Detecting Neglected Nonlinearity in Dynamic Panel Data with Time-Varying Conditional Heteroskedasticity

Yongmiao Hong
Department of Economics &
Department of Statistical Science
Cornell University
yh20@cornell.edu

Chihwa Kao
Department of Economics &
Center for Policy Research
Syracuse University
cdkao@maxwell.syr.edu

March 2004
Preliminary
Please Do Not Cite

We thank Yoon-Jin Lee for insightful comments and the National Science Foundation for support via NSF Grant SES–0111769.

ABSTRACT

A new class of specification tests is proposed to detect for neglected nonlinearity and dynamic misspecification in panel models. The tests can detect a wide range of model misspecifications while being robust to conditional heteroskedasticity and higher order time-varying moments of unknown form. They check a large number of lags so that they can capture dynamic misspecification at any lag order. The large number of lag orders does not cause loss of degrees of freedom because our tests naturally discount higher order lags, which is consistent with the stylized fact that economic behaviors are more affected by recent past events than by remote past events. No specific estimation method is required, and the tests have the appealing "nuisance parameter free" property that parameter estimation uncertainty has no impact on the limit distribution of the test. Simulations show the proposed tests have good finite sample properties. It is important to take into account conditional heteroskedasticity; failure to do so will cause overrejection of a correct linear panel model. Our tests have omnibus and robust power against a variety of alternatives relative to some existing tests for linearity in panel models.

Key Words: Conditional heteroskedasticity, Dynamic panel Model, Generalized spectral derivative, Hausman's test, Joint limit asymptotics, Linearity, Martingale.

1. Introduction

Panel/longitudinal data have been widely used in biology, economics and finance. They often provide insights not available in pure time-series or cross-sectional data. For example, panel data have the ability for control for unobservable individual/time effects which could be correlated with other regressors. There is relatively well-established econometric theory for the linear panel mode. For recent developments in the linear panel models, see Diggle et al. (2002), Baltagi (2002), Arellano (2003), and Hsiao (2003).

Most economic relationship are dynamic in nature. There have been many applications of dynamic panel models in economics and finance. The theoretical investigation of dynamic panel models has received a lot of attention recently. Related works are Arellano and Bond (1991), Kiviet (1995), Hahn and Kuersteiner (2002), Hahn et al. (2002), and Alvarez and Arellano (2003), to name just a few. On the other hand, unlike the nonlinear time series literature (e.g., Tong 1990, Granger and Terasvirta 1993), not much attention has been paid to the nonlinear panel models in the past. However, their usage has increased rapidly recently (e.g., Diggle 2002, Chs. 4 and 5, Li and Stengos 1996, Cheng and Wei 2000, Arellano and Honore, 2001, Honore 2002, Gourieroux and Jasiak, 1998, Stenseth et al. 1999, Yao et al. 2000, Rice 2003). For example, Paap et al. (2003) consider a panel smooth transition AR (STAR) model to capture a large panel of unemployment data in order to examine the common business cycle in unemployment and the importance of sector-specific variations. Stenseth et al. (1999) suggest a panel self-exciting threshold autoregressive (SETAR) model to study the Canada Lynx pollutions and Yao et al. (2000) also consider a panel SETAR model to study Canada mink-muskrat data.

Nonlinear dynamic modeling is rather challenging in the panel data context. It is known that the ability to difference out (e.g, using the conditional likelihood) the individual effects/time effects relies heavily on the linear specification for the panel models. For many nonlinear panel models, it is impossible to use the conditional likelihood or other methods, for example, to eliminate the individual/time effect (see Lancaster 2000 for a survey). This becomes even more challenging when the regressors include lagged dependent variables. For example, the initial condition problem for the nonlinear dynamic panel data with individual effect/time effect is a difficult issue (see Wooldridge 2002 for more discussion). Therefore it is important for researchers to determine first whether a panel model is linear before they go to a nonlinear specification. Before building a nonlinear panel model, it is advisable to find whether a linear panel model will adequately characterize the underlying economic relationships. If that is the case, there will be more econometric theory for building a reasonable model than if a nonlinear model is needed. On the other hand, it may happen, particularly when the time dimension if the panel data is short, that investigators successfully estimate a nonlinear model, although the true relationship is linear. The danger of unnecessarily complicating the model building is real, and this can be diminished by linearity testing. More generally, not much effort has been made to evaluate of dynamic panel models (Granger 1996). To fill this gap, we develop a new class of tests for linear specification in dynamic panel models with conditional heteroskedasticity of unknown form.

There are a variety of tests for various hypotheses on panel model specification, e.g., Hausman (1978), Hausman and Taylor (1981), Kang (1985), Holtz-Eakin (1988), Arellano (1990), Arellano and Bond (1991), Li and Stengos (1992), Baltagi (1997, 1999, 2001 Ch. 4), Metcalf (1996), Fu et al. (2002, FLF) and Andrews and Lu (2001). Hausman (1978) and Hausman and Taylor (1981) develop tests to check whether the unobserved individual/time effects are correlated to the regressors. Kang (1985) discusses how to extend Hausman's test to the two-way panel models. Holtz-Eakin (1988) develops a test for the presence of individual effects in dynamic panel models. Arellano (1990) develops minimum chi-square tests for various covariance restrictions. Arellano and Bond (1991) propose a test to the second-order serial correlation for the error terms. Li and Stengos (1992) propose a Hausman's type test using the semiparametric estimators. Arellano (1993) proposes an alternative variable addition test to Hausman's test which is robust to autocorrelation and heteroskedasticity of arbitrary form. Angrist and Newey (1991) propose an over-identification test to test the fixed effect specification. Baltagi (1997) applies the double-length artificial regression to test whether the panel model is linear or log-linear. Ahn and Low (1996) and Baltagi (1999) showed that Hausman's test can be obtained using artificial regressions. Metcalf (1996) shows can be obtained specification tests for correlated fixed effects developed by Hausman and Taylor (1981) can be extended to panel models with endogenous regressors. FLF extend the portmanteau time series test of Li and McLeod (1981) to dynamic panel models. Andrews and Lu (2002) develop consistent model and moment selection criteria (MMSC) for dynamic panel models to select the lag length for lagged dependent variables and determine the existence of correlation between regressors and the individual effect.

All the existing tests for panel model specification assume a known form of a common alternative for all individuals. This may not be desirable in terms of power because there may exist significant inhomogeneity across individuals (e.g., Choi 2002). For example, by assuming a common alternative model, existing tests (e.g., Hausman 1978, FLF, Andrews and Lu 2002) ignore inhomogeneity across individuals, and so have low or little power against some alternatives with significant inhomogeneity across individuals. Moreover, the existing tests may be powerful against some misspecifications in mean and may lose power in detecting certain nonlinear alternatives.

In this paper, we propose a class of generally applicable omnibus tests for neglected nonlinearity and dynamic misspecification in panel models, with no prior knowledge of possible alternatives (including functional forms, lag structures and inhomogeneity across individuals). We extend the generalized spectral analysis of Hong (1999) and Hong and Lee (2004) from the time series context to panel data contexts. The generalized spectrum was proposed by Hong (1999) as a basic frequency domain tool for nonlinear time series, in a spirit similar to the conventional spectral density as a basic tool for linear time series (e.g., Priestley 1981). Generalized spectrum can capture serial dependence in various conditional moments, and therefore is not suitable to test conditional mean dynamics. In a time series context, Hong and Lee (2004) use a suitable derivative of the generalized spectrum which focuses solely on the conditional mean specification of a time series model. We now generalize this method to test neglected nonlinearity and dynamic misspecification in a linear dynamic panel model with large numbers of both individuals and time observations. Panel data with large n and T have become increasingly available in practice. (Examples are the Penn World Tables, World Bank, NBER, etc.) They are particularly informative about the dynamic natures of a panel process. The associated asymptotic analysis and limited results are substantially different from those in pure time series analysis. We provide a joint limit analysis by allowing both n and T to grow to infinity simultaneously, giving a sensible alternative analytic approach to the existing joint limit analysis methods in the literature (e.g., Phillips and Moon 1999, Hahn et al. 2002). Distinct from the existing joint limit analysis, we allow relatively mild conditions on the relative speed of n and T. We allow n to grow faster or slower than or the same as T, whereas the existing joint limit theories usually assume n grows slower than T. Thus, our test is applicable to panel data with various combinations of n and T. In contrast to Hong and Lee (2003), our asymptotic results here are different. In pure time series contexts, Hong and Lee (2003) obtain the asymptotic normality of their nonparametric kernel-based tests by requiring the smoothing parameter (or lag order) to grow to infinity as T grows. This condition on the smoothing parameter is not required for the asymptotic normality of our nonparametric test statistics in panel contexts, thanks to the benefits of using panel data with large n and T. In time series analysis, it is well known (e.g., Skaug and Tjostheim 1993, 1996) that the asymptotic theory for nonparametric statistics usually provides poor finite sample approximation, because the asymptotically negligible higher order terms depend on the smoothing parameter and are very close in order of magnitude to the dominant term which determines the limit distribution. The fact that the asymptotic distribution of our nonparametric test statistics in panel contexts does not depend on the smoothing parameter indicates that test statistics may provide a reasonable approximation even in moderately small samples. This is an advantage of using panel data with large n and T. Indeed, our asymptotic theory works reasonably well for (n, T) as small as $(25, 25)$, and for different combinations of large n and small T, or small n and large T.

Because there exist infinite nonlinear alternatives and potentially significant inhomogeneity across individuals, we avoid assuming an alternative model in constructing our tests. Thanks to the use of the characteristic function, the generalized spectral derivative can capture both linear and nonlinear serial dependence in conditional mean. The latter can be subtle and difficult to detect using conventional techniques. In the meantime, the generalized spectrum enjoys the nice features of spectral analysis. In particular, it incorporates information on serial dependence from all lags and can characterize cyclical dynamics caused by linear or nonlinear serial dependence. Thus, our approach can detect a wide variety of misspecifications in both functional form and lag structure. Moreover, we treat each individual separately, and can capture inhomogeneity across individuals. This is distinct from the existing tests for panel conditional mean models. FLF and Andrews and Lu (2002) assume a linear functional form and focus only on lag order misspecification, with a common linear alternative for all individuals. One important feature of dynamic panel modelling similar to time series modelling is that the conditioning information set usually contains an infinite number of lags (i.e., the entire past history), unless Markovian assumption holds. In practice, most dynamic panel models usually employ a very small number of lags. As a consequence, it is important to check whether there exists dynamic misspecification due to the use of improper lag orders. Because we use a spectral approach, our tests check a large number of lags without suffering from the curse of dimensionality. When a large number of lags is used, chi-square tests for linearity usually have poor power in finite samples, due to the loss of a large number of degrees of freedom. This undesired feature, fortunately, is not shared by our generalized spectral approach, because it naturally discounts higher order lags, which is consistent with the stylized fact that economic behaviors are usually more influenced by recent events than by the remote past. Thus, our tests are particularly useful when the information set has a large dimension. We note that our tests can be used to test the martingale hypothesis for observed raw data with no modification.

Economic theory, while having implications on conditional mean dynamics, is usually silent about higher order conditional moment dynamics. Thus, it is important to develop tests of conditional mean models that are robust to conditional heteroskedasticity and other higher order time-varying moments of unknown form (e.g., Meghir and Windmeijer 1999, Cermeño and Grier 2001). Granger (1995) emphasizes the importance of testing linearity in the presence of conditional heteroskedasticity in a time series context, and this equally well applies to testing linearity in panel data contexts. Volatility clustering for economic time series is more a rule than an exception. Failure to accommodate conditional heteroskedasticity will lead to improper size for the tests, giving a misleading conclusion. For example, an ARCH process is similar to a bilinear autoregressive process in terms of autocorrelations in level and level-square respectively (e.g., Bera and Higgins 1997). An LM test for a bilinear alternative will likely be mistaken for an ARCH process if conditional homoskedasticity is assumed. As is well known (e.g., Diebold and Nason 1990, Meese and Rose 1991, Granger 1992, Sec. 8), the distinction between nonlinearity in mean and in higher order moments has important economic implications. For example, suppose an asset return follows a bilinear process. Then the level of asset return is predictable using its past history. In contrast, if the asset return follows an ARCH process, then its level is not predictable because it is a martingale difference sequence (*m.d.s.*). As an important feature, our tests for conditional mean models are robust to conditional heteroskedasticity and all other higher order conditional moments of unknown form. In contrast, all existing specification tests in panel models assume conditional homoskedasticity or *i.i.d.* errors (Baltagi 2001, Ch.4).

The vast literature on testing panel model specification is Hausman's (1978) test and its various extensions; there have been no discussion on tests for the conditional mean. Also Hausman's test is not robust to heteroskedasticity of unknown form with inhomogeneity across individuals. Because we compare a nonparametric (inefficient) generalized spectral derivative estimator with a restricted (efficient) counterpart implied by correct conditional mean specification, our tests can be viewed as a generalization of the methods of Hausman (1978) and Hausman and Taylor (1981) on panel models from a parametric context to a nonparametric context. Our tests are generally applicable. We only require the estimated residuals as inputs. No specific estimation method is required, and the tests have the appealing "nuisance parameter free" property that parameter estimation uncertainty has no impact on the limit distribution of the test. Our panel model can be static or dynamic, and one-way or two-way; both balanced and unbalanced panel data are covered; individual and time effects can be fixed or random; regressors can contain lagged dependent variables or deterministic/stochastic trending variables; and no specific estimation method is required.

Section 2 sets up the model. Section 3 describes test statistics. Section 4 introduces the assumptions and derives the asymptotic distribution, and Section 5 investigates the asymptotic power. Section 6 discusses the data-driven lag order selection. Section 7 reports the simulation results. Section 8 concludes, and all proofs are given in the appendix. A GAUSS code for implementing our tests is available from the authors upon request. Throughout, we use C to denote a generic bounded constant, $||\cdot||$ the Euclidean norm, and A^* the complex conjugate of A.

2. Model

We consider a general linear dynamic panel model

$$Y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + \varepsilon_{it},, \qquad t = 1, 2, \cdots T_i, i = 1, ..., n, \tag{2.1}$$

where α_i is an individual-specific effect, λ_t is a time-specific effect, and X_{it} may contain lagged dependent variables $\{Y_{i,t-j},\ j > 0\}$, and current and lagged exogenous variables $\{Z_{i,t-j},\ j \geq 0\}$, and $\beta \in B$ is a finite-dimensional parameter. Both α_i and λ_t can be fixed or random. We allow for unbalanced panel data, which are often the case in practice. For convenience, we assume $T_i = c_i T$ for some integer T, where $c_i \in [c, C]$, where $0 < c < C < \infty$, and c and C do not depend on i.

Throughout, we make the following assumption about the data generating process.

Assumption A.1: (i) For each i, $\{X_{it} - EX_{it}, \varepsilon_{it}\}$ is a stationary α-mixing process with

$$\max_{1 \leq i \leq n} \sum_{j=0}^{\infty} \alpha_i(j)^{\nu/(\nu-1)} \leq C,$$

and $\max_{1 \leq i \leq n} E||X_{it} - EX_{it}||^4 \leq C$, $E(\varepsilon_{it}^2) = \sigma_i^2 \in [c, C]$ for all i, and $\max_{1 \leq i \leq n} E(\varepsilon_{it}^4) \leq C$.

We impose strict stationarity. only on $\{X_{it} - EX_{it}, \varepsilon_{it}\}$. Thus, we allow X_{it} to contain some deterministic trends. For example, $X_{it} = \sum_{j=0}^{d} \alpha_{ij} t^j + Z_{it}$, where $\{Z_{it}\}$ is a stationary α-mixing process. However, Assumption A.1 rules out unit root processes for X_{it}, though it seems plausible to extend our analysis to these cases. This will be pursued in subsequent research. Recall the definition of α-mixing. Let F_{it}^s be the σ-field generated by $\{X_{i\tau} - EX_{i\tau}, \varepsilon_{i\tau}\}_{\tau=t}^s$. Then $\{X_{it} - EX_{it}, \varepsilon_{it}\}$ is mixing if $\sup_{A \in F_{i1}^{\tau}, F_{i\tau+j}^{\infty}} [P(A \cap B) - P(A)P(B)] \leq \alpha_i(j)$ and $\alpha_i(j) \to 0$ as $j \to \infty$. The condition on $\alpha_i(j)$ thus characterizes the rate at which serial dependence decays to zero asymptotically. The mixing condition is rather convenient for nonlinear time series analysis and nonlinear dynamic panel analysis.

Model (2.1) covers most of the popular econometric linear dynamic panel models in the literature (e.g., Holtz-Eakin et al. 1988, Arellano and Bond 1991, Arellano 2003, Hsiao 2003), which are often specified as

$$Y_{it} = \alpha_i + X_{it}'\beta + \sum_{j=1}^{L} \rho_j Y_{i,t-j} + \varepsilon_{it},$$

where L is the lag length of the lagged Y_{it}, and there is no time-specific effect λ_t. In contrast, Hjellvik and Tjostheim (1999) consider

$$Y_{it} = \lambda_t + \sum_{j=1}^{L} \rho_j Y_{i,t-j} + \varepsilon_{it}$$

where the time-specific effect λ_t is used to capture the intercorrelation (i.e., cross-sectional dependence) across i and λ_t and ε_{it} are assumed to be iid. In the econometrics literature, the time effect λ_t is usually removed for simplicity. However, Hjellvik and Tjostheim (1999) point out that the removal of λ_t will lead to ignorance of the intercorrelation across i and they discuss the consequence of neglecting the intercorrelation.

In a dynamic panel context, one is often interested in knowing whether the linear dynamic model is adequate in capturing the dynamics of Y_{it}. Conditional mean modelling has been the primary interest in panel data models, because $E(Y_{it}|I_{i,t-1})$ is the optimal predictor for Y_{it} using $I_{i,t-1}$ in terms of the mean squared error criterion, where $I_{i,t-1}$ is the information available to individual i at time $t-1$. The regressor X_{it} is a subset of $I_{i,t-1}$. We say that the linear dynamic panel model is correctly specified for $E(Y_{it}|I_{i,t-1})$ if

$$\mathbb{H}_0 : \Pr[E(Y_{it}|I_{i,t-1}) = X_{it}'\beta^0] = 1 \text{ for some } \beta^0 \in B. \tag{2.2}$$

The null hypothesis H_0 is a joint hypothesis of (i) $E(Y_{it}|I_{i,t-1}) = E(Y_{it}|X_{it})$ for some $X_{it} \in I_{i,t-1}$ and (ii) $E(Y_{it}|X_t) = X_{it}'\beta_0$ for some β_0. All existing tests for linearity in mean in time series contexts only focus on testing (ii) and ignore testing (i); they can easily miss conditional mean misspecification that occurs only at higher lag orders. Moreover, even for any given lag order, existing linearity tests except White's (1989) neural network test cannot detect all departures from (i).

Alternatively, the linear dynamic panel model is misspecified for $E(Y_{it}|I_{i,t-1})$ if

$$\mathbb{H}_A : \sup_{\beta \in B} \Pr[E(Y_{it}|I_{i,t-1}) = X_{it}'\beta] < 1. \tag{2.3}$$

In this case, we say that model (2.2) suffers from dynamic misspecification and/or neglected nonlinearity in mean. Model misspecification can arise when there exist omitted variables, improper specification of a lag structure, and misspecification of functional form.

Our definition of linearity in mean H_0 differs from the usual notion of a linear dynamic panel model in the literature, which is a weighted sum of current and past shocks $\{\varepsilon_{i,t-j},\ j \geq 0\}$, where $\{\varepsilon_{it}\}$ is serially uncorrelated (e.g., Andrews and Lu 2001, p.144). Such a process may not be linear in mean in the sense of (2.2), because a white noise $\{\varepsilon_{it}\}$ may not be a *m.d.s.* A white noise process with non-zero conditional mean is predictable in mean. Only when $\{\varepsilon_{it}\}$ is a *m.d.s.*, the notion of a linear panel time series coincides with our definition of linearity in mean. The latter concept is more useful for modeling conditional mean dynamics.

We are interested in testing $\mathbb{H}_0$ vs. $\mathbb{H}_A$. This is obviously a chandelling job, because there are infinite number of nonlinear alternatives, and because $I_{i,t-1}$ is possibly infinite-dimensional (dating back to the infinite past), as is the case for non-Markovian processes. There have been plenty of linearity tests in the time series contexts; see Granger and Terasvirta (1993, Ch.6), Hansen (1999), and Hong and Lee (2004) for partial surveys. Most of the linearity tests in time series assume an alternative model, and it is not clear how they can be extended to test nonlinearity in panel contexts, particularly in light of potential inhomogeneity across individuals, where the assumption of a common alternative model is inappropriate. In panel models, testing linearity in mean has not attracted much attention, which should be the first stage in any nonlinear modelling. For example, in the dynamic panel model with unobserved effects, (α_i, λ_t) , the treatment of the initial observations is a difficult theoretical and practical problem (e.g., Wooldridge 2002). One of the reasons is that there are no known transformations that can eliminate the unobserved effects. On the other hand, the linear dynamic panel models with unobserved effects are well developed in the literature, e.g., Baltagi (2001, Ch.8) and Hsiao (2003, Ch.4). Hence, before we build a nonlinear panel model it is advisable to find out if linear models would adequately characterize the data.

In the panel literature, the most important and popular specification test is the test pioneered by Hausman (1978) and Hausman and Taylor (1981). Hausman's test is essentially testing the null that

$$E\left(u_{it}|I_{i,t-1}\right)=0 \ a.s.$$

where $u_{it}=\alpha_i+\lambda_t+\varepsilon_{it}$. Hausman's test is constructed by comparing the generalized least squares (GLS) and within estimators, both of which are consistent under the null $E\left(u_{it}|I_{i,t-1}\right)=0$, but which have different limits (for a large n and a fixed T) if the null is not true. It is clear to see that Hausman's test statistic is model-dependent. For example, Hausman's test will have no power against the alternative when there exists heterokedasticity across individuals since GLS is misspecified and not consistent under both the null and alternative. That is the failure of a variety of specification, e.g., heterokedasticity, neglected nonlinearity, will also cause the usual Hausman's test to have a nonstandard limiting distribution, which means the resulting test will have size distortion (see Wooldridge 2002, p.289 for more discussion). Ahn and Moon (2001) pointed out that the GLS and within estimators are asymptotically equivalent not only under the null, but also under the alternatives if both n and T are large. Ahn and Moon also find that the convergence rate of Hausman's test is sensitive to data generating process (DGP), i.e., again Hausman's test is model dependent especially if both n and T are large. For example, the cross-sectional heterogeneity has a big impact on the performance of Hausman's test. Ahn and Moon suggest the future studies should pay more attention to the cross-sectional heterogeneity. This paper offers answers to this call. In our paper, we are testing

$$E[\varepsilon_{it}|I_{i,t-1}]=0 \ a.s.$$

Hence, unlike Hausman's test and other existing specification tests whose constructions are model-dependent, our tests are generally applicable. Our tests are applicable to test the hypothesis that $E(u_{it}|I_{i,t-1})=0$ and hence can detect a wide range of model misspecifications in mean (including misspecification of functional form and lag order), and are robust to conditional heteroskedasticity and higher order time-varying moments of unknown form and to the inhomogeneity across individuals.

In addition, as discussed earlier, most economic theories have implications on and only on the conditional mean dynamics of the underlying economic variable. The popular method for estimating the dynamic panel models is the generalized method of moments (GMM). Andrews and Lu (2001) recently extend the moment selection criteria of Andrews (1999) to suggest an MMSC using GMM to select the lag length, to detect the number of locations of structural breaks, to determine the exogeneity of regressors, and/or to determine the existence of correlation between the regressors and the individual effect. The MMSC is based on the J-test statistic used for testing over-identifying restrictions. In the context of dynamic panel models in econometrics, such that $u_{it}=\alpha_i+\varepsilon_{it}$, the J-test statistic of MMSC is testing

$$E\left(\triangle\varepsilon_{it}|I_{i,t-1}\right)=0,$$

where $\triangle\varepsilon_{it}=\varepsilon_{it}-\varepsilon_{i,t-1}=u_{it}-u_{i,t-1}$. Note Andrews-Lu's MMSC only works in the GMM context and only when finitely many moment conditions are imposed. On the other hand, our procedures only require the existence of an $\sqrt{nT}$-consistent parameter estimate and allow an infinite set of moment conditions. Again our tests are generally applicable and Andrews-Lu's MMSC can be seen as a special case of our tests.

In a recent development, FLF give several definitions of residual autocorrelations and derive their joint asymptotic distribution for the panel time series model of Hjellvik and Tjostheim (1999). FLF propose a portmanteau test to test the model specification on the lag order. The FLF test statistic is defined as

$$Q(m) = nr\left(\widehat{\theta}\right)' \widehat{V}^{-1} r\left(\widehat{\theta}\right)$$

where Ω is an $m \times m$ matrix with the ith diagonal element $(T-i-1)^{-1}$,

$$\hat{V} = \Omega - (T-1)^{-1}\left(1-\widehat{\theta}^2\right)\hat{M},$$

$$\hat{M} = \begin{pmatrix} 1 & \hat{\theta} & \cdots & \hat{\theta}^{m-1} \\ \hat{\theta} & \hat{\theta}^2 & \cdots & \hat{\theta}^m \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}^{m-1} & \hat{\theta}^m & \cdots & \hat{\theta}^{2m-2} \end{pmatrix},$$

$$r\left(\widehat{\theta}\right) = \left(r_1\left(\widehat{\theta}\right), ..., r_m(\widehat{\theta})\right)',$$

$$r_l\left(\widehat{\theta}\right) = \frac{\sum_{i=1}^n \sum_{t=1}^T \widehat{\varepsilon}_{i,t+l}\widehat{\varepsilon}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \widehat{\varepsilon}_{it}^2},$$

$$\widehat{\theta} = \frac{\sum_{i=1}^n \sum_{t=1}^T \widetilde{Y}_{i,t+1}\widetilde{Y}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \widetilde{Y}_{it}^2},$$

and

$$\widehat{\varepsilon}_{it} = \widetilde{Y}_{it} - \widehat{\theta}\widetilde{Y}_{i,t-1}$$

$l = 1, ..., m$, where

$$\widetilde{Y}_{it} \equiv Y_{it} - \overline{Y}_{\cdot t}.$$

FLF show that the FLF test statistic has an asymptotic χ_m^2 distribution under $\mathbb{H}_0$ and has reasonable size and power from a simulation study. A limit distribution of the FLF test is that it is derived under $i.i.d.$ $\text{N}(0, \sigma^2)$. Also its robust versions under conditional heteroskedasticity of unknown form are not available.

Compared to the existing tests in the panel literature, our tests can detect dynamic misspecification at any unknown lag order and can check generic nonlinearity at each lag, and our tests are robust to heteroskedasticity of unknown form over time and to inhomogeneity across individuals

Meghir and Windmeijer (1999) argue the importance to model the higher order moments, such as ARCH, of the dynamic panel models. For example, the ARCH effects could come from that the fact that persons at different levels of the income distribution face a different variance of their time-income profile when one models the income dynamic and uncertainty. As we have emphasized, it is important to distinguish nonlinearity in different moments, which have different economic implications. More importantly, the rejection of the linear specification according the conventional sense could be due the existence of the ARCH effect in the error term. To our knowledge, our paper is the first work to allow the $m.d.s$ errors $\{\varepsilon_{it}\}$ in the panel literature.

3. Test Statistics

3.1 Generalized Spectral Analysis

Our approach to testing a correct conditional mean specification ($\mathbb{H}_0$ in (2.2)) is based on Hong's (1999) generalized spectrum, which is an analytic tool for nonlinear time series, just as a power spectrum is an analytic tool for linear time series (Priestley 1981).

Recall the model error $\varepsilon_{it} = Y_{it} - X_{it}'\beta - \mu_i - \lambda_t$ has the property that $E[\varepsilon_{it}(\beta_0)|I_{i,t-1}] = 0$ $a.s.$ for some $\theta_0 \in \Theta$. This implies

$$E[\varepsilon_{it}(\theta_0)|I^{\varepsilon}_{i,t-1}] = 0 \quad a.s., \tag{3.1}$$

where $I^{\varepsilon}_{i,t-1} \equiv \{\varepsilon_{i,t-1}(\theta_0), \varepsilon_{i,t-2}(\theta_0), \cdots\}$. Thus, to test $\mathbb{H}_0$, we can check if $E[\varepsilon_{it}(\theta_0)|I^{\varepsilon}_{i,t-1}] = 0$ $a.s.$[1] Still, we have the curse of the dimensionality problem because $I^{\varepsilon}_{i,t-1}$ has an infinite dimension. Fortunately, the generalized spectral approach provides a sensible way to tackle this difficulty.

For notational economy, we put $\varepsilon_{it} \equiv \varepsilon_{it}(\theta)$. Suppose $\{\varepsilon_{it}\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) \equiv E(e^{\mathbf{i}u\varepsilon_{it}})$ and pairwise joint characteristic function $\varphi_j(t,s) \equiv E(e^{\mathbf{i}u\varepsilon_{it}+\mathbf{i}v\varepsilon_{i,t-|j|}})$, where $\mathbf{i} \equiv \sqrt{-1}$, $u, v \in \mathbb{R}$, and $j = 0, \pm1, \cdots$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{\mathbf{i}u\varepsilon_{it}}\}$. It is defined as

$$f_i(\omega, u, v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_{ij}(u,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi, \pi], \tag{3.2}$$

where ω is the frequency, and $\sigma_{ij}(u, v)$ is the covariance function of the transformed series:

$$\sigma_{ij}(u,v) \equiv \text{cov}(e^{\mathbf{i}u\varepsilon_{it}}, e^{\mathbf{i}v\varepsilon_{i,t-|j|}}), \quad j = 0, \pm1, \cdots.$$

The function $f_i(\omega, u, v)$ can capture any type of pairwise serial dependence in $\{\varepsilon_{it}\}$, i.e., dependence between ε_{it} and $\varepsilon_{i,t-j}$ for any nonzero lag j, including that with zero autocorrelation. This is analogous to higher order spectra (Brillinger and Rosenblatt 1966a, 1966b) in the sense that $f_i(\omega, u, v)$ can capture serial dependence in higher order moments. However, unlike higher order spectra, $f_i(\omega, u, v)$ does not require existence of any moment of $\{\varepsilon_{it}\}$. This is important in economics and finance because it has been argued that the higher order moments of many economic/financial time series may not exist.

[1] For a univariate time series, the knowledge of $I_{t-1} \equiv \{Y_{t-1}, Y_{t-2}, \cdots\}$ is equivalent to the knowledge of $I^{\varepsilon}_{t-1} \equiv \{\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}$ under certain regularity conditions. However, when I_{t-1} contains other current and lagged exogeneous variables, these two information sets generally differ. Below, we will first develop tests based on I^{ε}_{t-1}, and then consider extensions to the more general information set in Section 5. (??)

When $\sigma_i^2 \equiv E(\varepsilon_{it}^2)$ exists, we can obtain power spectrum as a derivative of $f_i(\omega,u,v)$:

$$-\frac{\partial^2}{\partial u \partial v} f_i(\omega,u,v)\,|_{(u,v)=(0,0)} = h_i(\omega) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \mathrm{cov}(\varepsilon_{it},\varepsilon_{i,t-|j|})e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi].$$

For this reason, we call $f_i(\omega,u,v)$ the generalized spectrum of $\{\varepsilon_{it}\}$.

As is well known, the interpretation of spectral analysis is much more difficult for nonlinear time series than for linear time series. For example, the bispectrum $B_i(\omega_1,\omega_2)$ has no physical (e.g., energy decomposition over frequencies) interpretation as the power spectrum $h_i(\omega)$. This is also the case for the generalized spectrum $f_i(\omega,u,v)$. However, the basic idea of characterizing cyclical dynamics still applies: $f_i(\omega,u,v)$ has useful interpretations when searching for linear or nonlinear cycles. A strong cyclicity of data may be linked with a strong serial dependence in $\{\varepsilon_{it}\}$, which is not necessarily measurable by autocorrelation. The generalized spectrum $f_i(\omega,u,v)$ can capture such nonlinear cyclical patterns by displaying distinct spectral peaks. For example, suppose an ARCH process has a stochastic cyclical dynamics in volatility. Then power spectrum $h_i(\omega)$ will be flat and miss the volatility cycles. In contrast, $f_i(\omega,u,v)$ can effectively capture such cycles. More generally, the generalized spectrum can capture cyclical dynamics caused by linear and nonlinear dependence. The latter includes dynamics in volatility, skewness, and other higher order conditional moments.[2]

The generalized spectrum $f_i(\omega,u,v)$ itself is not suitable for testing $\mathbb{H}_0$ in (2.2), because it can capture serial dependence in mean and in higher order moments. An example is an ARCH process. The generalized spectrum $f_i(\omega,u,v)$ can capture this process, although it is a *m.d.s.* However, just as the characteristic function can be differentiated to generate various moments of $\{\varepsilon_{it}\}$, $f_i(\omega,u,v)$ can be differentiated to capture serial dependence in various moments. To capture (and only capture) dependence in conditional mean, one can use the derivative

$$f_i^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \sigma_{ij}^{(1,0)}(0,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \tag{3.3}$$

where

$$\sigma_{ij}^{(1,0)}(0,v) \equiv \frac{\partial}{\partial u}\sigma_{ij}(u,v)\,|_{u=0} = \mathrm{cov}(\mathbf{i}\varepsilon_{it}, e^{\mathbf{i}v\varepsilon_{i,t-|j|}}).$$

The measure $\sigma_{ij}^{(1,0)}(0,v)$ checks whether the autoregression function $E(\varepsilon_{it}|\varepsilon_{i,t-j})$ at lag j is zero. Under appropriate conditions, $\sigma_{ij}^{(1,0)}(0,v) = 0$ for all $v \in \mathbb{R}$ if and only if $E(\varepsilon_{it}|\varepsilon_{i,t-j}) = 0$ *a.s.*[3] The autoregression function can capture linear and nonlinear serial dependence in mean, including the processes with zero autocorrelation. Examples are a panel bilinear autoregressive process $\varepsilon_{it} = \alpha z_{i,t-1}\varepsilon_{i,t-2} + z_{it}$ and a panel nonlinear moving-average process $\varepsilon_{it} = \alpha z_{i,t-1}z_{i,t-2} + z_{it}$, where $\{z_{it}\}$ is *i.i.d.*$(0,\sigma^2)$. These

[2] A potentially useful application is the investigation of possible nonlinear business cycles by $f_i(\omega,u,v)$. Power spectrum $h_i(\omega)$, when applied to macroeconomic time series such as the U.S. GDP growth rate, often produces a flat spectrum. However, some nonlinear time series experts (e.g., Tong 1990, p.) believe that business cycles are related to nonlinear cyclical dynamics. It will be interesting to examine whether $f_i(\omega,u,v)$ can capture and identify such nonlinear cycles.

[3] See Bierens (1982) and Stinchcombe and White (1998) for discussion in a different but related context with *i.i.d.* samples.

processes are white noises, but they are not *m.d.s.,* because their conditional means are time-varying. Thus, $E(\varepsilon_{it}|\varepsilon_{i,t-j})$ is a natural tool to test $\mathbb{H}_0$, whereas $\mathrm{cov}(\varepsilon_{i,t}, \varepsilon_{i,t-j})$ will miss such subtle nonlinear processes as those with zero autocorrelation. Nevertheless, $E(\varepsilon_{it}|\varepsilon_{i,t-j})$ has not been as widely used as $\mathrm{cov}(\varepsilon_{it}, \varepsilon_{i,t-j})$. An exception is Hjellvik and Tjøstheim (1996), who consider testing linearity in mean for observed raw data using a kernel estimator for the autoregression function. Teräsvirta *et al.* (1994) and Tong (1990) also discuss smoothed nonparametric estimation of the autoregression function.

Although $E(\varepsilon_{it}|\varepsilon_{i,,t-j})$ and $\sigma_{ij}^{(1,0)}(0,v)$ are equivalent measures, the use of $\sigma_{ij}^{(1,0)}(0,v)$ avoids smoothed nonparametric estimation. The measure $\sup_{v\in\mathbb{R}} |\sigma_{ij}^{(1,0)}(0,v)|$ can be viewed as an operational version of the maximum mean correlation, $\max_{f(\cdot)} |\mathrm{corr}[\varepsilon_{it}, f(\varepsilon_{i,t-j})]|$, which is proposed by Granger and Teräsvirta (1993, p.23) as a measure for nonlinearity in mean. Similarly, the supremum generalized spectral derivative modulus

$$m_i(\omega) \equiv \sup_{v\in(-\infty,\infty)} |f_i^{(0,1,0)}(\omega,0,v)|, \quad \omega \in [-\pi,\pi], \tag{3.4}$$

can be viewed as the maximum dependence in mean at frequency ω. It can be used to search cycles in mean that are caused by linear or nonlinear serial dependence in mean. An example of the latter is the well-known ARCH-in-mean effect (Engle, Lilien and Robins 1987).

The hypothesis of $E(\varepsilon_{it}|I_{i,t-1}^{\varepsilon}) = 0$ *a.s.* is not the same as the hypothesis of $E(\varepsilon_{it}|\varepsilon_{i,t-j}) = 0$ *a.s.* for all $j > 0$. The former implies the latter but not vice versa. This is the price we have to pay for dealing with the difficulty of the "curse of dimensionality". One example that is not *m.d.s.* but has $E(\varepsilon_{it}|\varepsilon_{i,t-j}) = 0$ *a.s.* for all $j > 0$ is a panel nonlinear moving-average process

$$\varepsilon_{it} = \alpha z_{i,t-2} z_{i,t-3} + z_{it}, \quad z_{it} \sim i.i.d.(0,\sigma^2). \tag{3.5}$$

Obviously, there are many such examples.[4]

It seems to be extremely difficult to formally characterize the gap between $E(\varepsilon_{it}|I_{i,t-1}^{\varepsilon}) = 0$ *a.s.* and $E(\varepsilon_{it}|\varepsilon_{i,t-j}) = 0$ *a.s.* for all $j > 0$. However, these two hypotheses coincide under some special cases. The first case is when $\{\varepsilon_{it}\}$ is a stationary Gaussian process, which can have a long memory. The second case is when $\{\varepsilon_{it}\}$ is an Markovian process. This covers both linear and nonlinear Markovian processes. The examples of the latter are a panel nonlinear autoregressive process $\varepsilon_{it} = g(\varepsilon_{i,t-1}) + z_{it}$ and a panel bilinear autoregressive process $\varepsilon_{it} = \alpha\varepsilon_{i,t-1} + \beta z_{it}\varepsilon_{i,t-1} + z_{it}$, where $\{z_{it}\} \sim i.i.d.(0,\sigma^2)$. The third case is when $\{\varepsilon_{it}\}$ follows an additive panel time series process:

$$\varepsilon_{it} = \alpha_{i0} + \sum_{j=1}^{\infty} g_{ij}(\varepsilon_{t-j}) + z_{it},$$

where $g_{ij}(\cdot)$ is not a zero function at least for some $j > 0$.

3.2 Generalized Spectral Derivative Estimation

In the present context, ε_{it} is not observed. Suppose we have a random sample $\{Y_{it}\}_{t=1}^{T}$ which is used to

[4]It is well-known (e.g., Granger and Teräsvirta 1993, Tong 1990, Priestley 1988) that the class of nonlinear moving averages processes is genenrally not invertible, and as a consequence, has found little empirical applications in practice.

estimate model (2.1). For two way panel models, we obtain the "within" type estimated model residual

$$\hat{\varepsilon}_{it} \equiv \hat{u}_{it} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t} + \bar{u}, \tag{3.6}$$

$t = 1, ..., T_i$, $i = 1, ..., n$, where $\hat{u}_{it} \equiv Y_{it} - X_{it}'\hat{\beta}$,

$$\bar{u}_{i\cdot} \equiv T_i^{-1} \sum_{t=1}^{T_i} \hat{u}_{it}, \bar{u}_{\cdot t} \equiv n^{-1} \sum_{i=1}^{n} \hat{u}_{it}, \bar{u} \equiv (nT_i)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \hat{u}_{it},$$

and $\hat{\beta}$ is a consistent estimator for β^0.

Assumption A.2: $(nT)^{1/2}(\hat{\beta} - \beta^*) = O_P(1)$, where $\beta^* = \beta^0$ under H_0.

We only require that $\hat{\beta}$ is $\sqrt{nT}$-consistent for β^0. The estimator $\hat{\beta}$ needs not be asymptotically most efficient. Examples of $\hat{\beta}$ are GMM, within estimator, limited information maximum likelihood (LIML) (see Alvarez and Arellano 2003 for more discussion about the estimation issues when both T and n tend to infinity). We do not require the knowledge of the asymptotic structure of $\hat{\beta}$ because as will be shown below, parameter estimation uncertainty in $\hat{\beta}$ has no impact on the limit distribution of our tests.

We can estimate $f_i^{(0,1,0)}(\omega, 0, v)$ by a smoothed kernel estimator

$$\hat{f}_i^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=1-T_i}^{T_i-1} (1 - |j|/T_i)^{1/2} k(j/p_i) \hat{\sigma}_{ij}^{(1,0)}(0, v) e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi, \pi], \tag{3.7}$$

where $\hat{\sigma}_{ij}^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_{ij}(u, v)|_{u=0}$, $\hat{\sigma}_{ij}(u, v) = \hat{\varphi}_{ij}(u, v) - \hat{\varphi}_{ij}(u, 0)\hat{\varphi}_{ij}(0, v)$, and

$$\hat{\varphi}_{ij}(u, v) = \frac{1}{T_i - |j|} \sum_{t=|j|+1}^{T_i} e^{\mathbf{i}u\hat{\varepsilon}_{it} + \mathbf{i}v\hat{\varepsilon}_{i,t-|j|}}.$$

Here, The factor $(1 - |j|/T)^{1/2}$ is a finite-sample correction. It could be replaced by unity. The parameter $p_i \equiv p_i(T_i)$ is a bandwidth. In our theory, we allow p_i to differ across individuals so as to pick up potential inhomogeneity across individuals. k is a kernel function satisfying the following condition.

Assumption A.3: (i) $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel function that is continuous at zero and all points except a finite number of points on $\mathbb{R}$ such that $k(0) = 1$, $\int_0^\infty |zk(z)|dz \leq C$ and $|k(z)| \leq C|z|^{-b}$ as $z \to \infty$ for some $b > \frac{1}{2}$. (ii) There exists a monotonically decreasing function $\bar{k} : \mathbb{R} \to [-1, 1]$ such that $|k(\cdot)| \leq \bar{k}(\cdot)$ and $\int_0^\infty z\bar{k}(z)dz < \infty$.

Assumption A.3 is a regularity condition on the kernel $k(\cdot)$. It includes all commonly used kernels in practice. The condition of $k(0) = 1$ ensures that the asymptotic bias of the smoothed kernel estimator $\hat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.7) vanishes as $T_i \to \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decays to zero sufficiently fast as $|z| \to \infty$. It implies $\int_0^\infty (1 + z)k^2(z)dz < \infty$. For kernels with bounded support, such as the Bartlett and Parzen kernels, $b = \infty$. For the Daniell and Quadratic-spectral kernels, $b = 1$ and 2, respectively. These two kernels have unbounded support, and thus all $T_i - 1$ lags contained in the sample are used in constructing our test statistics. Under certain conditions, $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$ is consistent for $f_i^{(0,1,0)}(\omega, 0, v)$ (See Theorem 2 below).

Under $\mathbb{H}_0$, the generalized spectral derivative $f_i^{(0,1,0)}(\omega, 0, v)$ becomes a "flat" spectrum:

$$f_{i0}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\sigma_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi], \tag{3.8}$$

which can be consistently estimated by

$$\hat{f}_{i0}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\hat{\sigma}_{i0}^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi]. \tag{3.9}$$

To test $\mathbb{H}_0$, we can compare $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$ with $\hat{f}_{i0}^{(0,1,0)}(\omega, 0, v)$. Any significant difference between them will indicate the rejection of H_0.

3.3 Tests under Conditional Heteroskedasticity

There is a growing consensus among economists that volatilities of most high-frequency economic and financial data are time-varying. It is well-known that the asymptotic variances of test statistics for autocorrelations and conditional mean models depend on the type and degree of heteroskedasticity present. Ignoring it will invalidate the limit distribution of the test statistics (e.g., Diebold 1986, Lo and Mackinlay 1988, Wooldridge 1990, 1991, Whang 1998). This is also true for our tests. In fact, for our tests, it is also important to take into account other higher order time-varying moments. Recent studies (e.g., Hansen 1994, Harvey and Siddique 1999, 2000, Jondeau and Rockinger 2003) find that the conditional skewness and kurtosis of asset returns are time-varying. Below, we will propose tests of testing $\mathbb{H}_0$ that are robust to conditional heteroskedasticity and other time-varying higher order moments of unknown form. This is one of the most important contributions of our paper in terms of empirical relevance and asymptotic analysis. The asymptotic analysis is non-trivial because of the need to take care of serial dependence in higher order moments.

Our test statistic that is robust to conditional heteroskedasticity and other time-varying higher order conditional moments of unknown form is given as follows:

$$\hat{M}_1 \equiv \left[\sum_{i=1}^{n}\sum_{j=1}^{T_i-1} k^2(j/p_i)(T_i - j)\int \left|\hat{\sigma}_{ij}^{(1,0)}(0, v)\right|^2 dW(v) - \sum_{i=1}^{n}\hat{C}_{1i}\right] / \sqrt{\sum_{i=1}^{n}\hat{D}_{1i}}, \tag{3.10}$$

where $W : \mathbb{R} \longrightarrow \mathbb{R}^+$ is a nondecreasing function that weighs set about zero equally,

$$\hat{C}_{1i} = \sum_{j=1}^{T_i-1} k^2(j/p_i)\frac{1}{T_i - j}\sum_{t=j+1}^{T_i-1} \hat{\varepsilon}_i^2 \int \left|\hat{\psi}_{i,t-j}(v)\right|^2 dW(v),$$

$$\hat{D}_{1i} = 2\sum_{j=1}^{T_i-2}\sum_{l=1}^{T_i-2} k^2(j/p_i)k^2(l/p_i)\iint \left| \frac{1}{T_i - \max(j,l)} \sum_{t=\max(j,l)+1}^{T_i} \hat{\varepsilon}_{it}^2 \hat{\psi}_{i,t-j}(v)\hat{\psi}_{i,t-l}(v') \right|^2 dW(v)dW(v'),$$

and $\hat{\psi}_{it}(v) = e^{\mathbf{i}v\hat{\varepsilon}_{it}} - \hat{\varphi}_i(v)$, and $\hat{\varphi}_i(v) = T_i^{-1}\sum_{t=1}^{T_i} e^{\mathbf{i}v\hat{\varepsilon}_t}$. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$.[5] An example of $W(\cdot)$ is the $N(1,0)$ CDF, which is commonly used in the empirical characteristic function literature. The centering and scaling factors $\hat{C}_{1i}$ and $\hat{D}_{1i}$ are approximately the mean and variance of $T_i \int \int_{-\pi}^{\pi} \left| \hat{f}_i^{(0,1,0)}(\omega,0,v) - \hat{f}_{i0}^{(0,1,0)}(\omega,0,v) \right|^2 d\omega dW(v)$. They have taken into account the impact of conditional heteroskedasticity and other time-varying higher order conditional moments. Alternatively, we could also consider the following test statistic

$$\hat{M}_1^+ \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\sum_{j=1}^{T_i-1} k^2(j/p_i)(T_i-j)\int \left|\hat{\sigma}_{ij}^{(1,0)}(0,v)\right|^2 dW(v) - \hat{C}_{1i}\right] / \sqrt{\hat{D}_{1i}}$$

Intuitively, $\hat{M}_1^+$ can be viewed as a heteroskedasticity-corrected test while $\hat{M}_1$ is a heteroskedasticity-consistent test, where heteroskedasticity arises from different variances $\hat{D}_{1i}$ and bandwidth parameters p_i. In $\hat{M}_1^+$, these two forms of heteroskedasticity are corrected first for each i. As is shown below, $\hat{M}_1$ and $\hat{M}_1^+$ are asymptotically $\mathrm{N}(0,1)$ under $\mathbb{H}_0$, but their power properties generally differ. Both $\hat{M}_1$ and $\hat{M}_1^+$ apply to one-way or two-way error component models. For one-way component models, however, one can use $\hat{\varepsilon}_{it} \equiv \hat{u}_{it} - \bar{u}_{i\cdot}$ if one knows $\lambda_t = 0$, and use $\hat{\varepsilon}_{it} \equiv \hat{u}_{it} - \bar{u}_{\cdot t}$ if one knows $\mu_i = 0$. To test the null hypothesis of Hausman's test, we use $\hat{\varepsilon}_{it} = \hat{u}_{it} - \bar{u}_{\cdot t}$. The limit distribution of the test statistics is unchanged.

3.4 Tests under Conditional Homoskedasticity

To examine why it is important to take into account conditional heteroskedasticity and higher order time-varying moments in testing $\mathbb{H}_0$, we now derive generalized spectral tests for $\mathbb{H}_0$ under conditional homoskedasticity and under *i.i.d.* for $\{\varepsilon_{it}\}$ respectively. Suppose $\{\varepsilon_{it}\}$ is conditionally homoskedastic (i.e., $E(\varepsilon_{it}^2|I_{i,t-1}) = \sigma_i^2 \ a.s.$). Then we can simplify our test statistics as follows:

$$\hat{M}_2 \equiv \left[\sum_{i=1}^{n}\sum_{j=1}^{T_i-1} k^2(j/p_i)(T_i-j)\int \left|\hat{\sigma}_{ij}^{(1,0)}(0,v)\right|^2 dW(v) - \sum_{i=1}^{n}\hat{C}_{2i}\right] / \sqrt{\sum_{i=1}^{n}\hat{D}_{2i}}$$

where

$$\hat{C}_{2i} = \hat{s}_i^2 \int |\hat{\sigma}_{i0}(v,-v)|^2 dW(v) \sum_{j=1}^{T_i-1} k^2(j/p_i),$$

$$\hat{D}_{2i} = 2\hat{s}_i^4 \sum_{j=1}^{T_i-2}\sum_{l=1}^{T_i-2} k^2(j/p_i)k^2(l/p_i)\iint \left|\hat{\sigma}_{i,j-l}(v,v')\right|^2 dW(v)dW(v'),$$

[5] To compute $\hat{M}_1(p)$, there is no need to integrate frequency ω but numerical integration over transform parameter v is needed. The latter can be implemented by using Gauss-Legendre quadrature, which is available in most statistical software. Alternatively, it can be approximated to any desired accuracy by simulation.

and $\hat{s}_i^2 \equiv T_i^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2$ is the sample variance of $\{\varepsilon_{it}\}_{t=1}^{T}$. Both centering and scale factors $\hat{C}_{2i}$ and $\hat{D}_{2i}$ have been simplified, by exploiting the implication of conditional homoskedasticity. The form of $\hat{D}_{2i}$ still takes into account the impact of possibly time-varying third order and higher order moments.

3.5 Tests under the i.i.d. Case

In the panel literature, $\{\varepsilon_{it}\}$ is generally assumed to be an i.i.d.$(0, \sigma^2)$ sequence. When $\{\varepsilon_{it}\}$ is $i.i.d.(0, \sigma_i^2)$, which implies $\mathbb{H}_0$, our test statistic can be further simplified:

$$\hat{M}_3 \equiv \left[\sum_{i=1}^{n} \sum_{j=1}^{T_i - 1} k^2(j/p_i)(T_i - j) \int \left| \hat{\sigma}_{ij}^{(1,0)}(0, v) \right|^2 dW(v) - \sum_{i=1}^{n} \hat{C}_{3i} \right] / \sqrt{\sum_{i=1}^{n} \hat{D}_{3i}}$$

where $\hat{C}_{3i} = \hat{C}_{2i}$ and $\hat{D}_{3i} = 2\hat{s}_i^4 \iint |\hat{\sigma}_{i0}(v, v')|^2 \, dW(v) dW(v') \sum_{j=1}^{T-2} k^4(j/p_i)$.

The scale factor $\hat{D}_{3i}$ has been greatly simplified. Interestingly, the $\hat{M}_2$ test derived under conditional homoskedasticity differs from the $\hat{M}_3$ test derived under *i.i.d.* This is because $\hat{M}_2$ still takes into account possibly time-varying higher order moments (e.g., skewness and kurtosis).

4. Asymptotic Distribution

To derive the null limit distribution of our tests, we provide some regularity conditions:

Assumption A.4: (i) $\{\varepsilon_{it}\}$ is spatially independent across different individuals i. (ii) There exists a sequence of $\{\varepsilon_{q,it}\}$ such that $E(\varepsilon_{q,it} | \mathcal{F}_{i,t-1}) = 0$ and as $q \to \infty$, $\varepsilon_{q,it}$ is independent of $\mathcal{F}_{i,t-q-1}$, and $\max_{1 \le i \le n} E(\varepsilon_{q,it}^4) \le C$.

Assumption A.5: $W : \mathbb{R} \to R^+$ is a weighting function with $\int v^4 dW(v) \le C$ and $W(\cdot)$ weighs set about zero equally.

Assumption A.4 will be required only under $\mathbb{H}_0$. Part (i) implies that $\{\varepsilon_{it}\}$ and $\{\varepsilon_{js}\}$ are mutually independent for all t, s whenever $i \ne j$. It assumes that the *m.d.s.* $\{\varepsilon_{it}\}$ can be approximated by a q dependent *m.d.s.* process $\{\varepsilon_{it}\}$ arbitrarily well when q is sufficiently large. Because $\{\varepsilon_{it}\}$ is a *m.d.s.*, Assumption A.4 essentially imposes restrictions on serial dependence in higher order moments of ε_{it}. Among other things, it implies ergodicity for $\{\varepsilon_{it}\}$. It holds trivially when $\{\varepsilon_{it}\}$ is a $q-$dependent process with an arbitrarily large but finite order q. It also covers many non-Markovian processes. To appreciate this, we consider as an example an infinite order conditionally heteroskedastic error process $\{\varepsilon_{it}\}$:

$$\begin{cases} \varepsilon_{it} = h_{it}^{1/2} z_{it}, \\ h_{it} = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j z_{i,t-j}^2, \\ \{z_{it}\} \sim i.i.d.(0,1), \ \alpha_0 > 0, \ \alpha_j \ge 0 \ \ \forall j > 0. \end{cases} \tag{4.1}$$

Define $\varepsilon_{iq,t} \equiv h_{iq,t}^{1/2} z_{it}$, where $h_{iq,t} \equiv \alpha_0 + \sum_{j=1}^{q} \alpha_j z_{i,t-j}^2$. Then we have

$$\begin{aligned} E(\varepsilon_{it} - \varepsilon_{iq,t})^2 &= E\left[\sqrt{\alpha_0 + \sum_{j=1}^{\infty} \alpha_j z_{i,t-j}^2} - \sqrt{\alpha_0 + \sum_{j=1}^{q} \alpha_j z_{i,t-j}^2}\right]^2 \\ &\leqslant \frac{1}{\alpha_0} E\left(\sum_{j=q+1}^{\infty} \alpha_j z_{i,t-j}^2\right)^2 \leqslant C\left(\sum_{j=q+1}^{\infty} \alpha_j\right)^2, \end{aligned}$$

provided $E(z_{it}^4) < \infty$. Thus, Assumption A.2 holds if $\sum_{j=q+1}^{\infty} \alpha_j \leqslant Cq^{-\kappa}$. A sufficient condition is $\alpha_j \leq Cj^{-\kappa-1}$ for $j \to \infty$, which rules out long-memory volatility processes (i.e., the processes with $\text{cov}(\varepsilon_{it}^2, \varepsilon_{i,t-j}^2)$ not summable over j).

We now state the asymptotic distribution of the $\hat{M}_c, c = 1, 2, 3$, tests under $\mathbb{H}_0$.

Theorem 1: *Suppose Assumptions A.1–A.5 hold, and $p = cT^{\lambda}$ (needs to be changed) for $\lambda \in (0, (2b-1)/(4b-1))$ and $c \in (0, \infty)$. (i) $\hat{M}_1 \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$, (ii) If in addition $E\left(\varepsilon_{it}^2 | I_{i,t-1}\right) = \sigma_i^2$ a.s., then $\hat{M}_2 \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$. (iii) If $\{\varepsilon_{it}\}$ is i.i.d.$(0, \sigma_i^2)$, then $\hat{M}_3 \xrightarrow{d} N(0,1)$.*

As an important feature of $\hat{M}_c(p)$ the use of *estimated* model residuals $\{\hat{\varepsilon}_{it}\}$ in place of true unobservable errors $\{\varepsilon_{it}\}$ has no impact on the limit distribution of $\hat{M}_c(p)$. One can proceed as if the true parameter value θ_0 were known and equal to $\hat{\theta}$. The reason is that the convergence rate of parametric parameter estimator $\hat{\theta}$ to θ_0 is faster than that of nonparametric kernel estimator $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$ to $f_i^{(0,1,0)}(\omega, 0, v)$. Consequently, the limit distribution of $\hat{M}_c(p)$ is solely determined by $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$, and replacing θ_0 by $\hat{\theta}$ has no impact asymptotically. This delivers a convenient procedure, because no specific estimation method for θ_0 is required. Of course, parameter estimation uncertainty in $\hat{\theta}$ may have impact on the small sample distribution of $\hat{M}_c(p)$. In small samples, one can use a bootstrap procedure similar to Hansen (1996) to obtain more accurate size of the tests.

Because parameter estimation uncertainty in $\hat{\theta}$ has no impact on the limit distribution of $\hat{M}_c(p)$, $\hat{M}_c(p)$ can be readily used to test the *m.d.s.* hypothesis for observed raw data with conditional heteroskedasticity of unknown form. No modification to the test statistic $\hat{M}_1$ or its limit distribution is needed.

A very important feature of our tests is that they are asymptotically N(0,1) even if the lag orders $\{p_i\}$ do not grow with the sample sizes $\{T_i\}$. This is expected to enhance the size performance of our tests, as is confirmed in our simulation studies. Indeed, our simulation shows that sizes of $\hat{M}_c(p)$ are close to the nominal sizes even the sample is as small as $(n, t) = (25, 25)$.

5. Asymptotic Power

Our tests are derived without assuming an alternative model. To gain insight into the nature of the alternatives that our tests are able to detect, we now examine the asymptotic behavior of $\hat{M}_c$ under $\mathbb{H}_A$ in (2.3). For this purpose, we impose an additional condition on $k(\cdot)$ and a condition on serial dependence in $\{\varepsilon_{it}\}$.

Assumption A.6: There exists some $q \in (0, \infty)$ such that $k_q \lim_{z \to 0}[1 - k(z)]/|z|^q \in (0, \infty)$.

Assumption A.7: $\max_{1\le i\le n}\sum_{j=1}^{\infty} j^{\max(1,q)}[\sup_{v\in\mathbb{R}}|\sigma_{ij}^{(1,0)}(0,v)|]^2 \le C$.

Theorem 2: *Suppose Assumptions A.1 and A.3–A.7 hold, and* $p = cT^{\lambda}$ (needs to be changed) *for* $\lambda \in (0, \frac{1}{2})$ *and* $c \in (0, \infty)$. *Then as for* $c = 1, 2, 3$,

$$(p^{\frac{1}{2}}/nT_i)\hat{M}_c(p) \xrightarrow{p} \left[2D\int_0^{\infty} k^4(z)dz\right]^{-\frac{1}{2}} \pi \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\int\int_{-\pi}^{\pi}\left|f_i^{(0,1,0)}(\omega,0,v) - f_{i0}^{(0,1,0)}(\omega,0,v)\right|^2 d\omega dW(v)$$

$$= \left[2D\int_0^{\infty} k^4(z)dz\right]^{-\frac{1}{2}} \lim_n n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{\infty}\int\left|\sigma_j^{(1,0)}(0,v)\right|^2 dW(v)$$

where

$$\begin{aligned} D &\equiv \lim_n n^{-1}\sum_{i=1}^{n}\sigma_i^4\sum_{j=-\infty}^{\infty}\int\int|\sigma_{ij}(u,v)|^2 dW(u)dW(v) \\ &= 2\pi\lim_n\sum_{i=1}^{n}\sigma_i^4\int\int\int_{-\pi}^{\pi}|f_i(\omega,u,v)|^2 d\omega dW(u)dW(v). \end{aligned}$$

The constant D takes into account serial dependence in conditioning variables $\{e^{\mathbf{i}v\varepsilon_{i,t-j}}, j > 0)\}$, which generally exists even under $\mathbb{H}_0$, due to the presence of serial dependence in the conditional variance and higher order moments of $\{\varepsilon_{it}\}$. This differs from the *i.i.d.* case, where

$$D = \lim_n n^{-1}\sum_{i=1}^{n}\sigma_i^4\int\int|\sigma_{i0}(v,v')|^2 dW(v)dW(v')$$

depends only on the marginal distribution of ε_{it}.

Suppose the autoregression function $E(\varepsilon_{it}|\varepsilon_{i,t-j}) \ne 0$ at some lag $j > 0$. Then we have

$$\int|\sigma_{ij}^{(1,0)}(0,v)|^2 dW(v) > 0$$

for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. As a consequence, $\lim_{T\to\infty}\mathrm{P}[\hat{M}_c > C(T)] = 1$ for any constant $C(T) = o(T/p^{\frac{1}{2}})$(needs to be changed). Therefore, $\hat{M}_c$ has asymptotic unit power at any given significance level, whenever $E(\varepsilon_{it}|\varepsilon_{i,t-j})$ is nonzero at some lag $j > 0$.[6] We thus expect that $\hat{M}_c$ has relatively omnibus power against a wide variety of linear and nonlinear alternatives with unknown lag structure, as is confirmed in our simulation below. It should be emphasized that the omnibus power property does not mean that $\hat{M}_c$ is more powerful than *any* other existing tests against *every* alternative. In fact, just because $\hat{M}_c$ has to take care of a wide range of possible alternatives, it may be less powerful

[6] Since $\int\int_{-\pi}^{\pi}\left|fLi^{(0,1,0)}(\omega,0,v) - f_{i0}^{(0,1,0)}(\omega,0,v)\right|^2 d\omega dW(v)$ is strictly positive whenever $E(\varepsilon_{it}|\varepsilon_{i,t-j}) \ne 0$ for some lag $j > 0$, upper-tailed asymptotic critical values (e.g., 1.645 at the 5% level) should be used.

against certain specific alternatives than a parametric test. Nevertheless, the main advantage of $\hat{M}_c$ is that it can eventually detect all possible model misspecifications that render $E(\varepsilon_{it}|\varepsilon_{i,t-j})$ nonzero at some lag $j > 0$. This avoids the blindness of searching for different alternatives when one has no prior information.

Because existing tests for linearity in mean only consider a fixed order lag, they can easily miss misspecifications at higher lag orders. Of course, these tests could be used to check a large number of lags when a large sample is available. However, they are not expected to be powerful when the number of lags is too large, due to the loss of a large number of degrees of freedom. This power loss is greatly alleviated for our tests due to the role played by $k^2(\cdot)$. Most non-uniform kernels discount higher order lags. This enhances good power against the alternatives whose serial dependence decays to zero as lag order j increases. Thus, our tests can check a large number of lags without losing too many degrees of freedom. This feature is not available for popular χ^2-type tests with a large number of lags, which essentially give equal weighting to each lag. Equal weighting is not fully efficient when a large number of lags is considered.

Once the linear dynamic model is rejected by our omnibus test $\hat{M}_c$, one may like to go further to explore possible sources of model misspecification in mean. For this purpose, we can further differentiate the generalized spectral derivative $f_i^{(0,1,0)}(\omega, 0, v)$ with respect to v and construct corresponding tests in a similar spirit to our $\hat{M}_c$ tests. In particular, the derivatives $\sigma_{ij}^{(1,l)}(0,0)$ with $l = 1, 2, 3, 4$ yield $\text{cov}(\varepsilon_{it}, \varepsilon_{i,t-j})$, $\text{cov}(\varepsilon_{it}, \varepsilon_{i,t-j}^2)$, $\text{cov}(\varepsilon_{it}, \varepsilon_{i,t-j}^3)$ and $\text{cov}(\varepsilon_{it}, \varepsilon_{i,t-j}^4)$ respectively. Tests based on these derivatives can thus tell us whether there exists linear correlation, ARCH-in-mean, skewness-in-mean or kurtosis-in-mean effects respectively. ARCH-in-mean effects are important in finance (Engle, Lilien and Robins 1987), and the recent literature also finds time-varying skewness and kurtosis and their economic relevance in finance (Harvey and Siddique 1999, 2000).

6. Data-Driven Lag Order

A practical issue in implementing our tests is the choice of lag order $\{p_i\}$. As an advantage, our smoothing generalized spectral approach can provide a data-driven method to choose p_i, which, to some extent, let data themselves speak for a proper p_i. Theorems 1 and 2 allow use of different lag orders for different individuals, which may enhance power of our tests when substantial degrees in neglected nonlinearity and dynamic misspecification are present. However, substantially different data-driven lag orders may distort the size of the tests in finite samples. To control the size of our tests, we consider a data-driven method for a common lag order p in this section. Such a common lag order will converge to zero under H_0 but automatically grows to infinity under H_A, thus ensuring the power of the tests.

Before discussing any specific method, we first justify the use of a data-driven lag order, $\hat{p}$, say. For this, we impose a Lipschitz continuity condition on $k(\cdot)$.

Assumption A.8: *For any $x, y \in \mathbb{R}$, $|k(x) - k(y)| \leq C|x - y|$ for some constant $C < \infty$.*

This condition rules out the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function, but it still includes most commonly used non-uniform kernels.

Theorem 3: *Suppose Assumptions A.1–A.6 and A.8 hold, and $\hat{p}$ is a data-driven bandwidth such that $\hat{p}/p = 1 + O_P(p^{-(\frac{3}{2}\beta-1)})$ for some $\beta > (2q - \frac{1}{2})/(2q-1)$, where q is as in Assumption A.6, and p is a nonstochastic bandwidth with $p = cT^{\lambda}$ for $\lambda \in (0, (2q-1)/(4q-1))$ and $c \in (0, \infty)$. Let $\hat{M}_c(p)$ be* defined as $\hat{M}_c$ where p is used for all individuals i. *Then (i) $\hat{M}_1(\hat{p}) - \hat{M}_1(p) \xrightarrow{p} 0$ and $\hat{M}_1(\hat{p}) \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$. (ii) If in addition $E(\varepsilon_{it}^2|I_{i,t-1}) = \sigma^2$ a.s., then $\hat{M}_2(\hat{p}) - \hat{M}_2(p) \xrightarrow{p} 0$ and $\hat{M}_2(\hat{p}) \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$. (iii) If $\{\varepsilon_{it}\}$ is i.i.d.$(0,\sigma^2)$, then $\hat{M}_3(\hat{p}) - \hat{M}_3(p) \xrightarrow{p} 0$ and $\hat{M}_3(\hat{p}) \xrightarrow{d} N(0,1)$.*

Thus, as long as $\hat{p}$ converges to p sufficiently fast, the use of $\hat{p}$ instead of p has no impact on the limit distribution of $\hat{M}_c(\hat{p})$. This is an additional "nuisance parameter-free" property.

Theorem 3 allows for a wide range of admissible rates for $\hat{p}$. One plausible choice of $\hat{p}$ is the nonparametric plug-in method similar to Hong (1999, Thm. 2.2). It minimizes an average asymptotic integrated mean squared error (IMSE) criterion for the estimator $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$ in (3.7). Consider some "pilot" generalized spectral derivative estimators based on a preliminary bandwidth $\bar{p}$:

$$\bar{f}_i^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=1-T_i}^{T_i-1} (1 - |j|/T_i)^{\frac{1}{2}} \bar{k}_0(j/\bar{p}) \hat{\sigma}_{ij}^{(1,0)}(0, v) e^{-\mathbf{i}j\omega}, \tag{6.1}$$

$$\bar{f}_i^{(q,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=1-T_i}^{T_i-1} (1 - |j|/T_i)^{\frac{1}{2}} \bar{k}_0(j/\bar{p}) \hat{\sigma}_{ij}^{(1,0)}(0, v) |j|^q e^{-\mathbf{i}j\omega}, \tag{6.2}$$

where the kernel $k_0(\cdot)$ needs not be the same as the kernel $k(\cdot)$ used in (3.7). For example, $k_0(\cdot)$ can be the Bartlett kernel while $k(\cdot)$ is the Daniell kernel. Note that $\bar{f}_i^{(0,1,0)}(\omega, 0, v)$ is an estimator for $f_i^{(0,1,0)}(\omega, 0, v)$ and $\bar{f}_i^{(q,1,0)}(\omega, 0, v)$ is an estimator for the generalized spectral derivative

$$f_i^{(q,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_{ij}^{(1,0)}(0, v) |j|^q e^{-\mathbf{i}j\omega}. \tag{6.3}$$

Then we define the plug-in bandwidth

$$\hat{p}_0 \equiv \hat{c}_0 T^{\frac{1}{2q+1}}, \tag{6.4}$$

where the tuning parameter estimator

$$\begin{aligned}
\hat{c}_0 &\equiv \left[\frac{2qk_q^2}{\int_{-\infty}^{\infty} k^2(z)dz} \frac{\sum_{i=1}^n \int \int_{-\pi}^{\pi} |\bar{f}_i^{(q,1,0)}(\omega, 0, v)|^2 d\omega dW(v)}{\sum_{i=1}^n \int_{-\pi}^{\pi} [\int \bar{f}_i^{(0,1,0)}(\omega, v, -v) dW(v)]^2 d\omega} \right]^{\frac{1}{2q+1}} \\
&= \left[\frac{2qk_q^2}{\int_{-\infty}^{\infty} k^2(z)dz} \frac{\sum_{i=1}^n \sum_{j=1-T}^{T-1} (T_i - |j|) \bar{k}_0^2(j/\bar{p}) |j|^{2q} \int |\hat{\sigma}_{ij}^{(1,0)}(0, v)|^2 dW(v)}{\sum_{i=1}^n \sum_{j=1-T}^{T-1} (T_i - |j|) \bar{k}_0^2(j/\bar{p}) \hat{R}_i(j) \int \hat{\sigma}_{ij}(v, -v) dW(v)} \right]^{\frac{1}{2q+1}},
\end{aligned}$$

and $\hat{R}_i(j) \equiv (T_i - |j|)^{-1} \sum_{t=|j|+1}^{T_i} \hat{\varepsilon}_{it} \hat{\varepsilon}_{i,t-|j|}$ and k_q is given in Assumption A.6. Note that $\hat{p}_0$ is real-valued. One can take its integer part, and the impact of integer-clipping is expected to be negligible.

The data-driven $\hat{p}_0$ in (6.4) involves the choice of a preliminary bandwidth $\bar{p}$, which can be fixed or grow with sample size T. If $\bar{p}$ is fixed, $\hat{p}_0$ still generally grows at rate $T^{\frac{1}{2q+1}}$ under $\mathbb{H}_A$, but $\hat{c}_0$ does not converge to the optimal tuning constant c_0 (say) that minimizes the average IMSE of $\hat{f}_i^{(0,1,0)}(\omega, 0, v)$

over all i. This is a parametric plug-in method. Alternatively, following Hong (1999), we can show that when $\bar{p}$ grows with T properly, the data-driven bandwidth $\hat{p}_0$ in (6.4) will minimize an asymptotic IMSE of $\hat{f}_i^{(0,1,0)}(\omega,0,v)$. The choice of $\bar{p}$ is somewhat arbitrary, but we expect that it is of secondary importance. This is confirmed in our simulation below.[7]

We emphasize that from a theoretical point of view, the data-driven $\hat{p}$ based on the IMSE criterion generally will not maximize the power of $\hat{M}_c$. A more sensible alternative would be to develop a data-driven $\hat{p}$ using a power criterion, or a criterion that trades off size distortion and power loss. This will necessitate higher order asymptotic analysis and is beyond the scope of this paper. We are content with the IMSE criterion here. Our simulation suggests that the power of our tests seems to be relatively flat in the neighborhood of the optimal lag order that maximizes the power, and $\hat{p}_0$ in (6.4) performs reasonably well in finite samples. Nevertheless, the issue of the optimal data-driven $\hat{p}$ for our tests is far from being resolved from a theoretical perspective.

7. Simulations

We now investigate the finite sample performance of $\hat{M}_c(\hat{p}_0), c = 1,2,3$, tests. While our tests can be used to test nonlinear conditional mean models, we focus on testing linearity in mean, which is still the main choice for most researchers in the panel literature. For example, in Hsiao (2003) only two chapters out of eleven devoted to nonlinear models. Even for these two chapters (there are limited dependent panel models) the true latent structures are still linear. Because our tests are derived without specifying an alternative, we will compare them with a linearity test of FLF portmanteau test.

7.1 Simulation Design

7.1.1 Size

To examine the size of the tests under $\mathbb{H}_0$, we consider the following DGPs

$$\text{DGP S.1 [AR(1)-}i.i.d.\text{(0,1)]:} \qquad Y_{it} = 0.3Y_{i,t-1} + \eta_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim i.i.d.\ N(0,1),$$

$$\text{DGP S.2 [AR(1)-ARCH(1)]:} \qquad \begin{cases} Y_{it} = 0.3Y_{i,t-1} + \eta_t + \varepsilon_{it}, \\ \varepsilon_{it} = h_{it}^{1/2} z_{it},\ h_{it} = 0.43 + 0.57\varepsilon_{i,t-1}^2, \\ z_{it} \sim i.i.d.\ N(0,1). \end{cases}$$

where η_t is generated from some fixed constant by following FLF. Under DGPs, the dynamic panel linear model

$$E(Y_{it}|I_{i,t-1}) = \beta^0 Y_{i,t-1} \tag{7.1}$$

is correctly specified for $E(Y_{it}|I_{i,t-1}) = \beta^0 Y_{i,t-1}$. The parameter $\beta^0 \equiv 0.3$ can be estimated consistently by an consistent estimator $\hat{\beta}$. The model error $\{\varepsilon_{it}(\beta^0)\}$ is conditionally homoskedastic under DGP S.1; all tests considered are asymptotically valid under this DGP. Under DGP S.2, $\{\varepsilon_{it}(\beta^0)\}$ is conditionally heteroskedastic; only $\hat{M}_1(\hat{p}_0)$ has a valid limit distribution. This allows us to examine the importance of

[7] The tuning parameter estimator $\hat{c}_0$ will converge to zero under $\mathbb{H}_0$. To ensure $\hat{p}_0 \to \infty$, we can use formula (6.4) supplemented with a slow-growing lower bound (say $\ln T$) such that $\tilde{p}_0 = \max(\ln T, \hat{p}_0)$. The choice of the slow-growing lower bound $\ln T$ is arbitrary, but it will not affect the IMSE-optimal rate $\hat{p}_0$ under $\mathbb{H}_A$, when T is sufficiently large.

taking into account conditional heteroskedasticity. We have chosen parameter values in DGP S.2 such that $E[\varepsilon_{it}^4(\beta^0)] < \infty$.[8] To examine the size performance, we consider nine sample sizes: $(n,T) = (25,25)$, $(25,50)$, $(25,100)$, $(50,25)$, $(50,50)$, $(50,100)$, $(100,25)$, $(100,50)$, $(100,100)$. These cover various combinations of relative sizes of n and T.

7.1.2 Power

Next, we examine the power of the tests for neglected nonlinearity or dynamic misspecification in mean. Because FLF is not valid under conditional heteroskedasticity, we will focus on homoskedastic errors for power comparison. We consider the following DGPs:

DGP P.1 [AR(2)]:	$Y_{it} = 0.3Y_{i,t-1} + 0.05Y_{i,t-2} + \eta_t + \varepsilon_{it}$,
DGP P.2 [ARMA(1,1)]:	$Y_{it} = 0.3Y_{i,t-1} + \eta_t + \varepsilon_{it}$, $\varepsilon_{it} = \rho_i\varepsilon_{i,t-1} + v_{it}$, and $\rho_i \sim U[-0.3, 0.3]$.
DGP P.3 [STAR(1)]:	$Y_{it} = 0.5Y_{i,t-1} - (4 + 0.4Y_{i,t-1})G(-2Y_{i,t-1}) + \eta_t + \varepsilon_{it}$, where $G(z) = [1+\exp(-z)]^{-1}$,
DGP P.4 [SETAR(1)]:	$Y_{it} = \begin{cases} 0.5Y_{i,t-1} + \eta_t + \varepsilon_{it} & \text{if } Y_{i,t-1} \leq 0, \\ -0.5Y_{i,t-1} + \eta_t + \varepsilon_{it} & \text{if } Y_{i,t-1} > 0, \end{cases}$

where $\{\varepsilon_{it}\}$ is $i.i.d.$N(0,1).

DGP P.1, AR(2), is taken from FLF. DGP P.2, a panel random-coefficient model, has been used extensively in literature (see Chapter 6 in Hsiao 2002 for a survey). DGP P.3 is a panel STAR model. Paap et al. (2003) consider a similar panel STAR model to examine the common business cycle in unemployment and the importance of sector-specific variations. DGPs P.3, panel SETAR model was suggested by Stenseth et al (1999) and Yao et al. (2000) to model a panel of ecological time series.

We choose sample size: $(n,T) = (25,25)$, $(50,50)$ respectively for the power comparison.

7.2 Computation of Test Statistics.

To compute FLF portmanteau test one has to determine how many lags, m, to be used in the test vector

$$r\left(\widehat{\theta}\right) = \left(r_1\left(\widehat{\theta}\right), ..., r_m(\widehat{\theta})\right)'.$$

We choose $m = 2, 4$, and 6. FLF test statistic is computed as follows: (i) regress $\widetilde{Y}_{it}$ on 1 and $\widetilde{Y}_{i,t-1}$ and save the estimated residuals $\{\hat{\varepsilon}_{it}\}$; (ii) let Ω is an $m \times m$ matrix with the ith diagonal element $(T-i-1)^{-1}$; (iii) regress $\hat{\varepsilon}_{it}$ on $(\hat{\varepsilon}_{it}, \hat{\varepsilon}_{i,t-1}, \cdots, \hat{\varepsilon}_{i,t-l})$, $l = 1, ..., m$, and save the estimates $r\left(\widehat{\theta}\right)$; (iv) compute FLF statistic

$$Q(m) = nr\left(\widehat{\theta}\right)' \widehat{V}^{-1} r\left(\widehat{\theta}\right),$$

where

$$\hat{V} = \Omega - (T-1)^{-1}\left(1 - \widehat{\theta}^2\right)\hat{M},$$

[8] We also consider a GARCH process with an infinite fourth order moment. The size performance of the generalized spectreal derivative tests is similar.

$$\hat{M} = \begin{pmatrix} 1 & \hat{\theta} & \cdots & \hat{\theta}^{m-1} \\ \hat{\theta} & \hat{\theta}^2 & \cdots & \hat{\theta}^{m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}^{m-1} & \hat{\theta}^{m} & \cdots & \hat{\theta}^{2m-2} \end{pmatrix}$$

and compare it to the $N(0,1)$ distribution.

To compute $\hat{M}_c(\hat{p}_0)$ we use the $N(1,0)$ CDF truncated on $[-3,3]$ for the weighting function $W(\cdot)$. We also use the Bartlett kernel $k_B(z) = (1-|z|)\mathbf{1}(|z| \leq 1)$ for $k(\cdot)$, which has bounded support and is computationally efficient. Our simulation suggests that the choices of $W(\cdot)$ and $k(\cdot)$ have little impact on both the size and power of our tests.[9] We choose a data-driven $\hat{p}_0$ via the plug-in method in (6.4), with the Bartlett kernel for $\bar{k}(\cdot)$ used in the preliminary generalized spectral derivative estimators in (6.1) and (6.2). To examine the impact of the choice of preliminary bandwidth $\bar{p}$, we consider $\bar{p} = 1, 2, 3, \cdots, 10$.

7.3 Monte Carlo Results

Table 1 reports the empirical sizes of the tests under DGP S.1 (homoskedastic errors) at the 10% and 5% levels, using asymptotic theory. When $n > T$, $\hat{M}_1(\hat{p}_0)$ overrejects $\mathbb{H}_0$, while $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$ underreject $\mathbb{H}_0$ but not excessively.

When $n = T$ and $n < T$, FLF and $\hat{M}_c(\hat{p}_0)$ all perform well under $\mathbb{H}_0$, and they do even better when $n = T$ increases.

Over all the tests $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$ derived under conditional homoskedasticity and *i.i.d.* respectively have better sizes than $\hat{M}_1(\hat{p}_0)$. There is some tendency that a larger preliminary lag order $\bar{p}$ gives a better size for $\hat{M}_1(p)$ test.

Under DGP S.2 (ARCH errors) in Table 2, $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$ display very strong overrejection, as is expected. In contrast, $\hat{M}_1(\hat{p}_0)$ performs very well for all cases also as expected.

Next, we turn to the size of FLF test which is valid test under. Under DGP S.1, the FLF test has better levels than the $\hat{M}_c(\hat{p}_0)$ tests, though not in every case. Under DGP S.2, FLF shows mild overrejections.

Table 3 reports the size-corrected power at the 10% and 5% levels under DGPs P.1–P.4. The empirical critical values are obtained under DGP S.1. We compare $\hat{M}_c(\hat{p}_0)$ with FLF test, $Q(m)$, which is derived under conditional homoskedasticity. Under DGP P.1, a panel AR2, FLF test is more powerful than $\hat{M}_c(\hat{p}_0)$. Under DGP P.2, a panel random coefficient ARMA (1,1), FLF test fails to detect the model. In contrast, $\hat{M}_c(\hat{p}_0)$ are very powerful against the random coefficient model, indicating that generalized spectral tests, $\hat{M}_c(\hat{p}_0)$, are rather effective in capturing inhomogeneous serial correlations across individuals. The power of $Q(m)$ is relatively sensitive to the choice of m. On the other hand, the power of $\hat{M}_c(\hat{p}_0)$ is robust to the choice of preliminary lag order $\bar{p}$.

Under DGP P.3, a SETAR, again FLF test has no power for all cases. All $\hat{M}_c(\hat{p}_0)$ tests are equally powerful. Under DGP P.4 (TAR), $\hat{M}_c(\hat{p}_0)$ tests are also powerful.

The relatively omnibus and robust power performance of the $\hat{M}_c(\hat{p}_0)$ tests under DGP P.2–P.4 is

[9] We have also used the Parzen kernel (not reported). Although the data-driven lag order $\hat{p}_0$ is substantially smaller, the test statistics are rather similar to those based on the Bartlett kernel in most cases.

encouraging given the fact that DGP P.2–P.4 are all first order dynamic processes whereas the $\hat{M}_c(\hat{p}_0)$ employ several or many lags. Such omnibus and robust power apparently comes from the use of the characteristic function and downward weighting kernel $k(\cdot)$ for lags, which highlights the advantages of the generalized spectrum.

In summary, we observe:

1. The empirical sizes of the $\hat{M}_c(\hat{p}_0)$ tests are close the nominal levels. Under homoskedastic errors, the homoskedasticity-specific tests, $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$, have better size than the heteroskedasticity-robust test $\hat{M}_1(\hat{p}_0)$. Under ARCH errors, $\hat{M}_1(\hat{p}_0)$ remains to have reasonable size, but all homoskedasticity-specific tests strongly overreject the correct model.

2. The powers of FLF test, $Q(m)$, are rather sensitive to the choice of lag order m especially when the sample size is small. In contrast, the $\hat{M}_c(\hat{p}_0)$ tests have relatively robust power with respect to the choice of preliminary lag order $\bar{p}$, and they require no knowledge of the lag structure of the potential alternative.

3. The $\hat{M}_c(\hat{p}_0)$ tests are not always the most powerful in detecting each of the four DGPs. However, they have relatively omnibus power against all four DGPs provided the sample size is sufficiently large. $Q(m)$ can be very powerful in detecting some DGPs but may have little power against others even when the sample size increases.

4. The heteroskedasticity-robust generalized spectral test has similar power to the homoskedasticity-specific generalized tests in most cases.

9. Conclusion

Using a generalized spectral derivative approach, we develop a class of residual-based, generally applicable specification tests for neglected nonlinearity and dynamic misspecification in dynamic panel models. The tests can detect a wide range of model misspecification in mean while being robust to conditional heteroskedasticity and other higher order time-varying moments of unknown form, and inhomogeneity across individuals. They check a large number of lags but naturally discount higher order lags, which alleviates the power loss due to the loss of a large number of degrees of freedom. Our test statistics have a convenient limit N(0,1) distribution even if the lag order of the generalized spectral density estimator is fixed, as is the case under the null hypothesis. The tests enjoy the appealing "nuisance parameter free" property that parameter estimation uncertainty has no impact on the limit distribution of the tests. Simulations show that it is important to take into account conditional heteroskedasticity to ensure a proper size. The tests have omnibus and robust power against a variety of dynamic misspecification and nonlinear alternatives in mean relative to an existing test.

References

Ahn, S. C., and S. Low (1996), "A Reformulation of the Hausman Test for Regression Models with Pooled Cross-Section Time-Series Data," *Journal of Econometrics*, 71, 309 - 319.

Ahn, S. C., and H. R. Moon (2001), "Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test," Working Paper, USC.

Alvarez, J., and M. Arellano (2003), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica,* 71, 1121-1160.

Andrews, D.W.K. (1999), "Consistent Moment Selection Procedures for generalized method of moments estimation," *Econometrica,* 67, 543-654.

Andrews, D.W.K., and B. Lu (2001), "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics,* 101, 123-164.

Angrist, J. D., and W. K. Newey (1991), "Over-Identification Tests in Earnings Functions with Fixed Effects," *Journal of Business and Economic Statistics,* 9, 317 - 323.

Arellano, M. (2003), *Panel Data Econometrics,* Oxford University Press, Oxford.

Arellano, M. (1990), "Testing for Autocorrelation in Dynamic Random Effects Models," *Review of Economic Studies,* 57, 124-134.

Arellano, M. (1993), "On the Testing of Correlated Effects with Panel Data," *Journal of Econometrics,* 59, 87 - 97.

Arellano, M., and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies,* 58, 277-297.

Arellano, M., and B. Honore (2001), "Panel Data Models: Some Recent Developments," in Heckman, J. J. ,and E. Leamer (eds), *Handbook of Econometrics,* Vol. 5, Chapter 53, North-Holland.

Baltagi, B. (1997), "Testing for Linear and loglinear Error Component Regression Against Box-Cox Alternatives," *Statistics and Probability Letters* 33, 63-68.

Baltagi, B. (1999), "Specification Tests in Panel Data Models Using Artificial Regressions," *Annales D'Economie et de Statistique* 55-56, 277-297.

Baltagi, B. (2001), *Econometric Analysis of Panel Data,* Wiley, New York.

Barnett, W., Gallant, R., Hinich, M., Jungeilges, J., Kaplan, D., and Jensen, M. (1997), "A Single-blind Controlled Competition Among Tests for Nonlinearity and Chaos," *Journal of Econometrics* 82, 157-192.

Bera, A.K., and M.L. Higgins (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics* 15, 43-50.

Bierens, H. (1982), "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134.

Brillinger, D., and Rosenblatt, M. (1966a), "Asymptotic Theory of Estimates of k-th Order Spectra,"

Brillinger, D., and Rosenblatt, M. (1966b), "Computation and Interpretation of k-th Order Spectra,"

Brock, W., D. Hsieh, and B. LeBaron (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* MIT Press: Cambridge, MA.

Brock, W., D. Dechert, J. Scheinkman, and B. LeBaron (1996), "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews* 15, 197-235.

Brown, B. M. (1971), "Martingale Central Limit theorems," *Annals of Mathematical Statistics* 42, 59-66.

Cermeño, R. (2002), "Growth Convergence Clubs: Evidence from Markov-Switching Modeles Using Panel Data" Working Paper, CIDE, Mexico.

Cermeño, R., and K. B. Grier (2001), "Modeling GARCH Processes in Panel Data: Theory, Simulations and Examples" Working Paper, CIDE, Mexico.

Cheng, S., and C. L. Wei (2000), "Inferences for a Semiparametric Models with Panel Data," *Biometrika* 87, 89-97.

Cochrane, J. (2001), *Asset Pricing*, Princeton University Press: Princeton, New Jersey.

Dahl, C. (2002), "An Investigation of Tests for Linearity and the Accuracy of Likelihood Based Inference Using Random Fields," *Econometric Journal* 5, 263-284.

Dahl, C., and G. Gonzalez-Rivera (2003), "Testing for Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields," *Journal of Econometrics* 114, 141-164.

Diebold, F. X. (1986), "Testing for Serial Correlation in the Presence of ARCH," *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 323-328.

Diebold, F. X., and J. Nason (1990), "Nonparametric Exchange Rate Prediction" *Journal of International Economics* 28, 315-332.

Diggle, P. J., P. Heagerty, K. Y. Liang, and S. L. Zager (2002), *Analysis of Longitudinal Data*, Oxford: Oxford.

El-Gamal, M. A., and D. Ryu (2003), "Short-Memory and the PPP-Hypothesis," Working Paper, Department of Economics, Rice University.

Eitrheim, Ø., and T. Teräsvirta (1996), "Testing the Adequacy of Smooth Transition Autoregressive Models," *Journal of Econometrics* 74, 59-76.

Engle R., D., and L. R. Robins (1987), "Estimating Time Varying Risk Premium in the Term Structure: The ARCH-M model," *Econometrica* 55, 391-407.

Epps, T. W. (1987), "Testing that a Stationary Time Series is Gaussian", *Annals of Statistics* 15, 1683-1698.

Epps, T. W. (1988), "Testing that a Gaussian Process is Stationary," *Annals of Statistics* 16, 1667-1683.

Gourieroux, C., and J. Jasiak (1998), "Nonlinear Panel Data Models with Dynamic Heterogeneity", Working paper, York University.

Granger, C.W.J. (1992), "Forecasting Stock Market Prices: Lessons for Forecasters", *Macroeconomic Dynamics* 5, 466-481.

Granger, C.W.J. (1996), "Evaluation of Panel Models: Some Suggestions from Time Series", Working Paper, Department of Economics, University of California, San Diego.

Granger, C.W.J. (2001), "Overview of Nonlinear Macroeconometric Empirical Models", *International Journal of Forecasting* 8, 3-13.

Granger, C.W.J., and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships.* Oxford University Press: New York.

Griliches, Z., and J. A. Hausman (1999), "Errors in Variables in Panel Data", *Journal of Econometrics,* 31, 93-118.

Hahn, J., J. Hausman, and G. Kuersteiner (2002), "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects", Working Paper, MIT.

Hahn, J., and G. Kuersteiner (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T are Large", *Econometrica* 70, 1639-1658.

Hamilton, J. D. (1989), "A New Approach to Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica* 57, 357-84.

Hamilton, J. D. (2001), "A Parametric Approach to Flexible Nonlinear Inference," *Econometrica* 69, 537-73.

Hansen, B. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review* 35, 705-730.

Hansen, B. (1996), "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica* 64, 413-430.

Hansen, B. (1999), "Testing for Linearity," *Journal of Economic Surveys* 13, 551-576.

Hansen, B. (2000), "Sample Splitting and Threshold Estimation," *Econometrica* 68, 575-603.

Harvey, C., and A. Siddque (1999), "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis* 34, 465-487.

Harvey, C., and A. Siddque (2000), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance* 55, 1263-1296.

Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.

Hausman, J., and W. E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49, 1377-1398.

Hinich, M. (1982), "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis* 3, 169-176.

Hjellvik, V., and D. Tjøstheim (1996), "Nonparametric Statistics for Testing of Linearity and Serial Independence," *Journal of Nonparametric Statistics* 6, 223-251.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association* 84, 1201-1220.

Hong, Y., and Y.-J. Lee (2004), "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form," *Review of Economic Studies,* forthcoming.

Honore, B. E. (2002), "Nonlinear Models with Panel Data," Cemmap Working paper CWP13/02.

Hsiao, C. (2003), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Jondeau, E., and M. Rockinger (2003), "Testing for Differences in the Tails of Stock-Market

Returns," *Journal of Empirical Finance* 10, 559-581.

Jiang, G. J., and J. L. Knight (2002), "Estimation of Continuous Time Processes via the Empirical Characteristic Function," *Journal of Business and Economic Statistics* 20, 198-212.

Kang, S. (1985), "A Note on the Equivalence of Specification Tests in the Two-factor Multivariate Variance Component Model," *Journal of Econometrics* 28, 193-203.

Keenan, D. M. (1985), "A Tukey Non-Additivity-Type Test for Time Series Nonlinearity," *Biometrika* 72, 39-44.

Kiviet, J. F. (1995), "On Bias, Inconsistency and Efficiency of Some Estimators in Dynamic Panel Data," *Journal of Econometrics* 68, 53-78.

Knight, J. L., and S.E. Satchell (1997), "The Cumulant Generating Function Estimation Method," *Econometric Theory* 13, 170-184.

Knight, J. L., and J. Yu (2002), "Empirical Characteristic Function in Time Series Estimation," *Econometric Theory* 18, 691-721.

Lancaster, T. (2000), "The Incidental Parameter Problem Since 1948," *Journal of Econometrics* 95, 391-413.

Lee, T.-H., H. White and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics* 56, 269-290.

Li, Q. (1999), "Consistent Specification Tests for Time Series Econometric Models," *Journal of Econometrics* 92, 101-147.

Li, Q., and T. Stengos (1996), "Semiparametric Estimation of Partially Linear Panel Data Models," *Journal of Econometrics* 71, 389-397.

Lobato, I. (2002), "A Consistent Test for the Martingale Difference Hypothesis," Forthcoming in *Econometric Reviews.*

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988 a), "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika* 75, 491-499.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988 b), "Testing Linearity in Univariate Time Series Models," *Scandinavian Journal of Statistics* 15, 161-75.

Meese, R., and A. Rose (1991), "An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination," *Review of Economic Studies* 58, 603-619.

Meghir, C., and F. Windmeijer (1999), "Moment Conditions for Dynamic Panel Data Models with Multiplicative Individual Effects in the Conditional Variance," *Annales D'Economie et de Statistique* 55-56, 317-330.

Metcalf, G. E. (1996), "Specification Testing in Panel Data with Instrumental Variables," *Journal of Econometrics* 71, 291-307.

Paap, R., P.H. Franses, and D. J. C. van Dijk (2003)**,** "Do African Countries Grow Slower Than Asian Countries", Econometric Institute Report EI 2003-07. Erasmus University Rotterdam.

Park, J., and Y.J. Whang (2003), "Testing for the Martingale Hypothesis," Working Paper, Department of Economics, Rice University and Department of Economics, Korea University.

Peña, D., Tiao, G. and R.S. Tsay (2001), *A Course in Time Series Analysis,* Wiley: New York.

Pinkse, Y. (1998), "Consistent Nonparametric Testing for Serial Independence," *Journal of Econometrics* 84, 205-231.

Potter, S. (1995) "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics* 10, 109-125.

Priestley, M. B. (1981), *Spectral Analysis and Time Series.* Academic press: London.

Priestley, M. B. (1988), *Non-Linear and Non-Stationary Time Series Analysis.* Academic Press: London.

Rice, J. (2003),"Functional and Longitudinal Data Analysis: Perspectives on Smoothing," Working Paper, Department of Statistics, UC Berkeley.

Saikkonen, P., and Luukkonen, R. (1988), "Lagrange Multiplier Tests for Testing Non-linearities in Time Series Models," *Scandinavian Journal of Statistics* 15, 55-68.

Sargent, T., and L. Ljungqvist (2002), *Recursive Macroeconomic Theory.* MIT Press: Cambridge, MA.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: London.

Singleton, K. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics* 102, 111-141.

Stenseth, N. C. (and 10 others) (1999), "Common Dynamic Structure of Canada Lynx Populations within three Climatic Regions," *Science* 285, 1071-1073.

Stinchcombe M., and H. White (1998), "Consistent Specification Testing with Nuisance Parameters Present only Under the Alternative," *Econometric Theory* 14, 295-324.

Subba R. T. and M. Gabr (1980), "A Test for Linearity of Stationary Time Series," *Journal of Time Series Analysis* 1, 145-158.

Teräsvirta, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association* 89, 208-218.

Teräsvirta, T., C. J. Lin and C.W.J. Granger (1994), "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis* 14, 209-220.

Teräsvirta, T., D. Tjøstheim and C.W.J. Granger (1994), "Aspects of Modelling Nonlinear Time Series," *Handbook of Econometrics,* Vol. IV, Chapter 48, pp. 2917-2957, R.F. Engle and D.L. McFadden (ed.), North Holland, Elsevier Science.

Tiao, G., and R. S. Tsay (1994), "Some Advances in Nonlinear and Adaptive Modelling in Time Series," *Journal of Forecasting* 13, 109-131.

Tjøstheim, D. (1994), "Non-Linear Time Series: A Selective Review," *Scandinavian Journal of Statistics* 21, 97-130.

Tong, H. (1990), *Nonlinear Time Series: A Dynamic System Approach*, Clarendon Press: Oxford.

Tsay, R. S. (1986), "Nonlinearity Tests for Time Series," *Biometrika*, 73, 461-6.

Tsay, R. S. (1989), "Testing and Modelling Threshold Autoregressive Processes," *Journal of the American Statistical Association* 84, 231-240.

Tsay, R. S. (2001), "Nonlinear Time Series Models: Testing and Applications," in D. Peña, G. Tiao and R.S. Tsay (eds.), *A Course in Time Series Analysis*, Chapter 10, pp. 267-285, John Wiley: New York.

Whang, Y.-J. (1998), "A Test of Autocorrelation in the Presence of Heteroskedasticity of Unknown Form," *Econometric Theory* 14, 87-122.

White, H. (1981), "Consequences and Detection of Misspecified Nonlinear Regression Models," *Journal of the American Statistical Association* 76, 419-433.

White, H. (1989), "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks, " in *Proceeding of the International Joint Conference on Neural Networks*, IEEE Press, New York, NY, Washington, DC, pp. 451-455.

Wooldridge, J. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6, 17-43.

Wooldridge, J. (1991), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47, 5-46.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.

Wooldridge, J. (2002), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," Working Paper, Department of Economics, Michigan State University.

Yao, Q., H. Tong, B. Finkenstadt, and N. C. Stenseth (2000). "Common Structure in Panels of Short Ecological Time-Series," *Proceeding Royal Society of London Series B.*, 267, 2459-2467.

Table 1: Empirical Sizes of Tests Under DGP S1

(n,T)	(25,25)		(25,50)		(25,100)		(50,25)		(50,50)		(50,100)		(100,25)		(100,50)		(100,100)	
Level	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$\hat{M}_1^a(2)$	9.3	4.8	11.6	6.1	10.3	5.6	11.5	6.4	10.8	5.6	13.0	6.2	16.0	8.4	13.3	7.6	10.1	4.9
$\hat{M}_1^a(4)$	9.2	4.5	11.4	6.0	11.0	6.0	11.4	6.3	10.9	6.2	13.0	6.2	15.8	8.1	13.6	7.0	10.3	5.0
$\hat{M}_1^a(6)$	8.9	3.5	10.3	5.8	10.7	5.7	10.9	5.3	9.8	5.5	11.5	6.1	15.4	6.8	12.4	5.5	9.8	4.8
$\hat{M}_2^a(2)$	8.0	3.5	10.6	6.0	10.0	6.1	6.8	4.3	9.0	4.4	11.7	5.9	8.3	3.9	10.1	5.8	8.5	4.6
$\hat{M}_2^a(4)$	7.7	3.4	10.4	5.9	10.1	6.2	6.7	3.8	8.8	4.4	11.3	6.1	7.5	3.5	9.8	5.4	8.9	4.8
$\hat{M}_2^a(6)$	6.3	2.7	9.2	5.4	9.9	6.0	5.4	2.8	8.0	4.2	10.5	5.6	4.7	2.0	7.9	3.7	8.5	4.1
$\hat{M}_3^a(2)$	8.0	3.5	10.5	5.8	10.0	6.1	7.0	4.3	9.0	4.4	11.9	5.9	8.2	4.1	10.2	5.8	8.5	4.7
$\hat{M}_3^a(4)$	8.2	3.7	10.8	5.9	10.2	6.3	7.5	4.2	9.0	4.6	11.5	6.1	8.1	4.1	10.1	5.5	9.0	4.8
$\hat{M}_3^a(6)$	8.2	3.3	10.0	5.9	10.7	6.5	8.5	3.8	9.3	5.1	11.4	6.1	8.0	3.3	10.0	5.0	9.4	4.6
$\hat{M}_1^b(2)$	10.2	5.4	11.6	6.4	10.3	5.6	11.6	6.4	10.8	5.6	13.0	6.4	16.1	8.8	14.3	7.3	10.0	5.0
$\hat{M}_1^b(4)$	9.7	5.1	11.8	6.5	10.6	5.7	12.1	6.3	11.1	5.8	13.4	6.8	16.3	9.1	14.6	7.2	10.4	5.0
$\hat{M}_1^b(6)$	9.8	5.0	10.6	5.9	10.9	6.3	13.2	5.9	10.7	5.7	12.0	6.5	18.2	9.0	13.7	5.8	11.1	5.0
$\hat{M}_2^b(2)$	8.3	4.0	10.9	6.0	9.9	5.9	7.4	4.3	8.9	4.6	11.9	5.9	8.0	3.9	10.4	5.9	8.2	4.5
$\hat{M}_2^b(4)$	8.1	4.1	10.9	6.0	10.6	6.1	7.4	4.2	9.1	4.6	11.4	6.1	8.0	4.0	10.1	5.8	9.2	4.4
$\hat{M}_2^b(6)$	7.7	3.3	10.2	5.7	10.5	6.6	7.4	3.3	9.1	4.8	11.4	5.8	7.3	3.2	9.5	4.4	9.0	4.7
$\hat{M}_3^b(2)$	8.3	4.0	10.9	5.8	9.9	5.9	7.4	4.2	9.2	4.9	11.9	5.9	8.0	3.9	10.4	5.9	8.2	4.5
$\hat{M}_3^b(4)$	8.3	4.2	10.9	6.2	10.6	6.4	7.5	4.2	9.2	4.9	11.4	6.3	8.2	4.1	9.9	5.1	9.2	4.4
$\hat{M}_3^b(6)$	8.9	3.9	10.7	6.1	10.7	6.7	8.4	3.6	9.4	5.1	11.6	5.9	8.4	3.7	10.0	4.8	9.3	4.7
$Q(2)$	11.0	4.2	10.7	5.4	10.9	5.7	10.7	5.9	9.9	5.7	9.2	3.6	9.5	4.3	11.9	6.5	9.7	4.5
$Q(4)$	10.0	4.3	11.4	6.1	11.7	6.6	11.5	6.2	10.8	5.7	11.0	5.3	9.6	4.0	11.2	4.9	8.5	4.5
$Q(6)$	11.0	5.2	10.3	5.6	13.4	6.6	11.0	5.7	11.1	6.7	10.6	5.3	9.2	4.9	10.1	5.3	10.0	5.8

Table 2: Empirical Sizes of Tests: DGP S2

(n,T)	(25,25)		(25,50)		(25,100)		(50,25)		(50,50)		(50,100)		(100,25)		(100,50)		(100,100)	
Level	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$\hat{M}_1^a(2)$	11.1	4.1	10.5	5.7	10.7	5.9	12.2	5.1	10.0	5.3	10.2	5.5	18.4	9.6	14.0	5.9	10.1	5.0
$\hat{M}_1^a(4)$	10.2	4.0	10.3	5.8	10.5	5.6	11.8	4.7	9.5	5.0	10.2	5.2	18.5	9.3	13.7	5.4	10.0	5.0
$\hat{M}_1^a(6)$	8.4	3.2	9.5	5.2	10.1	5.1	11.3	4.3	8.2	3.8	9.7	5.1	17.1	8.1	12.7	4.9	9.3	4.5
$\hat{M}_2^a(2)$	64.5	53.1	77.9	66.1	82.4	72.7	88.8	79.4	94.3	91.0	97.4	94.9	98.9	96.6	100	99.8	100	99.8
$\hat{M}_2^a(4)$	63.3	51.2	76.9	661.	82.8	72.3	87.3	77.7	94.6	90.7	97.3	94.5	98.7	96.3	100	99.8	99.9	99.8
$\hat{M}_2^a(6)$	54.2	41.4	70.5	58.1	78.0	66.7	78.0	65.2	92.3	86.4	95.9	90.8	94.6	89.5	99.6	98.9	99.8	99.4
$\hat{M}_3^a(2)$	64.6	53.1	77.8	66.1	82.6	72.8	88.7	79.6	94.5	91.1	97.4	94.8	99.0	96.6	100	99.8	100	99.8
$\hat{M}_3^a(4)$	64.2	52.5	77.3	66.5	82.9	72.4	87.8	78.8	94.6	90.7	97.3	94.5	98.9	96.5	100	99.8	99.9	99.3
$\hat{M}_3^a(6)$	57.9	46.5	72.6	60.7	79.1	67.7	82.0	71.6	92.3	86.4	96.1	91.8	96.1	92.9	100	99.0	99.8	99.4
$\hat{M}_1^b(2)$	8.6	3.5	10.0	5.3	9.5	5.3	9.5	3.9	8.7	4.1	9.2	5.2	15.3	7.0	11.3	4.3	8.8	4.3
$\hat{M}_1^b(4)$	8.4	3.5	9.9	5.4	9.4	5.1	9.7	3.6	8.2	4.2	9.1	5.1	15.4	7.2	1.8	3.9	9.1	4.5
$\hat{M}_1^b(6)$	7.4	2.8	9.4	4.8	9.6	5.0	10.1	2.8	7.2	3.5	8.4	4.8	15.6	6.6	11.5	4.3	8.3	4.2
$\hat{M}_2^b(2)$	60.2	47.8	74.7	62.8	9.9	81.4	84.0	75.5	93.1	89.4	96.9	93.6	97.8	94.9	99.9	99.6	99.9	99.8
$\hat{M}_2^b(4)$	59.6	47.2	75.1	61.9	10.6	81.8	83.1	73.8	93.1	88.9	96.7	93.4	97.5	94.4	99.8	99.5	99.9	99.8
$\hat{M}_2^b(6)$	51.3	39.4	67.4	55.8	10.5	76.5	75.3	63.0	89.9	82.5	94.7	89.6	93.8	88.5	99.6	98.4	99.7	99.4
$\hat{M}_3^b(2)$	60.2	47.8	74.7	62.6	9.9	81.4	83.9	75.5	93.1	89.5	96.9	93.5	97.8	94.9	99.9	99.6	99.9	99.8
$\hat{M}_3^b(4)$	59.8	47.8	75.3	62.0	10.6	81.8	83.3	74.2	93.1	89.2	96.7	93.5	97.5	94.6	99.9	99.5	99.8	99.8
$\hat{M}_3^b(6)$	54.3	41.2	68.5	56.8	10.7	77.2	77.1	66.0	90.2	83.7	95.0	90.0	94.0	89.4	99.6	98.7	99.7	99.4
$Q(2)$	19.8	12.2	18.7	12.4	19.4	11.9	18.9	12.3	18.7	11.2	17.5	10.2	17.7	11.5	19.9	13.1	17.9	10.8
$Q(4)$	19.0	10.6	18.5	10.7	17.8	11.7	17.3	9.9	16.5	10.7	16.0	9.7	16.8	10.5	17.8	10.9	15.3	8.0
$Q(6)$	18.2	10.8	17.1	9.1	18.5	10.4	15.4	9.2	16.7	10.7	15.9	9.2	16.3	9.3	17.1	9.6	14.4	7.6

Table 3: Size-Adjusted Power of Tests:

	DGP1: AR2				DGP2: ARMA				DGP3: SETAR				DGP4: TAR			
(n,T)	(25,25)		(50,50)		(25,25)		(50,50)		(25,25)		(50,50)		(25,25)		(50,50)	
Level	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$\hat{M}_1^a(2)$	18.9	9.5	21.6	11.9	38.9	25.2	93.7	88.6	54.0	37.7	76.7	62.7	41.9	27.0	94.3	89.1
$\hat{M}_1^a(4)$	18.6	9.8	21.9	12.4	38.5	26.2	94.0	89.1	53.5	39.1	77.2	62.3	40.9	27.0	94.0	88.2
$\hat{M}_1^a(6)$	15.6	8.8	21.2	11.1	36.9	25.3	93.6	86.5	47.7	34.2	72.5	55.1	33.3	20.8	90.7	81.4
$\hat{M}_2^a(2)$	17.3	8.6	21.0	13.0	40.4	30.0	95.8	91.9	59.4	45.1	87.3	78.6	46.6	31.2	98.0	95.6
$\hat{M}_2^a(4)$	17.2	8.8	22.2	12.9	40.3	29.3	96.1	91.6	58.6	44.2	87.1	79.3	45.4	30.5	97.8	95.3
$\hat{M}_2^a(6)$	14.8	8.4	20.8	11.7	39.3	27.1	94.9	88.9	53.3	38.9	83.7	72.0	40.1	24.8	96.0	92.2
$\hat{M}_3^a(2)$	17.4	8.9	21.1	13.0	40.0	30.0	95.8	91.9	58.9	45.5	87.2	78.5	46.7	31.2	97.9	95.5
$\hat{M}_3^a(4)$	17.6	9.0	22.1	13.6	41.1	29.5	96.0	91.9	58.7	44.6	86.2	76.5	46.3	30.6	97.6	95.3
$\hat{M}_3^a(6)$	14.7	8.3	21.1	11.8	38.9	28.2	94.9	89.2	52.9	39.0	83.6	72.5	38.9	24.9	96.0	92.4
$\hat{M}_1^b(2)$	17.7	9.0	21.2	11.7	39.8	24.1	95.1	89.5	53.1	35.8	78.4	62.3	37.9	21.7	94.0	89.0
$\hat{M}_1^b(4)$	17.5	9.1	21.8	12.7	39.5	24.4	94.8	89.8	53.1	36.8	77.9	62.3	38.0	21.7	93.9	89.0
$\hat{M}_1^b(6)$	16.4	7.9	20.5	12.1	38.7	23.9	93.6	88.1	48.1	31.7	73.0	56.8	32.3	17.7	90.5	81.6
$\hat{M}_2^b(2)$	18.8	9.0	21.2	12.3	41.9	29.4	95.9	91.5	59.6	43.6	86.6	76.6	46.3	30.6	97.9	95.3
$\hat{M}_2^b(4)$	17.8	8.5	22.1	12.3	42.4	28.5	95.6	91.4	58.0	41.9	86.6	76.3	44.6	29.0	97.7	95.0
$\hat{M}_2^b(6)$	15.0	8.3	21.6	12.4	39.2	26.7	95.0	89.8	51.8	37.5	83.8	70.8	39.1	24.0	95.9	91.8
$\hat{M}_3^b(2)$	18.8	9.0	21.2	12.5	41.6	29.6	95.9	91.6	59.5	43.9	86.6	76.6	46.0	30.6	97.9	95.2
$\hat{M}_3^b(4)$	17.7	8.5	22.1	12.3	42.2	28.6	95.6	91.3	57.9	41.9	86.7	76.3	45.6	29.0	97.7	95.0
$\hat{M}_3^b(6)$	15.5	8.2	21.6	12.4	39.3	26.8	95.0	89.6	51.9	37.3	83.7	70.5	39.1	23.6	95.8	91.7
$Q(2)$	23.5	17.1	66.8	53.1	11.5	7.1	13.3	7.0	16.8	12.4	30.2	20.8	10.8	7.6	11.8	5.5
$Q(4)$	22.6	15.1	56.8	42.1	13.6	8.2	11.6	7.1	18.0	12.4	37.4	26.7	12.3	9.2	11.2	5.6
$Q(6)$	19.0	11.3	48.7	33.4	12.2	6.6	11.4	6.3	15.4	9.1	30.0	19.6	12.0	7.0	10.6	6.0