.4500	0.6035	0.9465	0.9985
	0.95	0.2715	0.4940	0.6670	0.9745	1.0000
$\rho = 0.50$	0.50	0.3910	0.6345	0.8210	0.9965	1.0000
	0.95	0.3555	0.6045	0.8090	0.9975	1.0000
$\rho = 0.10$	0.50	0.3845	0.6255	0.8595	0.9990	1.0000
	0.95	0.3655	0.6485	0.8390	0.9980	1.0000
$\rho = 0$	0.50	0.3690	0.6280	0.8395	1.0000	1.0000
	0.95	0.3950	0.6570	0.8490	0.9995	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.6175	0.8830	0.9690	0.9980	0.9965
	0.95	0.6060	0.8495	0.9515	0.9900	0.9915
$\rho = 0.99$	0.50	0.6025	0.8700	0.9425	0.9825	0.9895
	0.95	0.5240	0.7850	0.8875	0.9645	0.9785
$\rho = 0.95$	0.50	0.5030	0.6790	0.8070	0.9545	0.9950
	0.95	0.4690	0.6455	0.7600	0.9425	0.9900
$\rho = 0.90$	0.50	0.4115	0.6520	0.8035	0.9820	0.9985
	0.95	0.3935	0.6270	0.7840	0.9865	1.0000
$\rho = 0.80$	0.50	0.4310	0.7165	0.9105	0.9995	1.0000
	0.95	0.4555	0.7505	0.9250	0.9990	1.0000
$\rho = 0.50$	0.50	0.6335	0.9200	0.9850	1.0000	1.0000
	0.95	0.6020	0.9025	0.9820	1.0000	1.0000
$\rho = 0.10$	0.50	0.7110	0.9465	0.9970	1.0000	1.0000
	0.95	0.7215	0.9500	0.9960	1.0000	1.0000
$\rho = 0$	0.50	0.6715	0.9430	0.9945	1.0000	1.0000
	0.95	0.6730	0.9330	0.9950	1.0000	1.0000

Table 3.12: Maynard and Shimotsu (2009) t-statistic: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.2440	0.3590	0.5535	0.9205	0.9925
	0.95	0.1985	0.3525	0.4830	0.8380	0.9680
$\rho = 0.99$	0.50	0.2080	0.3640	0.5490	0.9000	0.9930
	0.95	0.2015	0.3265	0.4750	0.8235	0.9705
$\rho = 0.95$	0.50	0.2825	0.4310	0.5935	0.9380	0.9950
	0.95	0.2555	0.3775	0.5165	0.8765	0.9850
$\rho = 0.90$	0.50	0.2335	0.3730	0.5290	0.9040	0.9925
	0.95	0.2215	0.4015	0.5260	0.8655	0.9845
$\rho = 0.80$	0.50	0.2110	0.3470	0.5145	0.8750	0.9945
	0.95	0.2280	0.3685	0.5210	0.8795	0.9855
$\rho = 0.50$	0.50	0.2200	0.3535	0.5235	0.8875	0.9910
	0.95	0.2135	0.3695	0.5000	0.8905	0.9935
$\rho = 0.10$	0.50	0.1855	0.3145	0.4765	0.8850	0.9965
	0.95	0.1930	0.3285	0.5255	0.9140	0.9990
$\rho = 0$	0.50	0.2220	0.3830	0.5495	0.9165	0.9970
	0.95	0.1990	0.3400	0.5125	0.9195	0.9970
			$T = 200$			
$\rho = 1$	0.50	0.3330	0.5675	0.7780	0.9915	0.9995
	0.95	0.3470	0.5330	0.7440	0.9780	1.0000
$\rho = 0.99$	0.50	0.3640	0.6195	0.8090	0.9955	1.0000
	0.95	0.3265	0.5485	0.7450	0.9690	0.9985
$\rho = 0.95$	0.50	0.3650	0.5995	0.8160	0.9900	1.0000
	0.95	0.2925	0.5120	0.6960	0.9695	0.9990
$\rho = 0.90$	0.50	0.3620	0.5950	0.7970	0.9915	1.0000
	0.95	0.3305	0.5175	0.7065	0.9740	0.9995
$\rho = 0.80$	0.50	0.3520	0.6110	0.7685	0.9895	1.0000
	0.95	0.3345	0.5685	0.7480	0.9875	1.0000
$\rho = 0.50$	0.50	0.3730	0.5855	0.7800	0.9950	1.0000
	0.95	0.3180	0.5525	0.7560	0.9925	1.0000
$\rho = 0.10$	0.50	0.3440	0.5865	0.8220	0.9980	1.0000
	0.95	0.3360	0.6050	0.8090	0.9985	1.0000
$\rho = 0$	0.50	0.3260	0.5955	0.8155	0.9990	1.0000
	0.95	0.3425	0.5905	0.8040	1.0000	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.6445	0.9030	0.9845	1.0000	1.0000
	0.95	0.6060	0.8610	0.9640	1.0000	1.0000
$\rho = 0.99$	0.50	0.6500	0.9205	0.9785	1.0000	1.0000
	0.95	0.5660	0.8460	0.9405	1.0000	1.0000
$\rho = 0.95$	0.50	0.6750	0.9080	0.9875	1.0000	1.0000
	0.95	0.6125	0.8610	0.9720	1.0000	1.0000
$\rho = 0.90$	0.50	0.6130	0.8870	0.9760	1.0000	1.0000
	0.95	0.6035	0.8485	0.9575	1.0000	1.0000
$\rho = 0.80$	0.50	0.5540	0.8420	0.9705	1.0000	1.0000
	0.95	0.5630	0.8385	0.9640	1.0000	1.0000
$\rho = 0.50$	0.50	0.5865	0.8975	0.9810	1.0000	1.0000
	0.95	0.5720	0.8730	0.9745	1.0000	1.0000
$\rho = 0.10$	0.50	0.6780	0.9320	0.9940	1.0000	1.0000
	0.95	0.6910	0.9425	0.9940	1.0000	1.0000
$\rho = 0$	0.50	0.6480	0.9340	0.9945	1.0000	1.0000
	0.95	0.6455	0.9205	0.9945	1.0000	1.0000

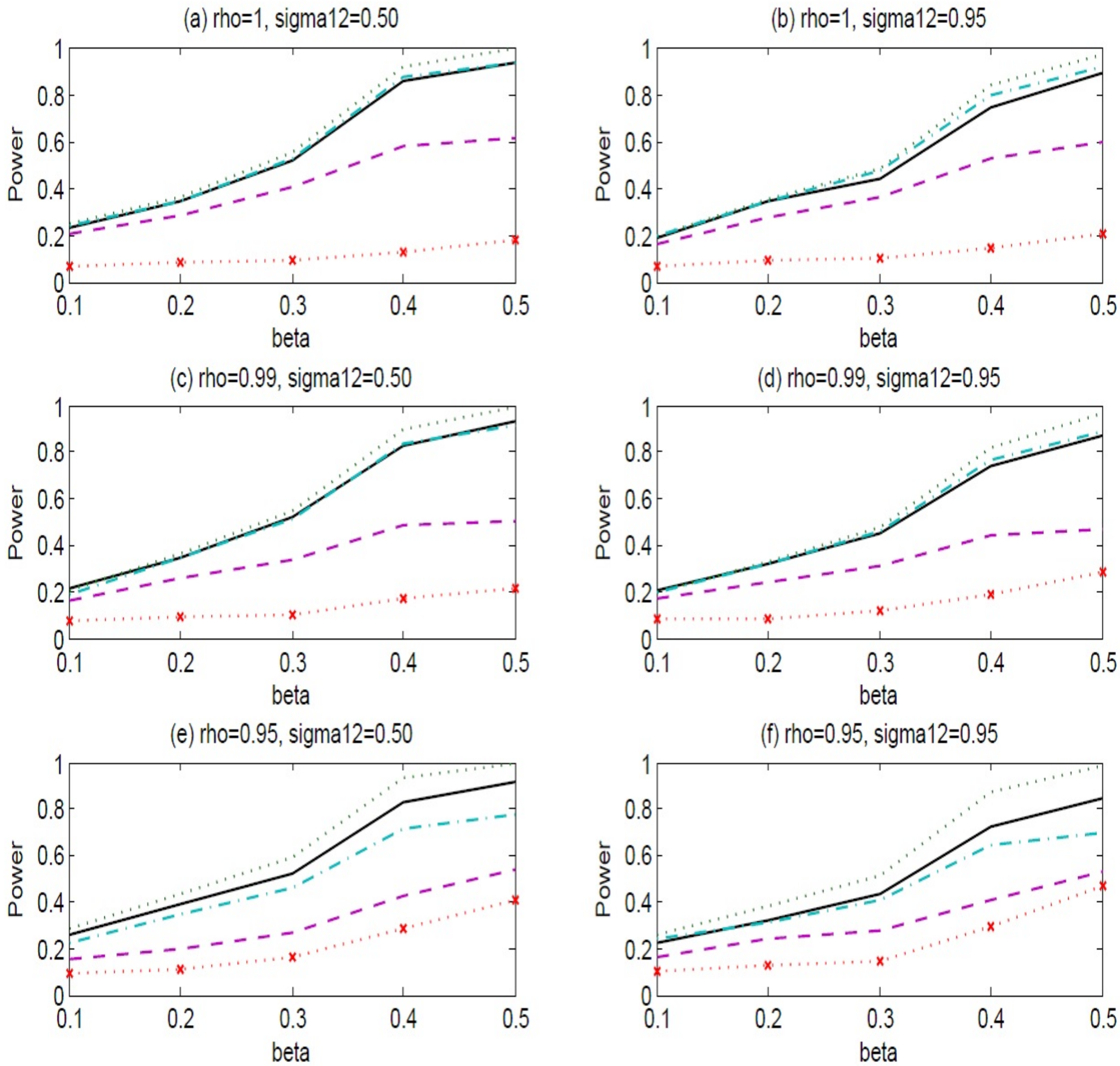


Figure 3.4: Finite sample power against equation (3.12), T = 100

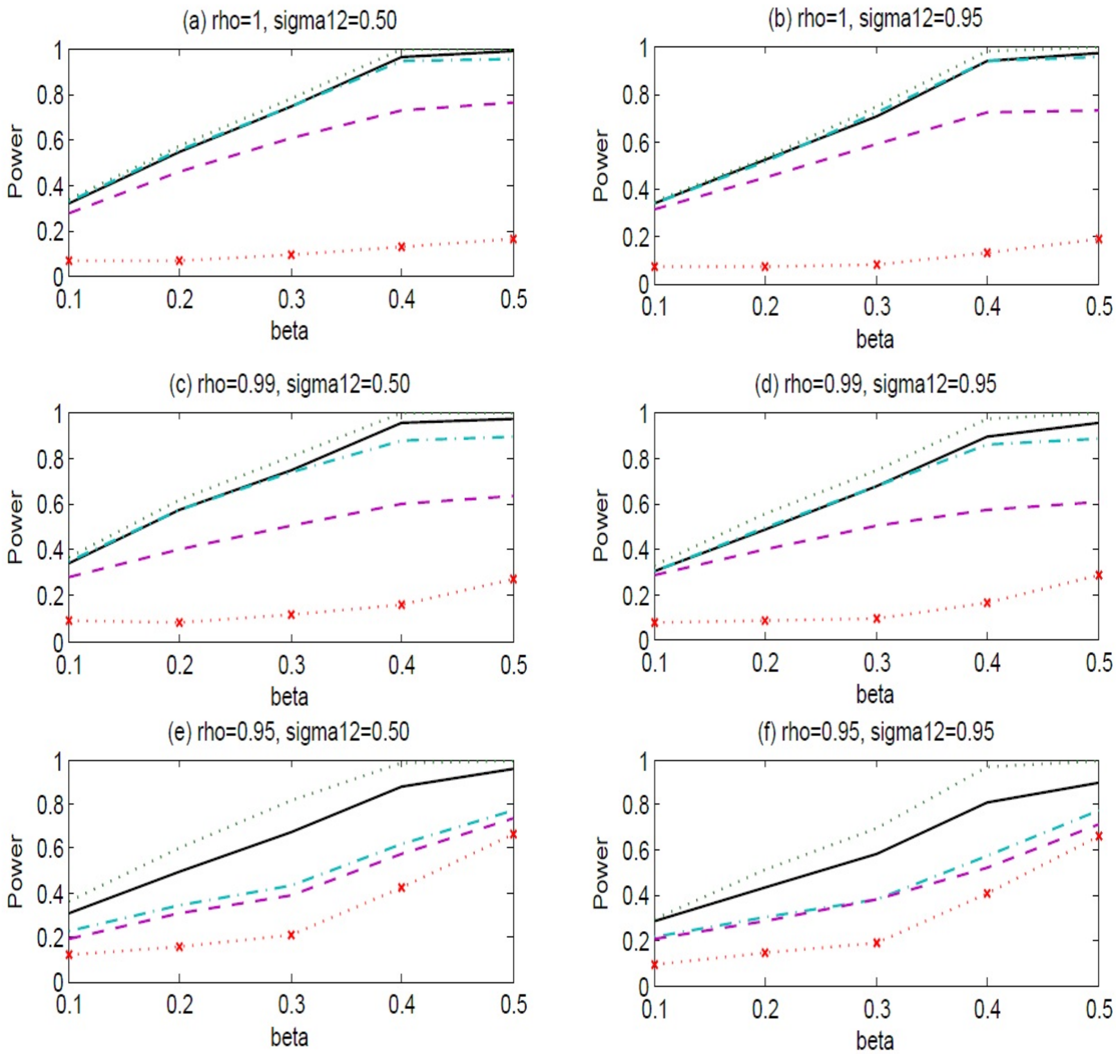


Figure 3.5: Finite sample power against equation (3.12), T = 200

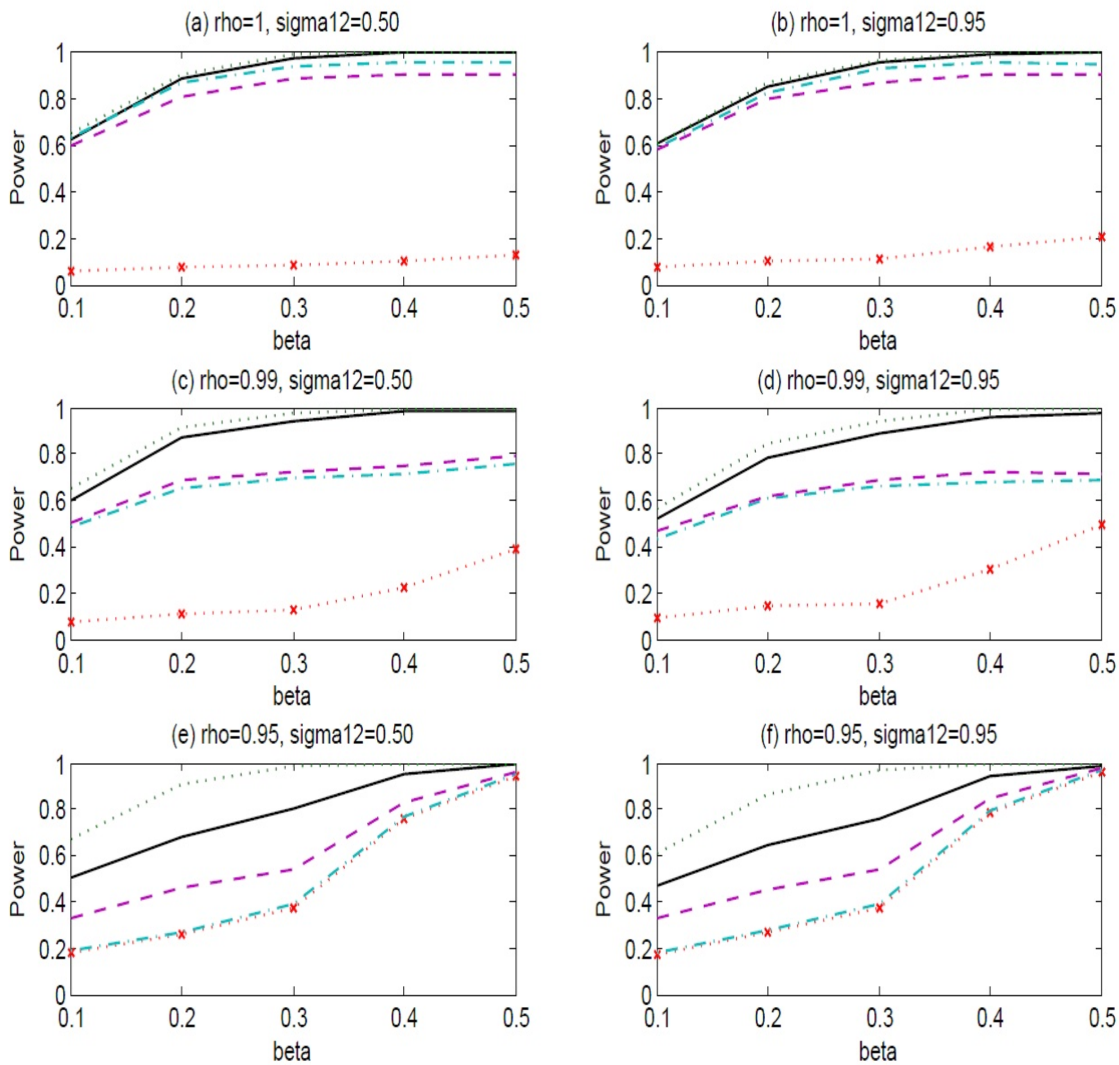


Figure 3.6: Finite sample power against equation (3.12), T = 500

Table 3.13: Regression t-statistic: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.0655	0.0790	0.0955	0.1255	0.1815
	0.95	0.0645	0.0920	0.0970	0.1455	0.2045
$\rho = 0.99$	0.50	0.0710	0.0895	0.1030	0.1670	0.2190
	0.95	0.0840	0.0850	0.1160	0.1905	0.2810
$\rho = 0.95$	0.50	0.0915	0.1110	0.1585	0.2815	0.4090
	0.95	0.1015	0.1265	0.1475	0.2960	0.4735
$\rho = 0.90$	0.50	0.0935	0.1325	0.1790	0.3885	0.6225
	0.95	0.1165	0.1590	0.2275	0.4465	0.7240
$\rho = 0.80$	0.50	0.1475	0.2340	0.3035	0.6695	0.8920
	0.95	0.1540	0.2620	0.3490	0.7215	0.9430
$\rho = 0.50$	0.50	0.2310	0.3625	0.5530	0.9135	0.9930
	0.95	0.1960	0.3640	0.5075	0.9210	0.9980
$\rho = 0.10$	0.50	0.2490	0.4090	0.5940	0.9480	1.0000
	0.95	0.2420	0.4210	0.6145	0.9645	0.9995
$\rho = 0$	0.50	0.2415	0.4190	0.6135	0.9515	0.9985
	0.95	0.2205	0.3935	0.5895	0.9460	1.0000
			$T = 200$			
$\rho = 1$	0.50	0.0670	0.0680	0.0905	0.1295	0.1600
	0.95	0.0720	0.0725	0.0830	0.1270	0.1900
$\rho = 0.99$	0.50	0.0870	0.0830	0.1150	0.1630	0.2720
	0.95	0.0765	0.0810	0.0890	0.1625	0.2795
$\rho = 0.95$	0.50	0.1185	0.1590	0.2080	0.4230	0.6695
	0.95	0.0920	0.1450	0.1885	0.4055	0.6645
$\rho = 0.90$	0.50	0.1500	0.2210	0.3240	0.6470	0.8995
	0.95	0.1670	0.2190	0.3315	0.6875	0.9325
$\rho = 0.80$	0.50	0.1825	0.3330	0.4735	0.8805	0.9930
	0.95	0.2035	0.3590	0.5270	0.9350	0.9985
$\rho = 0.50$	0.50	0.3550	0.6010	0.8065	0.9950	1.0000
	0.95	0.3305	0.5750	0.7935	0.9960	1.0000
$\rho = 0.10$	0.50	0.4085	0.6650	0.8825	0.9995	1.0000
	0.95	0.3745	0.6780	0.8640	0.9995	1.0000
$\rho = 0$	0.50	0.3985	0.6760	0.8770	1.0000	1.0000
	0.95	0.4265	0.7020	0.8845	1.0000	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.0530	0.0745	0.0860	0.0995	0.1245
	0.95	0.0755	0.1015	0.1125	0.1580	0.2075
$\rho = 0.99$	0.50	0.0760	0.1130	0.1235	0.2235	0.3935
	0.95	0.0925	0.1425	0.1570	0.3055	0.4980
$\rho = 0.95$	0.50	0.1765	0.2565	0.3765	0.7595	0.9460
	0.95	0.1725	0.2660	0.3730	0.7890	0.9640
$\rho = 0.90$	0.50	0.2505	0.4180	0.6110	0.9505	0.9975
	0.95	0.2550	0.4285	0.6015	0.9605	0.9995
$\rho = 0.80$	0.50	0.3575	0.6225	0.8380	0.9995	1.0000
	0.95	0.3860	0.6805	0.8815	0.9990	1.0000
$\rho = 0.50$	0.50	0.6100	0.9050	0.9815	1.0000	1.0000
	0.95	0.6185	0.9135	0.9820	1.0000	1.0000
$\rho = 0.10$	0.50	0.7370	0.9555	0.9985	1.0000	1.0000
	0.95	0.7595	0.9640	0.9975	1.0000	1.0000
$\rho = 0$	0.50	0.6930	0.9500	0.9970	1.0000	1.0000
	0.95	0.7035	0.9490	0.9975	1.0000	1.0000

Table 3.14: DFGLS pre-test: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.2370	0.3435	0.5275	0.8740	0.9340
	0.95	0.1945	0.3445	0.4720	0.8000	0.9175
$\rho = 0.99$	0.50	0.1905	0.3445	0.5160	0.8400	0.9155
	0.95	0.1980	0.3175	0.4620	0.7635	0.8910
$\rho = 0.95$	0.50	0.2255	0.3455	0.4565	0.7120	0.7760
	0.95	0.2375	0.3150	0.4060	0.6445	0.6975
$\rho = 0.90$	0.50	0.1700	0.2715	0.3505	0.6435	0.8070
	0.95	0.1575	0.2205	0.3070	0.5580	0.7775
$\rho = 0.80$	0.50	0.1595	0.2680	0.3585	0.7300	0.9290
	0.95	0.1785	0.3130	0.4110	0.8025	0.9685
$\rho = 0.50$	0.50	0.2215	0.3415	0.5185	0.8910	0.9920
	0.95	0.1845	0.3395	0.4605	0.8820	0.9935
$\rho = 0.10$	0.50	0.2110	0.3365	0.4990	0.8995	0.9955
	0.95	0.2095	0.3540	0.5335	0.9240	0.9970
$\rho = 0$	0.50	0.2310	0.3945	0.5570	0.9165	0.9965
	0.95	0.1805	0.3180	0.4850	0.8860	0.9940
			$T = 200$			
$\rho = 1$	0.50	0.3290	0.5505	0.7470	0.9445	0.9520
	0.95	0.3390	0.5150	0.7170	0.9340	0.9515
$\rho = 0.99$	0.50	0.3480	0.5770	0.7390	0.8710	0.8960
	0.95	0.2990	0.4950	0.6730	0.8560	0.8860
$\rho = 0.95$	0.50	0.2305	0.3455	0.4300	0.6170	0.7765
	0.95	0.2160	0.3015	0.3780	0.5715	0.7760
$\rho = 0.90$	0.50	0.1730	0.2725	0.3890	0.7075	0.9175
	0.95	0.1810	0.2515	0.3785	0.7265	0.9405
$\rho = 0.80$	0.50	0.2025	0.3665	0.5280	0.9015	0.9955
	0.95	0.2235	0.3975	0.5740	0.9450	0.9985
$\rho = 0.50$	0.50	0.3365	0.5740	0.7745	0.9935	1.0000
	0.95	0.3410	0.5760	0.7820	0.9940	1.0000
$\rho = 0.10$	0.50	0.3735	0.6100	0.8365	0.9985	1.0000
	0.95	0.3420	0.6070	0.8130	0.9965	1.0000
$\rho = 0$	0.50	0.3410	0.5855	0.8060	0.9985	1.0000
	0.95	0.3685	0.6120	0.8095	0.9990	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.6285	0.8675	0.9370	0.9500	0.9545
	0.95	0.5870	0.8250	0.9285	0.9520	0.9415
$\rho = 0.99$	0.50	0.4885	0.6525	0.6985	0.7135	0.7625
	0.95	0.4375	0.6130	0.6620	0.6800	0.6880
$\rho = 0.95$	0.50	0.1850	0.2705	0.3895	0.7650	0.9475
	0.95	0.1835	0.2805	0.3890	0.7960	0.9685
$\rho = 0.90$	0.50	0.2555	0.4315	0.6155	0.9525	0.9975
	0.95	0.2575	0.4375	0.6075	0.9625	0.9995
$\rho = 0.80$	0.50	0.3665	0.6250	0.8450	0.9995	1.0000
	0.95	0.3950	0.6845	0.8870	0.9990	1.0000
$\rho = 0.50$	0.50	0.6190	0.9105	0.9795	1.0000	1.0000
	0.95	0.5925	0.8990	0.9745	1.0000	1.0000
$\rho = 0.10$	0.50	0.7000	0.9380	0.9950	1.0000	1.0000
	0.95	0.6870	0.9340	0.9925	1.0000	1.0000
$\rho = 0$	0.50	0.6375	0.9220	0.9925	1.0000	1.0000
	0.95	0.6340	0.9115	0.9920	1.0000	1.0000

Table 3.15: KPSS pre-test: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.2035	0.2830	0.4060	0.5780	0.6095
	0.95	0.1615	0.2780	0.3620	0.5295	0.5985
$\rho = 0.99$	0.50	0.1640	0.2565	0.3390	0.4840	0.5080
	0.95	0.1720	0.2425	0.3150	0.4405	0.4650
$\rho = 0.95$	0.50	0.1520	0.2015	0.2645	0.4270	0.5385
	0.95	0.1645	0.2395	0.2750	0.4075	0.5305
$\rho = 0.90$	0.50	0.1280	0.1910	0.2660	0.5150	0.7225
	0.95	0.1370	0.1945	0.2790	0.5030	0.7475
$\rho = 0.80$	0.50	0.1510	0.2415	0.3200	0.6805	0.8975
	0.95	0.1730	0.2815	0.3810	0.7505	0.9505
$\rho = 0.50$	0.50	0.2280	0.3540	0.5475	0.9060	0.9920
	0.95	0.1875	0.3575	0.4955	0.9185	0.9980
$\rho = 0.10$	0.50	0.2495	0.4045	0.5900	0.9455	0.9995
	0.95	0.2395	0.4200	0.6080	0.9620	0.9990
$\rho = 0$	0.50	0.2415	0.4175	0.6130	0.9495	0.9980
	0.95	0.2150	0.3870	0.5780	0.9410	1.0000
			$T = 200$			
$\rho = 1$	0.50	0.2750	0.4590	0.6060	0.7235	0.7620
	0.95	0.3115	0.4480	0.5855	0.7215	0.7255
$\rho = 0.99$	0.50	0.2770	0.4040	0.5060	0.5955	0.6320
	0.95	0.2840	0.3940	0.5040	0.5720	0.6045
$\rho = 0.95$	0.50	0.1925	0.3100	0.3925	0.5765	0.7400
	0.95	0.2100	0.2855	0.3790	0.5215	0.7120
$\rho = 0.90$	0.50	0.1865	0.2800	0.4035	0.6940	0.9120
	0.95	0.1900	0.2670	0.3830	0.7230	0.9380
$\rho = 0.80$	0.50	0.1985	0.3645	0.4980	0.8950	0.9940
	0.95	0.2245	0.3865	0.5570	0.9405	0.9990
$\rho = 0.50$	0.50	0.3565	0.6020	0.8020	0.9950	1.0000
	0.95	0.3425	0.5840	0.7995	0.9955	1.0000
$\rho = 0.10$	0.50	0.4035	0.6570	0.8720	0.9995	1.0000
	0.95	0.3760	0.6695	0.8550	0.9990	1.0000
$\rho = 0$	0.50	0.4005	0.6645	0.8760	0.9995	1.0000
	0.95	0.4240	0.6975	0.8785	0.9985	1.0000
			$T = 500$			
$\rho = 1$	0.50	0.5950	0.8025	0.8800	0.8975	0.9000
	0.95	0.5805	0.7925	0.8660	0.8995	0.8965
$\rho = 0.99$	0.50	0.5045	0.6920	0.7200	0.7510	0.7970
	0.95	0.4665	0.6205	0.6925	0.7235	0.7185
$\rho = 0.95$	0.50	0.3330	0.4630	0.5400	0.8270	0.9595
	0.95	0.3250	0.4480	0.5380	0.8435	0.9760
$\rho = 0.90$	0.50	0.3005	0.4905	0.6725	0.9570	0.9980
	0.95	0.3135	0.4975	0.6645	0.9670	1.0000
$\rho = 0.80$	0.50	0.3735	0.6330	0.8470	0.9995	1.0000
	0.95	0.3985	0.6895	0.8950	0.9990	1.0000
$\rho = 0.50$	0.50	0.6200	0.9095	0.9825	1.0000	1.0000
	0.95	0.6155	0.9095	0.9830	1.0000	1.0000
$\rho = 0.10$	0.50	0.7340	0.9530	0.9980	1.0000	1.0000
	0.95	0.7505	0.9595	0.9970	1.0000	1.0000
$\rho = 0$	0.50	0.6985	0.9500	0.9970	1.0000	1.0000
	0.95	0.7015	0.9480	0.9980	1.0000	1.0000

Finally, we also compare our new test with the IVX test which bases on the the predictive regression and is expected to have good power against regression alternatives when x_t is mildly integrated, local-to-unity or $I(1)$ and y_t and x_{t-1} share the same order of integration. Using the same DGP as Tables 3.1, 3.6 and 3.11, Tables 3.16–3.18 present size and finite sample power for the IVX test. For the values of C_z and δ, we follow Magdalinos and Phillips (2009b) and Kostakis, Magdalinos and Stamatogiannis (2010)'s recommendations and choose $c_z = -1$, $\delta = 5/6$. Table 3.16 shows the rejection rates for the IVX tests are quite reliable when ρ are not that far from one. However, the size distortion problem still exists as ρ approaches to zero, which may confirm that the IVX method is invalid in stationary cases.

For the power part, although the IVX test exhibits very good power against balanced alternative (3.11) that both y_t and x_t are set to be (near) integrated, as we showed in Theorem 1, it has little power against the unbalanced alternative (3.12) which is more reasonable in practice. Table 3.18 confirms this point. Moreover, when ρ is close or equal to one, the power does not seem to improve as we move further into the alternative, nor as the sample size becomes larger. For example, even for a sample size of 500 and $\gamma = 0.5$, the power is only about 5.6% for $\rho = 1$ and $\sigma_{12} = 0.50$.

Table 3.16: The IVX method: Size

ρ	$\sigma_{12} = 0$	0.25	0.50	0.75	0.95
			$T = 100$		
1.000	0.0595	0.0520	0.0585	0.0570	0.0510
0.990	0.0685	0.0555	0.0610	0.0590	0.0655
0.950	0.0610	0.0685	0.0465	0.0695	0.0535
0.900	0.0590	0.0745	0.0655	0.0565	0.0665
0.800	0.0700	0.0700	0.0620	0.0515	0.0470
0.500	0.0760	0.0685	0.0655	0.0715	0.0595
0.100	0.1185	0.1270	0.1225	0.1090	0.1130
0.000	0.1455	0.1290	0.1250	0.1090	0.1220
			$T = 200$		
1.000	0.0575	0.0440	0.0590	0.0520	0.0570
0.990	0.0475	0.0575	0.0640	0.0565	0.0565
0.950	0.0695	0.0615	0.0650	0.0625	0.0545
0.900	0.0710	0.0555	0.0580	0.0690	0.0580
0.800	0.0610	0.0610	0.0450	0.0555	0.0575
0.500	0.0760	0.0735	0.0810	0.0640	0.0750
0.100	0.1470	0.1415	0.1485	0.1275	0.1395
0.000	0.1580	0.1520	0.1665	0.1405	0.1520
			$T = 500$		
1.000	0.0485	0.0540	0.0495	0.0495	0.0610
0.990	0.0485	0.0590	0.0510	0.0435	0.0530
0.950	0.0660	0.0660	0.0520	0.0650	0.0490
0.900	0.0740	0.0565	0.0635	0.0520	0.0540
0.800	0.0495	0.0510	0.0535	0.0565	0.0485
0.500	0.1080	0.1160	0.1035	0.0985	0.1105
0.100	0.1985	0.1915	0.2010	0.2055	0.1955
0.000	0.2145	0.2220	0.2235	0.2120	0.1945

Table 3.17: The IVX method: Power (against equation 3.11)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.9000	0.9775	0.9960	0.9995	1.0000
	0.95	0.8820	0.9615	0.9930	1.0000	1.0000
$\rho = 0.99$	0.50	0.8850	0.9735	0.9940	1.0000	1.0000
	0.95	0.8510	0.9505	0.9850	0.9995	1.0000
$\rho = 0.95$	0.50	0.7370	0.9265	0.9755	0.9990	1.0000
	0.95	0.7425	0.9085	0.9640	0.9995	1.0000
$\rho = 0.90$	0.50	0.6390	0.8570	0.9565	0.9990	1.0000
	0.95	0.5910	0.8030	0.9360	0.9990	1.0000
$\rho = 0.80$	0.50	0.4730	0.7025	0.8810	0.9965	1.0000
	0.95	0.4560	0.7105	0.8705	0.9920	1.0000
$\rho = 0.50$	0.50	0.3010	0.4910	0.6840	0.9715	0.9990
	0.95	0.3140	0.5055	0.6950	0.9590	0.9990
$\rho = 0.10$	0.50	0.2215	0.3740	0.5055	0.8850	0.9880
	0.95	0.2130	0.3620	0.5075	0.8475	0.9825
$\rho = 0$	0.50	0.1875	0.3175	0.4705	0.8340	0.9795
	0.95	0.2075	0.3155	0.4905	0.8325	0.9785
			$T = 200$			
$\rho = 1$	0.50	0.9935	0.9995	1.0000	1.0000	1.0000
	0.95	0.9875	0.9990	1.0000	1.0000	1.0000
$\rho = 0.99$	0.50	0.9910	0.9990	1.0000	1.0000	1.0000
	0.95	0.9805	0.9970	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	0.9440	0.9955	1.0000	1.0000	1.0000
	0.95	0.9220	0.9920	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.8380	0.9725	0.9990	1.0000	1.0000
	0.95	0.8505	0.9675	0.9955	1.0000	1.0000
$\rho = 0.80$	0.50	0.6880	0.9210	0.9890	1.0000	1.0000
	0.95	0.7110	0.9245	0.9825	1.0000	1.0000
$\rho = 0.50$	0.50	0.4625	0.7325	0.9000	0.9990	1.0000
	0.95	0.4595	0.7095	0.8775	0.9980	1.0000
$\rho = 0.10$	0.50	0.2980	0.4840	0.6665	0.9765	0.9995
	0.95	0.2820	0.4760	0.6525	0.9660	0.9995
$\rho = 0$	0.50	0.2540	0.4350	0.6405	0.9625	0.9990
	0.95	0.2720	0.4475	0.6350	0.9680	0.9970
			$T = 500$			
$\rho = 1$	0.50	1.0000	1.0000	1.0000	1.0000	1.0000
	0.95	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.99$	0.50	1.0000	1.0000	1.0000	1.0000	1.0000
	0.95	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.95$	0.50	1.0000	1.0000	1.0000	1.0000	1.0000
	0.95	0.9980	1.0000	1.0000	1.0000	1.0000
$\rho = 0.90$	0.50	0.9930	1.0000	1.0000	1.0000	1.0000
	0.95	0.9890	1.0000	1.0000	1.0000	1.0000
$\rho = 0.80$	0.50	0.9535	1.0000	1.0000	1.0000	1.0000
	0.95	0.9400	0.9985	1.0000	1.0000	1.0000
$\rho = 0.50$	0.50	0.7140	0.9335	0.9910	1.0000	1.0000
	0.95	0.7030	0.9205	0.9895	1.0000	1.0000
$\rho = 0.10$	0.50	0.3800	0.6635	0.8715	0.9995	1.0000
	0.95	0.4400	0.7150	0.8895	0.9985	1.0000
$\rho = 0$	0.50	0.4255	0.6665	0.8745	0.9995	1.0000
	0.95	0.4270	0.6895	0.8645	0.9970	1.0000

Table 3.18: The IVX method: Power (against equation 3.12)

ρ	σ_{12}	0.10	0.15	0.20	0.35	0.50
			$T = 100$			
$\rho = 1$	0.50	0.0580	0.0630	0.0620	0.0550	0.0720
	0.95	0.0680	0.0620	0.0650	0.0805	0.0815
$\rho = 0.99$	0.50	0.0520	0.0570	0.0585	0.0595	0.0790
	0.95	0.0725	0.0635	0.0690	0.0650	0.0830
$\rho = 0.95$	0.50	0.0705	0.0925	0.1025	0.1350	0.1625
	0.95	0.0560	0.0865	0.0885	0.1035	0.1380
$\rho = 0.90$	0.50	0.0870	0.1030	0.1325	0.2025	0.3125
	0.95	0.0595	0.0920	0.1020	0.1635	0.2340
$\rho = 0.80$	0.50	0.1200	0.1725	0.2430	0.5040	0.7745
	0.95	0.1280	0.1745	0.2430	0.5035	0.8575
$\rho = 0.50$	0.50	0.2455	0.4125	0.6140	0.9505	0.9995
	0.95	0.2495	0.4290	0.6520	0.9720	1.0000
$\rho = 0.10$	0.50	0.1870	0.3155	0.5275	0.8760	0.9890
	0.95	0.2265	0.3595	0.5325	0.8925	0.9935
$\rho = 0$	0.50	0.2100	0.3280	0.4650	0.8540	0.9815
	0.95	0.1925	0.3230	0.4500	0.8105	0.9825
			$T = 200$			
$\rho = 1$	0.50	0.0525	0.0710	0.0605	0.0605	0.0750
	0.95	0.0485	0.0490	0.0550	0.0465	0.0595
$\rho = 0.99$	0.50	0.0515	0.0575	0.0570	0.0625	0.0610
	0.95	0.0545	0.0615	0.0645	0.0635	0.0660
$\rho = 0.95$	0.50	0.0845	0.1050	0.1075	0.1645	0.2190
	0.95	0.0665	0.0765	0.0885	0.1115	0.1565
$\rho = 0.90$	0.50	0.1065	0.1365	0.1935	0.3490	0.5420
	0.95	0.0955	0.1235	0.1565	0.2900	0.4785
$\rho = 0.80$	0.50	0.1890	0.3230	0.4605	0.8365	0.9845
	0.95	0.1460	0.2410	0.3725	0.8495	0.9995
$\rho = 0.50$	0.50	0.3825	0.6735	0.8345	0.9995	1.0000
	0.95	0.4040	0.6790	0.8960	0.9995	1.0000
$\rho = 0.10$	0.50	0.3240	0.5285	0.7410	0.9825	0.9995
	0.95	0.3050	0.5105	0.7370	0.9835	0.9995
$\rho = 0$	0.50	0.2360	0.4225	0.6095	0.9565	0.9975
	0.95	0.2445	0.4170	0.6220	0.9540	0.9985
			$T = 500$			
$\rho = 1$	0.50	0.0600	0.0575	0.0505	0.0535	0.0555
	0.95	0.0570	0.0580	0.0595	0.0630	0.0640
$\rho = 0.99$	0.50	0.0505	0.0695	0.0705	0.0840	0.0845
	0.95	0.0565	0.0590	0.0545	0.0840	0.0840
$\rho = 0.95$	0.50	0.0880	0.1300	0.1550	0.2490	0.3560
	0.95	0.0885	0.1220	0.1455	0.2110	0.3100
$\rho = 0.90$	0.50	0.1590	0.2270	0.3615	0.7090	0.9275
	0.95	0.1375	0.2160	0.2925	0.6725	0.9565
$\rho = 0.80$	0.50	0.3745	0.6445	0.8710	0.9995	1.0000
	0.95	0.3615	0.6375	0.8770	1.0000	1.0000
$\rho = 0.50$	0.50	0.6550	0.9260	0.9950	1.0000	1.0000
	0.95	0.6685	0.9270	0.9945	1.0000	1.0000
$\rho = 0.10$	0.50	0.4560	0.7530	0.9215	0.9995	1.0000
	0.95	0.4420	0.7365	0.9050	1.0000	1.0000
$\rho = 0$	0.50	0.4375	0.6885	0.8970	0.9995	1.0000
	0.95	0.3920	0.6545	0.8620	0.9975	1.0000

3.5.4 Sensitivity to Parameter Choice

In the preceding simulations, we set the parameter c in the weight function $\lambda(U, S)$ to be 0.006. As shown by the large sample distribution theory in Section 3, the variations in the finite sample behaviour of our tests because of the precise choice of c should diminish as the sample size is increased. When we increase the sample size to 1000 (this is very close to the sample size used in the empirical application.), Table 3.19 confirms that the choice of the parameter c has little effect on the size of the test. All results using our test t are quite reliable in either the nonstationary case or the stationary case, and much better than t_0 or t_1.

3.6 Application

To test the market efficiency/constant risk premium hypothesis, we apply our test to monthly excess stock returns, employing a variety of commonly used predictors including dividend price ratios, earnings price ratios, book-to-market ratios, T-bill rates and other nine variables. Both single and multiple predictors cases are presented. We also compare our test with other tests– the standard regression-based test (t_0 or W_0), the covariance-based test (t_1 or W_1), two pretest-based tests and the IVX test (t_{IVX} or W_{IVX}). Different methods have different conclusions on regressors' predictive power. More specifically, for the dividend-price ratio, previous literature such as Viceira (1997), Torous, Valkanov, and Yan (2004) and Maynard and Shimotsu (2009) show only modest evidence of predictability. However, some other previous studies including Campbell and Yogo (2006) and Camponovo, Scaillet and Trojani (2013) find dividend-price ratio is a strong predictive variable for stock returns.

Table 3.19: Comparison of different c

						t						t_0	t_1
ρ	$c = 0.0005$	0.0010	0.0030	0.0060	0.0120	0.0360	0.0600	0.1200	0.3000	0.6000	1.0000		
0.999	0.0470	0.0470	0.0470	0.0465	0.0435	0.0430	0.0410	0.0430	0.0455	0.0460	0.0500	0.0940	0.0475
0.500	0.0355	0.0435	0.0480	0.0515	0.0545	0.0570	0.0585	0.0600	0.0610	0.0610	0.0610	0.0610	0.0230

Note: The table shows rejection rates under the null hypothesis for a norminal 5% test using t. Different choices of the parameter c in the weight function are considered. We also report rejecion rates for t_0 and t_1 for comparison. $T = 1000$, $\sigma_{12} = 0.50$.

3.6.1 Data

Excess Return: The excess return on the market is calculated as $R_m - R_f$, where R_m is the market return and R_f is the risk-free rate. In particular, R_m is proxied by the CRSP value-weighted return on all NYSE, AMEX, and NASDAQ stocks, and R_f is 1-month Treasury bill rate. This source is from Kenneth French's online data library. The sample preiod is from January 1927 to December 2012. We also report test results for two subsamples: 1927–1951 and 1952–2012[5].

The excess return is always our dependent variable. Our set of 13 independent variables comes from Amit Goyal's website[6], which is an updated version of the dataset used in Goyal and Welch (2008). Following Cenesizoglu and Timmermann (2008), we classify these predictors into four categories.

(1). Valuation ratios

Dividend Price Ratio (D/P): The dividend price ratio is the different between the log of dividends and the log of stock prices. It is calculated as $\ln((d_t + \cdots + d_{t-11})/p_t)$ where d_t are dividends paid on the S&P 500 index and p_t are prices.

Earnings Price Ratio (E/P): The earnings price ratio is the different between the log of earnings and the log of stock prices. It is calculated as $\ln((e_t + \cdots + e_{t-11})/p_t)$ where e_t are earnings on the S&P 500 index and p_t are prices.

Book-to-market Ratio (B/M): The book-to-market ratio is calculated by dividing the book value of the previous year by the current level of the DJIA and taking logarithm.

(2). Bond yield measures

Treasury Bill Rate (TBL): The three-month Treasury bill rate.

[5]We follow the previous literature in breaking the sample at 1951 (see Campbell, Lo, and MacKinlay, 1997).

[6]The data are available at http://www.hec.unil.ch/agoyal/.

Long-term Yield (LTY): The long-term government bond yield.

Term Spread (TMS): The difference between the long-term government bond yield and the three-month Treasury-bill rate.

Default Yield Spread (DFY): The DFY is the difference between BAA and AAA- rated corporate bond yields and taking logarithm.

Default Return Spread (DFR): The DFR is the difference between long-term corporate bond returns and long-term government bond returns.

(3). Estimates of equity risk

Long-term Return (LTR): The return on long-term government bonds.

Stock Variance (SVAR): The SVAR is the sum of squared daily returns on S&P 500 index.

Cross-sectional equity premium (CSP): The CSP measures the relative valuations of high- and low-beta stocks. The available data for this variable is from May 1937 to December 2002.

(4). Corporate fiance variables

Net Equity Expansion (NTIS): The ratio of one year moving sums of net issues by NYSE listed stocks to the total end-of-year market capitalization of NYSE stocks.

Dividend Payout Ratio (D/E): The difference between the log of dividends and the log of earnings.

3.6.2 Single-Predictor Model

Table 3.20 presents the test results using the lagged value of each of the 13 variables as the predictor, these results are calculated in the same way as in the simulations [7].

[7]For better comparison with the predictive model literature such as the IVX method, the demeaned version of the estimator was considered.

Table 3.20: Tests on monthly excess stock returns (single-predictor model)

Sample		Monthly excess Return												
		D/P	E/P	B/M	TBL	LTY	TMS	DFY	DFR	LTR	SVAR	CSP	NTIS	D/E
1927-2012[1]	t	-0.486	0.518	-2.172**	-1.695*	-2.459**	0.391	1.077	0.997	0.615	-0.773	12.914***	0.590	-1.118
	pvalue	0.627	0.604	0.030	0.090	0.014	0.696	0.281	0.319	0.539	0.439	0	0.555	0.264
	$\lambda(U,S)$	0.999	0.971	0.993	0.990	0.999	0.889	0.866	*0.000*	*0.797*	*0.765*	0.999	0.992	0.978
	t_1	-0.487	0.487	-2.197**	-1.694*	-2.461**	0.246	1.129	0.988	0.280	-0.956	12.916***	0.609	-1.149
	t_0	1.630	1.590	1.704*	-1.717*	-1.218	1.547	0.744	0.997	1.926*	-0.178	2.877***	-1.709*	0.240
	t_{DFGLS}	-0.487	1.590	1.704*	-1.717*	-2.461**	1.547	0.744	0.997	1.926*	-0.178	12.916***	0.609	0.240
	$DFGLS$	-0.900	-2.502**	-2.031**	-1.945**	-0.916	-2.569**	-2.969***	-6.129***	-6.301***	-3.928***	-0.409	-1.271	-3.578***
	t_{KPSS}	-0.487	0.487	-2.197**	-1.694*	-2.461**	0.246	1.129	0.997	0.280	-0.956	12.916***	0.609	-1.149
	$KPSS$	2.915***	1.135***	1.931***	1.528***	2.176***	0.579**	0.607**	0.146	1.023***	0.588**	2.617***	1.083***	1.838***
	t_{IVX}	2.154**	1.896*	2.377**	-1.186	-0.522	1.668*	1.310	1.930*	-2.248**	-1.561	0.686	-2.853***	0.531
1927-1951[2]	t	0.581	0.213	-0.882	1.205	0.573	-1.063	0.875	0.428	-0.755	0.744	3.898***	1.197	-1.378
	pvalue	0.561	0.831	0.378	0.228	0.567	0.288	0.382	0.669	0.450	0.457	0	0.231	0.168
	$\lambda(U,S)$	*0.299*	0.977	0.969	0.992	0.999	0.944	0.978	*0.137*	*0.753*	0.948	*0.009*	0.948	0.962
	t_1	-0.252	0.169	-0.964	1.218	0.573	-1.132	0.883	0.849	-0.812	0.767	4.860***	1.378	-1.407
	t_0	0.937	2.066**	1.648*	-0.426	-0.696	0.098	0.495	0.361	-0.580	0.322	3.890***	-2.100**	-0.625
	t_{DFGLS}	0.937	0.169	-0.964	1.218	0.573	-1.132	0.883	0.361	-0.580	0.767	3.890***	1.378	-0.625
	$DFGLS$	-2.431**	-1.900*	-1.083	-0.778	-0.539	-1.066	-1.544	-3.493***	-2.445**	-1.797*	-4.311***	-1.382	-2.082**
	t_{KPSS}	0.937	0.169	-0.964	1.218	0.573	0.098	0.883	0.361	-0.580	0.767	3.890***	1.378	-1.407
	$KPSS$	0.171	0.958***	0.472**	0.672**	1.615***	0.344	0.808***	0.192	0.356*	0.605**	0.154	0.463**	0.820***
	t_{IVX}	1.055	1.605	2.044**	-0.233	-0.703	-0.038	0.423	2.413**	-1.537	0.561	3.242***	-2.658***	-0.930
1952-2012[3]	t	-1.396	-0.156	-2.417**	-2.333**	-2.543**	0.820	-0.680	0.857	1.287	-3.003***	11.638***	-2.439**	-0.218
	pvalue	0.163	0.876	0.016	0.020	0.011	0.412	0.497	0.391	0.198	0.003	0	0.015	0.827
	$\lambda(U,S)$	0.999	0.989	0.999	0.967	0.996	0.949	0.936	*0.354*	*0.748*	0.896	0.999	0.984	*0.775*
	t_1	-1.397	-0.166	-2.421**	-2.352**	-2.550**	0.764	-0.781	0.449	0.840	-3.067***	11.642***	-2.473**	-0.526
	t_0	1.752*	0.673	0.757	-1.762*	-0.988	1.855*	0.794	1.081	2.614***	-2.451**	2.427**	-0.296	0.843
	t_{DFGLS}	-1.397	-0.166	-2.421**	-2.352**	-2.550**	1.855*	0.794	0.449	2.614***	-2.451**	11.642***	-2.473**	0.843
	$DFGLS$	-0.498	-1.589	-0.957	-1.672*	-0.778	-3.294***	-2.210**	-1.075	-5.200***	-3.830***	-0.326	-1.773*	-3.051***
	t_{KPSS}	-1.397	-0.166	-2.421**	-2.352**	-2.550**	0.764	-0.781	1.081	0.840	-3.067***	11.642***	-2.473**	-0.526
	$KPSS$	2.035***	1.164***	2.047***	0.711**	0.925***	1.110***	0.665**	0.082	0.748***	0.896***	1.219***	1.095***	0.468**
	t_{IVX}	1.968**	0.933	1.156	-1.516	-0.673	2.037**	0.916	-0.108	-1.472	-5.368***	0.287	0.039	1.370

1: The sample period for CSP is from 1937.5 to 2002.12.
2: The sample period for CSP is from 1937.5 to 1951.12.
3: The sample period for CSP is from 1952.1 to 2002.12.
*: The statistic is significant at 10% level. **: The statistic is significant at 5% level. ***: The statistic is significant at 1% level.

We first focus on the results using the whole sample period (the first panel of Table 3.20). Test statistics based on our method are presented in the first row. Our approach would point to the conclusion that the null hypothesis of no predictability can be rejected (at a level lower than 5%) when the lagged series of the book-to-market ratio, the long-term yield and the cross-sectional premium are alternatively used as predictors. In particular, the strongest evidence for predictability is documented for the cross-sectional premium (at a level much lower than 1%). There is some weak evidence in favour of the predictive ability of the T-bill rate (at a level around 10%), while there is no such evidence for other variables including the dividend price ratio and the earning price ratio to be employed as predictors of next month excess market returns. These results demonstrate that the overall evidence on short-term predictability is very weak. The seventh and ninth rows present the time series properties of the data applying two different tests: the DF-GLS test and the KPSS test. Based on the two tests, strong evidence of stationarity appear for the series of the default return spread. D/P, LTY, CSP and NTIS show strong nonstationary properties. For the rest of the data series, the conclusions are ambiguous. This confirms previous studies concerning the uncertainty about the time series properties of these predictors and shows the importance of the use of local to unity frameworks. Row six and row eight show results using two pretest-based methods. Because of the ambiguous conclusions about the time series properties of the data, t_{DFGLS} based on the DF-GLS tests and t_{kpss} based on the KPSS often choose different values. Moreover, the results show the selections of the pretests depend much on the choice of the significance level. For example, if 1% siginificance level was taken, the DF-GLS test should not reject the unit root hypothesis and select t_1 for B/M, changing the test result from positive 1.704 to negative 2.172, and the KPSS test will accept the stationarity hypothesis and select t_0 for TMS, increasing t_{kpss} from

0.246 to 1.547. All of these demonstrate the restrictions of pretest-based methods.

When using the whole sample period, the covariance-based test t_1 has the same conclusions with our test due to the large weights of t_1 when calculate t for most of these predictors. In the third row, eight of the thirteen predictors have $\lambda(U, S)$ that are more than 0.95, in accordance with the fact that most of these predictors are highly persistent. The default return spread is an exception. Because of its stationary property, the $\lambda(U, S)$ for DFR is almost zero and t is approximately equal to t_0. This is very important in practice considering the restrictions of the covariance-based tests in stationary cases, although in this particular case, the difference between t_1 and t_0 is not big.

Comparing our results with the ones applying the IVX test (the last column of the panel), the IVX test shows overall stronger evidence against the null hypothesis of no predictability than ours when the whole sample period is used. For example, the IVX test is able to reject the null hypothesis using D/P, B/M, LTR and NTIS respectively at the 5% level or lower. The ability of E/P, TMS or DFR to predict future stock returns is on the borderline of statistical significance. Since DFR appears strong stationary properties based on the DFGLS and the KPSS tests, its ability of prediction shown by the IVX methodology is suspectable because of the size distortion problem in stationary cases. On the other hand, the previous evidence on the signifiance of CSP and LTY (and TBL) as a predictors is overturned when the IVX approach is employed. Since strong evidence of nonstationarity are shown for both of CSP and LTY, this may support the point that the IVX test has little power against the unbalanced fixed alternatives.

To check if the degree of predictability has changed through time, we split the whole sample period into two halves. The first sub-period is from January 1927 to

December 1951 and the second one is from January 1952 to December 2012[8]. The results are provided in the second and the third panels of Table 3.20.

For the subsample of 1927-1951, there is no significant evidence in favour of predictability for excess market returns in the case of single predictor, except for CSP. The numbers of $\lambda(U,S)$ again confirm most of the vatiables are highly persistent. However, D/P reports quite small $\lambda(U,S)$ (0.299) along with the significant DFGLS statistic and the small KPSS statistic, which means D/P appears stationary in this sub-period.

We next examine the sub-period of 1952-2012. A large number of predictive regression literature has reported that the stock return predictability becomes much weaker for post-1952 data (for example, Campbell and Yogo, 2006). They argue that the disappearance of predictability is due to structural change in financial markets or improved market efficiency. While our tests show totally different conclusions and much stronger evidence in favour of predictability is documented. Six predictors (book-to-market ratio, T-bill rate, long-term yield, stock variance, cross-sectional premium and net equity expansion) are found to predict excess market returns at the 5% level or lower. On the other hand, only three predictors (D/P, TMS and SVAR) are shown to have predictive power when the IVX approach is employed. This finding is in line with Kostakis, Magdalinos and Stamatogiannis (2010), who reported much weaker evidence in favour of predictability in their post-1967 sample using the IVX method.

Overall, when employing only one variable as the predictor, seven variables (D/P, E/P, TMS, DFY, DFR, LTR and D/E) are found to have no predictive power in either the whole sample period or any of the two sub-periods we consider. Therefore,

[8]When CSP is employed, the first sub-period is from May 1937 to December 1951 and the second one is from January 1952 to December 2002.

the inability of these variables to predict next month excess market returns is robust to the choice of the sample period of analysis and it cannot be solely attributed to parameter instability. Since the dividend price ratio is one of the most commonly used variables and our test casts doubt on its predictive power, researchers who use it for conditional asset pricing tests and conditional performance evaluation should be aware of this problem. On the other hand, our test finds cross-sectional premium (CSP) is the most reliable predictor of future excess stock returns, which supports the arguments of Polk et al. (2006). The predictable pattern persists through time and reflects time-varying risk premia rather than mispricings (Fama, 1991). Regarding the rest of the variables, B/M and LTY are also strong predictors and found to predict next month's excess returns when using the whole sample period and the post-1952 sub-period.

3.6.3 Multivariate Model

Since multivariate models are widely employed in practice, for example, the tests of the semi-strong form of market efficiency, we also present test results using selective predictor combinations. The excess market return is treated as the dependent variable and a subset of the previously regressors are used as predictors. It should be mentioned that we exclude from this exercise the cross-sectional premium due to the lack of data. We examine various combinations. The selection scheme is as follows: First, We classify these 12 predictors into four categories—valuation ratios, bond yield measures, estimates of equity risk, and corporate fiance variables. Second, we separate significant predictors from those that were not significant (for example, when the whole sample period is employed, B/M and LTY are separated from other variables.). Third, within each of these two subsets, we select one predictor from each categories to produce a bivariate predictor and run tests. For the whole sample period, we exclude the default return spread from this exercise since it is the only one stationary variable and consider 31 combinations. The test results are omited to present to save space. The combination of the two individually significant variables B/M and LTY are found to be jointly significant at 1% level, showing very strong evidence in favour of predictability. On the other hand, there is no such evidence for all combinations of the individually insignificant predictors (TBL and D/E are jointly significant at about 10% level). In summary, when using the whole sample period, no strong evidence supporting the predictive ability of D/P, E/P, TBL, TMS, DFY, LTR, SVAR, NTIS and D/E is found, neither individually nor jointly.

For the first sub-period, we select combinations following the previous sheme. it should be mentioned that we combine D/P and DFR due to their strong stationary properties. There are no significant combinations.

When the second sub-period is employed, all of the two-predictor combinations within the subset of individually significant variables are found to be jointly significant at 1% level or even lower. Most strong evidence for the joint significance are documented when combinations of B/M and LTY, B/M and SVAR, TBL and SVAR, LTY and SVAR, and LTY and NTIS are employed in the model. On the other hand, for no combination of the individually insignificant predictors (DFR is excluded from this exercise) can we reject the null hypothesis of no predictability even at the 10% level.

4. CONCLUSION

In the first essay, we use a mixture copula to model the temporal dependence of a univariate time series (copula-based stationary Markov model). To handle misspecification cases, we estimate the mixture copula by a model average approach. Our theorem shows that the model average approach can generate an asymptotically optimal estimator in the sense of achieving the infeasible lowest possible squared estimation losses. Simulations show that compared with competing methods, our model average approach can generate estimation results with the smallest mean square errors, especially when the copula model is misspecified. Extreme conditional quantiles estimated by the model average estimators are also more accurate than those estimated by other methods. In empirical studies, we apply the model average method to estimate the temporal dependence of the daily returns of several equity indexes. Estimation and predicting results support the superiority of our method in capturing the temporal dependence structures of financial returns.

In the second essay, we propose a new test that is a data-dependent weighted average of the regression-based test and the covariance-based test. This test can automatically select the regression-based test statistic when predictors are stationary, and select the covariance-based test statistic when predictors are (near) nonstationary. Our test yields a test statistic that has a standard χ^2 limiting distribution regardless of whether the regressors are stationary, local-to-unity or $I(1)$, i.e., no prior knowledge of the orders of integration is required. The new test does not force the dependent variable and predictors to share the same order of integration under the alternative hypothesis. This property is useful because in practice, the dependent variable (such as financial returns) usually appears to be stationary while predictors (such as dividend yields or interest rates) may be (near) nonstationary. Simulations show that this test could successfully control the size while maintaining a strong power in either balanced or unbalanced cases.

REFERENCES

Amihud, Y. and Hurvich, M., 2004. Predictive regression: A reduced-bias estimation method. Journal of Financial and Quantitative Analysis 39, 813-841.

Amihud, Y., Hurvich, M. and Wang, Y., 2009. Multiple-predictor regressions: hypothesis testing. The Review of Financial Studies 22, 413-434.

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance estimation. Econometrica 59, 817-858.

Bouyé, E., Gaussel, N. and Salmon, M., 2002. Investigating dynamic dependence using copulae. Manuscript, Financial Econometrics Research Center.

Bouyé, E. and Salmon, M., 2009. Dynamic copula quantile regressions and tail area dynamic dependence in forex markets. The European Journal of Finance 15, 721-750.

Cai, Z. and Wang, Y., 2014. Testing predictive regression models with nonstationary regressors. Journal of Econometrics 178, 4-14.

Campbell, J., Lo, A. and MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.

Campbell, J. and Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.

Camponovo, L., 2012. Differencing transformations and robust inference in predictive regression models. Unpublished manuscript.

Camponovo, L., Scaillet, O. and Trojani, F., 2013. Predictability hidden by anomalous observations. Research Paper Series N°13-05, Swiss Finance Institute.

Cavanagh, C., Elliott, G. and Stock, J., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11, 1131-1147.

Cenesizoglu, T. and Timmermann, A., 2008. Is the distribution of stock returns predictable? Unpublished manuscript, HEC Montreal and UCSD.

Chen, W. and Deo, R., 2009. Bias reduction and likelihood based almost-exactly sized hypothesis testing in predictive regressions using the restricted likelihood. Econometric Theory 25, 1143-1179.

Chen, W., Deo, R. and Yi, Y., 2013. Uniform inference in predictive regression models. Unpublished manuscript.

Chen, X., and Fan, Y., 2006. Estimation of Copula-based Semiparametric Time Series Models, Journal of Econometrics 130, 307-335.

Chen, X., Fan, Y. and Patton, A., 2003. Simple tests for models of dependence between financial time series: with applications to US equity returns and exchange rates. Unpublished manuscript.

Chen, X., Wu, W. and Yi, Y., 2009. Efficient estimation of copula-based semiparametric Markov models. The Annals of Statistics 37, 4214-4253.

Cherubini, U., Vecchiato, W. and Luciano, E., 2004. Copula methods in finance. John Wiley & Sons, New York.

Chollete, L., Heinen, A. and Valdesogo, A., 2009. Modeling international financial returns with a multivariate regime-switching copula. Journal of Financial Econometrics 7, 437-480.

Chollete, L., Peña, V. and Lu, C., 2005. Comovement of international financial markets. Unpublished manuscript.

Darsow, W., Nguyen, B. and Olsen, E., 1992. Copulas and Markov processes. Illinois Journal of Mathematics 36, 600-642.

Fama, E., 1991. Efficient capital markets: II. Journal of Finance 46, 1575-1617.

Fan, Y. and Patton, A., 2014. Copulas in econometrics. Annual Review of Economics 6, 179-200.

Fermanian, J.D., 2005. Goodness of Fit Tests for Copulas. Journal of Multivariate Analysis 95, 119-152.

Frey, R. and McNeil, A., 2001. Modeling dependent defaults. ETH E-Collection, http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=273, ETHZ.

Genest, C. and Rivest, L., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034-1043.

Goyal, A. and Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455-1508.

Hannan, E.J., 1970. Multiple Time Series. John Wiley & Sons, New York.

Hansen, B., 2007. Least squares model averaging. Econometrica 75, 1175-1189.

Harvey, D., Leybourne, S. and Taylor, A., 2007. A simple, robust and powerful test of the trend hypothesis. Journal of Econometrics 141, 1302-1330.

Hu, L., 2006. Dependence patterns across financial markets: A mixed copula approach. Applied Financial Economics 16, 717-729.

Ibragimov, R., 2009, Copulas-based characterizations and higher-order Markov processes. Econometric Theory 25, 819-846.

Joe, H., 1997. Multivariate models and dependence concepts. Chapman & Hall, London.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M., 2010. Robust econometric inference for stock return predictability. Unpublished manuscript, University of Nottingham.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54, 159-178.

Li, D., 2000. On default correlation: A copula function approach. Journal of

Fixed Income 9, 43-54.

Long, W., Liu, G., Zhang, X. and Li, Q., 2015. Detecting Financial Data Dependent Structure by Averaging Mixture Copula. Unpublished manuscript.

Magdalinos, T. and Phillips, P., 2009a. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory 25, 482-526.

Magdalinos, T. and Phillips, P., 2009b. Econometric inference in the vicinity of unity. CoFie Working Paper (7), Singapore Management University.

Manner, H. and Reznikova, O., 2012. A survey on time-varying copulas: Specification, simulations and application. Econometric Reviews 31, 654-687.

Maynard, A. and Shimotsu, K., 2009. Covariance-based orthogonality tests for regressors with unknown persistence. Econometric Theory 25, 63-116.

Nelsen, R., 2006. An introduction to copulas. Springer, New York.

Newey, W. and West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Ng, S. and Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

Patton, A., 2006. Modelling asymmetric exchange rate dependence. International Economic Review 47, 527-556.

Patton, A., 2012. A review of copula models for economic time series. Journal of Multivariate Analysis 110, 4-18.

Phillips, P., 2014. On confidence intervals for autoregressive roots and predictive regression. Econometrica 82, 1177-1195.

Phillips, P. and J.H. Lee, 2013. Predictive regression under various degrees of persistence and robust long-horizon regression. Journal of Econometrics 177, 250-264.

Polk, C., Thompson, S. and Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. Journal of Financial Economics 81, 101-141.

Scaillet, O., 2007. Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters, Journal of Multivariate Analysis, 98, 533-543. Shao, J., 1997. An asymptotic theory for linear model selection. Statistica Sinica 7, 221-264.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publication de l'Institut de Statistique de l'Universite de Paris 8, 229-231.

Stock, J., 1991. Confidence intervals for the largest autoregressive root in U.S. economic time series. Journal of Monetary Economics 28, 435-460.

Torous, W., Valkanov, R. and Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937-966.

Valkanov, R., 2003. Long-horizon regressions: theoretical results andapplications. Journal of Financial Economics 68, 201-232.

Viceira, L., 1997. Testing for structural change in the predictability of asset returns. Unpublished manuscript.

Victor, H., Ibragimov, R. and Sharakhmetov, S., 2006. Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. Optimality, 183-209. Institute of Mathematical Statistics.

APPENDIX A

We denote $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\theta}}_1^\top, \ldots, \widehat{\boldsymbol{\theta}}_K^\top\right)^\top$, assuming K is fixed. Define the fixed dimension of $\widehat{\boldsymbol{\theta}}$ by $\varkappa$. We also define $\boldsymbol{\theta}_k^*$ as the pseudo true value

$$\boldsymbol{\theta}_k^* \;= \arg\max_{\boldsymbol{\theta}_k} E\left[\log c_k(G_0(Y_{t-1}), G_0(Y_t); \boldsymbol{\theta}_k)\right], \qquad k = 1, ..., K,$$

where $c_k(G_0(Y_{t-1}), G_0(Y_t); \boldsymbol{\theta}_k)$ is the copula density of $C_k(G_0(Y_{t-1}), G_0(Y_t); \boldsymbol{\theta}_k)$. If the copula model is correctly specified, then $\boldsymbol{\theta}_k^*$ equals the true copula parameter $\boldsymbol{\theta}_0$. Otherwise, $c_k(\cdot, \cdot; \boldsymbol{\theta}_k^*)$ is the closest to the true copula density in terms of KLIC. Denote $\boldsymbol{\theta}^* = \left(\boldsymbol{\theta}_1^{*\top}, \ldots, \boldsymbol{\theta}_K^{*\top}\right)^\top$. Let $\mathbf{C}^*(\mathbf{w}) = \widehat{\mathbf{C}}(\mathbf{w}) \mid_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*}$, $\nu_t(\mathbf{w}) = \partial C(\widetilde{U}_t; \widehat{\boldsymbol{\theta}}_k, \mathbf{w})/\partial\widehat{\boldsymbol{\theta}}|_{\widehat{\boldsymbol{\theta}}=\bar{\boldsymbol{\theta}}_t}$, for $t = 1, \ldots, T$, where $\bar{\boldsymbol{\theta}}_t$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$, $\mathbf{Q}(\mathbf{w}) = \{\nu_1(\mathbf{w}), \ldots, \nu_T(\mathbf{w})\}^\top$, $L_T^*(\mathbf{w}) = \|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\|^2$, and $\xi_T = \inf_{\mathbf{w}\in\mathcal{W}} L_T^*(\mathbf{w})$.

Furthermore, for the k^{th} copula, let $C_k\left\{v_1, v_2; \boldsymbol{\theta}_k\right\} \equiv C_k(\mathbf{v}; \boldsymbol{\theta}_k)$ and $c_k\left\{v_1, v_2; \boldsymbol{\theta}_k\right\} \equiv c_k(\mathbf{v}; \boldsymbol{\theta}_k)$. Denote $l_k(\mathbf{v}; \boldsymbol{\theta}_k) \equiv \log c_k(\mathbf{v}; \boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta},k}(\mathbf{v}; \boldsymbol{\theta}_k) \equiv \partial l_k(\mathbf{v}; \boldsymbol{\theta}_k)/\partial\, \boldsymbol{\theta}_k$, $l_{\boldsymbol{\theta}\boldsymbol{\theta},k}(\mathbf{v}; \boldsymbol{\theta}_k) \equiv \partial^2 l_k(\mathbf{v}; \boldsymbol{\theta}_k)/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_k^\top$, $l_{\boldsymbol{\theta}\boldsymbol{j},k}(\mathbf{v}; \boldsymbol{\theta}_k) \equiv \partial^2 l_k(\mathbf{v}; \boldsymbol{\theta}_k)/\partial u_j\partial\boldsymbol{\theta}_k^\top$ for $j = 1, 2$, and
$U_t \equiv (F(Y_{t-1}), F(Y_t))$.

The following regularity conditions are needed to prove the asymptotic optimality as stated in Theorem 1.

Condition C.1: $\boldsymbol{\theta}^*$ is a finite dimensional vector with constant components. $\boldsymbol{\theta}^*$ takes value in a compact subset of $\mathcal{R}^\varkappa$, and $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = O_p(T^{-1/2})$.

Condition C.2: The elements of $T \times \varkappa$ matrix $\mathbf{Q}(\mathbf{w})$ are uniformly bounded.

Condition C.3 There exists a sequence $c_T \to 0$ such that $T\xi_T^{-2} \leq c_T$ almost surely.

Remark: Condition C.1 requires that $\boldsymbol{\theta}^*$ takes values in a compact set. Therefore, C.1 rules out some cases such as the true distribution is normal, while one fits a t_ν-distribution model and estimate the degree of freedom ν (because the true value $\nu^* = \infty$). Condition C.1 also requires that the convergence rate of $\widehat{\boldsymbol{\theta}}$ to the pseudo true value $\boldsymbol{\theta}^*$ is $O_p(T^{-1/2})$. Proposition 4.3 in Chen and Fan (2006) shows that Condition C.1 holds true when the copula model is correctly specified. When the copula model is misspecified, Condition C.1 can be proved in the same way as that for Proposition 4.3 in Chen and Fan (2006), except to replace their true copula parameter value by the pseudo-true value $\boldsymbol{\theta}_k^*$. The assumptions for showing Condition C.1 are quite general, including: (i) $\boldsymbol{\theta}_k^*$ are in the interior of the parameter space for $k = 1, ..., K$, (ii) $\{Y_t\}_{t=1}^{\mathrm{T}}$ is stationary β mixing with the appropriate decay rate, (iii) $l_{\boldsymbol{\theta},k}(\mathbf{u}; \boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta\theta},k}(\mathbf{u}; \boldsymbol{\theta}_k)$ and $l_{\boldsymbol{\theta j},k}(\mathbf{u}; \boldsymbol{\theta}_k)$ satisfy some standard smooth conditions for $k = 1, ..., K$ and $j = 1, 2$, and (iv) $l_{\boldsymbol{\theta},k}(U_t; \boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta\theta},k}(U_t; \boldsymbol{\theta}_k)$ and $l_{\boldsymbol{\theta j},k}(U_t; \boldsymbol{\theta}_k)$ satisfy some appropriate moment conditions for $k = 1, ..., K$ and $j = 1, 2$.

With Conditions C.1 and C.2, we have that uniformly for $w \in \mathcal{W}$,

$$\begin{aligned} & T^{-1/2} \left\| \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \right\|^2 \\ \leq \ & T^{-1/2} \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\|^2 \lambda_{\max} \left\{ \mathbf{Q}^{\mathrm{T}}(\mathbf{w}) \mathbf{Q}(\mathbf{w}) \right\} \\ = \ & T^{-1/2} O_p(T^{-1}) O_p(T) \\ = \ & O_p(T^{-1/2}), \end{aligned} \tag{A.1}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix (since C.2 implies that $\lambda_{\max}(\cdot) = O_p(T)$), and

$$T^{-1/2} \left\{ \mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0 \right\}^\top \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \ \boldsymbol{\theta}^*) = T^{-1/2} O_p(T^{1/2}) = O_p(1), \tag{A.2}$$

where we also used the fact that the elements of vector $|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0|$ are uniformly bounded by 2.

In Condition C.3, we imposes a limitation on the situation to apply our asymptotic results. Condition C.3 requires that ξ_T grow at a rate faster than $T^{1/2}$, which implies all candidate copulas are misspecified. This condition is similar to condition 7 of Ando and Li (2014). Considering the fact that empirical researchers usually take a small copula candidate set, the misspecification problem should be common. Also please note that the assumption that all candidate models are misspecified is a common condition used in proving optimality properties of model average estimators.

We first present a Lemma which will be used in the proof of Theorem 1.

Lemma 1: If $CV_J(\mathbf{w})$ can be written as $CV_J(\mathbf{w}) = L_T(\mathbf{w}) + a_T(\mathbf{w}) + b_T$, where $L_T(\mathbf{w}) = \|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\|^2$ is defined in (2.12), $L_T^*(\mathbf{w}) = \|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\|^2$ with $\mathbf{C}^*(\mathbf{w}) = \widehat{\mathbf{C}}(\mathbf{w}) \mid_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*}$, $a_T(\mathbf{w})$ is a function satisfying (A.3) below, and the last term b_T is not related to $\mathbf{w}$. If, as $T \to \infty$,

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})} = o_p(1), \tag{A.3}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} \left| \frac{L_T(\mathbf{w})}{L_T^*(\mathbf{w})} - 1 \right| = o_p(1), \tag{A.4}$$

and there exists a positive constant c such that

$$\xi_T \geq c \qquad \text{almost surely}, \tag{A.5}$$

then Theorem 1 holds true.

Proof: See the proof of Lemma 1 of Long et al. (2015).

One can easily see that

$$\begin{aligned}
CV(\mathbf{w}) &= \left\|\widehat{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}}\right\|^2 \\
&= \left\|\left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\right\} + \left(\mathbf{C}_0 - \widetilde{\mathbf{C}}\right)\right\|^2 \\
&= L_T(\mathbf{w}) + \left\|\mathbf{C}_0 - \widetilde{\mathbf{C}}\right\|^2 + 2\left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\right\}^\top \left(\mathbf{C}_0 - \widetilde{\mathbf{C}}\right) \\
&= L_T(\mathbf{w}) + 2\widehat{\mathbf{C}}(\mathbf{w})^\top(\mathbf{C}_0 - \widetilde{\mathbf{C}}) - (\mathbf{C}_0 + \widetilde{\mathbf{C}})^\top(\mathbf{C}_0 - \widetilde{\mathbf{C}}) \\
&\equiv L_T(\mathbf{w}) + \Xi_T(\mathbf{w}) - (\mathbf{C}_0 + \widetilde{\mathbf{C}})^\top(\mathbf{C}_0 - \widetilde{\mathbf{C}}),
\end{aligned}$$

where the last term has nothing to do with the weight vector $\mathbf{w}$, and

$$\begin{aligned}
L_T(\mathbf{w}) &= \left\|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\right\|^2 \\
&= \left\|\left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\} + \left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}\right\|^2 \\
&= \left\|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\|^2 + \|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\|^2 \\
&+ 2\left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}^\top \left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\} \\
&= L_T^*(\mathbf{w}) + \left\|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\|^2 + 2\left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}^\top \left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\} \\
&\equiv L_T^*(\mathbf{w}) + \Pi_T(\mathbf{w}).
\end{aligned}$$

First we obtain equation (A.5) directly from Condition C.3. Hence from Lemma 1, Theorem 1 is valid if the following hold:

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Xi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} = o_p(1) \tag{A.6}$$

and

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Pi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} = o_p(1). \tag{A.7}$$

Using Taylor expansion,

$$\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w}) = \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*). \tag{A.8}$$

where $\mathbf{Q}(\mathbf{w}) = \mathbf{Q}(\mathbf{w}; \overline{\boldsymbol{\theta}}_1, \ldots, \overline{\boldsymbol{\theta}}_T)$ with $\overline{\boldsymbol{\theta}}_t$'s being between the line segment of $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$ (the detailed definition of $\mathbf{Q}(\mathbf{w})$ can be found in the first paragraph of Appendix A).

From (A.8), (A.1), (A.2), and Condition C.3 we can show

$$\begin{aligned}
& \sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Pi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \\
\leq\ & \xi_T^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \left\| \widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w}) \right\|^2 + 2\xi_T^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \left| \{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\}^\top \left\{ \widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w}) \right\} \right| \\
=\ & \frac{T^{1/2}}{\xi_T} T^{-1/2} \sup_{\mathbf{w}\in\mathcal{W}} \left\| \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \right\|^2 \\
+\ & 2\frac{T^{1/2}}{\xi_T} T^{-1/2} \sup_{\mathbf{w}\in\mathcal{W}} \left| \{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\}^\top \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \right| \\
=\ & o_p(1),
\end{aligned}$$

which is (A.7).

Define $F_0(\cdot)$ as the true distribution function of (Y_{t-1}, Y_t). It is a standard result that

$$\sqrt{T} \sup_{\mathbf{y}\in\mathcal{R}^2} \mid F_0(\mathbf{y}) - \widetilde{C}(\mathbf{y}) \mid = O_p(1). \tag{A.9}$$

From $c_T \to 0$, (A.9), and the fact that any element of vectors $|\widehat{\mathbf{C}}_k|$ are bounded by 1, we have

$$
\begin{aligned}
& c_T^{1/2} T^{-1/2} \left| \sum_{t=1}^{T} C_k(\widetilde{U}_t; \widehat{\boldsymbol{\theta}}_k) \left\{ C_0(U_{0t}; \boldsymbol{\theta}_0) - \widetilde{C}(Y_t, Y_{t-1}) \right\} \right| \\
\leq \;& c_T^{1/2} T^{1/2} T^{-1} \sum_{t=1}^{T} \left| C_0(U_{0t}; \boldsymbol{\theta}_0) - \widetilde{C}(Y_t, Y_{t-1}) \right| \\
= \;& c_T^{1/2} T^{1/2} T^{-1} \sum_{t=1}^{T} \left| F_0(Y_t, Y_{t-1}) - \widetilde{C}(Y_t, Y_{t-1}) \right| \\
= \;& o_p(1), \qquad \text{(A.10)}
\end{aligned}
$$

which, along with Condition C.3 and the assumption that K is fixed, implies that

$$
\begin{aligned}
& \xi_T^{-1} \sup_{\mathbf{w} \in \mathcal{W}} \left| \widehat{\mathbf{C}}(\mathbf{w})^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \right| \\
= \;& \xi_T^{-1} \sup_{\mathbf{w} \in \mathcal{W}} \left| \sum_{k=1}^{K} w_k^\top \widehat{\mathbf{C}}_k^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \right| \\
\leq \;& \sum_{k=1}^{K} \frac{T^{1/2}}{\xi_T} T^{-1/2} \left| \widehat{\mathbf{C}}_k^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \right| \\
= \;& \sum_{k=1}^{K} \frac{T^{1/2}}{\xi_T} \left| T^{-1/2} \sum_{t=1}^{T} C_k(\widetilde{U}_t; \widehat{\boldsymbol{\theta}}_k) \left\{ C_0(U_{0t}; \boldsymbol{\theta}_0) - \widetilde{C}(Y_t, Y_{t-1}) \right\} \right| \\
\leq \;& \sum_{k=1}^{K} c_T^{1/2} \left| T^{-1/2} \sum_{t=1}^{T} C_k(\widetilde{U}_t; \widehat{\boldsymbol{\theta}}_k) \left\{ C_0(U_{0t}; \boldsymbol{\theta}_0) - \widetilde{C}(Y_t, Y_{t-1}) \right\} \right| \\
= \;& \sum_{k=1}^{K} c_T^{1/2} T^{-1/2} \left| \sum_{t=1}^{T} C_k(\widetilde{U}_t; \widehat{\boldsymbol{\theta}}_k) \left\{ C_0(U_{0t}; \boldsymbol{\theta}_0) - \widetilde{C}(Y_t, Y_{t-1}) \right\} \right| \\
= \;& o_p(1), \qquad \text{(A.11)}
\end{aligned}
$$

where the second '$\leq$' holds almost surely. Therefore, we obtain (A.6) from (A.11).

APPENDIX B

Proof of Theorem 1: We denote the long run covariance matrices associated with $\mathbf{u}_t$ by:

$$\begin{aligned}
\Omega &= \begin{bmatrix} \Omega_{yy} & \Omega_{yv} \\ \Omega_{vy} & \Omega_{vv} \end{bmatrix} = F(1)\Sigma F(1)' = \sum_{h=-\infty}^{\infty} E\mathbf{u}_t\mathbf{u}'_{t-h}, \\
\Lambda &= \begin{bmatrix} \Lambda_{yy} & \Lambda_{yv} \\ \Lambda_{vy} & \Lambda_{vv} \end{bmatrix} = \sum_{h=1}^{\infty} E\mathbf{u}_t\mathbf{u}'_{t-h}.
\end{aligned}$$

To apply Theorem 1 of Kostakis et al. (2010), we first need to show

$$T^{\frac{1-(\alpha\wedge\delta)}{2}}(\widehat{\Lambda}_{\varepsilon v} - \Lambda_{yv}) \to_p 0 \tag{B.1}$$

Notice that

$$\begin{aligned}
\widehat{\Lambda}_{\varepsilon v} &= \frac{1}{T}\sum_{h=1}^{M}(1-\frac{h}{M+1})\sum_{t=h+1}^{T}\widehat{\varepsilon}_t\widehat{v}'_{t-h} \\
&= \frac{1}{T}\sum_{h=1}^{M}(1-\frac{h}{M+1})\sum_{t=h+1}^{T}(y_t - \widehat{\beta}_{OLS}\mathbf{x}_{t-1})[v_{t-h} - (\widehat{R}_{OLS} - R)\mathbf{x}_{t-h-1}]'
\end{aligned}$$

Using Lemma 7.2 of Kostakis et al. (2010) and the facts that $\widehat{\beta}_{OLS} = O_p(T^{-\alpha})$ as well as by equation (11) of Magdalinos and Phillips (2009a), $\widehat{R}_{OLS} - R = O_p(T^{-\alpha})$, we have

$$\widehat{\Lambda}_{\varepsilon v} = \frac{1}{T}\sum_{h=1}^{M}(1-\frac{h}{M+1})\sum_{t=h+1}^{T} y_t v'_{t-h} + O_p(\frac{M}{T^{\alpha}})$$

Then we follow the proof of Lemma 1 (i) of Kostakis et al. (2010) and obtain

$$\begin{aligned} & \frac{1}{T}\sum_{h=1}^{M}(1-\frac{h}{M+1})\sum_{t=h+1}^{T} y_t \upsilon'_{t-h} \\ = & \Lambda_{y\upsilon} + O_p(\frac{M}{T^{1/2}}) + O_p(\frac{1}{M}) \end{aligned}$$

Since $\alpha > 1/2$, we get the following equation

$$\widehat{\Lambda}_{\varepsilon\upsilon} - \Lambda_{y\upsilon} = O_p(\max\left\{\frac{M}{T^{1/2}}, \frac{1}{M}\right\}).$$

Finally, applying the conclusion of Lemma 1 (ii) of Kostakis et al. (2010): let $M = L(T)T^{\gamma}$ for some slowly varying function L and $\gamma > 0$. When $\delta \in (\frac{2}{3}, 1)$, a choice of $\gamma = 1/4$ guarantees the validity of (B.1).

Having the condition of (B.1), we now can apply Theorem 1 of Kostakis et al. (2010)

$$\begin{aligned} (i)\ T^{\frac{1+\delta}{2}} vec(\widehat{\beta}_{IVX}) & \rightarrow {}_d MN\left(0, \Sigma_{\widehat{\beta}_{IVX}}\right) \text{ if } \frac{2}{3} < \delta < \min(\alpha, 1) \\ (ii)\ T^{\frac{1+\delta}{2}} vec(\widehat{\beta}_{IVX}) & \rightarrow {}_d N\left(0, \Upsilon_{\widehat{\beta}_{IVX}}\right) \text{ if } \frac{1}{2} < \alpha \leq \delta, \end{aligned}$$

and Theorem 2 of Kostakis et al. (2010)

$$\begin{aligned} W_{IVX} & = vec(\widehat{\beta}_{IVX})'[(X'P_{\tilde{Z}}X)^{-1} \otimes \widehat{\Omega}_{\varepsilon\varepsilon}]^{-1} vec(\widehat{\beta}_{IVX}) \\ & \rightarrow {}_d\chi^2(p). \end{aligned}$$

Proof of Lemma 1:

(i). $\lim_{T\to\infty} m^q E(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) = -k_q \sum_{h=1}^{\infty} \Gamma_{\Delta\mathbf{x}y}(h)h^q$ since Lemma 1 in MS has shown $\lim_{T\to\infty} m^q E(\widehat{\lambda}_{y,\Delta x_i} - \lambda_{y,\Delta x}) = -k_q \sum_{h=1}^{\infty} \Gamma_{\Delta x_i y}(h)h^q$ for each i.

(ii). $\widehat{\lambda}_{y,\Delta\mathbf{x}} = (\widehat{\lambda}_{y,\Delta x_1}, \widehat{\lambda}_{y,\Delta x_2}, \cdots \widehat{\lambda}_{y,\Delta x_p})' \underset{p}{\longrightarrow} \lambda_{y,\Delta\mathbf{x}}$ as $T \to \infty$ since Lemma 1 in MS has shown $\widehat{\lambda}_{y,\Delta x_i} \underset{p}{\longrightarrow} \lambda_{y,\Delta x_i}$ for each i.

(iii). MS (2009) gives the proof of $\lim_{T\to\infty} Tm^{-1}var(\widehat{\lambda}_{y,\Delta x_i})$= $4\pi^2 f_{yy}(0) f_{\Delta x_i \Delta x_i}(0) \int_0^\infty k^2(x)dx$ for each i. What we need to show is $\lim_{T\to\infty} Tm^{-1}cov(\widehat{\lambda}_{y,\Delta x_i}, \widehat{\lambda}_{y,\Delta x_j})$=$4\pi^2 f_{yy}(0) f_{\Delta x_i \Delta x_j}(0) \int_0^\infty k^2(x)dx$ for $i \neq j$.

Observe that

$$Tm^{-1}cov(\widehat{\lambda}_{y,\Delta x_i}, \widehat{\lambda}_{y,\Delta x_j}) = \frac{T}{m} \sum_{h'=1}^{T-1} \sum_{h=1}^{T-1} k(\frac{h'-1}{m}) k(\frac{h-1}{m}) cov(\widehat{\Gamma}_{\Delta x_i y}(h'), \widehat{\Gamma}_{\Delta x_j y}(h)) \tag{B.2}$$

Hannan (1970, p.313) gives

$$cov(\widehat{\Gamma}_{\Delta x_i y}(h'), \widehat{\Gamma}_{\Delta x_j y}(h)) = T^{-1} \sum_{u=-\infty}^{\infty} \{\Gamma_{\Delta x_i \Delta x_j}(u) \Gamma_{yy}(u+h-h')$$

$$+\Gamma_{\Delta x_i y}(u+h) \Gamma_{y \Delta x_j}(u-h') + k_{\Delta x_i y \Delta x_j y}(0, h', u, u+h)\} \phi_T(u, h', h)$$

where $k_{\Delta x_i y \Delta x_j y}(0, h', u, u{+}h)$ is the fourth cumulant of $(y_t, \Delta x_{it}, \Delta x_{jt})$ and $\phi_T(u, h', h)$ is given by

$$\phi_T(u, h', h) = \begin{cases} = 0, \qquad u \leq -T + h'; = 1 - \frac{h'-u}{T}, -T + h' \leq u \leq 0; \\ = 1 - \frac{h'}{T}, 0 \leq u \leq h' - h; = 1 - \frac{h+u}{T}, h' - h \leq u \leq T - h; \\ = 0, \qquad u \leq T - h \end{cases}$$

Therefore, following MS (Proof of Lemma 1), (B.2) is equal to

$$\frac{1}{m} \sum_{h'=1}^{T-1} \sum_{h=1}^{T-1} k(\frac{h'-1}{m}) k(\frac{h-1}{m}) \tag{B.3}$$

$$\times \sum_{u=-\infty}^{\infty} [\Gamma_{\Delta x_i \Delta x_j}(u) \Gamma_{yy}(u+h-h') \phi_T(u, h', h)] + o(1) + o(1)$$

Let $v = h' - h$ and we can write (B.3) as

$$\sum_{v=-T+2}^{T-2} \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x_i \Delta x_j}(u)\Gamma_{yy}(u-v)\{\frac{1}{m}\sum\nolimits_h' \phi_T(u,h',h)k(\frac{h+v-1}{m})k(\frac{h-1}{m})\} \quad \text{(B.4)}$$

where $\sum_h'$ runs only for $\{h : 1 \leq h \leq T-1 \text{ and } 1 \leq h+v \leq T-1\}$.

Following Hannan (1970, pp.314-315), the expression in braces converges to $\int_0^\infty k^2(x)dx$. And $\sum_{v=-T+2}^{T-2}\sum_{u=-\infty}^{\infty} \Gamma_{\Delta x_i \Delta x_j}(u)\Gamma_{yy}(u-v) \to 4\pi^2 f_{yy}(0) f_{\Delta x_i \Delta x_j}(0)$ as $T \to \infty$. Finally we have $\lim_{T\to\infty} Tm^{-1}cov(\hat{\lambda}_{y,\Delta x_i}, \hat{\lambda}_{y,\Delta x_j}) = 4\pi^2 f_{yy}(0) f_{\Delta x_i \Delta x_j}(0)$ $\int_0^\infty k^2(x)dx$ for $i \neq j$.

Proof of Lemma 2: Following Theorem 2 of MS, we have $\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta x_i} - \lambda_{y,\Delta x_i}) \to_d$ $N(0, V_i)$ for each i. To use Cramer-Wold device, we need to prove that for any real $p \times 1$ vector a such that $a'a = 1$,

$$a'\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) \to_d N(0, a'\mathbf{V}a).$$

Observe

$$\begin{aligned}
a'\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,\Delta\mathbf{x}} - \lambda_{y,\Delta\mathbf{x}}) &= \sqrt{\frac{T}{m}}(a'\widehat{\lambda}_{y,\Delta\mathbf{x}} - a'\lambda_{y,\Delta\mathbf{x}}) \\
&= \sqrt{\frac{T}{m}}[\sum_{h=1}^{T-1} k(\frac{h-1}{m})\frac{1}{T}\sum_{t=h+1}^{T} y_t(a'\Delta\mathbf{x}_{t-h}) \\
&\quad -\sum_{h=1}^{\infty} cov(y_t, a'\Delta\mathbf{x}_{t-h})] \\
&= \sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,a'\Delta\mathbf{x}} - \lambda_{y,a'\Delta\mathbf{x}})
\end{aligned}$$

Then we can simply apply Theorem 2 of MS to $z_t^* = (y_t, a'\Delta\mathbf{x})'=$

$\begin{bmatrix} 1 : \mathbf{0}_{1\times p} \\ 0 : a' \end{bmatrix} F(L)\varsigma_t = F^*(L)\varsigma_t$ and obtain

$$\sqrt{\frac{T}{m}}(\widehat{\lambda}_{y,a'\Delta\mathbf{x}} - \lambda_{y,a'\Delta\mathbf{x}}) \to_d N(0,a'\mathbf{V}a).$$

Proof of Lemma 7: We know that $\widehat{\beta} = O_p(T^{-1})$ when $\mathbf{x}_t$ are $I(1)$ or local to unity. For simplicity, we calculate $\widehat{Avar(\widehat{\beta})}$ as

$$\widehat{Avar(\widehat{\beta})}] = (\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}'_{t-1})^{-1}s_T^2$$

where

$$s_T^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}'_{t-1}\widehat{\beta})^2 = \frac{1}{T}\sum_{t=1}^{T}y_t^2 - 2(\frac{1}{T}\sum_{t=1}^{T}y_t\mathbf{x}'_{t-1})\widehat{\beta} + \widehat{\beta}'(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}'_{t-1})\widehat{\beta} \quad \text{(B.5)}$$

The first term of (B.5) is $O_p(1)$, the second term is $O_p(T^{-1})$ since $\sum_{t=1}^{T} y_t\mathbf{x}'_{t-1} = O_p(T)$ and $\widehat{\beta} = O_p(T^{-1})$, and the third term is $O_p(T^{-1})O_p(T)O_p(T^{-1}) = O_p(T^{-1})$. Therefore, $s_T^2 = O_p(1)$, and $\widehat{Avar(\widehat{\beta})}] = O_p(T^{-1})O_p(1) = O_p(T^{-1})$. Finally, we have $W_0 = T\widehat{\beta}'[\widehat{Avar(\widehat{\beta})}]^{-1}\widehat{\beta} = O_p(1)$.

Proof of Lemma 8: Lemma 4 in MS has shown that if Assumption 5, 2, and 3 hold, and $k(x)$ is *Lipschitz(1)*, then for each $i = 1, \cdots, p$,

$$\sqrt{T}(\widehat{\lambda}_{y,\Delta x_i} - \lambda_{y,\Delta x_i}) = O_e(1) + B_T + o_p(1)$$

where B_T satisfying

$$B_T = \begin{cases} 0, & \text{if } Ey_t x_{it-h} = 0 \text{ for all } h \geq 2 \\ O(T^{1/2} m^{-q}), & \text{otherwise.} \end{cases}$$

Therefore, $\sqrt{T}(\widehat{\lambda}_{y,\Delta\mathbf{x}}) = O_e(1)$ under the null, and $\sqrt{T}(\widehat{\lambda}_{y,\Delta\mathbf{x}}) = O_p(T^{1/2})$ under the alternative.

Define $\widetilde{\mathbf{V}}$ as

$$\begin{aligned} \widetilde{\mathbf{V}} &= \frac{1}{m}\sum_{h'=1}^{T-1}\sum_{h=1}^{T-1} k(\frac{h'-1}{m})k(\frac{h-1}{m}) \\ &\quad \times \sum_{u=-\infty}^{\infty} \widetilde{k}(\frac{u}{\widetilde{m}})\widehat{\Gamma}_{\Delta\mathbf{x}\Delta\mathbf{x}}(u)\widetilde{k}(\frac{u+h-h'}{\widetilde{m}})\widehat{\Gamma}_{yy}(u+h-h'), \\ \widehat{\Gamma}_{\Delta\mathbf{x}\Delta\mathbf{x}}(u) &= [T^{-1}\sum_{t=u+1}^{T} \Delta x_{jt}\Delta x_{i,t-u}]_{p\times p}. \end{aligned}$$

Next, we show the order of $\widetilde{\mathbf{V}}$. The proof closely follows MS. In views of equations (B.3) and (B.4) in the proof of Lemma 1, each element in $\widetilde{\mathbf{V}}$ reduces to

$$\sum_{v=-T+2}^{T-2}\sum_{u=-\infty}^{\infty} \widetilde{k}(\frac{u}{\widetilde{m}})\widehat{\Gamma}_{\Delta x_i \Delta x_j}(u)\widetilde{k}(\frac{u-v}{\widetilde{m}})\widehat{\Gamma}_{yy}(u-v)\left\{\int_0^{\infty} k^2(x)dx + o(1)\right\}.$$

Since $\widetilde{m}/T \to 0$ and $\widetilde{k}(x) = 0$ for all $\mid x \mid > 1$, the above equation can be simplifies to

$$\sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}(\frac{u}{\widetilde{m}})\widehat{\Gamma}_{\Delta x_i \Delta x_j}(u)\sum_{u-v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}(\frac{u-v}{\widetilde{m}})\widehat{\Gamma}_{yy}(u-v)\left\{\int_0^{\infty} k^2(x)dx + o(1)\right\}. \quad \text{(B.6)}$$

The order of magnitudes for the variance and bias of $\sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}(\frac{u}{\widetilde{m}})\widehat{\Gamma}_{\Delta x_i \Delta x_j}(u)$ are given in Theorem 9 and 10 respectively of Hannan (1970, pp. 280-283), and

$f_{\Delta x_i \Delta x_j}(0) = 0$ under Assumption 5. Therefore, $\widetilde{\mathbf{V}} = O_p(\widetilde{m}^{-\widetilde{q}} + \widetilde{m}^{1/2}T^{-1/2})$.