Essays on Heterogeneity and Non-Linearity in Panel Data and Time Series Models

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Nazarii SALISH
aus Lviv, Ukraine

Bonn 2016

Dekan:	Prof. Dr. Daniel Zimmer
Erstreferent:	Prof. Dr. Matei Demetrescu
Zweitreferent:	Prof. Dr. Joerg Breitung

Tag der mündlichen Prüfung: 21.11.2016

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online/elektronisch publiziert.

Acknowlegment

A number of people have contributed to this thesis and I owe them a great amount of gratitude. First of all I would like to thank to my supervisors Matei Demetrescu and Joerg Breitung. This work would not have been possible without their guidance, encouragement and relentless support in the past five years. Their continuous feedback and insightful comments paved the way to this thesis and my interest in econometric research. I am also gratefully indebted to Alois Kneip for his extensive advice and support of the third chapter of this thesis.

Furthermore, I would like to thank to my coauthors Alexander Gleim, Thomas Nebeling and Christoph Roling with whom I spent many hours examining some of the chapters presented here. It was a great experience and a lot of hard work. Also special thanks go to my fellow PhD students at the Bonn Graduate School of Economics for many interesting and entertaining conversations or for simply having a good time.

I would like to thank my wife Mirjam Salish, for many useful comments and suggestions on how to write this thesis, and my parents, who provided me with a carefree childhood, a great early education and constant support.

Finally I wold like to thank the Bonn Graduate School of Economics and Econometrics Department at University of Cologne for providing an excellent research environment. The financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft) through grant DE 1617/2-1 is gratefully acknowledge.

Contents

List of Figures

List of Tables

Overview

In recent years advances in data collection and storage allow us to observe and analyze many financial, economic or environmental processes with higher precision. This in turn reveals new features of the underlying processes and creates a demand for the development of new econometric techniques. The aim of this thesis is to tackle some of these challenges in the filed of panel data and time series models. In particular, the first and the last chapters contribute to the issue of testing and estimating heterogeneous panel models with random coefficients. The second chapter discusses a generalization of the classical linear time series models to asymmetric ones and presents a test statistic to help empirical researchers to choose the appropriate modeling framework in this context. Finally, the objective of the third chapter is to extend the available (nonlinear) time series techniques on big data sets or functional data.

In more detail, Chapter1, which is joint work with Joerg Breitung and Christoph Roling, employs the Lagrange Multiplier (LM) principle to test parameter homogeneity across cross-section units in panel data models. The test can be seen as a generalization of the Breusch-Pagan test against random individual effects to all regression coefficients. While the original test procedure assumes a likelihood framework under normality, several useful variants of the LM test are presented to allow for non-normality, heteroskedasticity and serially correlated errors. Moreover, the tests can be conveniently computed via simple artificial regressions. We derive the limiting distribution of the LM test and show that if the errors are not normally distributed, the original LM test is asymptotically valid if the number of time periods tends to infinity. A simple modification of the score statistic yields an LM test that is robust to non-normality if the number of time periods is fixed. Further adjustments provide versions of the LM test that are robust to heteroskedasticity and serial correlation. We compare the local power of our tests and the statistic proposed by Pesaran and Yamagata. The results of the Monte Carlo experiments suggest that the LM-type test can be substantially more powerful, in particular, when the number of time periods is small.

Chapter 2, which is joint work with Thomas Nebeling, develops a Lagrange multiplier test statistic and its variants to test for the null hypothesis of no asymmetric effects of shocks on time series. In asymmetric time series models that allow for different responses to positive and negative past shocks the likelihood functions are, in general, non-differentiable. By making use of the theory of generalized functions Lagrange multiplier type tests and the resulting asymptotics are derived. The test statistics possess standard asymptotic limiting behavior under the null hypothesis. Monte Carlo experiments illustrate the accuracy of the asymptotic approximation and show that conventional model selection criteria can be used to estimate the required lag length. We provide an empirical application to the U.S. unemployment rate.

In Chapter 3, written in collaborative work with Alexander Gleim, statistical tools for forecasting functional times series are developed, which for example can be used to analyze big data sets. To tackle the issue of time dependence we introduce the notion of functional dependence through scores of the spectral representation. We investigate the impact of time dependence thus quantified on the estimation of functional principal components. The rate of mean squared convergence of the estimator of the covariance operator is derived under long range dependence of the functional time series. After that, we suggest two forecasting techniques for functional time series satisfying our measure of time dependence and derive the asymptotic properties of their predictors. The first is the functional autoregressive model which is commonly used to describe linear processes. As our notion of functional dependence covers a broader class of processes we also study the functional additive autoregressive model and construct its forecasts by using the k-nearest neighbors approach. The accuracy of the proposed tools is verified through Monte Carlo simulations. Empirical relevance of the theory is illustrated through an application to electricity consumption in the Nordic countries.

In Chapter 4, which was jointly done with Joerg Breitung, three main estimation procedures for the panel data models with heterogeneous slopes are discussed: pooling, generalized LS and mean-group estimator. In our analysis we take an explicit account of the statistical dependence that may exists between regressors and the heterogeneous effects of the slopes. It is shown that under systematic slope variations: (i) pooling gives inconsistent and highly misleading estimates, and (ii) generalized LS in general is not consistent even in settings when N and T are large, (iii) while mean-group estimator always provide consistent result at a price of higher variance. We contribute to the literature by suggesting a simple robustified version of the pooled based on Mundlak type corrections. This estimator provides consistent results and is asymptotically equivalent to the mean-group estimator for large N and T. Monte Carlo experiments confirm our theoretical findings and show that for large N and fixed T new estimator can be an attractive option when compare to the competitors.

Chapter 1

LM-type Tests for Slope Homogeneity in Panel Data Models

1.1 Introduction

In classical panel data analysis it is assumed that unobserved heterogeneity is captured by individual-specific constants, whether they are assumed to be fixed or random. In many applications, however, it cannot be ruled out that slope coefficients are also individual-specific. For instance, heterogenous preferences among individuals may result in individual-specific price or income elasticities. Ignoring this form of heterogeneity may result in biased estimation and inference. Therefore, it is important to test the assumption of slope homogeneity before applying standard panel data techniques such as the least-squares dummy-variable (LSDV) estimator for the fixed effect panel data model.

If there is evidence for individual-specific slope parameters, economists are interested in estimating a population average like the mean of the individual-specific coefficients. Pesaran and Smith (1995) advocate mean group estimation, where in a first step the model is estimated separately for each cross-section unit. In a second step, the unit-specific estimates are averaged to obtain an estimator for the population mean of the parameters. Alternatively, Swamy (1970) proposes a generalized least squares (GLS) estimator for the random coefficients model, which assumes that the individual regression coefficients are randomly distributed around a common mean.

In this paper we derive a test for slope homogeneity by employing the LM principle within a random coefficients framework, which allows us to formulate the null hypothesis of slope homogeneity in terms of K restrictions on the variance parameters. Hence, the LM approach substantially reduces the number of restrictions to be tested compared to the set of $K(N-1)$ linear restrictions on the coefficients implied by the test proposed by Pesaran and Yamagata (2008), henceforth referred to as PY. This does not mean, however, that our test is confined to detect random deviation in the coefficients. In fact our test is optimal against the alternative of random coefficients but it is also powerful against any systematic variations of the regression coefficients.

Our approach is related but not identical to the conditional LM test recently suggested by Juhl and Lugovskyy (2014) which is referred to as the JL test. The main difference is that the latter test is derived for a more restrictive alternative, where it is assumed that the individual-specific slope coefficients attached to the K regressors have identical variances. In contrast, our test focuses on the alternative that the coefficients have different variances which allows us to test for heterogeneity in a subset of the regression coefficients. Furthermore, the derivation of our test follows the original LM principle involving the information matrix, whereas the JL test employs the outer product of the scores as an estimator of the information matrix. Our simulation study suggest that both non-standard features of the latter test may result in size distortions in small samples and a sizable loss in power. An important advantage of the JL test is however that it is robust against non-Gaussian and heteroskedastic errors. We therefore propose variants of the original LM test that share the robustness against non-Gaussian and heteroskedastic errors. Furthermore, we also suggest a modified LM test that is robust to serially correlated errors. Another contribution of the paper is the analysis of the local power of the test that allows us to compare the power properties of the LM and PY tests. Specifically, we find that the location parameter of the LM test depends on the cross-section dispersion of the regression variances, whereas the location parameter of the PY test only depends on the mean of the regressor variances. Thus, if the regressor variances differ across the panel groups, the gain in power from using the LM test may be substantial.

The outline of the paper is as follows. In Section 1.2 we compare two tests for slope heterogeneity recently proposed in the literature. We introduce the random coefficients model in 1.3 and lay out the (standard) assumptions for analyzing the large-sample properties. In Section 1.4 we derive the LM statistic and establish its asymptotic distribution. Section 1.5 discusses several variants of the proposed test. First, we relax the normality assumption and extend the result of the previous section to this more general setting. Second, we propose a regression-based version of the LM test. Section 1.6 investigates the local asymptotic power of the LM test. Section 1.7 describes the design of our Monte Carlo experiments and discusses the results. Section 1.8 concludes.

1.2 Existing tests

To prepare the theoretical discussion in the following sections, we briefly review the random coefficients model and existing tests. Following Swamy (1970), consider a linear panel data model

$$y_{it} = x_{it}'\beta_i + \epsilon_{it},$$

for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$, where y_{it} is the dependent variable for unit i at time period t, x_{it} is a $K \times 1$ vector of explanatory variables and ϵ_{it} is an idiosyncratic error with zero mean and variance $\mathbb{E}(\epsilon_{it}^2) = \sigma_i^2$. For the slope coefficient β_i we assume

$$\beta_i = \beta + v_i \ ,$$

where β is a fixed $K \times 1$ vector and v_i is a i.i.d. random vector with zero mean and $K \times K$ covariance matrix Σ_v.[1]

The null hypothesis of slope homogeneity is

$$\beta_1 = \beta_2 = \cdots = \beta_N = \beta, \tag{1.1}$$

which is equivalent to testing $\Sigma_v = 0$. To test hypothesis (1.1), Swamy suggests the statistic

$$\widehat{S}^* = \sum_{i=1}^{N} \left(\widehat{\beta}_i - \widehat{\beta}_{\text{WLS}}\right)' \left(\frac{X_i' X_i}{s_i^2}\right) \left(\widehat{\beta}_i - \widehat{\beta}_{\text{WLS}}\right),$$

with $X_i = (x_{i1}, \ldots, x_{iT})'$ and $\widehat{\beta}_i = (X_i' X_i)^{-1} X_i' y_i$ is the ordinary least squares (OLS) estimator of (1.2) for panel unit i, and $t = 1, \ldots, T$. The common slope parameter β is estimated by the weighted least-squares estimator

$$\widehat{\beta}_{\text{WLS}} = \left(\sum_{i=1}^{N} \frac{X_i' X_i}{s_i^2}\right)^{-1} \left(\sum_{i=1}^{N} \frac{X_i' y_i}{s_i^2}\right),$$

where s_i^2 denotes the standard OLS estimator of σ_i^2.

Intuitively, if the regression coefficients are identical, the differences between the individual estimators and the pooled estimator should be small. Therefore, Swamy's test rejects the null hypothesis of homogenous slopes for large values of this statistic, which possesses a limiting χ^2 distribution with $K(N-1)$ degrees of freedom as N is fixed and $T \to \infty$.

Pesaran and Yamagata (2008) emphasize that in many empirical applications N is large relative to T and the approximation by a χ^2 distribution is unreliable. PY adapt the test to a setting in which N and T jointly tend to infinity. In particular, they assume

[1] For more details and extensions of the basic random coefficient model see Hsiao and Pesaran (2008). As pointed out by a referee, this specification may be replaced by some systematic variation of the coefficients that depends on observed variables. For example, we may specify the deviations as $\beta_i - \beta = \Gamma z_i + \eta_i$, where z_i is some vector of observed variables possibly correlated with x_{it}. The corresponding variant of the LM test (which is different from our LM test based assuming that v_i and x_{it} are independent) will be optimal against this particular form of systematic variation. In general, our test assuming independent variation with $\Gamma = 0$ will also have power against systematic variations but admittedly our test is not optimal against alternative with systematically varying coefficients.

individual-specific intercepts and derive a test for the hypothesis $\beta_1 = \cdots = \beta_N = \beta$ in

$$y_{it} = \alpha_i + x_{it}'\beta_i + \epsilon_{it}. \tag{1.2}$$

The analogue of the pooled weighted least squares estimator above eliminates the unobserved fixed effects,

$$\widehat{\beta}_{\text{WFE}} = \left(\sum_{i=1}^{N} \frac{X_i' M_\iota X_i}{\widehat{\sigma}_i^2}\right)^{-1} \left(\sum_{i=1}^{N} \frac{X_i' M_\iota y_i}{\widehat{\sigma}_i^2}\right),$$

where $M_\iota = I_T - \iota_T \iota_T' / T$, and ι_T is a $T \times 1$ vector of ones. A natural estimator for σ_i^2 is

$$\widehat{\sigma}_i^2 = \frac{\left(y_i - X_i \widehat{\beta}_i\right)' M_\iota \left(y_i - X_i \widehat{\beta}_i\right)}{T - K - 1},$$

where $\widehat{\beta}_i = \left(X_i' M_\iota X_i\right)^{-1} \left(X_i' M_\iota y_i\right)$ and the test statistic becomes

$$\widehat{S} = \sum_{i=1}^{N} \left(\widehat{\beta}_i - \widehat{\beta}_{\text{WFE}}\right)' \left(\frac{X_i' M_\iota X_i}{\widehat{\sigma}_i^2}\right) \left(\widehat{\beta}_i - \widehat{\beta}_{\text{WFE}}\right).$$

Employing a joint limit theory for N and T, PY obtain the limiting distribution as

$$\widehat{\Delta} = \frac{\widehat{S} - NK}{\sqrt{2NK}} \xrightarrow{d} \mathcal{N}(0, 1), \tag{1.3}$$

provided that $N \to \infty$, $T \to \infty$ and $\sqrt{N}/T \to 0$. Thus, by appropriately centering and standardizing the test statistic, inference can be carried out by resorting to the standard normal distribution, provided the time dimension is sufficiently large relative to the cross-section dimension. PY propose several modified versions of this test, which for brevity we shall refer to as the Δ tests or statistics. In particular, to improve the small sample properties of the test, PY suggest the adjusted statistic under normally distributed errors (see Remark 2 in PY),

$$\widetilde{\Delta}_{\text{adj}} = \sqrt{N(T+1)} \left(\frac{N^{-1}\widetilde{S} - K}{\sqrt{2K(T - K - 1)}}\right), \tag{1.4}$$

where $\widetilde{S}$ is computed as $\widehat{S}$ but replacing $\widehat{\sigma}_i^2$ by the variance estimator

$$\widetilde{\sigma}_i^2 = \frac{\left(y_i - X_i \widetilde{\beta}_{\text{FE}}\right)' M_\iota \left(y_i - X_i \widetilde{\beta}_{\text{FE}}\right)}{T - 1}, \tag{1.5}$$

where $\widetilde{\beta}_{\mathrm{FE}} = \left(\sum_{i=1}^{N} X_i' M_\iota X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' M_\iota y_i\right)$ is the standard 'fixed effects' (within-group) estimator. Note that this asymptotic framework does not seem to be well suited for typical panel data applications where N is large relative to T. Therefore, it will be of interest to derive a test statistic that is valid when T is small (say $T = 10$) and N is very large (say $N = 1000$), which, for instance, is encountered in microeconomic panels.

The test statistic proposed by Juhl and Lugovskyy (2014) is based on the individual scores

$$\mathcal{S}_i = \widehat{u}_i' M_\iota X_i X_i' M_\iota \widehat{u}_i - \widehat{\sigma}_i^2 \operatorname{tr}(X_i' M_\iota X_i),$$

where $\widehat{u}_i = y_i - X_i \widetilde{\beta}_{\mathrm{FE}}$ and $\operatorname{tr}(A)$ denotes the trace of the matrix A. The (conditional) LM statistic results as

$$\mathrm{CLM} = \sum_{i=1}^{N} \mathcal{S}_i' \left(\sum_{i=1}^{N} \mathcal{S}_i \mathcal{S}_i'\right)^{-1} \sum_{i=1}^{N} \mathcal{S}_i \ . \tag{1.6}$$

It is interesting to compare this test statistic to the PY test which is based on the sum $\widehat{S} = \sum_{i=1}^{N} \widehat{S}_i$ with

$$\begin{aligned} \widehat{S}_i &= \left(\widehat{\beta}_i - \widehat{\beta}_{\mathrm{WFE}}\right)' \left(\frac{X_i' M_\iota X_i}{\widehat{\sigma}_i^2}\right) \left(\widehat{\beta}_i - \widehat{\beta}_{\mathrm{WFE}}\right) \\ &= \frac{1}{\sigma_i^2} u_i' M_\iota X_i (X_i' M_\iota X_i)^{-1} X_i' M_\iota u_i + o_p(1) \end{aligned}$$

if N and T tend to infinity. Note that $\lim_{N\to\infty} \mathbb{E}(\widehat{S}_i) = K$. The main difference between the JL and the PY statistics is that the statistic $\mathcal{S}_i$ neglects the additional inverse $(\sigma_i^2 X_i' M_\iota X_i)^{-1}$ in the statistic $\widehat{S}_i$. Thus, although these two test statistics are derived from different statistical principles, the final test statistics are essentially testing the independence of u_i and $M_\iota X_i$ or $\mathbb{E}(u_i' M_\iota X_i W_i X_i' M_\iota u_i) = \sigma_i^2 \mathbb{E}(\operatorname{tr}[M_\iota X_i W_i X_i' M_\iota])$ with $W_i = I_K$ for the JL test and $W_i = (\sigma_i^2 X_i' M_\iota X_i)^{-1}$ for the PY test.

1.3 Model and Assumptions

Consider a linear panel data model with random coefficients,

$$y_i = X_i \beta_i + \epsilon_i, \tag{1.7}$$

$$\beta_i = \beta + v_i, \tag{1.8}$$

for $i = 1, 2, \ldots, N$, where y_i is a is a $T \times 1$ vector of observations on the dependent variable for cross-section unit i, and X_i is a $T \times K$ matrix of possibly stochastic regressors. To simplify the exposition we assume a balanced panel with the same number of observation in each panel unit (see also Remark 1 of Lemma 1). The vector of random coefficients is decomposed into a common non-stochastic vector β and a vector of individual-specific disturbances v_i. Let $X = [X_1', X_2', \dots, X_N']'$.

In order to construct the LM test statistic for slope homogeneity we start with model (1.7)-(1.8) under stylized assumptions. However, in Section 5 these assumptions will be relaxed to accommodate more general and empirically relevant setups. The following assumptions are imposed on the errors and the regressor matrix:

Assumption 1 *The error vectors are distributed as* $\epsilon_i|X \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_T)$ *and* $v_i|X \overset{iid}{\sim} \mathcal{N}(0, \Sigma_v)$, *where* $\Sigma_v = diag\left(\sigma^2_{v,1}, \dots, \sigma^2_{v,K}\right)$. *The errors* ϵ_i *and* v_j *are independent from each other for all* i *and* j.

Assumption 2 *For the regressors we assume* $\mathbb{E}|x_{it,k}|^{4+\delta} < C < \infty$ *for some* $\delta > 0$, *for all* $i = 1, 2\dots, N$, $t = 1, 2, \dots, T$ *and* $k = 1, 2\dots, K$. *The limiting matrix* $\lim_{N\to\infty} N^{-1}\mathbb{E}(X'X)$ *exists and is positive definite for all* N *and* T.

In Assumption 1, the random components of the slope parameters are allowed to have different variances but we assume that there is no correlation among the elements of v_i. Note that this framework is more general than the one considered by Juhl and Lugovskyy (2014) who assume $\mathbb{E}(v_i v_i') = \overline{\sigma}^2_v I_K$. The latter assumption seems less appealing if there are sizable differences in the magnitudes of the coefficients. Furthermore, the power of the test depends on the scaling of regressors, whereas the (local) power of our test is invariant to a rescaling of the regressors (see Theorem 5). The alternative hypothesis can be further generalized by allowing for a correlation among the elements of the error vector v_i. However, this would increase the dimension of the null hypothesis to $K(K+1)/2$ restrictions and it is therefore not clear whether accounting for the covariances helps to increase the power of the test. Obviously, if all variances are zero, then the covariances are zero as well.[2]

Let $u_i = X_i v_i + \epsilon_i$. Stacking observations with respect to i yields

$$y = X\beta + u, \tag{1.9}$$

where $y = (y_1', \dots, y_N')'$ and $u = (u_1', \dots, u_N')'$. The $NT \times NT$ covariance matrix of u is given by

$$\Omega \equiv \mathbb{E}[uu'|X] = \begin{bmatrix} X_1 \Sigma_v X_1' + \sigma^2 I_T & & 0 \\ & \ddots & \\ 0 & & X_N \Sigma_v X_N' + \sigma^2 I_T \end{bmatrix}.$$

[2]We also conducted Monte Carlo simulations allowing for non-zero diagonal elements in the matrix Σ_v. We found that the results are quite similar to the setting where Σ_v is diagonal.

The hypothesis of fixed homogeneous slope coefficients, $\beta_i = \beta$ for all i, corresponds to testing

$$H_0: \ \sigma_{v,k}^2 = 0, \text{ for } k = 1, ..., K,$$

against the alternative

$$H_1: \ \sum_{k=1}^{K} \sigma_{v,k}^2 > 0, \tag{1.10}$$

that is, under the alternative at least one of the variance parameters is larger than zero.

1.4 The LM Test for Slope Homogeneity

Let $\theta = \left(\sigma_{v,1}^2, ..., \sigma_{v,K}^2, \sigma^2\right)'$. Under Assumption 1 the corresponding log-likelihood function results as

$$\ell\left(\beta, \theta\right) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\log\left|\Omega\left(\theta\right)\right| - \frac{1}{2}\left(y - X\beta\right)'\Omega\left(\theta\right)^{-1}\left(y - X\beta\right). \tag{1.11}$$

The restricted ML estimator of β under the null hypothesis coincides with the pooled OLS estimator $\widetilde{\beta} = (X'X)^{-1}X'y$ and the corresponding residual vector and estimated residual variance are denoted by $\widetilde{u}_i = y_i - X_i\widetilde{\beta}$ and $\widetilde{\sigma}^2$. The following lemma presents the score and the information matrix derived from the log-likelihood function in (1.11).

Lemma 1 *The score vector evaluated under the null hypothesis is given by*

$$\widetilde{\mathcal{S}} \equiv \left.\frac{\partial \ell}{\partial \theta}\right|_{H_0} = \frac{1}{2\widetilde{\sigma}^4}\begin{bmatrix} \sum_{i=1}^{N}\left(\widetilde{u}_i' X_i^{(1)} X_i^{(1)\prime}\widetilde{u}_i - \widetilde{\sigma}^2 X_i^{(1)\prime} X_i^{(1)}\right) \\ \vdots \\ \sum_{i=1}^{N}\left(\widetilde{u}_i' X_i^{(K)} X_i^{(K)\prime}\widetilde{u}_i - \widetilde{\sigma}^2 X_i^{(K)\prime} X_i^{(K)}\right) \\ 0 \end{bmatrix}, \tag{1.12}$$

where $X^{(k)}$ is the k-th column of X for $k = 1, 2, \ldots, K$.

The information matrix evaluated under the null hypothesis is

$$\mathcal{I}(\widetilde{\sigma}^2) \equiv -\,\mathbb{E}\left[\frac{\partial^2 \ell}{\partial\theta\partial\theta'}\right]\Bigg|_{H_0}$$

$$= \frac{1}{2\widetilde{\sigma}^4}\begin{bmatrix} \sum_{i=1}^{N}\left(X_i^{(1)\prime}X_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(X_i^{(1)\prime}X_i^{(K)}\right)^2 & X^{(1)\prime}X^{(1)} \\ \sum_{i=1}^{N}\left(X_i^{(2)\prime}X_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(X_i^{(2)\prime}X_i^{(K)}\right)^2 & X^{(2)\prime}X^{(2)} \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^{N}\left(X_i^{(K)\prime}X_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(X_i^{(K)\prime}X_i^{(K)}\right)^2 & X^{(K)\prime}X^{(K)} \\ X^{(1)\prime}X^{(1)} & \cdots & X^{(K)\prime}X^{(K)} & NT \end{bmatrix}, \quad (1.13)$$

where $X_i^{(k)}$ *denotes the* k*-th column of the* $T \times K$ *matrix* X_i*,* $k = 1, 2, \ldots, K$ *and* $i = 1, ..., N$*.*

Remark 1 It is straightforward to extend Lemma 1 to unbalanced panel data, where observations are assumed to be missing at random. Let X_i be a $T_i \times K$ matrix and $\widetilde{u}_i$ be a conformable $T_i \times 1$ vector. The score vector is given by

$$\widetilde{\mathcal{S}} = \frac{1}{2\widetilde{\sigma}^4}\begin{bmatrix} \sum_{i=1}^{N}\left(\widetilde{u}_i' X_i^{(1)} X_i^{(1)\prime}\widetilde{u}_i - \widetilde{\sigma}^2 X_i^{(1)\prime}X_i^{(1)}\right) \\ \vdots \\ \sum_{i=1}^{N}\left(\widetilde{u}_i' X_i^{(K)} X_i^{(K)\prime}\widetilde{u}_i - \widetilde{\sigma}^2 X_i^{(K)\prime}X_i^{(K)}\right) \\ 0 \end{bmatrix},$$

where

$$\widetilde{\sigma}^2 = \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\widetilde{u}_i'\widetilde{u}_i \,.$$

The information matrix is computed accordingly.

Remark 2 If individual-specific constants α_i are included in the regression, then a conditional version of the test is available (cf. Juhl and Lugovskyy (2014)). The individual effects can be "conditioned out" by considering the transformed regression

$$M_\iota y_i = M_\iota X_i \beta + M_\iota u_i \,, \quad (1.14)$$

with M_ι as defined in Section 2. The typical elements of the corresponding score vector result as

$$\frac{1}{\widetilde{\sigma}^4}\left(\widetilde{u}_i' M_\iota X_i^{(j)} X_i^{(j)\prime} M_\iota \widehat{u}_i - \widetilde{\sigma}^2 X_i^{(j)\prime} M_\iota X_i^{(j)}\right), \quad j = 1, \ldots, K,$$

where $\widetilde{u}_i = M_\iota y_i - M_\iota X_i \widetilde{\beta}$ and $\widetilde{\beta}$ is the pooled OLS estimator of the transformed model (1.14), and $\widetilde{\sigma}^2$ is the corresponding estimated residual variance. It follows that we just have to replace the vector $X_i^{(j)}$ by the mean-adjusted vector $M_\iota X_i^{(j)}$ in Theorem 1.

Remark 3 It is easy to see that under the more restrictive alternative $\mathbb{E}(v_i v_i') = \overline{\sigma}_v^2 I_K$ of Juhl and Lugovskyy (2014), where $\sigma_{v,1}^2 = \cdots = \sigma_{v,k}^2 = \overline{\sigma}_v^2$, the score is simply the sum of all elements of $\widetilde{S}$.

Remark 4 Notice also that the LM-type statistics do not require the restriction $K < T$, which is important for the PY approach. This is of course not an issue for the asymptotic framework, where $T \to \infty$, however, it can be a substantive restriction in many empirical applications when T is small.

In the following theorem it is shown that when T is fixed, the LM statistic possesses a χ^2 limiting null distribution with K degrees of freedom as $N \to \infty$.

Theorem 1 *Under Assumptions 1, 2 and the null hypothesis*

$$LM = \widetilde{\mathcal{S}}' \mathcal{I}(\widetilde{\sigma}^2)^{-1} \widetilde{\mathcal{S}} = \widetilde{s}' \, \widetilde{V}^{-1} \widetilde{s} \xrightarrow{d} \chi_K^2, \tag{1.15}$$

as $N \to \infty$ and T is fixed, where $\widetilde{s}$ is defined as the $K \times 1$ vector with typical element

$$\widetilde{s}_k = \frac{1}{2\widetilde{\sigma}^4} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \widetilde{u}_{it} x_{it,k} \right)^2 - \frac{1}{2\widetilde{\sigma}^2} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it,k}^2, \tag{1.16}$$

and the (k,l) element of the matrix $\widetilde{V}$ is given by

$$\widetilde{V}_{k,l} = \frac{1}{2\widetilde{\sigma}^4} \left[\sum_{i=1}^{N} \left(\sum_{t=1}^{T} x_{it,k} x_{it,l} \right)^2 - \frac{1}{NT} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it,k}^2 \right) \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it,l}^2 \right) \right]. \tag{1.17}$$

Remark 5 If T is fixed, normality of the regression disturbances is required. If we relax the normality assumption, an additional term enters the variance of the score vector and the information matrix becomes an inconsistent estimator. Theorem 2 discusses this issue in more details and derives the asymptotic distribution of the LM test if the errors are not normally distributed.

Remark 6 It may be of interest to restrict attention to a subset of coefficients. For example, in the classical panel data model it is assumed that the constants are individual-specific and, therefore, the respective parameters are not included in the null hypothesis. Another possibility is that a subset of coefficients is assumed to be constant across all panel units. To account for such specifications the model is partitioned as

$$y_{it} = \beta_{1i}' X_{it}^a + \beta_2' X_{it}^b + \beta_{3i}' X_{it}^c + u_{it} \ .$$

The $K_1 \times 1$ vector X_{it}^a includes all regressors that are assumed to have individual-specific coefficients stacked in the vector β_{1i}. The $K_2 \times 1$ vector X_{it}^b comprises all regressors that are supposed to have homogenous coefficients. The null hypothesis is that the coefficient vector β_{3i} attached to the $K_3 \times 1$ vector of regressors X_{it}^c is identical for all panel units, that is, $\beta_{3i} = \beta_3$ for all i, where $\beta_{3i} = \beta_3 + v_{3i}$. The null hypothesis implies $\Sigma_{v_3} = 0$. Let

$$Z = \begin{bmatrix} X_1^a & 0 & \cdots & 0 & X_1^b & X_1^c \\ 0 & X_2^a & \cdots & 0 & X_2^b & X_2^c \\ \vdots & & \ddots & \vdots & & \\ 0 & 0 & \cdots & X_N^a & X_N^b & X_N^c \end{bmatrix},$$

where $X_i^a = [X_{i1}^a, \ldots, X_{iT}^a]'$ and the matrices X_i^b and X_i^c are defined accordingly. The residuals are obtained as $\widetilde{u} = (I - Z(Z'Z)^{-1}Z')y$ and the columns of the matrix X^c are used to compute the LM statistic. Some caution is required if a set of individual-specific coefficients are included in the panel regression since in this case the ML estimator $\widetilde{\sigma}^2 = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \widetilde{u}_{it}^2$ is inconsistent for fixed T and $N \to \infty$. This implies that the expectation of the score vector (1.12) is different from zero. Accordingly, the unbiased estimator

$$\widehat{\sigma}^2 = \frac{1}{NT - K_1 - K_2 - K_3} \sum_{i=1}^N \sum_{t=1}^T \widetilde{u}_{it}^2 \tag{1.18}$$

must be employed. As a special case, assume that the constant is included in X_i^c, whereas all other regressors are included in the matrix X_i^b, and X_i^a is dropped. This case is equivalent to the test for random individual effects as suggested by Breusch and Pagan (1980). The LM statistic then reduces to

$$LM = \frac{NT}{2(T-1)} \left[1 - \frac{\widetilde{u}' (I_N \otimes \iota_T \iota_T') \widetilde{u}}{\widetilde{u}'\widetilde{u}} \right]^2,$$

where ι_T is a $T \times 1$ vector of ones, which is identical to the familiar LM statistic for random individual effects.

1.5 Variants of the LM Test

In this section we generalized the LM test statistic by allowing for non-normally distributed, heteroskedastic and serially dependent errors. First we show in Section 1.5 that the proposed LM test is robust against non-normally distributed errors once we assume $N, T \to \infty$ jointly and specific restrictions on the existence of higher-order moments. Moreover, the variants of the test with non-normally distributed errors are proposed for the settings when $N \to \infty$ and T is fixed. Second, in Section 1.5.2 we propose a variant of the LM test that is robust to heteroskedastic errors. Finally, Section 1.5.3 discusses how to robustify the LM test, when the errors are serially correlated.

1.5.1 The LM statistic under non-normality

In this section we consider useful variants of the original LM statistic under the assumption that the errors are not normally distributed. Therefore, we replace Assumptions 1 and 2 by:

Assumption 1′ *ϵ_{it} is independently and identically distributed with $\mathbb{E}(\epsilon_{it}|X) = 0$, $\mathbb{E}(\epsilon_{it}^2|X) = \sigma^2$ and $\mathbb{E}\left(|\epsilon_{it}|^6|X\right) < C < \infty$ for all i and t. Furthermore, ϵ_{it} and ϵ_{js} are independently distributed for $i \neq j$ and $t \neq s$.*

Assumption 2′ *For the regressors we assume $\mathbb{E}|x_{it,k}|^6 < C < \infty$ for some $\delta > 0$, for all $i = 1, 2 \ldots, N$, $t = 1, 2, \ldots, T$ and $k = 1, 2 \ldots, K$. Further, $\lim\limits_{T\to\infty} T^{-1} \sum\limits_{t=1}^{T} \mathbb{E}\left[x_{it}x_{it}'\right]$ tend to a positive definite matrix Q_i and the limiting matrix $Q := \lim\limits_{N,T\to\infty} (NT)^{-1} \sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} \mathbb{E}\left[x_{it}x_{it}'\right]$ exists and is positive definite.*

Assumption 3 *The error vector v_i is independently and identically distributed with $\mathbb{E}(v_i|X) = 0$, $\mathbb{E}(v_iv_i'|X) = \Sigma_v$, where $\Sigma_v = diag\left(\sigma_{v,1}^2, \ldots, \sigma_{v,K}^2\right)$ and $\mathbb{E}\left(|v_{ik}|^{2+\delta}\,|X\right) < C < \infty$ for some $\delta > 0$, for all i and $k = 1, ..., K$. Further, v_i and ϵ_j are independent from each other for all i and j.*

Notice that, as in Section 1.3 under the null hypothesis Σ_v or $v_i = 0$ for all i. Hence, Assumption 3 is not required for the derivation of the asymptotic null distribution. To study the behaviour of the LM test statistic under (local) alternatives, Assumption 3 will be used in Section 1.6.

With these modifications of the previous setup, the limiting distribution of the LM statistic is given in

Theorem 2 *Under Assumptions 1′, 2′ and the null hypothesis,*

$$LM \xrightarrow{d} \chi_K^2, \tag{1.19}$$

as $N \to \infty$, $T \to \infty$ jointly.

Generalizing the model to allow for non-normally distributed errors introduces a new term into the variance of the score: the (k, l) element of the covariance matrix now becomes (see equation (A.5) in appendix A.2)

$$V_{k,l}+\left(\frac{\mu_u^{(4)}-3\sigma^4}{\left(2\sigma^4\right)^2}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it,k}^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it,k}^2\right)\left(x_{it,l}^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it,l}^2\right), \tag{1.20}$$

where $\mu_u^{(4)}$ denotes the fourth moment of the error distribution, and $V_{k,l}$ is as in (1.17) with $\widetilde{\sigma}^4$ replaced by σ^4. The additional term depends on the excess kurtosis $\mu_u^{(4)}-3\sigma^4$. Clearly, for normally distributed errors, this term disappears, but it deviates from zero in the more general setup. Under Assumptions $1'$ and $2'$, the first term $V_{k,l}$ is of order NT^2, while the new component is of order NT, such that, when the appropriate scaling underlying the LM statistic is adopted, it vanishes as $T\to\infty$. Therefore, the LM statistic as presented in the previous section continues to be χ_K^2 distributed asymptotically. By incorporating a suitable estimator of the second term in (1.20), however, a test statistic becomes available that is valid in a framework with non-normally distributed errors as $N\to\infty$, whether T is fixed or $T\to\infty$. Therefore, denote the adjusted LM statistic by

$$LM_{\text{adj}}=\widetilde{s}'\left(\widetilde{V}_{\text{adj}}\right)^{-1}\widetilde{s},$$

where $\widetilde{V}_{\text{adj}}$ is as in (1.20) with $V_{k,l}$, σ^4 and $\mu_u^{(4)}$ replaced by the consistent estimators $\widetilde{V}_{k,l}$ defined in (1.17), $\widetilde{\sigma}^4$ and $\widetilde{\mu}_u^{(4)}=(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{u}_{it}^4$ for $k,l=1,...,K$. As a consequence of Theorem 2 and the preceding discussion, we obtain the following result.

Corollary 1 *Under Assumptions $1'$, 2 and the null hypothesis*

$$LM_{\text{adj}}\xrightarrow{d}\chi_K^2,$$

as $N\to\infty$ and T is fixed. Furthermore,

$$LM_{\text{adj}}-LM\xrightarrow{p}0,$$

as $N\to\infty$, $T\to\infty$ jointly.

As mentioned above, once the regression disturbances are no longer normally distributed, the fourth moments of the error distribution enter the variance of the score. It is insightful to identify exactly which terms give rise to this new form of the covariance matrix. According to Lemma 1, the contribution of the i-th panel unit to the k-th element of the score vector is

$$\widetilde{u}_i'X_i^{(k)}X_i^{(k)\prime}\widetilde{u}_i-\widetilde{\sigma}^2X_i^{(k)\prime}X_i^{(k)}=\left(\sum_{t=1}^{T}x_{it,k}^2\left(\widetilde{u}_{it}^2-\widetilde{\sigma}^2\right)\right)+\sum_{t=1}^{T}\sum_{s\neq t}\widetilde{u}_{it}\widetilde{u}_{is}x_{it,k}x_{is,k}. \tag{1.21}$$

The variance of the first term on the right hand side depends on the fourth moments of the errors. Since the contribution of this term vanishes if T gets large, it can be dropped without any severe effect on the power whenever T is sufficiently large. Hence, we consider a modified score vector as presented in the following theorem.

Theorem 3 *Under Assumptions 1′, 2 and the null hypothesis, the modified LM statistic*

$$LM^* = \tilde{s}^{*\prime} \left(\widetilde{V}^*\right)^{-1} \tilde{s}^* \xrightarrow{d} \chi^2_K,$$

as $N \to \infty$ and T fixed, where $\tilde{s}^*$ is $K \times 1$ vector with contributions for panel unit i

$$\tilde{s}^*_{i,k} = \frac{1}{\widetilde{\sigma}^4} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \widetilde{u}_{it} \widetilde{u}_{is} x_{it,k} x_{is,k}, \tag{1.22}$$

for $i = 1, ..., N$, $k = 1, ..., K$, and the (k, l) element of $\widetilde{V}^*$ is given by

$$\widetilde{V}^*_{k,l} = \frac{1}{\widetilde{\sigma}^4} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} x_{it,k} x_{it,l} x_{is,k} x_{is,l}, \tag{1.23}$$

for $k, l = 1, ..., K$.

Remark 7 It is important to note that this version of the LM test is invalid if the panel regression allows for individual-specific coefficients (cf. Remark 3). Consider for example the regression

$$y_{it} = \alpha_i + x'_{it}\beta_i + u_{it} \tag{1.24}$$

where α_i are fixed individual effects and we are interested in testing $H_0 : var(\beta_i) = 0$. The residuals are obtained as

$$\widetilde{u}_{it} \;=\; y_{it} - \overline{y}_i - (x_{it} - \overline{x}_i)'\widetilde{\beta} \;=\; u_{it} - \overline{u}_i - (x_{it} - \overline{x}_i)'(\widetilde{\beta} - \beta).$$

It follows that in this case $E(\widetilde{u}_{it}\widetilde{u}_{is}x_{it,k}x_{is,k}) \neq 0$ and, therefore, the modified scores (1.22) result in a biased test. To sidestep this difficulty, orthogonal deviations (e.g. Arellano and Bover (1995)) can be employed to eliminate the individual-specific constants yielding

$$y^*_{it} = \beta' x^*_{it} + u^*_{it} \quad t = 2, 3, \ldots, T,$$
$$\text{with } y^*_{it} = \sqrt{\frac{t-1}{t}} \left[y_{it} - \frac{1}{t-1} \left(\sum_{s=1}^{t-1} y_{is} \right) \right],$$

where x^*_{it} and u^*_{it} are defined analogously. It is well known that if u_{it} is i.i.d. so is u^*_{it}. It follows that the modified LM statistic can be constructed by using the OLS residuals $\widetilde{u}^*_{it}$

instead of $\widetilde{u}_{it}$. This approach can be generalized to arbitrary individual-specific regressors x^a_{it}. Let $X^a_i = [x^a_{i1}, \ldots, x^a_{iT}]'$ denote the individual-specific $T \times K_1$ regressor matrix in the regression

$$y_i = X^a_i \beta_{1i} + X^b_i \beta_2 + X^c_i \beta_{3i} + u_i, \tag{1.25}$$

(see Remark 3). Furthermore, let

$$M^a_i \;=\; I_T - X^a_i (X^{a\prime}_i X^a_i)^{-1} X^{a\prime}_i,$$

and let $\widetilde{M}^a_i$ denote the $(T - K_1) \times T$ matrix that results from eliminating the last K_1 rows from M^a_i such that $(M^a_i M^{a\prime}_i)$ is of full rank. The model (1.25) is transformed as

$$y^*_i = X^{b*}_i \beta_2 + X^{c*}_i \beta_{3i} + u^*_i \;, \tag{1.26}$$

where $y^*_i = \Xi^a_i y_i$ and $\Xi^a_i = (\widetilde{M}^a_i \widetilde{M}^{a\prime}_i)^{-1/2} \widetilde{M}^a_i$. It is not difficult to see that $E(u^*_i u^{*\prime}_i) = \sigma^2 I_{T-K_1}$ and, thus, the modified scores (1.22) can be constructed by using the residuals of (1.26), where the time series dimension reduces to $T - K_1$. Note that orthogonal deviations result from letting X^a_i be a vector of ones.

To review the results of this section, the important new feature in the model without assuming normality is that the fourth moments of the errors enter the variance of the score. The information matrix of the original LM test derived under normality does not incorporate higher order moments, but the test remains applicable as $T \to \infty$. To apply the LM test in the original framework when T is fixed and errors are no longer normal we can proceed in two ways. A direct adjustment of the information matrix to account for higher order moments yields a valid test. Alternatively, we can adjust the score itself and restrict attention to that part of the score that does not introduce higher order moments into the variance. In the next section, we further pursue the second route of dealing with non-normality and thereby robustify the test against heteroskedasticity and serial correlation.

1.5.2 The regression-based LM statistic

In this section we offer a convenient way to compute the proposed LM statistic via a simple artificial regression. Moreover, the regression-based form of the LM test is shown to be robust against heteroskedastic errors. Following the decomposition of the score contribution in (1.21) and the discussion thereafter, we construct the "Outer Product of Gradients" (OPG) variant of the LM test based on the second term in (1.21). Rewriting the corresponding elements of the score contributions of panel unit i as

$$\tilde{s}^*_{i,k} \;=\; \sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{u}_{it}\widetilde{u}_{is}x_{it,k}x_{is,k}, \tag{1.27}$$

for $k = 1, ..., K$. Note that we dropped the factor $1/\widetilde{\sigma}^4$ as this factor cancels out in the final test statistic. This gives the usual LM-OPG variant

$$LM_{\text{opg}} = \left(\sum_{i=1}^{N}\tilde{s}^*_i\right)'\left(\sum_{i=1}^{N}\tilde{s}^*_i\tilde{s}^{*\prime}_i\right)^{-1}\left(\sum_{i=1}^{N}\tilde{s}^*_i\right), \tag{1.28}$$

where $\tilde{s}^*_i = \left[\tilde{s}^*_{i,1}, ..., \tilde{s}^*_{i,K}\right]'$. An asymptotically equivalent form of the LM-OPG statistic can be formulated as a Wald-type test for the null hypothesis $\varphi = 0$ in the auxiliary regression

$$\widetilde{u}_{it} \;=\; \sum_{k=1}^{K}\widetilde{z}_{it,k}\varphi_k + e_{it}, \quad \text{for } i = 1, \ldots, N, \quad t = 1, \ldots, T \tag{1.29}$$

where

$$\widetilde{z}_{it,k} = x_{it,k}\sum_{s=1}^{t-1}\widetilde{u}_{is}x_{is,k}$$

for $k = 1, ..., K$. Therefore, with the Eicker-White heteroskedasticity-consistent variance estimator, the regression based test statistic results as

$$LM_{\text{reg}} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\widetilde{u}_{it}\widetilde{z}_{it}\right)'\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\widetilde{u}^2_{it}\widetilde{z}_{it}\widetilde{z}'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\widetilde{u}_{it}\widetilde{z}_{it}\right), \tag{1.30}$$

It follows from the arguments similar as in Theorem 3 that M_{reg} test statistic is asymptotically χ^2 distributed but it turns out to be robust against heteroskedasticity:

Corollary 2 *Under Assumption 1′ but allowing for heteroscedastic errors such that* $\mathbb{E}[\epsilon^2_{it}|X] = \sigma^2_{it} < C < \infty$, *Assumption 2 and the null hypothesis*

$$LM_{reg} \xrightarrow{d} \chi^2_K, \tag{1.31}$$

as $N \to \infty$ *and* T *is fixed.*

It is important to note that the LM-OPG variant cannot be applied to residuals from a fixed effect regression, see Remark 7. Furthermore, the replacement of the residuals by orthogonal forward deviation will not fix this problem since orthogonal forward deviations are no longer serially uncorrelated if the errors are heteroskedastic. Therefore, a version of the test is required that is robust against autocorrelated errors.

1.5.3 The LM statistic under serially dependent errors

In this section we propose a variant of the LM test statistic that accommodates serially correlated errors, that is, we relax Assumptions $1'$ as follows:

Assumption $1''$ *The $T \times 1$ error vector ϵ_i is independently and identically distributed with $\mathbb{E}(\epsilon_i|X) = 0$, $\mathbb{E}(\epsilon_i\epsilon_i'|X) = \mathbb{E}(\epsilon_i\epsilon_i') = \Sigma$ and $\mathbb{E}\left[|\epsilon_{it}|^{4+\delta}|X\right] < C < \infty$ for some $\delta > 0$ and all i and t. The $T \times T$ matrix Σ is positive definite with typical element σ_{ts} for $t, s = 1, \ldots, T$.*

Note that Assumption $1''$ allows for heteroscedasticity and serial dependence across time, however, it restricts the error vector ϵ_i to be iid across individuals.

Under this assumption the expectation of the score vector (1.22) is under the null hypothesis

$$\mathbb{E}[u_{it}u_{is}x_{it,k}, x_{is,k}] = \sigma_{ts}\mathbb{E}[x_{it,k}x_{is,k}].$$

We therefore suggest a modification for autocorrelated errors based on the adjusted $K \times 1$ score vector $\tilde{s}^{**}$ with typical element $\tilde{s}^{**}_k = \sum_{i=1}^N \tilde{s}^{**}_{i,k}$ for $k = 1, \ldots, K$ and

$$\tilde{s}^{**}_{i,k} = \sum_{t=2}^{T}\sum_{s=1}^{t-1} \left(\widetilde{u}_{it}\widetilde{u}_{is} - \widetilde{\sigma}_{ts}\right) x_{it,k}x_{is,k} \ , \tag{1.32}$$

where $\widetilde{\sigma}_{ts} = \frac{1}{N}\sum_{i=1}^N \widetilde{u}_{it}\widetilde{u}_{is}$. The asymptotic properties of the LM statistic based on the modified score vector are presented in

Theorem 4 *Let*

$$LM_{\text{ac}} = \tilde{s}^{**\prime}\left(\widetilde{V}^{**}\right)^{-1}\tilde{s}^{**},$$

where $\widetilde{V}^{**}$ is a $K \times K$ matrix with typical element

$$\widetilde{V}^{**}_{k,l} = \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{\tau=2}^{T}\sum_{s=1}^{t-1}\sum_{q=1}^{t-1} \widehat{\delta}_{ts\tau q}x_{it,k}x_{is,k}x_{i\tau,l}x_{iq,l} \tag{1.33}$$

$$\text{and} \quad \widehat{\delta}_{ts\tau q} = \frac{1}{N}\left(\sum_{j=1}^{N}\widetilde{u}_{jt}\widetilde{u}_{js}\widetilde{u}_{j\tau}\widetilde{u}_{jq} - \widetilde{\sigma}_{ts}\widetilde{\sigma}_{\tau q}\right).$$

Under Assumptions $1''$, 2, the null hypothesis (1.1) and as $N \to \infty$ with T fixed the LM_{ac} statistic has a χ^2_K limiting distribution.

Note that this version of the test has a good size control irrespective of serial dependence in errors. However, the test involves some power loss relative to the original test statistics when errors are serially uncorrelated, which is not surprising given a more general setup of this variant of the test. The respective asymptotic power results are analyzed in the next section (see Remark 10). Section 1.7 elaborates in detail on the size-power properties of the LM_{ac} in finite samples.

1.6 Local Power

The aim of this section is twofold. First, we investigate the distributions of the LM-type test under suitable sequences of local alternatives. Two cases are of interest, $N \to \infty$ with T fixed and $N, T \to \infty$ jointly, which are presented in the respective theorems below. Second, we adopt the results of PY to our model in order to compare the local asymptotic power of the two tests. To formulate an appropriate sequence of local alternatives, we specify the random coefficients in (1.8) in a setup in which T is fixed. The error term v_i is as in Assumption 1 with elements of Σ_v given by

$$\sigma_{v,k}^2 = \frac{c_k}{\sqrt{N}}, \tag{1.34}$$

where $c_k > 0$ are fixed constants for $k = 1, \dots, K$. The asymptotic distribution of the LM statistic results as follows.

Theorem 5 *Under Assumptions 1, 2 and the sequence of local alternatives (1.34),*

$$LM \xrightarrow{d} \chi_K^2(\mu),$$

as $N \to \infty$ and T fixed, with non-centrality parameter $\mu = c'\Psi c$, where $c = (c_1, \dots, c_K)'$ and Ψ is a $K \times K$ matrix with (k,l) element

$$\Psi_{k,l} = \frac{1}{2\sigma^4} \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} x_{it,k} x_{it,l} \right)^2 - \frac{1}{T} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it,k}^2 \right) \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it,l}^2 \right) \right].$$

In order to relax the assumption of normally distributed errors we adopt Assumption $1'$ for v_i, where the sequence of local alternatives is now given by

$$\sigma_{v,k}^2 = \frac{c_k}{T\sqrt{N}}, \tag{1.35}$$

for $k = 1, \dots, K$. Note that according to Theorem 2 we require $T \to \infty$.

Theorem 6 *Under Assumptions $1'$, $2'$, 3 and the sequence of alternatives (1.35),*

$$LM \xrightarrow{d} \chi_K^2(\mu),$$

as $N \to \infty$, $T \to \infty$, with non-centrality parameter $\mu = c'\Psi c$, where $c = (c_1, \dots, c_K)'$ and Ψ is a $K \times K$ matrix with (k,l) element

$$\Psi_{k,l} = \frac{1}{2\sigma^4} \operatorname*{plim}_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} x_{it,k} x_{it,l} \right)^2.$$

Remark 8 As in Section 1.5.1 above, when the normality assumption is relaxed, local power can be studied for LM^* under Assumptions $1'$, 2 and 3 when T is fixed. The specification of local alternatives as in Theorem 5 applies. The non-centrality parameter of the limiting non-central χ^2 distribution results as $\mu^* = c'\Psi^* c$ with

$$\Psi^*_{k,l} = \frac{1}{\sigma^4} \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} x_{it,k} x_{it,l} x_{is,k} x_{is,l} \ ,$$

for $k, l = 1, ..., K$.

Remark 9 Given the results for the modified statistic LM^* in remark 8, and the fact that $\tilde{s}^* = \sum_{i=1}^{N} \tilde{s}_i^* = \left(\widetilde{Z}'\widetilde{u}^*\right)$, we expect a similar result for the regression-based LM statistic LM_{reg} to hold. Recall that LM^* uses $N^{-1}\widetilde{V}^*$ as an estimator of the variance of $\tilde{s}^*$ (see (1.23)), while LM_{reg} employs $\left(N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} \widetilde{u}_{it}^2 \widetilde{z}_{it} \widetilde{z}_{it}'\right)$. Under the null hypothesis, it is not difficult to see that these two estimators are asymptotically equivalent. Under the alternative, when studying the (k, l) element of the variance of LM_{reg}, we obtain (see appendix A.2 for details)

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} \widetilde{u}_{it}^2 \widetilde{z}_{it,k} \widetilde{z}_{it,l} = {} & \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} \epsilon_{it}^2 x_{it,k} x_{it,l} \left(\sum_{s=1}^{t-1} \epsilon_{is} x_{is,k} \right) \left(\sum_{s=1}^{t-1} \epsilon_{is} x_{is,l} \right) \\ & + \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} \epsilon_{it}^2 v_i' B_{it}^X v_i + o_p(1) \,, \end{aligned} \tag{1.36}$$

with the $K \times K$ matrix $B_{it}^X = \left(x_{it,k}\sum_{s=1}^{t-1} x_{is} x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1} x_{is}' x_{is,l}\right)$. The first term on the right-hand side in (1.36) has the same probability limit as $N^{-1}\widetilde{V}^*_{k,l}$, the limiting covariance matrix element $\Psi^*_{k,l}$. In contrast to LM^*, however, the variance estimator of the regression-based test involves additional quadratic forms such as $v_i' B_{it}^X v_i$, contributing to the estimator. Since, in a setup with fixed T and the local alternatives $\sigma^2_{v,k} = \frac{c_k}{\sqrt{N}}$,

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} \epsilon_{it}^2 v_i' B_{it}^X v_i = O_p\left(N^{-1/2}\right),$$

the variance estimator remains consistent. In small samples, however, the additional term results in a bias of the variance estimator and may deteriorate the power of the regression-based test. See the appendix for details about the above result and the Monte Carlo experiments in Section 1.7.

Remark 10 The arguments of Remark 8 can be used to derive the local power of the LM_{ac} statistic that accounts for serial correlation in errors. The same specification of local alternatives applies. The non-centrality parameter of the limiting non-central χ^2 distribution takes the quadratic form $\mu^{**} = c'\Psi^{**}c$ with

$$\Psi^{**}_{k,l} = \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{\tau=2}^{T}\sum_{q=1}^{t-1}\left(u_{it}u_{is}u_{i\tau}u_{iq} - \sigma_{ts}\sigma_{\tau q}\right)x_{it,k}x_{is,k}x_{i\tau,l}x_{iq,l},$$

for $k,l = 1,...,K$. In the absence of serial correlation it can be shown that the LM_{ac} test involve a loss of power. To illustrate this fact assume for simplicity that $K = 1$ (single regressor case). Further, the score vector in (1.32) can be equivalently written as

$$\hat{s}^{**} = \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\widetilde{u}_{it}\widetilde{u}_{is} - \widetilde{\sigma}_{ts}\right)x_{it,k}x_{is,k} = \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{u}_{it}\widetilde{u}_{is}\left(x_{it,k}x_{is,k} - \overline{C}_{ts}\right), \qquad (1.37)$$

where $\overline{C}_{ts} = \frac{1}{N}\sum_{i=1}^{N} x_{it}x_{is}$. Thus, demeaning of $\widetilde{u}_{it}\widetilde{u}_{is}$ is equivalent with demeaning of $x_{it,k}x_{is,k}$. In the case of no autocorrelation and (1.37) it follows that $\Psi^{*} - \Psi^{**}$ is positive semi-definite. Therefore, the modification (1.32) tends to reduce the power of the LM_{ac} test when compared to LM^{*}.

We now proceed to examine the local power of the Δ statistic of PY in model (1.7) and (1.8) under the sequence of local alternatives (1.35). In our homoskedastic setup, the dispersion statistic becomes

$$\widetilde{S} \;=\; \sum_{i=1}^{N}\left(\widetilde{\beta}_i - \widetilde{\beta}\right)'\left(\frac{X_i'X_i}{\widetilde{\sigma}^2}\right)\left(\widetilde{\beta}_i - \widetilde{\beta}\right),$$

with $\widetilde{\beta}$ as the OLS estimator in (1.9) as above. Using this expression, the $\widehat{\Delta}$ statistic is computed as in (1.3). The next theorem presents the asymptotic distribution of the $\widehat{\Delta}$ statistic under the local alternatives as specified above. This result follows directly from Section 3.2 in PY.

Theorem 7 *Under Assumptions 1′, 2′, 3 and the sequence of local alternatives (1.35)*

$$\widehat{\Delta} \xrightarrow{d} \mathcal{N}(\lambda, 1),$$

as $N\to\infty$, $T\to\infty$, *provided* $\sqrt{N}/T \to 0$, *where* $\lambda = \Lambda' c/\sqrt{2K}$ *and* Λ *is a* $K\times 1$ *vector with typical element*

$$\Lambda_k = \frac{1}{\sigma^2}\underset{N,T\to\infty}{\text{plim}}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x^2_{it,k},$$

for $k = 1,\ldots,K$.

In Theorem 7, the mean of the limiting distribution of $\widehat{\Delta}$ is slightly different from the result in Section 3.2 in PY. Here, v_i is random and independently distributed from the regressors and, therefore, the second term of the respective expression in PY is zero.

Remark 11 Consider for simplicity a scalar regressor x_{it} that is i.i.d. across i and t with uniformly bounded fourth moments. Let $\mathbb{E}\left[x_{it}\right] = 0$ and $\mathbb{E}\left[x_{it}^2\right] = \sigma_{i,x}^2$, that is, the regressor is assumed to have a unit-specific variation which is constant over time for a given unit. We obtain

$$\mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}^2\right)^2\right] = \left(\sigma_{i,x}^2\right)^2 + O\left(T^{-1}\right),$$

implying $\mu = c^2/2\sigma^4 \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\left(\sigma_{i,x}^2\right)^2$ in Theorem 6. To gain further insight, we think of $\left(\sigma_{i,x}^2\right)^2$ as being randomly distributed in the cross-section such that the non-centrality parameter results as

$$\mu = \frac{c^2}{2\sigma^4}\mathbb{E}\left[\left(\sigma_{i,x}^2\right)^2\right] = \frac{c^2}{2\sigma^4}\left(Var\left[\sigma_{i,x}^2\right] + \left(\mathbb{E}\left[\sigma_{i,x}^2\right]\right)^2\right). \tag{1.38}$$

Similarly, under these assumptions, we find

$$\lambda = \frac{c}{\sigma^2\sqrt{2}}\mathbb{E}\left[\sigma_{i,x}^2\right]. \tag{1.39}$$

Comparing the mean of the normal distribution of the Δ statistic in (1.39) with the non-centrality parameter of the asymptotic χ_1^2 distribution of the LM statistic in (1.38), we see that the main difference between the two tests is that the variance of $\sigma_{i,x}^2$ contributes to the power of the LM statistic but not to the power of the Δ test. If $Var\left[\sigma_{i,x}^2\right] = 0$ such that $\sigma_{i,x}^2 = \sigma_x^2$ for all i, the LM test and the Δ test have the same asymptotic power in this example. If, however, $Var\left[\sigma_{i,x}^2\right] > 0$, so that there is variation in the variance of the regressor in the cross-section, the LM test has larger asymptotic power. To illustrate this point, we examine the local asymptotic power functions of the LM and the Δ test for two cases, using the expressions in (1.38) and (1.39). Figure 1.1 (see appendix C) shows the local asymptotic power of the LM (solid line) and the Δ test (dashed line) as a function of c when $\sigma_{i,x}^2$ has a χ_1^2 distribution. Figure 1.2 repeats this exercise for $\sigma_{i,x}^2$ drawn from a χ_2^2 distribution. In both cases, the LM test has larger asymptotic power. The power gain is substantial for the first case, but diminishes for the second. This pattern is expected, as the variance of $\sigma_{i,x}^2$ contributes relatively more to the non-centrality parameter in the first specification.

This discussion exemplifies the difference between the LM-type tests and the Δ statistic in terms of the local asymptotic power in a simplified framework. The analysis suggests that the LM-type tests are particularly powerful in an empirically relevant setting in which there is non-negligible variation in the variances of the regressors between panel units. Having studied the large samples properties of the LM tests under the null and the alternative hypothesis in our model, we now evaluate the finite-sample size and power properties of the LM-type tests in a Monte Carlo experiment.

1.7 Monte Carlo Experiments

1.7.1 Design

After deriving LM-type tests in the random coefficient model, we now turn to study the small-sample properties of the proposed test and it variants. The aim of this section is to evaluate the performance of the tests in terms of their empirical size and power in several different setups, relating to the theoretical discussion of Sections 1.4 - 1.6. We consider the following test statistics: the original LM statistic presented in Theorem 1, the adjusted LM statistic that adjusts the information matrix to account for fourth moments of the error distribution (see Corollary 1), the score-modified LM statistics (see Theorem 3 and Theorem 4) and the regression-based, heteroskedasticity-robust LM statistic (see Section 1.5.2). As a benchmark, we consider PY's statistic $\widetilde{\Delta}_{\text{adj}}$ given in (1.4). Following the notes in Table 1 in PY, the test using $\widetilde{\Delta}_{\text{adj}}$ is carried out as a two-sided test. In addition, the CLM test in (1.6) is included, which is also a two-sided test. We consider the following data-generating process with normally distributed errors as the standard design:

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}'\beta_i + \epsilon_{it}, \\
\epsilon_{it} &\overset{iid}{\sim} \mathcal{N}(0,1), \qquad\qquad (1.40)\\
\alpha_i &\overset{iid}{\sim} \mathcal{N}(0,0.25), \\
x_{it,k} &= \alpha_i + \vartheta_{it,k}^x, \;\; k=1,2,3, \\
\vartheta_{it,k}^x &\overset{iid}{\sim} \mathcal{N}\left(0,\sigma_{ix,k}^2\right), \\
\beta_i &\overset{iid}{\sim} \mathcal{N}_3(\iota_3,\Sigma_v), \\
\text{under the null hypothesis: } \Sigma_v &= 0 \qquad\qquad (1.41)\\
\text{under the alternative: } \Sigma_v &= \begin{bmatrix} 0.03 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}, \qquad\qquad (1.42)
\end{aligned}
$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$. Hence, to simulate a model under the null the slope vector β_i is generated as a 3×1 vector of ones ι_3 for all i. As discussed in Section 1.6 the variances of the regressors play an important role. In our benchmark specification we generate the variances as

$$\begin{aligned} \sigma^2_{ix,k} &= 0.25 + \eta_{i,k} \\ \eta_{i,k} &\overset{iid}{\sim} \chi^2_1, \end{aligned} \tag{1.43}$$

The choice of the χ^2 distribution for $\sigma^2_{ix,k}$ is made analogous to the Monte Carlo experiment in PY. We then consider variations of this specification below. All results are based on 5,000 Monte Carlo replications. We choose

$$N \in \{10, 20, 30, 50, 100, 200\},$$
$$T \in \{10, 20, 30\},$$

as we would like to study the small sample properties of the test procedures when the time dimension is small. In our first set of Monte Carlo experiments the errors are normally distributed; therefore we focus on the standard LM test. We also include their respective heteroskedasticity-robust regression variants for this exercise.

1.7.2 Normally distributed errors

Panel A of Table 1.1 (see Appendix B) shows the rejection frequencies when the null hypothesis is true. The $\widetilde{\Delta}_{\text{adj}}$ test has rejection frequencies close to the nominal size of 5% for all combinations of N and T, while the CLM test rejects the null hypothesis too often, in particular for small N. Deviations from the nominal size for the the standard LM test and the regression-based test are small and disappear as N increases, as expected from Theorem 1. Panel B of Table 1.1 shows the corresponding rejections frequencies under the alternative hypothesis. The LM test outperforms the $\widetilde{\Delta}_{\text{adj}}$ and the CLM test in general. This observation holds in particular for $T = 10$ where the power gain is considerable. The LM_{reg} variant, although as powerful as the $\widetilde{\Delta}_{\text{adj}}$ test for $T = 10$, suffers from a power loss relative to the standard LM test. This power loss may be due to the small sample bias of the variance estimator, see Remark 9.

Following Remark 7 the variants of the LM tests are computed as follows. First, the individual-specifc fixed effects α_i are eliminated by transforming the data using orthogonal forward deviations (see Arellano and Bover (1995)). The LM statistics are then computed using the transformed data. The results presented in Panel A of Table 1.2 indicate that by employing forward orthogonalization all variants of the LM test have size reasonably close to the nominal level. By comparing panel B of Table 1.1 and the rejection rates under the alternative in panel B of Table 1.2 we see that the power is very similar in both setups confirming usefulness of the forward orthogonalization procedure for the LM tests.

1.7.3 Non-normal errors

We now investigate the LM test when the errors are no longer normally distributed, thereby building on the results of Section 1.5.1. The errors in (1.40) are generated from a t-distribution with 5 degrees of freedom, scaled to have unit variance. All other specifications of the standard design remain unchanged. In addition to the statistics already considered, we now include the adjusted LM statistic (see corollary 1) and the score-modified statistic (see Theorem 3). Panel A in Table 1.3 reports the rejection frequencies under the null hypothesis in this case. We notice that the LM test has substantial size distortions when T is fixed and N increases, which is expected from Theorem 2. However, the adjusted LM statistic LM_{adj} and the modified score statistic LM^* are both successful in controlling the type-I error.

Panel B of Table 1.3 shows rejection frequencies under the alternative hypothesis. The power gain of the LM test relative to the $\widetilde{\Delta}_{\text{adj}}$ test is noticeable when $T = 10$ or $T = 20$. We found similar results when the errors are χ^2 distributed with two degrees of freedom, centered and standardized to have mean zero and variance equal to one. Given the similarity of the results for t and χ^2 distributed errors, we do not present the latter results.

1.7.4 Serially correlated errors

To study the impact of serially correlated errors on the test statistics we adjust the DGP as follows:

$$\begin{aligned} y_{it} &= x_{it}'\beta_i + \epsilon_{it}, \\ \epsilon_{it} &= \rho\epsilon_{it-1} + \left(1-\rho^2\right)^{1/2} e_{it}, \end{aligned}$$

for $i = 1, 2, \dots, N$, $t = 1, 2, \dots, T$, where $e_{it} \overset{iid}{\sim} \mathcal{N}(0, 1)$. Under the null hypothesis $\beta_i = 1$ for all i while under the alternative β_i is generated as in (1.42). The regressors, $x_{it,k}$, $k = 1, 2, 3$ are generated as

$$\begin{aligned} x_{it,k} &= \phi_{i,k}x_{it-1,k} + \left(1-\phi_{i,k}^2\right)^{1/2} \vartheta_{it,k}^x, \\ \phi_{i,k} &\overset{iid}{\sim} U[0.05, 0.95], \\ \vartheta_{it,k}^x &\overset{iid}{\sim} \mathcal{N}\left(0, \sigma_{ix,k}^2\right), \end{aligned}$$

where $\sigma_{ix,k}^2 = 0.25 + \eta_{i,k}$ with $\eta_{i,k} \overset{iid}{\sim} \chi_1^2$. Parameters $\phi_{i,k}$ and $\sigma_{ix,k}$ are fixed across replications.

Results of this simulation experiment are reported in Table 1.4. Panel A and B show the rejection frequencies under the null hypothesis in case of “small” serial dependence (i.e., $\rho = 0.2$, Panel A) and “moderate” dependence (i.e., $\rho = 0.5$, Panel B). For all LM based test statistics, except the LM_{ac} test, we observe substantial size deviations from the nominal level. However, the LM_{ac} test is successful in controlling the type-I error. Further, size properties of PY test are also significantly affected by autocorrelated errors. Note that this fact is already documented and studied in Blomquist and Westerlund (2013).

Panel C of Table 1.4 reports power properties of the test under no serial correlation (i.e., $\rho = 0$), building on the discussion in Remark 10. We observe that the LM_{ac} test involve a $5 - 10\%$ power loss compared to the LM^* test. This relative power loss dies out if T increases.

1.8 Concluding remarks

In this paper we examine the problem of testing slope homogeneity in a panel data model. We develop testing procedures using the LM principle. Several variants are considered that robustify the original LM test with respect to non-normality, heteroscedasticity and serially correlated errors. By studying the local power we identify cases where the LM-type tests are particularly powerful relative to existing tests. In sum, our Monte Carlo experiments suggest that the LM test are powerful testing procedures to detect slope homogeneity in short panels in which the time dimension is small relative to the cross-section dimension. The LM approach suggested in this paper may be extended in future research by allowing for dynamic specifications with lagged dependent variables and cross sectionally or serially correlated errors.

A Appendix: Proofs

To economize on notation we use $\sum_i$ and $\sum_t$ instead of full expressions $\sum_{i=1}^{N}$ and $\sum_{t=1}^{T}$ throughout this appendix.

A.1 Preliminary results

We first present an important result concerning the asymptotic effect of the estimation error $\widetilde{\beta}-\beta$ on the test statistics. Define

$$A_i^{(k)} = X_i^{(k)} X_i^{(k)\prime} - \left(\frac{1}{NT} \sum_i X_i^{(k)\prime} X_i^{(k)} \right) I_T.$$

Lemma A.1 *Let* $R_{XAX}^{(k)} = \sum_i X_i' A_i^{(k)} X_i$ *and* $R_{XAu}^{(k)} = \sum_i X_i' A_i^{(k)} u_i$ *for* $k = 1, ..., K$. *Furthermore let*

$$R_N^{(k)} = \left(\frac{\widetilde{\sigma}^4}{\sigma^4} \right) \frac{1}{2\sigma^2} \left(\left(\widetilde{\beta} - \beta \right)' R_{XAX}^{(k)} \left(\widetilde{\beta} - \beta \right) - 2 \left(\widetilde{\beta} - \beta \right)' R_{XAu}^{(k)} \right),$$

for $k = 1, ..., K$. *Under Assumptions 1, 2 and the null hypothesis the following properties hold if* T *is fixed:*

(i) $R_{XAX}^{(k)} = O_p(N)$,

(ii) $R_{XAu}^{(k)} = O_p\left(N^{1/2}\right)$,

(iii) $R_N^{(k)} = O_p(1)$,

for $k = 1, ..., K$.

Proof. **(i)** Using the definition of $A_i^{(j)}$ yields

$$R_{XAX}^{(k)} = \sum_i X_i' \left(X_i^{(k)} X_i^{(k)\prime} \right) X_i - \frac{1}{NT} \left(\sum_i \sum_t x_{it,k}^2 \right) \left(\sum_i X_i' X_i \right).$$

The first term is a $K \times K$ matrix with typical (l, m) element

$$\sum_i \left(\sum_t x_{it,l} x_{it,k} \right) \left(\sum_t x_{it,m} x_{it,k} \right) = O_p(N),$$

as a consequence of Assumption 2, while $\sum_i \sum_t x_{it,k}^2 / NT = O_p(1)$ and $\sum_i X_i' X_i = O_p(N)$.

(ii) Recall that under the null hypothesis, $u_i = \epsilon_i$. Thus

$$R_{XAu}^{(k)} = \sum_i \left(X_i' X_i^{(k)} \right) \left(X_i^{(k)\prime} u_i \right) - \frac{1}{NT} \left(\sum_i \sum_t x_{it,k}^2 \right) \left(\sum_i X_i' u_i \right).$$

The first and the second term are $O_p\left(N^{1/2}\right)$ by a the central limit theorem (CLT) for independent random variables and Assumption 2. **(iii)** Combining **(i)** and **(ii)** together with the fact that $\sqrt{N}\left(\widetilde{\beta}-\beta\right)=O_p(1)$ yields the result. ■

Lemma A.2 *Under Assumptions 1′, 2′ and the null hypothesis the following properties hold for $N\to\infty$ and $T\to\infty$:*

(i) $R^{(k)}_{XAX}=O_p\left(NT^2\right)$,

(ii) $R^{(k)}_{XAu}=O_p\left(N^{1/2}T^{3/2}\right)$,

(iii) $R^{(k)}_{NT}=O_p\left(T\right)$, *which is defined as* $R^{(k)}_{N}$ *in Lemma A.1,*

for $k=1,...,K$.

Proof. Following the proof of Lemma A.1 the element of the first term of $R^{(k)}_{XAX}$ is $O_p\left(NT^2\right)$, whereas the second term is $O_p\left(NT\right)$ by Assumption 2′ which yields statement **(i)**. Notice in **(ii)** $R^{(k)}_{XAu}$ has two terms as in Lemma A.1, where the first one has zero mean and variance of order T^3. Therefore by Lemma 1 in Baltagi et al. (2011) we have that $X_i'X_i^{(j)}X_i^{(j)\prime}u_i=O_p(T^{3/2})$ and by Lemma 2 in PY that $\sum\limits_i\left(X_i'X_i^{(j)}\right)\left(X_i^{(j)\prime}u_i\right)=O_p\left(N^{1/2}T^{3/2}\right)$ and $\left(\sum\limits_i X_i'u_i\right)=O_p\left(N^{1/2}T^{1/2}\right)$. These results and the fact that $\sqrt{NT}\left(\widetilde{\beta}-\beta\right)=O_p\left(1\right)$ imply **(iii)**. ■

A.2 Proofs of the main results

Proof of Lemma 1

We use the following rules for matrix differentiations:

$$\frac{\partial\ell}{\partial\theta_k}=-\frac{1}{2}\mathrm{tr}\left[\Omega^{-1}\frac{\partial\Omega}{\partial\theta_k}\right]+\frac{1}{2}\left[u'\Omega^{-1}\frac{\partial\Omega}{\partial\theta_k}\Omega^{-1}u\right], \tag{A.1}$$

$$-\mathbb{E}\left[\frac{\partial\ell}{\partial\theta_k\partial\theta_l}\right]=\frac{1}{2}\mathrm{tr}\left[\Omega^{-1}\left(\frac{\partial\Omega}{\partial\theta_k}\right)\Omega^{-1}\left(\frac{\partial\Omega}{\partial\theta_l}\right)\right], \tag{A.2}$$

for $k,l=1,2,\dots,K+1$, see, e.g., Harville (1977) and Wand (2002). First,

$$X_i\Sigma_vX_i'=\sum_k\sigma^2_{v,k}X_i^{(k)}X_i^{(k)\prime}\ ,$$

with $X_i^{(k)}$ denoting the k-th column vector of X_i. Hence

$$\begin{bmatrix} X_1\Sigma_vX_1' & & 0\\ & \ddots & \\ 0 & & X_N\Sigma_vX_N'\end{bmatrix}=\sum_k\sigma^2_{v,k}A_k,$$

with the $NT \times NT$ matrix ,

$$A_k = \begin{bmatrix} X_1^{(k)} X_1^{(k)\prime} & & 0 \\ & \ddots & \\ 0 & & X_N^{(k)} X_N^{(k)\prime} \end{bmatrix},$$

for $k = 1, \dots, K$, and $X_i^{(k)}$ denotes the k-th column of the $T \times K$ matrix X_i. Thus,

$$\Omega = \sum_k \sigma_{v,k}^2 A_k + \sigma^2 I_{NT}$$

and

$$\frac{\partial \Omega}{\partial \theta_k} = \begin{cases} A_k, & \text{for } k = 1, 2, \dots, K, \\ I_{NT}, & \text{for } k = K+1. \end{cases}$$

Under the null hypothesis we have $\Omega = \sigma^2 I_{NT}$. Using (A.1) we obtain

$$\left.\frac{\partial \ell}{\partial \theta_k}\right|_{H_0} = \begin{cases} -\frac{1}{2\widetilde{\sigma}^2} \text{tr}\left[A_k\right] + \frac{1}{2\widetilde{\sigma}^4} \widetilde{u}' A_k \widetilde{u}, & \text{for } k = 1, 2, \dots, K \\ 0, & \text{for } k = K+1, \end{cases}$$

where

$$\widetilde{\sigma}^2 = \frac{1}{NT} \widetilde{u}' \widetilde{u},$$
$$\widetilde{u} = \left(I_{NT} - X \left(X'X \right)^{-1} X' \right) y.$$

The representation of the score vector follows from

$$\text{tr}\left[A_k\right] = \sum_i \sum_t X_{it,k}^2 = X^{(k)\prime} X^{(k)},$$

where $X^{(k)}$ denotes the k-th column of the $NT \times K$ matrix X. Similarly, (A.2) yields

$$-\mathbb{E}\left[\frac{\partial \ell}{\partial \theta_k \partial \theta_l}\right]\Bigg|_{H_0} = \begin{cases} \frac{1}{2\sigma^4} \text{tr}\left[A_k A_l\right], & \text{for } k, l = 1, 2, \dots, K, \\ \frac{1}{2\sigma^4} X^{(k)\prime} X^{(k)}, & \text{for } k = 1, 2, \dots, K, \text{ and } l = K+1, \\ \frac{NT}{2\sigma^4}, & \text{for } k = l = K+1, \end{cases}$$

Using the fact that A_k and A_l are block-diagonal,

$$\text{tr}\left[A_k A_l\right] = \sum_i \text{tr}\left[\left(X_i^{(k)} X_i^{(k)\prime}\right)\left(X_i^{(l)} X_i^{(l)\prime}\right)\right] = \sum_i \left(X_i^{(k)\prime} X_i^{(l)}\right)^2,$$

where $X_i^{(k)}$ denotes the i-th column of X_i, which yields the form of the information matrix presented in the lemma.

Proof of Theorem 1

Recall that

$$A_i^{(k)} = X_i^{(k)} X_i^{(k)\prime} - \left(\frac{1}{NT} \sum_i X_i^{(k)\prime} X_i^{(k)} \right) I_T,$$

and rewrite the elements of the scores as

$$\widetilde{s}_k = \left(\frac{\widetilde{\sigma}^4}{\sigma^4} \right) \frac{1}{2\sigma^4} \sum_i \widetilde{u}_i' A_i^{(k)} \widetilde{u}_i,$$

for $k = 1, ..., K$. Since $\widetilde{u}_i = u_i - X_i(\widetilde{\beta} - \beta)$ we have

$$\frac{1}{\sqrt{N}} \widetilde{s}_k = \frac{1}{\sqrt{N}} \left(\frac{\sigma^4}{\widetilde{\sigma}^4} \right) \frac{1}{2\sigma^4} \sum_i u_i' A_i^{(k)} u_i + \frac{1}{\sqrt{N}} R_N^{(k)},$$

where $R_N^{(k)} = O_p(1)$ from Lemma A.1. Since $\sum_i \mathrm{tr}\left[A_i^{(k)} \right] = 0$ it follows that $\mathbb{E}(u_i' A_i^{(k)} u_i) = 0$ and, therefore,

$$\lim_{N \to \infty} \mathbb{E} \left(\frac{1}{\sqrt{N}} \widetilde{s} \right) = 0.$$

The covariances are obtained as

$$\begin{aligned}
&Cov\left(u_i' A_i^{(k)} u_i, u_i' A_i^{(l)} u_i \mid X \right) = 2\sigma^4 \mathrm{tr}\left[A_i^{(k)} A_i^{(l)} \right] \\
&= 2\sigma^4 \left(X_i^{(k)\prime} X_i^{(l)} \right)^2 - \left(\frac{1}{NT} \sum_i X_i^{(k)\prime} X_i^{(k)} \right) \left(X_i^{(l)\prime} X_i^{(l)} \right) - \left(\frac{1}{NT} \sum_i X_i^{(l)\prime} X_i^{(l)} \right) \left(X_i^{(k)\prime} X_i^{(k)} \right) \\
&+ T \left(\frac{1}{NT} \sum_i X_i^{(k)\prime} X_i^{(k)} \right) \left(\frac{1}{NT} \sum_i X_i^{(l)\prime} X_i^{(l)} \right),
\end{aligned}$$

and since $u_i' A_i^{(k)} u_i$ is independent of $u_j' A_i^{(l)} u_j$ for all $i \neq j$ conditional on X,

$$\begin{aligned}
&\left(\frac{1}{2\sigma^4} \right)^2 Cov\left(\sum_i u_i' A_i^{(k)} u_i, \sum_i u_i' A_i^{(l)} u_i \,\middle|\, X \right) \\
&= \frac{1}{2\sigma^4} \left(\sum_i \left(X_i^{(k)\prime} X_i^{(l)} \right)^2 - \frac{1}{NT} \left(\sum_i X^{(k)\prime} X^{(k)} \right) \left(\sum_i X_i^{(l)\prime} X_i^{(l)} \right) \right) \\
&= V_{k,l}.
\end{aligned}$$

The Liapounov condition in the central limit theorem for independent random variables (see White (2001), Theorem 5.10) is satisfied by Assumption 2 and therefore

$$\left(\frac{1}{N} \widetilde{V} \right)^{-1/2} \left(\frac{1}{\sqrt{N}} \widetilde{s} \right) \xrightarrow{d} \mathcal{N}(0, I_K),$$

where $\widetilde{V}$ replaces σ^4 in V by $\widetilde{\sigma}^4$. By the formula for the partitioned inverse

$$\left\{ \mathcal{I}(\widetilde{\sigma}^2)^{-1} \right\}_{1:K,1:K} = \widetilde{V}^{-1},$$

where $\{\cdot\}_{1:K,1:K}$ denotes the upper-left $K \times K$ block of the matrix, it follows finally that

$$\widetilde{\mathcal{S}}'\mathcal{I}(\widetilde{\sigma}^2)^{-1}\widetilde{\mathcal{S}} = \widetilde{s}'\widetilde{V}^{-1}\widetilde{s} \xrightarrow{d} \chi^2_K.$$

Proof of Theorem 2

The proof proceeds in three steps: **(i)** we derive the covariance matrix of the score vector, **(ii)** we establish the asymptotic normality of the score vector and **(iii)** we use these results to establish the asymptotic distribution of the LM statistic.

(i) Define the $K \times 1$ vector $s = [s_1, ..., s_K]'$ with typical element

$$s_k = \frac{1}{2\sigma^4}\sum_i u_i' A_i^{(k)} u_i = \frac{1}{2\sigma^4}\sum_i s_{i,k}, \tag{A.3}$$

where $s_{i,k} = u_i' A_i^{(k)} u_i$ and $1 \leq k \leq K$. Using standard results for quadratic forms (see e.g., Ullah (2004), appendix A.5),

$$\mathbb{E}\left[s_{i,k}\middle|X\right] = \sigma^2 \mathrm{tr}\left[A_i^{(k)}\right]$$

$$\mathbb{E}\left[s_{i,k}s_{i,l}\middle|X\right] = 2\sigma^4 \mathrm{tr}\left[A_i^{(k)}A_i^{(l)}\right] + \sigma^4 \mathrm{tr}\left[A_i^{(k)}\right]\mathrm{tr}\left[A_i^{(l)}\right] + \left(\mu_u^{(4)} - 3\sigma^4\right) a_i^{(k)\prime} a_i^{(l)},$$

where $a_i^{(k)}$ is a vector consisting of the main diagonal elements of the matrix $A_i^{(k)}$ and $\mu_u^{(4)}$ denotes the fourth moment of u_{it}. Since

$$\mathbb{E}\left[s_{i,k}\middle|X\right]\mathbb{E}\left[s_{i,l}\middle|X\right] = \sigma^4 \mathrm{tr}\left[A_i^{(k)}\right]\mathrm{tr}\left[A_i^{(l)}\right],$$

we have

$$Cov\left(s_{i,k}, s_{i,l}\middle|X\right) = 2\sigma^4 \mathrm{tr}\left[A_i^{(k)}A_i^{(l)}\right] + \left(\mu_u^{(4)} - 3\sigma^4\right) a_i^{(k)\prime} a_i^{(l)}. \tag{A.4}$$

Due to the independence of $u_i' A_i^{(k)} u_i$ and $u_j' A_j^{(l)} u_j$ for $i \neq j$, it follows that

$$Cov\left(\sum_i s_{i,k}, \sum_i s_{i,l}\middle|X\right) = 2\sigma^4 \sum_i \mathrm{tr}\left[A_i^{(k)}A_i^{(l)}\right] + \left(\mu_u^{(4)} - 3\sigma^4\right)\sum_i a_i^{(k)\prime} a_i^{(l)}.$$

Let V_{NT} denote the covariance matrix of s. Inserting the expression for $tr\left[A_i^{(k)}A_i^{(l)}\right]$, we determine the (k,l) element of V_{NT} as

$$\begin{aligned} V_{k,l} &= \frac{1}{2\sigma^4}\left(\sum_i\left(\sum_t x_{it,k}x_{it,l}\right)^2 - \frac{1}{NT}\left(\sum_i\sum_t x_{it,k}^2\right)\left(\sum_i\sum_t x_{it,k}^2\right)\right) \\ &\quad + \left(\frac{\mu_u^{(4)} - 3\sigma^4}{(2\sigma^4)^2}\right)\sum_i\sum_t\left(x_{it,k}^2 - \frac{1}{NT}\sum_i\sum_t x_{it,k}^2\right)\left(x_{it,l}^2 - \frac{1}{NT}\sum_i\sum_t x_{it,l}^2\right) \\ &= V_{1,k,l} + V_{2,k,l}. \end{aligned} \tag{A.5}$$

(ii) To verify that a central limit theorem applies to s, let $\lambda \in \mathbb{R}^k$, $||\lambda|| = 1$ and $Z_{i,T} = \frac{1}{T}\lambda' s_i$, where s_i is a $K \times 1$ vector with elements $s_{i,k}$ for $1 \leq k \leq K$. Further, $\mathbb{E}\left[Z_{i,T}\right] = 0$

and $\mathbb{E}\left[Z_{i,T}^2\right] = \frac{1}{T^2}\lambda'\mathbb{E}\left[V_{i,T}\right]\lambda$, where $V_{i,T}$ is a $K \times K$ matrix with the typical (k,l) element defined in (A.4) for $1 \le k,l \le K$. From the Cramer-Wold device we conclude that it is sufficient to show that

$$\frac{1}{\sqrt{N}}\sum_i Z_{i,T} \xrightarrow{d} N\left(0,\mathcal{V}\right), \tag{A.6}$$

where $\mathcal{V} \equiv \lim_{N,T\to\infty} \frac{1}{NT^2}\sum_i \lambda'\mathbb{E}\left[V_{i,T}\right]\lambda$. Assumption $2'$ and (A.4) ensure that $\mathcal{V}$ exists and is positive definite.

The asymptotic normality result (A.6) follows from the central limit theorem for the double indexed process (see e.g., Phillips and Moon (1999), Theorem 2) if the following condition holds

$$\sum_i \mathbb{E}\left[\frac{Z_{i,T}^2}{\mathcal{V}_{NT}} 1\left(\left|\frac{Z_{i,T}^2}{\mathcal{V}_{NT}}\right| > \varepsilon\right)\right] \longrightarrow 0 \text{ for all } \varepsilon > 0, \tag{A.7}$$

where $\mathcal{V}_{NT} = \frac{1}{T^2}\sum_i \lambda'\mathbb{E}\left[V_{i,T}\right]\lambda$. In turn the Lindeberg condition (A.7) holds provided that

$$\sup_{i,T}\mathbb{E}\left\|Z_{i,T}\right\|^3 \le \sup_{i,T}\mathbb{E}\left\|\frac{s_i}{T}\right\|^3 < \infty. \tag{A.8}$$

To study wether s_i/T is uniformly L_3 bounded for all i and T it suffices to consider s_i/T elementwise. Furthermore, each element of s_i/T can be written in terms of quadratic forms i.e.,

$$\frac{1}{T}s_{i,k} = \frac{1}{T}\left(u_i'B_{i,k}u_i - \mathbb{E}\left[\left(u_i'B_{i,k}u_i\right)\middle|X\right]\right),$$

where $B_{i,k} = X_i^{(k)}X_i^{(k)\prime}$ for $1 \le k \le K$ and by the triangle inequality

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T}s_{i,k}\right|^3 &= \frac{1}{T^3}\mathbb{E}\left|u_i'B_{i,k}u_i - \mathbb{E}\left[\left(u_i'B_{i,k}u_i\right)\middle|X\right]\right|^3 \le \frac{1}{T^3}\mathbb{E}\left|u_i'B_{i,k}u_i\right|^3 \\
+& \frac{3}{T^3}\mathbb{E}\left|u_i'B_{i,k}u_i\right|^2\left|\mathbb{E}\left[\left(u_i'B_{i,k}u_i\right)\middle|X\right]\right| + \frac{3}{T^3}\mathbb{E}\left|u_i'B_{i,k}u_i\right|\left|\mathbb{E}\left[\left(u_i'B_{i,k}u_i\right)\middle|X\right]\right|^2 \\
+& \frac{1}{T^3}\left|\mathbb{E}\left[\left(u_i'B_{i,k}u_i\right)\middle|X\right]\right|^3.
\end{aligned} \tag{A.9}$$

For the first term on the r.h.s of (A.9) we make use of a formula for the third moment of a quadratic form (see e.g., Wiens (1992) or Ullah (2004), appendix A.5), the law of iterated expectations and uniform bounds $\mathbb{E}\left[|u_{it}|^6|X\right] < C < \infty$ and $\mathbb{E}|x_{it,k}|^6 < C < \infty$ given in Assumptions $1'$ and $2'$, i.e.,

$$\begin{aligned}
&\frac{1}{T^3}\mathbb{E}\left|u_i'B_{i,k}u_i\right|^3 = \frac{1}{T^3}\mathbb{E}\left[\mathbb{E}\left[\left|u_i'B_{i,k}u_i\right|^3\middle|X\right]\right] \\
= &\frac{1}{T^3}\mathbb{E}\left[\mathbb{E}\left[\left|\sum_{t_1}\sum_{t_2}\sum_{t_3}u_{it_1}^2u_{it_2}^2u_{it_3}^2x_{it_1,k}^2x_{it_2,k}^2x_{it_3,k}^2\right|\middle|X\right]\right] + O\left(T^{-1}\right).
\end{aligned} \tag{A.10}$$

Further, from Assumptions $1'$ we have $\mathbb{E}\left[\left|u_{it_1}^2u_{it_2}^2u_{it_3}^2\right|\middle|X\right] < C < \infty$ and from Assumptions $2'$ the term $\mathbb{E}\left[\left|x_{it_1,k}^2x_{it_2,k}^2x_{it_3,k}^2\right|\right]$ is uniformly bounded for all i and T. Then it follows from triangle inequality that the first term on the r.h.s. of (A.10) is uniformly bounded. The same reasoning applies to the rest of the terms in (A.9) to show their uniform boundedness. This concludes the proof of (A.8) and the asymptotic normality of the score vector s.

(iii) Rewrite the first K elements of the score as

$$\widetilde{s} = \left(\frac{\widetilde{\sigma}^4}{\sigma^4}\right) s + R_{NT},$$

where R_{NT} is given in Lemma A.2 and s has typical element as defined in (A.3). By **(ii)**,

$$s' \left(V_{NT}\right)^{-1} s \xrightarrow{d} \chi^2_K, \tag{A.11}$$

as $N \to \infty$, $T \to \infty$, where V_{NT} has (k,l) element $V_{k,l}$ as in (A.5). Under Assumptions 1′ and 2′

$$\begin{aligned} V_1 &= O_p\left(NT^2\right), \\ V_2 &= O_p\left(NT\right), \end{aligned}$$

where V_1 and V_2 are specified elementwise in (A.5). Given the expression for $\widetilde{V}$ in Theorem 1,

$$\frac{\widetilde{V}}{NT^2} - \frac{V_1}{NT^2} \xrightarrow{p} 0$$

and hence

$$s'\widetilde{V}^{-1}s - s' \left(V_{NT}\right)^{-1} s \xrightarrow{p} 0 \tag{A.12}$$

as $N \to \infty, T \to \infty$. The LM statistic can be expanded as

$$\begin{aligned} \text{LM} &= \widetilde{s}\,'\widetilde{V}^{-1}\widetilde{s} \\ &= \left(\left(\frac{\widetilde{\sigma}^4}{\sigma^4}\right) s + R_{NT}\right)' \widetilde{V}^{-1} \left(\left(\frac{\widetilde{\sigma}^4}{\sigma^4}\right) s + R_{NT}\right) \\ &= \left(\frac{\widetilde{\sigma}^4}{\sigma^4}\right)\left(s'\widetilde{V}^{-1}s\right) + O_p\left(N^{-1/2}\right). \end{aligned} \tag{A.13}$$

where the last line follows from Lemma A.2. The theorem follows by combining (A.11), (A.12) and (A.13).

Proof of Corollary 1

The result follows immediately from the proof of Theorem 2 and the fact that $\widetilde{\mu}_u^{(4)} = (NT)^{-1} \sum_i \sum_t \widetilde{u}_{it}$ is a consistent estimator of $\mu_u^{(4)}$.

Proof of Theorem 3

Using similar arguments as in the proof of Theorem 1,

$$\frac{1}{\sqrt{N}}\widetilde{s}^* = \frac{1}{\sqrt{N}}\left(\frac{\sigma^4}{\widetilde{\sigma}^4}\right)\begin{bmatrix} \frac{1}{\sigma^4}\sum_i\sum_t\sum_{s=1}^{t-1} x_{it,1}u_{it}x_{is,1}u_{is} \\ \vdots \\ \frac{1}{\sigma^4}\sum_i\sum_t\sum_{s=1}^{t-1} x_{it,K}u_{it}x_{is,K}u_{is} \end{bmatrix} + o_p(1). \tag{A.14}$$

Let $u^*_{it} = u_{it}/\sigma$ and $z^*_{itk} = x_{it,k} u^*_{it}$. Clearly, $\mathbb{E}\left[\sum_t \sum_{s=1}^{t-1} z^*_{itk} z^*_{isk}\right] = 0$. Since conditional on X, $\sum_t \sum_{s=1}^{t-1} z^*_{it,k} z^*_{is,k}$ and $\sum_t \sum_{s=1}^{t-1} z^*_{jtl} z^*_{jsl}$ are independent for $i \neq j$, the covariances for two elements k and l of the vector (A.14) are

$$
\begin{aligned}
\mathbb{E}\left[s^*_k s^*_l \;\middle|\; X \right] &= \frac{1}{\sigma^4} \sum_i \mathbb{E}\left[\left(\sum_{t=2}^{T} \sum_{s=1}^{t-1} z^*_{itk} z^*_{isk} \right) \left(\sum_{t=2}^{T} \sum_{s=1}^{t-1} z^*_{itl} z^*_{isl} \right) \middle| X \right] \\
&= \frac{1}{\sigma^4} \sum_{i=1} \left(\sum_{t=2}^{T} x_{it,k} x_{it,l} \right) \left(\sum_{s=1}^{t-1} x_{is,k} x_{is,l} \right) \\
&= V^*_{k,l}.
\end{aligned}
$$

since all cross terms have zero expectation and $\mathbb{E}\left[(u^*_{it})^2 \right] = 1$. The central limit theorem for independent random variables and Slutsky's theorem imply

$$
\left(\frac{1}{N} V^* \right)^{-1/2} \left(\frac{1}{\sqrt{N}} s^* \right) \xrightarrow{d} \mathcal{N}(0, I_K)
$$

and the result follows.

Proof of Corollary 2

Using the arguments in Theorem 1 and 3 (under Assumption 1′, 2 and allowing for $\mathbb{E}[\epsilon^2_{it}|X] = \sigma^2_{it}$), $\mathrm{LM}_{\mathrm{reg}}$ is asymptotically χ^2_K if the Liapounov condition is satisfied and the asymptotic covariance matrix of the score vector is equal to the limit of $\sum\limits_i \sum\limits_t \widetilde{u}^2_{it} \widetilde{z}_{it,k} \widetilde{z}_{it,l}$ as $N \to \infty$ and T is fixed.

Regarding the Liapounov condition it suffices to show that $\mathbb{E}\left| s^*_{i,k} \right|^{2+\delta} < C < \infty$ for $k = 1, ..., K$. By Minkowski inequality,

$$
\mathbb{E}\left| s^*_{i,k} \right|^{2+\delta} \leq \left(\sum_t \sum_{s=1}^{t-1} \left(\mathbb{E}\left| u_{it} u_{is} x_{it,k} x_{is,k} \right|^{2+\delta} \right)^{2+\delta} \right)^{\frac{1}{2+\delta}}.
$$

Further by the Cauchy-Schwartz inequality, the law of iterated expectations, Assumptions 1′ and 2,

$$
\begin{aligned}
\mathbb{E}\left| x_{it,1} u_{it} x_{is,1} u_{is} \right|^{2+\delta} &\leq \mathbb{E}\left[\sqrt{ \mathbb{E}\left[\left| u^2_{it} u^2_{is} \right|^{2+\delta} |X \right] } \left| x_{it,1} x_{is,1} \right|^{2+\delta} \right] \\
= \quad \mathbb{E}\left[\sqrt{ \mathbb{E}\left[|u_{it}|^{4+\delta} \right] \mathbb{E}\left[|u_{is}|^{4+\delta} |X \right] } \left| x_{it,1} x_{is,1} \right|^{2+\delta} \right] &< C < \infty.
\end{aligned}
$$

Hence the Liapounov condition holds.

Regarding the (k, l) element of the covariance matrix of $\tilde{s}^*$, note that

$$
\begin{aligned}
&\sum_i \mathbb{E}\left[\left(\sum_t x_{it,k} u_{it} \left(\sum_{s=1}^{t-1} x_{is,k} u_{is} \right) \right) \left(\sum_t x_{it,l} u_{it} \left(\sum_{s=1}^{t-1} x_{is,l} u_{is} \right) \right) \middle| X \right] \\
= \quad &\sum_i \sum_t \sum_{s=1}^{t-1} \sigma^2_{it} \sigma^2_{is} x_{it,k} x_{it,l} x_{is,k} x_{is,l}.
\end{aligned}
$$

Next let $z_{it,k} = x_{it,k} \sum_{s=1}^{t-1} u_{is} x_{is,k}$ and notice that

$$\mathbb{E}\left[\sum_i \sum_t u_{it}^2 z_{it,k} z_{it,l}' \middle| X\right] = \sum_i \sum_t \sum_{s=1}^{t-1} \sigma_{it}^2 \sigma_{is}^2 x_{it,k} x_{it,l} x_{is,k} x_{is,l}.$$

Furthermore

$$\frac{1}{N} \sum_i \sum_t \widetilde{u}_{it}^2 \widetilde{z}_{it,k} \widetilde{z}_{it,l} - \frac{1}{N} \sum_i \sum_t u_{it}^2 z_{it,k} z_{it,l} \xrightarrow{p} 0,$$

and result (1.31) follows.

Proof of Theorem 4

Consider the normalized scores

$$\frac{1}{\sqrt{N}} \tilde{s}^{**} = \frac{1}{\sqrt{N}} s^{**} + o_p(1)$$

where the k-element of the vector s^{**} is given by

$$s_k^{**} = \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \left(u_{it} u_{is} - \sigma_{ts}\right) x_{it,k} x_{is,k}$$

By construction $\mathbb{E}\left[s_k^{**}\right] = 0$ for $k = 1, \ldots, K$ under the null hypothesis. Same arguments as for result (1.31) apply to show the Liapounov condition and make use of the central limit theorem for independent and heterogeneously random variables . It remains to show that the (k, l) element of the covariance matrix of s^{**} takes form (1.33). Since conditional on X , contributions $s_{i,k}^{**}$ and $s_{j,k}^{**}$ are independent for $i \neq j$, the covariances for two elements k and l of the vector s^{**} normalized with N are

$$\begin{aligned}
&\mathbb{E}\left[s_k^{**} s_l^{**} \middle| X\right] = \\
&= \sum_i \mathbb{E}\left[\left(\sum_{t=2}^{T} \sum_{s=1}^{t-1} \left(u_{it} u_{is} - \sigma_{ts}\right) x_{it,k} x_{is,k}\right) \left(\sum_{\tau=2}^{T} \sum_{q=1}^{t-1} \left(u_{i\tau} u_{iq} - \sigma_{\tau q}\right) x_{i\tau,l} x_{iq,l}\right) \middle| X\right] \\
&= \sum_i \sum_{t=2}^{T} \sum_{s=1}^{t-1} x_{it,k} x_{is,k} \left(\sum_{\tau=2}^{T} \sum_{q=1}^{t-1} \mathbb{E}\left[\left(u_{it} u_{is} - \sigma_{ts}\right)\left(u_{i\tau} u_{iq} - \sigma_{\tau q}\right) |X\right] x_{i\tau,l} x_{iq,l}\right),
\end{aligned}$$

Further,

$$\mathbb{E}\left[\left(u_{it} u_{is} - \sigma_{ts}\right)\left(u_{i\tau} u_{iq} - \sigma_{\tau q}\right) |X\right] = \mathbb{E}\left[u_{it} u_{is} u_{i\tau} u_{iq} |X\right] - \sigma_{ts} \sigma_{\tau q},$$

and the (k, l) element can be written as

$$\begin{aligned}
\mathbb{E}\left[s_k^{**} s_l^{**} \middle| X\right] &= \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} x_{it,k} x_{is,k} \left(\sum_{\tau=2}^{T} \sum_{q=1}^{t-1} \delta_{ts\tau q} x_{i\tau,l} x_{iq,l}\right) \\
&= \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{\tau=2}^{T} \sum_{q=1}^{t-1} \delta_{ts\tau q} x_{it,k} x_{is,k} x_{i\tau,l} x_{iq,l},
\end{aligned}$$

where $\delta_{ts\tau q} = \mathbb{E}[u_{jt}u_{js}u_{j\tau}u_{jq}|X] - \sigma_{ts}\sigma_{\tau q}$. Substituting $\delta_{ts\tau q}$ by an appropriate consistent estimator $\widehat{\delta}_{ts\tau q} = 1/N\left(\sum_{j=1}^{N} \widetilde{u}_{jt}\widetilde{u}_{js}\widetilde{u}_{j\tau}\widetilde{u}_{jq} - \widetilde{\sigma}_{ts}\widetilde{\sigma}_{\tau q}\right)$ yields the limiting distribution of the modified test statistic.

Proof of Theorem 5

As in Honda (1985) the proof of the theorem proceeds in three steps: **(i)** first we show that $\widetilde{\sigma}^2$ remains consistent under the local alternative; **(ii)** second, we incorporate the local alternative into the score vector and **(iii)** establish the asymptotic distribution of the LM statistic. **(i)** Note first that with $M_X = I_{NT} - X\left(X'X\right)^{-1}X'$

$$\widetilde{u} = M_X u = M_X\left(D_X v + \epsilon\right),$$

where

$$D_X = \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_N \end{bmatrix}.$$

Hence,

$$\frac{\widetilde{u}'\widetilde{u}}{NT} = \frac{1}{NT}\left(\epsilon'\epsilon - \epsilon' P_X\epsilon + v'D_X' M_X D_X v + v'D_X' M_X\epsilon + v' M_X D_X \epsilon\right).$$

Using Assumptions $1'$, 2 and 3, it is straightforward to show that

$$\begin{aligned} \frac{1}{N}\epsilon' X\left(X'X\right)^{-1}X'\epsilon &= o_p(1), \\ \frac{1}{N}v'D_X' M_X D_X v &= o_p(1), \\ \frac{1}{N}v'D_X' M_X\epsilon &= o_p(1). \end{aligned}$$

and, thus, $\widetilde{\sigma}^2 = \sigma^2 + o_p(1)$.

(ii) Since $u_i = X_i v_i + \epsilon_i$ and

$$\widetilde{u}_i = X_i v_i + \epsilon_i - X_i\left(\widetilde{\beta} - \beta\right),$$

we obtain

$$\begin{aligned} \frac{1}{\sqrt{N}}\widetilde{s}_k &= \frac{1}{\sqrt{N}}\left(\frac{\sigma^4}{\widetilde{\sigma}^4}\right)\frac{1}{2\sigma^4}\sum_i \epsilon_i' A_i^{(k)}\epsilon_i \\ &+ \frac{1}{\sqrt{N}}\left(\frac{\sigma^4}{\widetilde{\sigma}^4}\right)\frac{1}{2\sigma^4}\sum_i v_i'\left(X_i' A_i^{(k)} X_i\right) v_i + o_p(1), \end{aligned} \qquad \text{(A.15)}$$

for $k = 1, ..., K$, where the order of the remainder term follows by similar arguments as in lemma A.1.

(iii) Using the same arguments as in the proof of Theorem 1, the first term of $\widetilde{s}/\sqrt{N}$ in

(A.15) is asymptotically normally distributed. Regarding the second term

$$\frac{1}{\sqrt{N}}\sum_i v_i'\left(X_i'A_i^{(k)}X_i\right)v_i = \frac{1}{N}\sum_i \left(N^{1/4}v_i\right)'\left(X_i'A_i^{(k)}X_i\right)\left(N^{1/4}v_i\right),$$

and by standard results for quadratic forms,

$$\mathbb{E}\left[\left(N^{1/4}v_i\right)'\left(X_iA_i^{(k)}X_i\right)\left(N^{1/4}v_i\right)\middle| X\right] = \mathrm{tr}\left[\left(X_iA_i^{(k)}X_i\right)D_c\right].$$

with $D_c = \mathrm{diag}(c_1, \ldots, c_K)$. Thus by the law of large numbers for sums of independent random variables,

$$\frac{1}{2\sqrt{N}\sigma^4}\sum_i v_i'\left(X_i'A_i^{(k)}X_i\right)v_i \xrightarrow{p} \lim_{N\to\infty}\frac{1}{2\sigma^4 N}\sum_i \mathrm{tr}\left[\left(X_i'A_i^{(k)}X_i\right)D_c\right].$$

Now

$$\sum_i \mathrm{tr}\left[\left(X_i'A_i^{(k)}X_i\right)D_c\right] = \sum_{l=1}^{K} c_l\left(\sum_i\left(\sum_t x_{it,k}x_{it,l}\right)^2 - \frac{1}{NT}\left(\sum_i\sum_t x_{it,k}^2\right)\left(\sum_i\sum_t x_{it,l}^2\right)\right)$$

Define the $K\times 1$ vector ψ elementwise by

$$\psi_k \equiv \sum_{l=1}^{K} c_l \operatorname*{plim}_{N\to\infty}\left(\frac{1}{N}\sum_i\left(\sum_t x_{it,k}x_{it,l}\right)^2 - \frac{1}{T}\left(\frac{1}{N}\sum_i\sum_t x_{it,k}^2\right)\left(\frac{1}{N}\sum_i\sum_t x_{it,l}^2\right)\right)$$

By Slutsky's theorem theorem we obtain

$$\left(\frac{1}{N}V\right)^{-1/2}\left(\frac{1}{\sqrt{N}}\widetilde{s}\right) \xrightarrow{d} \mathcal{N}\left(\psi, I_K\right),$$

and the theorem follows by the definition of the non-central χ^2 distributed random variable with $\psi = \Psi c$ and $c = (c_1, \ldots, c_K)'$.

Proof of Theorem 6

The proof is analogous to the proof of Theorem 5. To show that $\widetilde{\sigma}^2$ remains consistent under the sequence of alternatives we note that

$$\begin{aligned}
\epsilon' X\left(X'X\right)^{-1}X'\epsilon &= O_p(1),\\
v'D_X'M_XD_Xv &= O_p\left(N^{1/2}T\right),\\
v'D_X'M_X\epsilon &= O_p\left(NT^{1/2}\right) + O_p\left(T^{1/2}\right).
\end{aligned}$$

Using the same arguments as in the proof of Theorem 2, $\widetilde{s}/(T\sqrt{N})$ has a limiting normal distribution with nonzero mean which is determined by applying the law of large numbers to the second term in (A.15) with proper normalization.

Proof of Theorem 7 With the Swamy statistic as described in the text, the proof follows the steps outlined in Appendix A.6 in PY.

Details for Remark 9

We study the (k,l) element of $\left(N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\widetilde{u}_{it}^{2}\widetilde{z}_{it}\widetilde{z}_{it}'\right)$ under the sequence of alternatives in Theorem 5. Note that

$$\widetilde{u}_{it}^{2}=\left(\epsilon_{it}+x_{it}'v_i\right)^2+\left(\widetilde{\beta}-\beta\right)'x_{it}x_{it}'\left(\widetilde{\beta}-\beta\right)-2\left(\epsilon_{it}+x_{it}'v_i\right)x_{it}'\left(\widetilde{\beta}-\beta\right), \tag{A.16}$$

and

$$\widetilde{z}_{it,k}=x_{it,k}\sum_{s=1}^{t-1}\left(\epsilon_{is}+x_{is}'v_i+x_{is}'\left(\widetilde{\beta}-\beta\right)\right)x_{is,k}.$$

implying,

$$\begin{aligned}
\widetilde{z}_{it,k}\widetilde{z}_{it,l}&=x_{it,k}x_{it,l}\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,k}\right)\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,l}\right)+v_i'\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}x_{is}'x_{is,l}\right)v_i\\
&+\left(\widetilde{\beta}-\beta\right)'\left(x_{it,k}\left(\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\right)\left(x_{it,l}\left(\sum_{s=1}^{t-1}x_{is}'x_{is,l}\right)\right)\left(\widetilde{\beta}-\beta\right)\\
&+x_{it,k}\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}\left(x_{is}'v_i\right)x_{is,l}\right)+\left(x_{it,k}\sum_{s=1}^{t-1}\left(x_{is}'v_i\right)x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}\epsilon_{is}x_{is,l}\right)\\
&-x_{it,k}x_{it,l}\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,k}\right)\left(\sum_{s=1}^{t-1}x_{is,l}x_{is}'\right)\left(\widetilde{\beta}-\beta\right)-x_{it,k}x_{it,l}\left(\sum_{s=1}^{t-1}\left(x_{is}'v_i\right)x_{is,k}\right)\left(\sum_{s=1}^{t-1}x_{is,l}x_{is}'\right)\left(\widetilde{\beta}-\beta\right)\\
&-x_{it,k}x_{is,l}\left(\sum_{s=1}^{t-1}x_{is}'\left(\widetilde{\beta}-\beta\right)x_{is,k}\right)\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,l}\right)-x_{it,k}x_{is,l}\left(\sum_{s=1}^{t-1}x_{is}'\left(\widetilde{\beta}-\beta\right)x_{is,k}\right)\left(\sum_{s=1}^{t-1}x_{is}'v_ix_{is,l}\right)
\end{aligned} \tag{A.17}$$

First, from the first term on the right hand sides of (A.16) and (A.17), we obtain

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}^{2}x_{it,k}x_{it,l}\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,k}\right)\left(\sum_{s=1}^{t-1}\epsilon_{is}x_{is,l}\right).$$

Notice that this term has the same probability limit as $\widetilde{V}_{k,l}^{*}/N$, which is equal to $\Psi_{k,l}^{*}$. Next, from the first term on the right-hand side in (A.16) and the second term on the right-hand side in (A.17),

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}^{2}v_i'\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}x_{it}'x_{is,l}\right)v_i\\
&=\frac{1}{N^{1.5}}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}^{2}\left(N^{1/4}v_i\right)'\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}x_{it}'x_{is,l}\right)\left(N^{1/4}v_i\right)
\end{aligned}$$

Since ϵ_{it} and v_i are independent conditional on X,

$$\mathbb{E}\left[\epsilon_{it}^2\left(N^{1/4}v_i\right)'\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}x_{it}'x_{is,l}\right)\left(N^{1/4}v_i\right)\,\middle|\,X\right]=\sigma^2\text{tr}\left[B_{it}^X D_c\right]$$

with the $K\times K$ matrix $B_{it}^X=\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}x_{it}'x_{is,l}\right)$ such that

$$\frac{1}{N^{1.5}}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}^2\left(N^{1/4}v_i\right)'\left(x_{it,k}\sum_{s=1}^{t-1}x_{is}x_{is,k}\right)\left(x_{it,l}\sum_{s=1}^{t-1}X_{it}'x_{is,l}\right)\left(N^{1/4}v_i\right)=O_p\left(N^{-1/2}\right)$$

Using the properties of ϵ_{it}, v_i and the fact that $\left(\widetilde{\beta}-\beta\right)=o_p(1)$, it can be shown in a similar manner that all of the remaining terms are of lower order.

B Appendix: Tables

Table 1.1: Rejection frequencies for H_0 : all coefficients are homogenous

	A) Size				B) Power			
	$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM_{reg}	$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM_{reg}
$T = 10$								
$N = 10$	6.3	11.8	2.6	4.7	5.5	5.4	11.3	4.1
$N = 20$	5.6	12.0	3.2	4.3	8.4	4.2	24.8	7.0
$N = 30$	5.5	11.4	3.7	4.3	11.5	6.3	35.2	10.4
$N = 50$	5.2	8.9	3.7	4.1	18.9	15.9	51.2	19.6
$N = 100$	5.4	8.3	4.7	4.8	36.1	46.5	77.5	47.0
$N = 200$	4.6	7.5	4.9	4.6	65.6	87.3	96.9	83.1
$T = 20$								
$N = 10$	5.3	15.1	2.6	6.3	17.1	3.4	28.1	12.4
$N = 20$	5.7	14.2	3.4	5.7	35.0	9.5	53.0	25.8
$N = 30$	5.9	12.8	3.8	5.8	50.7	23.4	70.5	43.2
$N = 50$	5.1	10.8	4.1	5.3	74.6	53.5	88.9	71.8
$N = 100$	4.5	8.3	4.3	4.7	95.7	90.1	99.1	96.5
$N = 200$	5.1	7.1	5.2	5.5	99.9	98.8	100.0	100.0
$T = 30$								
$N = 10$	4.7	15.6	2.4	7.0	34.4	4.6	43.0	22.4
$N = 20$	4.5	14.5	3.5	6.2	64.8	19.7	74.3	50.9
$N = 30$	5.2	13.2	3.9	5.6	81.9	42.5	88.9	73.3
$N = 50$	5.3	11.6	4.5	5.9	96.3	76.9	98.2	94.5
$N = 100$	5.2	8.9	4.3	4.7	100.0	96.0	100.0	99.9
$N = 200$	5.2	7.4	5.0	5.8	100.0	99.3	100.0	100.0

Notes: Rejection frequencies (in %) for $K = 3$ under the null (panel A) and the alternative hypothesis (panel B). Nominal size is 5%.

Table 1.2: Rejection frequencies: model with individual effects

	A) Size				B) Power					
	LM	LM^*	LM_{reg}	LM_{ac}	$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM^*	LM_{reg}	LM_{ac}
$T = 10$										
$N = 10$	2.6	2.9	4.8	3.4	5.7	5.3	8.3	8.6	4.2	7.8
$N = 20$	3.3	4.0	5.2	3.8	8.1	4.3	21.6	20.2	6.8	17.8
$N = 30$	3.1	3.9	4.7	4.1	8.7	5.3	26.5	24.2	8.8	22.5
$N = 50$	3.8	4.6	4.3	4.5	13.1	10.8	42.0	38.9	15.7	35.8
$N = 100$	4.1	4.8	4.7	4.6	29.2	40.8	73.4	67.7	40.4	64.5
$N = 200$	4.2	4.8	4.7	4.7	49.0	79.5	91.8	87.9	72.9	84.9
$T = 20$										
$N = 10$	2.3	2.2	6.4	2.4	19.1	3.6	31.4	30.8	13.1	26.3
$N = 20$	3.4	3.5	6.0	3.9	39.1	9.4	59.3	57.8	31.7	53.6
$N = 30$	3.8	4.1	5.4	4.2	40.0	15.8	60.2	58.0	33.0	53.8
$N = 50$	3.8	4.2	5.1	3.8	72.5	53.9	89.0	88.1	73.3	84.7
$N = 100$	4.3	4.6	5.5	4.5	91.4	89.0	98.2	97.7	93.9	96.9
$N = 200$	4.8	4.8	5.3	4.7	99.8	98.4	100.0	100.0	99.9	100.0
$T = 30$										
$N = 10$	2.8	2.9	8.0	3.5	26.6	3.9	37.8	37.6	18.0	29.4
$N = 20$	3.6	3.5	6.2	3.8	66.8	24.6	78.0	77.3	56.3	72.2
$N = 30$	3.7	3.7	5.8	4.1	91.8	47.8	95.3	94.9	86.4	93.7
$N = 50$	4.7	4.5	5.8	4.8	94.6	67.1	97.0	96.9	91.8	96.1
$N = 100$	4.4	4.4	5.1	4.7	99.9	92.8	100.0	100.0	99.9	99.9
$N = 200$	4.6	4.6	4.8	4.6	100.0	98.4	100.0	100.0	100.0	100.0

Notes: Left panel: rejection frequencies (in %) under the null hypothesis with same design as in Table 1.1 and orthogonal forward orthogonalization to eliminate fixed effects. Right panel: rejection frequencies (in %) under the alternative hypothesis and forward orthogonalization to eliminate fixed effects. Nominal size is 5%.

Table 1.3: Size and power for t-distributed errors

A) Size		$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM_{adj}	LM^*	LM_{reg}
$T = 10$							
	$N = 10$	6.3	9.0	3.5	2.9	4.0	3.9
	$N = 20$	6.1	10.5	5.4	4.0	6.1	4.5
	$N = 30$	5.6	10.2	6.9	5.0	6.2	4.9
	$N = 50$	5.1	8.9	7.5	5.3	6.6	4.4
	$N = 100$	5.0	7.5	9.8	6.2	7.2	4.9
	$N = 200$	5.5	6.9	10.7	5.7	7.6	5.2
$T = 20$							
	$N = 10$	5.5	13.2	3.5	2.8	3.6	6.0
	$N = 20$	5.8	13.0	5.1	4.0	4.6	6.3
	$N = 30$	5.4	11.7	5.5	4.3	4.8	5.6
	$N = 50$	5.2	10.4	6.7	4.8	5.1	5.3
	$N = 100$	4.9	8.0	8.0	5.5	5.9	5.1
	$N = 200$	5.1	7.0	8.9	5.5	5.4	4.8
$T = 30$							
	$N = 10$	5.8	15.0	3.0	2.6	3.0	7.0
	$N = 20$	5.0	13.6	4.5	3.5	3.9	5.7
	$N = 30$	4.7	12.4	5.2	4.3	4.2	5.5
	$N = 50$	5.1	11.3	5.9	4.7	5.1	5.9
	$N = 100$	5.0	8.5	6.4	5.0	4.8	5.1
	$N = 200$	5.3	7.7	7.3	5.0	5.2	4.9

B) Power		$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM_{adj}	LM^*	LM_{reg}
$T = 10$							
	$N = 10$	5.9	3.4	12.4	10.9	11.8	4.9
	$N = 20$	9.7	3.9	25.6	22.7	22.9	7.5
	$N = 30$	13.6	6.9	36.1	31.8	32.4	11.9
	$N = 50$	24.1	16.0	52.4	46.7	47.9	22.6
	$N = 100$	45.3	44.4	78.9	71.8	73.9	49.4
	$N = 200$	76.1	83.1	95.7	92.6	93.8	83.2
$T = 20$							
	$N = 10$	20.0	3.2	30.5	28.6	29.4	13.7
	$N = 20$	39.5	10.1	53.6	50.0	52.5	29.7
	$N = 30$	57.3	22.8	70.5	67.3	68.5	46.7
	$N = 50$	80.0	51.5	88.3	85.5	86.7	72.2
	$N = 100$	97.8	89.5	99.0	98.4	99.0	96.4
	$N = 200$	100.0	98.8	100.0	100.0	100.0	100.0
$T = 30$							
	$N = 10$	37.5	3.7	45.8	43.8	45.1	25.4
	$N = 20$	67.7	19.6	74.3	71.9	73.3	54.3
	$N = 30$	85.4	43.0	88.6	86.6	87.4	74.4
	$N = 50$	97.3	76.2	98.0	97.2	97.7	93.8
	$N = 100$	100.0	95.9	100.0	100.0	100.0	99.9
	$N = 200$	100.0	99.4	100.0	100.0	100.0	100.0

Notes: Rejection frequencies (in %) for K=3 under the null (panel A) and the alternative hypothesis (panel B) when ϵ_{it} is drawn from a t-distribution with five degrees of freedom. Nominal size is 5%.

Table 1.4: Size and power in a model with autocorrelated errors

		$\widetilde{\Delta}_{\text{adj}}$	CLM	LM	LM^*	LM_{reg}	LM_{ac}
A) Size:	$\rho = 0.2$						
$T = 10$							
	$N = 10$	5.3	7.4	4.8	4.8	3.9	5.5
	$N = 50$	9.0	4.6	12.1	13.1	5.3	4.3
	$N = 100$	15.9	6.0	17.1	21.0	7.6	5.0
$T = 20$							
	$N = 10$	7.6	6.8	6.4	6.7	4.4	4.5
	$N = 50$	16.3	4.9	15.1	15.7	6.3	4.9
	$N = 100$	34.7	9.7	23.8	25.7	11.5	4.4
$T = 30$							
	$N = 10$	7.3	6.6	6.7	6.9	5.1	4.1
	$N = 50$	27.9	5.2	16.1	16.9	6.3	3.6
	$N = 100$	46.7	12.7	25.6	27.6	12.4	4.1
B) Size:	$\rho = 0.5$						
$T = 10$							
	$N = 10$	6.5	3.4	9.3	11.0	3.6	7.7
	$N = 50$	48.2	14.3	52.6	63.5	22.3	4.8
	$N = 100$	68.6	25.9	67.7	79.8	30.9	4.8
$T = 20$							
	$N = 10$	14.3	3.0	8.8	9.6	4.4	4.7
	$N = 50$	88.9	32.8	65.1	71.4	39.0	4.3
	$N = 100$	99.4	74.8	90.5	94.2	73.8	4.6
$T = 30$							
	$N = 10$	32.9	4.1	24.6	25.6	9.6	4.7
	$N = 50$	94.7	36.7	66.6	71.1	42.0	4.5
	$N = 100$	99.9	81.1	92.1	94.6	77.2	4.5
C) Power:	$\rho = 0$						
$T = 10$							
	$N = 10$	5.4	8.1	8.3	8.1	4.3	7.7
	$N = 50$	10.5	7.0	46.0	43.6	10.5	34.2
	$N = 100$	16.9	23.2	69.1	65.2	22.3	54.5
$T = 20$							
	$N = 10$	11.6	3.7	26.7	26.4	9.6	17.9
	$N = 50$	59.3	34.6	89.6	88.7	60.2	82.4
	$N = 100$	85.9	80.0	99.0	99.0	90.5	97.9
$T = 30$							
	$N = 10$	14.4	4.6	28.3	27.7	10.3	18.2
	$N = 50$	88.6	63.9	97.5	97.4	85.8	94.9
	$N = 100$	99.7	95.7	100.0	100.0	99.9	100.0

Notes: Panel A) and B) present rejection frequencies (in %) for K=3 under serial correlation of errors ($\rho = 0.2$ and $\rho = 0.5$, respectively) and the null hypothesis. Panel C) presents rejection frequencies (in %) under the alternative hypothesis without serial correlation in errors. Nominal size is 5%.

C Appendix: Figures

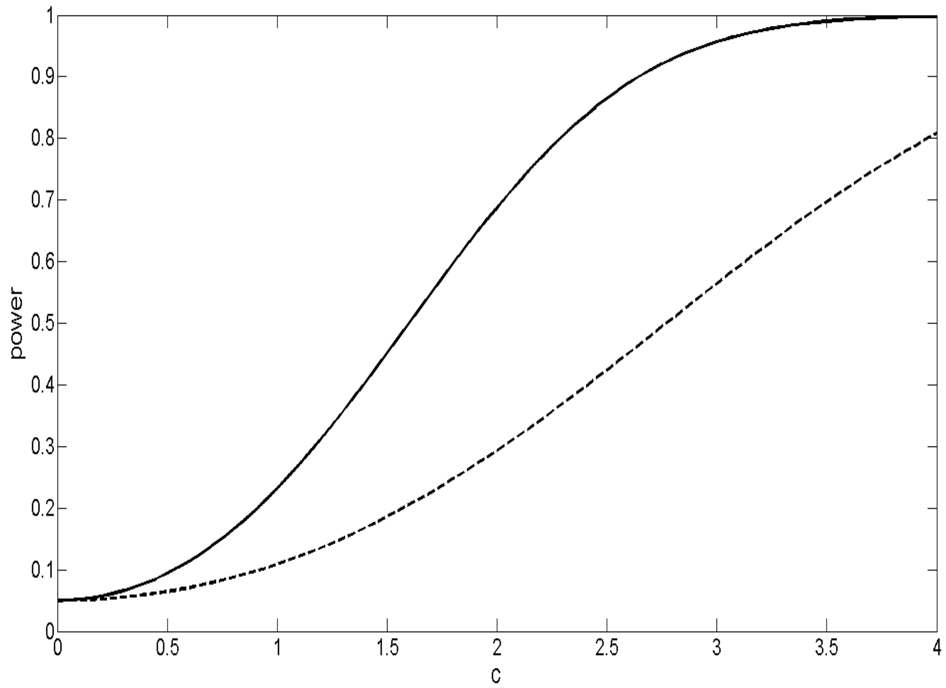


Figure 1.1: Asymptotic local power of the LM (solid line) and the Δ test (dahed line) when $\sigma^2_{i,x} \sim \chi^2_1$.

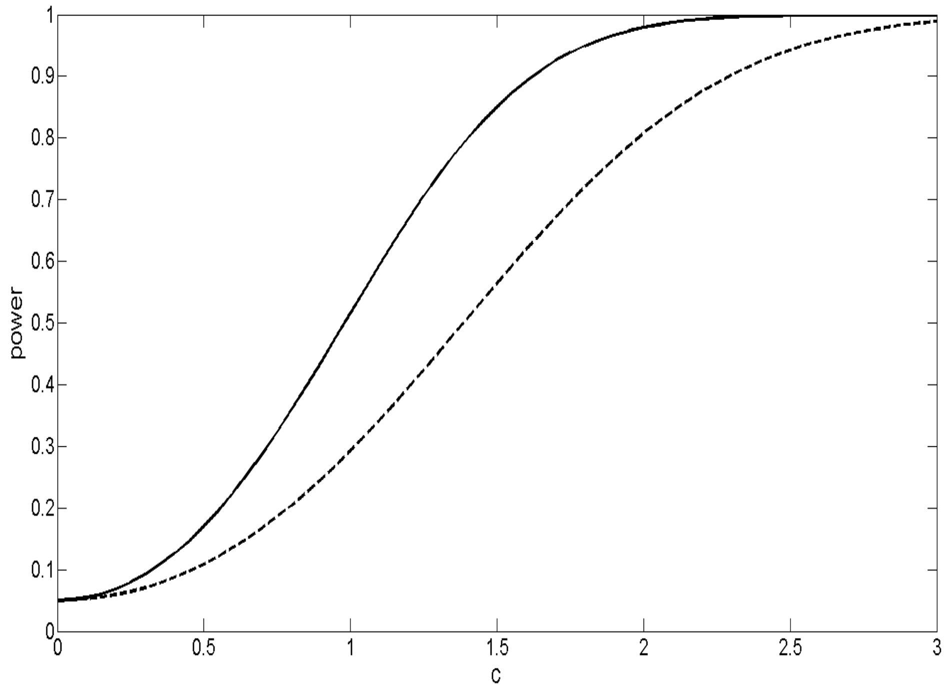


Figure 1.2: Asymptotic local power of the LM (solid line) and the Δ test (dashed line) when $\sigma^2_{i,x} \sim \chi^2_2$.

Bibliography

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51.

Baltagi, B., Q. Feng, and C. Kao (2011). Testing for sphericity in a fixed effects panel data model. *The Econometrics Journal 14*, 25–47.

Blomquist, J. and J. Westerlund (2013). Testing slope homogeneity in large panels with serial correlation. *Economics Letters 121*(3), 374 – 378.

Breusch, T. S. and A. R. Pagan (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies 47*, 239–253.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72*, 320–338.

Honda, Y. (1985). Testing the error components model with non-normal disturbances. *Review of Economic Studies 52*, 681–690.

Hsiao, C. and M. H. Pesaran (2008). Random coefficient models. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Chapter 6. Springer.

Juhl, T. and O. Lugovskyy (2014). A test for slope homogeneity in fixed effects models. *Econometric Reviews 33*, 906–935.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics 68*, 79–113.

Pesaran, M. H. and T. Yamagata (2008). Testing slope homogeneity in large panels. *Journal of Econometrics 142*, 50–93.

Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), 1057–1111.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica 38*, 311–323.

Ullah, A. (2004). *Finite-sample econometrics.* Oxford University Press.

Wand, M. P. (2002). Vector differential calculus in statistics. *The American Statistician 56*, 1–8.

White, H. (2001). *Asymptotic theory for econometricians.* Emerald.

Wiens , D. P. (1992). On moments of quadratic forms in non-spherically distributed variables. *Statistics 23*(3), 265–270.

Chapter 2

LM Tests for Shock Induced Asymmetries in Time Series

2.1 Introduction

"Losses loom larger than corresponding gains"

D. Kahneman and A. Tversky

In the last decades there has been a significant increase in findings from empirical studies in economics and finance indicating that processes react differently to positive and negative shocks. For instance, Koutmos (1999) tests and finds asymmetries in the conditional mean and the conditional standard deviation of the stock returns distribution of the G7 national stock markets. Karras and Stokes (1999) examine asymmetric effects of money-supply shocks in OECD countries and report that negative shocks have a stronger effect on output than positive ones. Other examples can be found in Elwood (1998), Kilian and Vigfusson (2011) and Brännäs et al. (2012) among others. In univariate time series settings this led to an asymmetric time series paradigm introduced by Wecker (1981). As the main framework to model asymmetries induced by the sign of innovations, Wecker (1981) suggests asymmetric moving average models (AsMA, hereafter). Complementary to the AsMA model we consider an extension of the autoregressive process to an asymmetric one (AsAR, hereafter). Note that the asymmetric time series models considered in this paper introduce a type of nonlinearity to the dynamics of the process, which differs from the one described by the threshold autoregressive model (TAR). In particular, the TAR model splits the sample into groups (regimes) based on the observed threshold variable and the unknown threshold parameter, while AsMA and AsAR models describe nonlinearity through the sign of shocks.

The potential presence of shock induced asymmetries raises the natural question of (pre)testing for the correct model specification. This testing problem has already been discussed in the literature. To test for the conventional moving average model against

AsMA, Wecker (1981) suggests a likelihood ratio test (LR, hereafter), while Brännäs and De Gooijer (1994) construct a Wald-type test to choose the correct model specification. Besides this, Brännäs et al. (1998) consider a test statistic based on the artificial regression constructed from the Lagrange multiplier (LM, hereafter) principle. However, the asymmetric nature of the AsMA model makes the corresponding likelihood function non-differentiable.[1] This in turn prevents the use of classical likelihood based tests, such as LM, LR and Wald test, since the standard approach of deriving the gradient and the Hessian from the likelihood function as well as the asymptotic behavior of these statistics are not valid anymore.

In this paper we contribute to the literature by constructing new test statistics based on the LM approach that account for the non-differentiability of the likelihood function. The tests are derived for AsMA and AsAR models. To deal with the absence of smoothness in the log-likelihood function we resort to the treatment of non-differentiability offered by Phillips (1991) for LAD estimators. The idea is to examine the problem in the space of generalized functions (distributions) whose derivatives do not exist in the classical sense, but can be accommodated by distributional derivatives. This approach allows us to operate with first order conditions and derive LM type test statistics. Moreover, with this generalization the asymptotic properties of the test statistics can be obtained. We show that the limiting distribution is a standard χ^2 distribution under the null hypothesis of no asymmetric effects. Further, by means of Monte Carlo simulations the finite sample properties of the new test statistics are explored in different setups. Finally, in order to make the testing procedures more accessible to potential users, it is shown via Monte Carlo experiments that the standard model selection criteria, such as BIC or HQ, applied to a linear model provide a reliable estimate of the lag length for the asymmetric counterpart model.

To illustrate the use of the proposed techniques, we apply the test to the U.S. unemployment rate. Our results show strong evidence that the growth of the unemployment rate is affected by an asymmetric impact of positive and negative shocks.

The remainder of this paper is as follows. Section 2.2 introduces the modelling framework for asymmetric time series. The construction of the LM type tests is described in Section 2.3. In Section 2.4 the asymptotic properties of the proposed statistics are investigated. In Section 2.5 we present results from a simulation study. An empirical example is discussed in Section 2.6. The final section contains concluding remarks. Proofs, figures and tables are relegated to the Appendix.

[1] Section 2.3 provides a detailed discussion on the type of non-differentiability present in log-likelihood functions obtained for asymmetric time series.

2.2 Preliminaries

This section lays out a basis of the asymmetric time series models as a counterpart to the usual linear moving average and autoregressive models. The main characteristic of this model class that distinguishes it from other well established nonlinear models (such as threshold AR models for instance) is that two different filters, one for positive and one for negative innovations are employed. In particular, Wecker (1981) advocated the use of the asymmetric moving average model which takes the form

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + ... + \alpha_p \varepsilon_{t-p} + \beta_1 \varepsilon_{t-1}^+ + ... + \beta_p \varepsilon_{t-p}^+, \tag{2.1}$$

where $\varepsilon_t^+ = \varepsilon_t 1(\varepsilon_t \geq 0)$ and $1(\cdot)$ defines an indicator function. We also complement Wecker's approach by considering the asymmetric autoregressive model defined as

$$y_t = \varepsilon_t - \alpha_1 y_{t-1} - ... - \alpha_p y_{t-p} - \beta_1 y_{t-1}^+ - ... - \beta_p y_{t-p}^+, \tag{2.2}$$

where $y_t^+ = y_t 1(\varepsilon_t \geq 0)$. In both models it is assumed that $y_t = 0$ for $t \leq 0$ and that the random disturbance term ε_t is a real *i.i.d.* sequence with $N(0, \sigma^2)$ distribution. The normality assumption is necessary only for the derivation of the LM statistics. For the application as well as for the derivation of the asymptotic results, this assumption is relaxed. In general, for the asymptotic analysis we require the process y_t to be stationary and invertible under the null hypothesis of no asymmetric effects. For this reason it is assumed that the roots of $\alpha(z) = 1 + \sum_{i=1}^{p} \alpha_i z^i$ lie outside the unit circle. We discuss the consequences of a violation of the stationarity assumption for the asymptotics in Remark 12 of Section 2.4.

To express model (2.1) and (2.2) in matrix notations, define $\mathbf{B}$ as a $T \times T$ backshift matrix with typical element $B_{ij} = 1$ if $i - j = 1$ and zero otherwise. As a convention $\mathbf{B}^0 = \mathbf{I}$ is set to be the identity matrix. Matrix $\mathbf{D}_{1(\varepsilon)} = diag\{1(\varepsilon_1 \geq 0), ..., 1(\varepsilon_T \geq 0)\}$ defines a $T \times T$ diagonal matrix and $\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_p)'$, $\boldsymbol{\beta} \equiv (\beta_1, ..., \beta_p)'$ are vectors of parameters. Then models (2.1) and (2.2) can be rewritten as

$$\mathbf{y} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right) \boldsymbol{\varepsilon}, \tag{2.3}$$

and

$$\left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right) \mathbf{y} = \boldsymbol{\varepsilon}, \tag{2.4}$$

respectively, where $\mathbf{M}_{\boldsymbol{\alpha}} = \sum_{i=0}^{p} \alpha_i \mathbf{B}^i$ and $\mathbf{M}_{\boldsymbol{\beta}} = \sum_{i=1}^{p} \beta_i \mathbf{B}^i$ with $\alpha_0 = 1$, $\mathbf{y} = (y_1, ..., y_T)'$ denotes a $T \times 1$ vector of observations and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ is a $T \times 1$ vector of error terms.

The representations (2.3) and (2.4) are convenient for our discussion since deviations from the conventional symmetric MA(p) or AR(p) models are now represented in both cases by matrix $\mathbf{M}_{\boldsymbol{\beta}}$. Therefore, the main question of interest can be formulated as

$$\mathrm{H}_0 : \mathbf{M}_{\boldsymbol{\beta}} = 0, \ \ (\text{or } \boldsymbol{\beta} = 0),$$

against the two alternatives that $\mathrm{H}_A : \{y_t\}$ is generated by (2.3) or $\mathrm{H}_B : \{y_t\}$ is generated by (2.4).

2.3 The Lagrange multiplier test

The corresponding log-likelihood function for time series processes (2.3) and (2.4) is given by

$$\mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2\right) = const - \frac{T}{2} \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}, \tag{2.5}$$

where $\boldsymbol{\varepsilon} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right)^{-1} \mathbf{y}$ for the AsMA(p) case and $\boldsymbol{\varepsilon} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right) \mathbf{y}$ for the AsAR(p) model. Denote $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{'}, \boldsymbol{\beta}^{'})'$ as the parameter vector of interest and $\widehat{\boldsymbol{\theta}}_0 = (\widehat{\boldsymbol{\alpha}}^{'}, \mathbf{0})'$ as the restricted ML estimator of $\boldsymbol{\theta}_0 = \left(\boldsymbol{\alpha}^{'}, \mathbf{0}\right)'$. The parameter σ^2 can be concentrated out. Furthermore, let $\mathbf{s}(\boldsymbol{\theta}) = \partial \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ denote the score and $\mathcal{H}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1} \partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ the asymptotic Hessian of the log-likelihood (2.5). It is convenient in this testing framework to use a partitioning of the score $\mathbf{s}(\boldsymbol{\theta}) = \left(\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})', \mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta})'\right)'$, with $\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \partial \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\alpha}$, and $\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \partial \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\beta}$. The asymptotic Hessian matrix can be expressed as

$$\mathcal{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{H}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}) & \mathcal{H}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta}) \\ \mathcal{H}_{\boldsymbol{\beta\alpha}}(\boldsymbol{\theta}) & \mathcal{H}_{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) \end{bmatrix}.$$

Here $\mathcal{H}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1} \partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'$, $\mathcal{H}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1}\, \partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}'$, etc. Then the usual form of the LM test for testing $\mathbf{H}_0$ can be written as,

$$\mathrm{LM}_T = \frac{1}{T} \mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)' \mathbf{V}_{\boldsymbol{\beta}}^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right) \mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right), \tag{2.6}$$

where $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ represents the variance of the score $\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and is taken from the appropriate diagonal block of the $\mathcal{H}(\boldsymbol{\theta})$ matrix, i.e., $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathcal{H}_{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) - \mathcal{H}_{\boldsymbol{\beta\alpha}}(\boldsymbol{\theta})\, \mathcal{H}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta})^{-1} \mathcal{H}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta})$.

Notice that the presence of the indicator functions in the likelihood function (2.5) makes it non-differentiable for ε_t at zero for all $t = 1, ..., T$. Therefore, the standard framework for deriving the LM test (and its asymptotics) with absence of smoothness is in general not applicable. We suggest here to resort to Phillips (1991), where a solution to nonregular problems like discontinuities in the criterion function is proposed on the example of the LAD estimator. In particular, if derivatives do not exist in the usual sense, these may be accomodated directly by the use of generalized functions or distributions (See, e.g., Gelfand and Shilov (1964) for more detailed overview of the theory of generalized functions). As presented below, this generalization of the classical approach does not only provide a justification of the derivation of the LM test but it also helps to develop generalized Taylor series expansions of the first order conditions which in turn are useful to extract the asymptotic theory.

We start with the derivative of the indicator function that can be written as the Dirac delta (generalized) function, i.e.,

$$\partial 1_{(x\geq 0)}/\partial x = \delta(x).$$

The required properties of the $\delta(x)$ function are given in Appendix A, Lemma A.5. Then for the convenience of the notations we define matrix $\mathbf{M}_{\boldsymbol{\alpha\beta}} \equiv \mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{1(\boldsymbol{\varepsilon})}$ which essentially represents the filtering (structure) of the processes (2.3) and (2.4). By proceeding in a purely formal way the derivative of $\mathbf{M}_{\boldsymbol{\alpha\beta}}$ with respect to $\boldsymbol{\theta}$ can be compactly written as

$$\frac{\partial \mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\partial \theta_i} = \begin{cases} \mathbf{B}^i + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} & \text{for } \theta_i = \alpha_i \\ \mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} & \text{for } \theta_i = \beta_i \end{cases}, \tag{2.7}$$

where $\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}$ is a $T \times T$ diagonal matrix defined as $diag\{\delta(\varepsilon_1), ..., \delta(\varepsilon_T)\}$ and $\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} = diag\,\{\partial\varepsilon_1/\partial\theta_i\ \ , ..., \partial\varepsilon_T/\partial\theta_i\}$. Further, under the null hypothesis $\mathbf{M}_{\boldsymbol{\beta}} = 0$ and $\partial\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}/\partial\theta_i$ takes a simple matrix form $\mathbf{B}^i$ or $\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}$. Finally, using standard results for matrix derivatives (see, e.g., Lütkepohl, 1996), the elements of the score vector $\mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)$ under the null hypothesis can be presented for process (2.3) in a quadratic form as

$$s_{\beta,i}\left(\widehat{\boldsymbol{\theta}}_0\right) = \frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{\varepsilon}}'\left(\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})}\right)'\widehat{\boldsymbol{\varepsilon}}, \tag{2.8}$$

and for process (2.4) as

$$s_{\beta,i}\left(\widehat{\boldsymbol{\theta}}_0\right) = -\frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{\varepsilon}}'\left(\mathbf{B}^i\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})}\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\right)'\widehat{\boldsymbol{\varepsilon}}, \tag{2.9}$$

where $i = 1, ..., p$, $\widehat{\boldsymbol{\varepsilon}}$ is the ML estimator of $\boldsymbol{\varepsilon}$ under H_0 and $\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})} = diag\{1_{(\widehat{\varepsilon}_1\geq 0)}, ..., 1_{(\widehat{\varepsilon}_T\geq 0)}\}$. The vector $\widehat{\boldsymbol{\varepsilon}}$ is estimated from the MA as $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{y}$ or from the AR process as $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}\boldsymbol{y}$, respectively, where $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}} = \sum_{i=0}^{p}\widehat{\alpha}_i\mathbf{B}^i$.

2.3.1 Variants of the LM test

There are as many different ways to compute the LM statistic (2.6) as there are asymptotically valid ways to estimate the covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_0\right)$. So far, we have assumed that $\mathbf{V}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_0\right)$ is derived from the asymptotic Hessian matrix evaluated under the null. However, any method that allows us to estimate $\mathbf{V}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_0\right)$ consistently can be used. In what follows, several different approaches that are commonly used in the literature are discussed.

Empirical Hessian and information matrix

The most straightforward method, based on (2.6), to compute the negative of the Hessian evaluated at the restricted vector of ML estimates $\widehat{\boldsymbol{\theta}}_0$, which is referred to as the empirical Hessian estimator, i.e.,

$$\mathbf{V}_{\boldsymbol{\beta}}^{(H)}\left(\widehat{\boldsymbol{\theta}}_0\right)=\frac{1}{T}\left(\mathbf{H}_{\boldsymbol{\beta\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)-\mathbf{H}_{\boldsymbol{\beta\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)\mathbf{H}_{\boldsymbol{\alpha\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)^{-1}\mathbf{H}_{\boldsymbol{\alpha\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)\right),$$

where $\mathbf{H}_{\boldsymbol{\alpha\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)=-\partial^2\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_0\right)/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'$, $\mathbf{H}_{\boldsymbol{\alpha\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)=\partial^2\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_0\right)/\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'$, etc. However, this estimator cannot be easily handled in practice due to the presence of the Dirac delta functions and its derivatives even under the null.

However, it can be shown that by taking the expectation the terms that include delta functions in the expression of the $\mathbf{V}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_0\right)$ can be eliminated. This follows from the definition of the delta function and so called sifting property (see Lemma A.5). Therefore, the information matrix approach can be used instead of the empirical Hessian to obtain an efficient and applicable estimator of $\mathbf{V}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_0\right)$. Hence, in what follows the estimator $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is constructed as

$$\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)=\frac{1}{T}\left(\mathbf{J}_{\boldsymbol{\beta\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)-\mathbf{J}_{\boldsymbol{\beta\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)\mathbf{J}_{\boldsymbol{\alpha\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)^{-1}\mathbf{J}_{\boldsymbol{\alpha\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)\right), \tag{2.10}$$

where $\mathbf{J}_{\boldsymbol{\alpha\alpha}}\left(\boldsymbol{\theta}\right)=\mathbb{E}\left[\mathbf{s}_{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)\mathbf{s}_{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)'\right]$, $\mathbf{J}_{\boldsymbol{\alpha\beta}}\left(\boldsymbol{\theta}\right)=\mathbb{E}\left[\mathbf{s}_{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)\mathbf{s}_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}\right)'\right]$, etc.

Finally, to derive an analytical expression for $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}$ we relax the Gaussian distributional assumption of ε_t for more specific restrictions on the existence of higher-order moments. This allows to robustify the estimator $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ to non-normal disturbances.

Assumption 2″

(i) $\{\varepsilon_t\}$ *is an i.i.d. sequence with zero mean and* $\mathbb{E}\left[\varepsilon_t^2\right]=\sigma^2>0$*;*

(ii) *There is a positive constant* $C>0$ *such that* $\mathbb{E}\left|\varepsilon_t\right|^{4+r}<C<\infty$ *for some* $r>0$ *and all* t*;*

(iii) *The density function of* ε_t*, defined as* $f_\varepsilon(\cdot)$*, is continuous and differentiable at zero.*

Assumption 3 constitutes sufficient conditions for the asymptotic results obtained in this paper. While part (i) and (ii) are standard identification assumptions in the time series literature, part (iii) restricts the analysis to innovations with a smooth density function at zero.

The matrix $\mathbf{J}_{\boldsymbol{\alpha\alpha}}\left(\boldsymbol{\theta}_0\right)$ is obtained by using standard results for quadratic forms (see, e.g., Ullah, 2004, Appendix A.5) and has the same shape for both H_A and H_B alternatives, with typical $J_{i,j}\left(\boldsymbol{\theta}_0\right)$ element

$$J_{i,j}\left(\boldsymbol{\theta}_0\right)\equiv\mathbb{E}\left[s_{\alpha,i}\left(\boldsymbol{\theta}_0\right)s_{\alpha,j}\left(\boldsymbol{\theta}_0\right)\right]=\operatorname{tr}\left[(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i)(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j)'\right], \tag{2.11}$$

for $i,j=1,...,p$. However, the results for other components $\mathbf{J}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta})$, $\mathbf{J}_{\boldsymbol{\beta\alpha}}(\boldsymbol{\theta})$ and $\mathbf{J}_{\boldsymbol{\beta\beta}}(\boldsymbol{\theta})$ differ depending on the modeling framework as presented below. Note, that in the following Lemmas we omit the argument $\boldsymbol{\theta}_0$ in $J_{i,j}$, $s_{\alpha,i}$ and $s_{\beta,i}$ to lighten the notational load.

Lemma A.3 *Let $\phi_k = \mathbb{E}\left(\varepsilon_t^+\right)^k$ for $k=1,2$. Then under the data generating process (2.3), assumption 3 and the null hypothesis,*

$$\begin{aligned}
\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] &= \gamma_1 J_{i,j}, &(2.12)\\
\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \left(\gamma_1-\gamma_2\right)J_{i,j}+\gamma_2 W_{i,j} &(2.13)
\end{aligned}$$

where $1\le i,j\le p$, $\gamma_1=\phi_2/\sigma^2$, $\gamma_2=\left(\phi_1\right)^2/\sigma^2$ and $W_{i,j}=\boldsymbol{l}'(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i)(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j)'\boldsymbol{l}$ with $\boldsymbol{l}$ being a $T\times 1$ vector of ones.

The invertibility of the process y_t ensures the existence of the inverse of $M_{\boldsymbol{\alpha}}$ under the null. Hence,

$$\mathbf{M}_{\boldsymbol{\alpha}}^{-1}=\left(\sum_{i=0}^{p}\alpha_i\mathbf{B}^i\right)^{-1}=\sum_{i=0}^{\infty}\psi_i\mathbf{B}^i, \qquad (2.14)$$

where $\psi_0=1$ and $\sum_{i=0}^{\infty}|\psi_i|<\infty$.

Lemma A.4 *Let $\phi_k = \mathbb{E}\left(\varepsilon_t^+\right)^k$ for $k=1,2$. Then under the data generating process (2.4), assumption 3 and the null hypothesis,*

$$\begin{aligned}
\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] &= \begin{cases}\mathcal{F}_0 J_{i,j} & \text{for } i>j\\ \mathcal{F}_0 J_{i,j}+\gamma_1\psi_{|i-j|} & \text{for } i\le j\end{cases}, &(2.15)\\
\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \begin{cases}\mathcal{F}_0 J_{i,j}+\gamma_1 & \text{for } i=j\\ \mathcal{F}_0^2 J_{i,j}+\gamma_2 & \text{for } i\ne j\end{cases}, &(2.16)
\end{aligned}$$

where $1\le i,j\le p$, $\mathcal{F}_0=(1-F_\varepsilon(0))$ and $F_\varepsilon(\cdot)$ denotes the distribution function of ε; $\gamma_1=(T-i)(\phi_2-\sigma^2\mathcal{F}_0)/\sigma^2$, $\gamma_2=\gamma_1\mathcal{F}_0\psi_{|i-j|}+\phi_1^2/\sigma^2\left(T-\max(i,j)\right)$.

Therefore, to test for no asymmetric effects of innovations it is sufficient to estimate parameter vector $\boldsymbol{\alpha}$ and error vector $\boldsymbol{\varepsilon}$ under the null and use them to construct the components of the LM test (2.6), i.e.,

$$\mathrm{LM}_T^{(IM)}=\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)'\left[\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)\right]^{-1}\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right), \qquad (2.17)$$

where $\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is given by (2.8) or (2.9) and $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is derived in (2.11) and Lemma A.3 or Lemma A.4 under the null hypothesis of the interest, respectively.

OPG variant

The second method is the one that can be most easily obtained. It is based on the outer product of the gradient and is referred to as the OPG estimator. First, recall that the inverse of $\mathbf{M}_{\boldsymbol{\alpha}}$ under the null is given by $\mathbf{M}_{\boldsymbol{\alpha}}^{-1} = \sum_{k=0}^{\infty} \psi_k \mathbf{B}^k = \sum_{k=0}^{T-1} \psi_k L^k$, where $\psi_0 = 1$ and $\sum_{k=0}^{\infty} |\psi_k| < \infty$. Then we can write the score vector $\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ as the sum of T contributions

$$s_{\theta,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{t=1}^{T} g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right), \tag{2.18}$$

where $i = 1, ..., p$, $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s \widehat{\psi}_{t-s-i}$ for $\theta_i = \alpha_i$, and if $\theta_i = \beta_i$ then $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s^{+} \widehat{\psi}_{t-s-i}$ for the AsMA model and $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s 1(\varepsilon_{t-1} \geq 0) \widehat{\psi}_{t-s-i}$ for the AsAR model. Define the $T \times 2p$ matrix $\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ with typical element $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$. Hence, if the OPG estimator is used in (2.6) the statistic becomes

$$\mathrm{LM}_T^{(OPG)} = \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)' \left[\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)' \mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\right]^{-1} \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right). \tag{2.19}$$

Furthermore, statistic (2.19) can readily be computed by use of an artificial regression, which has the form

$$\boldsymbol{l} = \mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) \boldsymbol{c} + \boldsymbol{u}, \tag{2.20}$$

where $\boldsymbol{l}$ is the vector of ones, $\boldsymbol{c}$ is a parameter vector and $\boldsymbol{u}$ is a residual vector. The explained sum of squares obtained from (2.20) is numerically equal to the OPG variant of the LM statistic (2.19).

This OPG variant has the advantage of being easy to calculate and is known to provide a heteroskedasticity robust version of the LM test (2.6). Nevertheless, it should be used with caution since there is evidence (see e.g., Davidson and MacKinnon, 1983 among others) suggesting that this form tends to be less reliable in finite samples. Section 2.5 provides a further discussion of this issue.

Other regression based variants

Other variants of the LM test presented in the from of artificial regressions can be used for our testing purpose. In this section we discuss one of the best known artificial regression forms of the LM test that is based on the Gauss-Newton regressions. For a review of other available regression based procedures see for instance Davidson and MacKinnon (2001). This approach simply involves regressing the disturbances from the restricted model on the derivatives of the criterion function with respect to all parameters of the unrestricted model.

More precisely, consider the following auxiliary test regression

$$\widehat{\boldsymbol{\varepsilon}} = \mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\boldsymbol{\rho}_{\alpha} + \mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\boldsymbol{\rho}_{\beta} + \boldsymbol{v}, \tag{2.21}$$

where $\mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \left[\frac{\partial\varepsilon}{\partial\alpha_1}\left(\widehat{\theta}_0\right), \ldots, \frac{\partial\varepsilon}{\partial\alpha_p}\left(\widehat{\theta}_0\right)\right]$ and $\mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \left[\frac{\partial\varepsilon}{\partial\beta_1}\left(\widehat{\theta}_0\right), \ldots, \frac{\partial\varepsilon}{\partial\beta_p}\left(\widehat{\theta}_0\right)\right]$. Both regression matrices $\mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ and $\mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ can be easily computed using the expressions for $\frac{\partial\varepsilon}{\partial\theta_i}$ derived in items (ii) and (iii) of Lemma A.5 (see Appendix A). Testing the null hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = 0$ is asymptotically equivalent to test whether $\boldsymbol{\rho}_{\beta} = 0$ in the test regression (2.21). Therefore, the test statistic can be computed as the standard Wald test from the Gauss-Newton regressions (2.21). In what follows we will refer to this variant of the LM test as regression based and denote it by $\mathrm{LM}_T^{(Reg)}$.

A careful inspection shows that this form of the statistic for the H_A alternative resembles closely the test proposed by Brännäs et al. (1998). Therefore, the arguments and the results obtained in this paper can be used to justify the derivation of the statistics in Brännäs et al. (1998) and establish its asymptotics.

2.4 Asymptotics

The difference between the LM-type test statistics discussed above lies in the estimation of $\mathbf{V}_{\boldsymbol{\beta}}$. Since all considered approaches are known to provide a consistent estimator for the covariance matrix of the score vector under the null, the $\mathrm{LM}_T^{(IM)}$, $\mathrm{LM}_T^{(OPG)}$ and $\mathrm{LM}_T^{(Reg)}$ are asymptotically equivalent and behave as χ^2 distribution with p degrees of freedom. This result is summarized in the following theorem.

Theorem 8 *For processes (2.3) and (2.4), under assumption 3 and the null hypothesis*

$$\mathrm{LM}_T \to \chi_p^2,$$

as $T \to \infty$.

Remark 12 *Notice that, if the stationarity assumption is violated under the null hypothesis the underlying asymptotics will differ from the ones obtained in Theorem 8. For instance, consider the underlying process y_t to be near integrated under the null, i.e.,*

$$y_t = \left(1 + \frac{c}{T}\right) y_{t-1} + \varepsilon_t. \tag{2.22}$$

Then, the LM test to test for AsAR(1) behaves asymptotically as

$$\mathrm{LM}_T \xrightarrow{p} \frac{\left(\int_0^1 J_c(r)\mathrm{d}W(r)\right)^2}{\int_0^1 J_c^2(r)\mathrm{d}r}, \tag{2.23}$$

where $J_c(r)$ is an Orhstein-Uhlenbeck process and $W(r)$ is a Brownian motion.[2] *However, at this point it is not clear how to discriminate nonstationarity from asymmetry. Therefore, pretesting for the unit root before applying the LM test for asymmetries might provide invalid results. We do not pursue this problem in this paper. However, this presents an interesting line of research for further investigation.*

2.5 MC simulations

After deriving LM-type tests for testing asymmetries induced by shocks in time series and their asymptotics, we now turn to study the small sample properties of the proposed test and its variants. The main aim of this section is to evaluate the performance of the tests in terms of their size and power in different empirically relevant setups.

2.5.1 Normally distributed errors

As a benchmark specification we consider two types of time series processes given as

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}^{-} + \beta\varepsilon_{t-1}^{+}, \quad (2.24)$$

$$y_t = \varepsilon_t + \alpha y_{t-1}^{-} + \beta y_{t-1}^{+}, \quad (2.25)$$

$$\text{with} \quad \varepsilon_t \sim N(0,1), \quad (2.26)$$

where (2.24) corresponds to the AsMA(1) and (2.25) to the AsAR(1) model. We examine different combinations of α and β selected from the set {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} and three sample sizes $T = 50$, $T = 100$ and $T = 200$. All Monte Carlo simulations are based on $N = 2000$ replications and are executed for tests of a nominal size of 10%, 5% and 1%. Only the results for the size of 5% are reported since no qualitative differences were observed.

The left panel of Table 2.1 (see Appendix B) shows rejection frequencies under the null hypothesis when the underlying processes are MA(1) and AR(1) (i.e., $\alpha = \beta$ in (2.24)) with a lag coefficient $\alpha \in \{0, ..., 0.9\}$. In the case of the MA(1) and $T = 50$ we observe moderate deviation from the nominal size for the $\text{LM}_T^{(OPG)}$ and the $\text{LM}_T^{(Reg)}$ test when the parameter α is close to unity. For the AR(1) process and $T = 50$ the obtained results show that the $\text{LM}_T^{(OPG)}$ and the $\text{LM}_T^{(Reg)}$ perform equally well, while $\text{LM}_T^{(IM)}$ slightly underrejects. The size properties of all tests approach the nominal level fast as T increases.

Figure 2.1 (see Appendix C) illustrates the corresponding rejection frequencies under the alternative. In particular, parameter β in (2.24) and (2.25) is fixed to zero, while α takes values from the set $\{0, ..., 0.9\}$ as described above. At this point we report that the

[2]The proof of this fact is almost identical to the proof presented in Phillips (1997).

setup with fixing α and allowing β to change will produce symmetric results and is omitted from the discussion. The left panel shows the results for the AsMA alternative and the right one for the AsAR alternative. All three tests performs equally well except for the case of $T = 50$ where the $\mathrm{LM}_T^{(IM)}$ test has marginally bigger power than its variants for the AsMA alternative and suffers slightly from a power loss relatively to the other tests in the case of the AsAR alternative.

2.5.2 Errors with skewed distribution

We now investigate the behavior of the LM tests when the errors are no longer normally distributed. Since we construct test statistics that are built to distinguish the contribution of positive and negative errors, it is of special interest to study if the obtained tests are robust to a skewed distribution of the underlying errors. For this reason the errors in (2.24) and (2.25) are generated from a beta distribution, i.e.,

$$\varepsilon_t \sim \mathcal{B}(\mu, \sigma, \xi, \kappa), \tag{2.27}$$

where the parameters $(\mu, \sigma, \xi, \kappa)$ are fixed to the values such that assumption 3 is satisfied. In particular, $\mu = 0$ and refers to the mean of the distribution, $\sigma = 1$ and refers to the standard deviation, $\xi = 0.8$ and $\kappa = 3$ refer to the skewness and to the kurtosis respectively. All other specifications of the MC design remain the same.

The middle panel of Table 2.1 (see Appendix B) shows the rejection frequencies under the null hypothesis for setups (2.24) and (2.25) with (2.27). The reported results have only marginal changes to the one obtained for the benchmark case where $\varepsilon_t \sim N(0, 1)$. This indicates that all three test statistics are robust in terms of their size property to setups where innovations are drawn from a non-normal skewed distribution.

Turning to the power analysis, Figure 2.2 (see Appendix C) illustrates the obtained rejection frequencies under the alternative. As a deviation point from the benchmark design each panel reports two setups, one with $\alpha = 0$ and $\beta \in \{0, ..., 0.9\}$ and one with $\beta = 0$ and $\alpha \in \{0, ..., 0.9\}$. It is clear from the Figure 2.2 that while the power properties of the $\mathrm{LM}_T^{(OPG)}$ and $\mathrm{LM}_T^{(Reg)}$ do not change qualitatively compared to the scenario with normal errors, a practical weakness of $\mathrm{LM}_T^{(IM)}$ is revealed. In particular, the power properties of the test are asymmetric with respect to the fixed α and fixed β setups. The problem vanishes fast as T increases. However, the $\mathrm{LM}_T^{(IM)}$ test seems to be less robust in small samples against skewed error distributions.

2.5.3 Conditional heteroskedasticity

To investigate the effect of conditional heteroskedasticity on the performance of the proposed LM type tests we use instead of (2.26) a GARCH(1,1) specification to generate errors of the processes (2.24) and (2.25), i.e.,

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t}\nu_t, & (2.28)\\ h_t &= \kappa + \delta h_{t-1} + \theta\varepsilon_{t-1}^2, & (2.29)\\ \nu_t &\sim N(0,1) & (2.30) \end{aligned}$$

with $\kappa = 0.01$, $\delta = 0.08$ and $\theta = 0.9$. In this simulation the chosen parameters are motivated by empirical results estimating a GARCH(1,1) on daily stock market returns (see Pelagatti and Lisis, 2009).

The right panel of Table 2.1 (see Appendix B) presents type I errors for this setup. As expected the OPG variant of the LM test shows the most conservative and close to the nominal level size performance, while the other two variants are oversized for all sample sizes.

Figure 2.3 reports the rejection frequencies under the alternative of the tests when errors are conditional heteroscedastic. In comparison to our benchmark specification we observe only marginal changes in power for all cases.

2.5.4 Model Selection

In practice knowledge of the lag length is required prior to the implementation of the LM test. Hence, in this section we study the estimation of the true order, which shall be called p_0, and its impact on the test statistics. Our primary aim is to establish the small sample behavior of $\widehat{p}$ estimated using a standard model selection approach within a linear time series model when the true underlying model is in fact a AsMA(p_0) or AsAR(p_0). Specifically, the lag length is estimated from a linear $MA(p)$ or $AR(p)$ model with $1 \leq p \leq P_{\max}$ where $P_{\max}$ is known a priori. The model selection criteria such as the AIC, BIC or HQ are used for the estimation of the p_0. The second aim of this section is to investigate the influence of the estimated lag length on the size-power properties of the LM test.

In a first step we investigate the performance of the three mentioned model selection criteria in two model setups each with two different parameterizations. In particular, we use the following specifications

$$\begin{aligned} y_t &= \varepsilon_t + \alpha_1\varepsilon_{t-1}^- + \alpha_2\varepsilon_{t-2}^- + \beta_1\varepsilon_{t-1}^+ + \beta_2\varepsilon_{t-2}^+ & (2.31)\\ y_t &= \varepsilon_t + \alpha_1 y_{t-1}^- + \alpha_2 y_{t-2}^- + \beta_1 y_{t-1}^+ + \beta_2 y_{t-2}^+ & (2.32) \end{aligned}$$

where the first corresponds to an AsMA(2) and the latter to an AsAR(2). We use the parameter combinations $\alpha_1 = 0.5$, $\alpha_2 = 0.4$, $\beta_1 = 0.3$, $\beta_2 = 0.2$ and $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\beta_1 = 0.1$, $\beta_2 = 0.1$. Further, we calculate the selected lag length frequencies up to a lag of six periods (i.e., $P_{\max} = 6$) for sample sizes $T = 100$, $T = 200$ and $T = 400$ using $N = 2000$ replications.

The results are given in Table 2.2 and are qualitatively similar for both model specifications. For $T = 100$ the BIC has a clear tendency to underselect the lag length for both parametrizations. However, this improves rapidly with an increase of T and furthermore BIC shows the highest percentage of correct lag selection (above 94%). Similar observations are made for the HQ criterion. As for the linear time series models, the AIC has a tendency to overselect for all sample sizes. For the first parameter specification, when $T = 400$, the AIC overselects in 25.6% cases for AsMA model and 28.95% for the AsAR model. When we compare to overselection rates of BIC and HQ it is 2.15% and 9.1% for theAsMA model, respectively, and 2.85% and 11.1% for the AsAR model. The same message holds for the second parameter specification.

Which criterion is preferable is nevertheless context specific and depends on the taste of the researcher. For our purposes it is important to note that standard criteria can be used to determine the lag length in finite samples, although one should be aware of a potential overselection of the AIC criterion.

In the second step, we turn to the influence of a preliminary model selection stage on the power of the LM test. For this reason we use the BIC in our baseline setup with normally distributed errors and compare outcomes with the benchmark model in Section 2.5.1. BIC values are calculated up to a lag of six periods. The results are shown in Figure 2.4. In this setup we only observe minor power deviations compared to the case with a known lag structure of the process.

2.6 Example: Growth of the U.S. unemployment rate

In this section we explore by using the AsAR model the presence of asymmetries in the growth of the U.S. unemployment. We use monthly, seasonally adjusted unemployment data of the U.S. population at the age of 16 and above, available from the Bureau of Labor Statistics. The sample runs from January 1958 to December 2014 and is plotted in Figure 2.5.

Based on BIC and HQ, with a maximum number of lags $P_{max} = 12$, the AR(4) model is selected as the appropriate test specification. We use the $\mathrm{LM}_T^{(OPG)}$ test which is robust to heteroscedasticity, since there is evidence of residual heteroscedasticity in the model under the null. The null hypothesis of no asymmetric effects of innovations on the growth of unemployment is rejected at the 1% significance level with $\mathrm{LM}_T^{(OPG)} = 19.35$. Furthermore, we can analyze the asymmetric effects lagwise. This can be simply done by using the same testing routine for the restricted asymmetric model. For instance, to test for asymmetry of innovations in the k-th lag, the LM test can be constructed for the model $y_t = \varepsilon_t - \alpha_1 y_{t-1} - ... - \alpha_p y_{t-p} - \beta_k y^+_{t-k}$ with $k = 1, ..., p$ in the same way as for the unrestricted model (2.2) in Section 2.3. The obtained results in our case are as follows

$$\text{LM}^{(OPG)}_{1,T} = 7.79,\ \text{LM}^{(OPG)}_{2,T} = 1.58,\ \text{LM}^{(OPG)}_{3,T} = 1.26,\ \text{LM}^{(OPG)}_{4,T} = 0.69,$$

where $\text{LM}^{(OPG)}_{k,T}$ is the OPG version of the test for asymmetry in the k-th lag. Only the first test rejects the null hypothesis (at the 1% level). This indicates that shock induced asymmetries are only present for the first lag of the series. Our findings suggest that the appropriate model specification for the growth of the unemployment rate takes the following form

$$y_t = \varepsilon_t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \alpha_4 y_{t-4} + \beta_1 y^+_{t-1}. \tag{2.33}$$

A thorough theoretical discussion of estimating asymmetric time series models goes beyond the scope of this paper. However, to illustrate how asymmetries can influence the dynamics of the process we complete this example by estimating model (2.33). To estimate asymmetric time series models, Wecker (1981) suggests the maximum likelihood approach. As argued in Section 2.3 the likelihood function is not differentiable (in a classical sense) and standard search techniques for the maximum can produce misleading or biased estimates. For this reason we suggest a simple iterative procedure:[3]

Step 1. First, the model (in our case AR(4)) is estimated under the null to obtain an estimation of the innovations $\left\{\widehat{\varepsilon}^{(1)}_t\right\}$. For this, standard OLS/GLS can be used. Estimates $\left\{\widehat{\varepsilon}^{(1)}_t\right\}$ are used to construct the asymmetric components $\widehat{y}^+_{t-i} = y_{t-i}1\left(\widehat{\varepsilon}^{(1)}_{t-i} \geq 0\right)$ for $i = 1, ..., p$;

Step 2. The AsAR model can be estimated with OLS/GLS approach by replacing the true asymmetric components y^+_{t-i} with estimated quantities $\widehat{y}^+_{t-i}$ for $i = 1, ..., p$. This step in turn will produce the estimated residuals from the asymmetric model $\left\{\widehat{\varepsilon}^{(2)}_t\right\}$;

Step 3. The innovations $\left\{\widehat{\varepsilon}^{(2)}_t\right\}$ from Step 2 are used to recalculate the asymmetric components, i.e. $\widehat{y}^+_{t-i} = y_{t-i}1\left(\widehat{\varepsilon}^{(1)}_{t-i} \geq 0\right)$ for $i = 1, ..., p$. Then Step 2 is repeated and new estimated residuals are produced $\left\{\widehat{\varepsilon}^{(3)}_t\right\}$.

Step 4. Step 2 and 3 can be repeated N times until the fit of the model does not change between iterations, i.e.,

$$\left\|\widehat{\sigma}^2_N - \widehat{\sigma}^2_{N-1}\right\| < \epsilon,$$

where $\widehat{\sigma}^2_N = \left(\widehat{\varepsilon}^{(N)}\right)' \widehat{\varepsilon}^{(N)}/T$ is the estimate of the fitted variance in iteration N and ϵ is the precision constant chosen by the researcher.

[3] The consistency of the suggested estimation procedure remains an open topic and the obtained estimates serve only for illustrative purposes.

In our example we choose $\epsilon = 10^{-4}$. Convergence of the estimation procedure is achieved after two iterations. Table 2.3 reports the parameter estimates and the respective t-statistics. In addition, we report that the residuals for the given AsAR process are not serially correlated, if we look at the Ljung-Box test for serial correlation up to 6 lags. The most noticeable result is that the first lag of y_t effects only through the asymmetric component y_{t-1}^{+} but not through the linear one y_{t-1}. Since it is difficult to assess the dynamics of the autoregressive process only through point estimates the corresponding impulse-response functions are constructed. To isolate the effects of positive and negative innovations we consider two shocks of one standard deviation, that is, $\varepsilon_0 = \widehat{\sigma}$ and $\varepsilon_0 = -\widehat{\sigma}$. In Figure 2.6, we plot the obtained impulse-response functions. The blue line with diamonds represents the impulse of the positive shock and the red line with triangles depicts the negative shock mirrored with respect to the time-axis for a better comparison. For completeness we also add the impulse of the standard AR(4) model (line with squares). This figure presents the difference between "positiv" and "negative" impulses that pertain in the first year after the shock. It becomes apparent that the positive shock affects immediately while the effect of the negative one is less pronounced and delayed.

This finding complements the existing literature on nonlinear behavior of the unemployment rate (see e.g., Hansen (1997), Yilanci (2008) and Caporale and Gil-Alana (2007) among others) and creates potentially a new discussion on what type of nonlinearity is present in the U.S. unemployment rate.

2.7 Conclusion

In this paper we used the theory of generalized functions to derive the Lagrange multiplier test when the likelihood function is not differentiable. In particular, we derived different variants of the LM test to detect asymmetries induced by positive and negative past shocks on time series. Further, we investigated the asymptotic properties of the test. In a simulation study, we examined the small sample properties of the LM_T test under different model specifications. It is also shown by means of Monte Carlo simulations that standard model selection criteria can be used for the implementation of the test. In an empirical example to the growth of the U.S. unemployment rate, we demonstrate the relevance of our testing procedure.

A Appendix: Proofs

First, some auxiliary results are collected in the following Lemma to simplify the exposition of the subsequent proofs.

Lemma A.5 ***(i)*** *Sifting property of delta functions*

$$\int_{\Omega} \delta(x) f(x)\, dx = f(0) \text{ and } \int_{\Omega} \dot{\delta}(x) f(x)\, dx = -\dot{f}(0),$$

where $\dot{\Delta}(x)$ defines the derivative of the delta function and $\dot{f}(x)$ is the derivative of $f(x)$;
(ii) *For process (2.3) it holds*

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i} = \begin{cases} -\widetilde{\mathbf{M}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\mathbf{B}^i\boldsymbol{\varepsilon}, & \text{if } \theta_i = \alpha_i \\ -\widetilde{\mathbf{M}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}\boldsymbol{\varepsilon}, & \text{if } \theta_i = \beta_i \end{cases}$$

where $\widetilde{\mathbf{M}}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}}\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{D}} = \mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\boldsymbol{\varepsilon}}$ for $i = 1, ..., p$;
(iii) *For process (2.4) it holds*

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i} = \begin{cases} \mathbf{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{B}^i\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon}, & \text{if } \theta_i = \alpha_i \\ \mathbf{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon}, & \text{if } \theta_i = \beta_i \end{cases}$$

where $\mathbf{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathbf{I} - \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\mathbf{y}}$ and $\mathbf{y} = \mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon}$ for $i = 1, ..., p$.

Proof. Sifting property **(i)** summarizes some of the features of delta functions (see, e.g., Gelfand and Shilov, 1964).

Property **(ii)** comes directly from differentiation of (2.3) and standard results for matrix derivatives (see, e.g., Lütkepohl, 1996), i.e.,

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}}{\partial \beta_i} &= -\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\left[\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\beta_i}\right]\boldsymbol{\varepsilon} \\ &= -\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}\boldsymbol{\varepsilon} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\boldsymbol{\varepsilon}}\partial\boldsymbol{\varepsilon}/\partial\beta_i. \end{aligned}$$

Solving the last equality for $\frac{\partial \boldsymbol{\varepsilon}}{\partial \beta_i}$ yields the required result. The same calculations are required for $\frac{\partial \boldsymbol{\varepsilon}}{\partial \alpha_i}$.

Finally, the last item **(iii)** follows from similar arguments, i.e.,

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}}{\partial \beta_i} &= \left[\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\beta_i}\right]\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon} \\ &= \mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\mathbf{y}}\partial\boldsymbol{\varepsilon}/\partial\beta_i \end{aligned}$$

where $\mathbf{y} = \mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\boldsymbol{\varepsilon}$. Again solving the last equation for $\frac{\partial \boldsymbol{\varepsilon}}{\partial \beta_i}$ yields item **(iii)**. The proof for $\frac{\partial \boldsymbol{\varepsilon}}{\partial \alpha_i}$ is identical. ■

Proof of Lemma A.3

Recall that invertibility of process y_t ensures the existence of the inverse of $\mathbf{M}_{\boldsymbol{\alpha}}$ under the null, i.e.,

$$\mathbf{M}_{\boldsymbol{\alpha}}^{-1} = \left(\sum_{k=0}^{p} \alpha_k \mathbf{B}^k\right)^{-1} = \sum_{l=0}^{\infty} \psi_l \mathbf{B}^l = \sum_{i=0}^{T-1} \psi_l \mathbf{B}^l,$$

where $\psi_0 = 1$ and $\sum_{k=0}^{\infty} |\psi_k| < \infty$.

(i) We have that

$$\begin{aligned} s_{\alpha,i} &= \frac{1}{\sigma^2}\boldsymbol{\varepsilon}' \left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\right)' \boldsymbol{\varepsilon} = \frac{1}{\sigma^2} \sum_{t=1+i}^{T} \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s \psi_{t-s-i}, \\ s_{\beta,j} &= \frac{1}{\sigma^2} \left(\boldsymbol{\varepsilon}^+\right)' \left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j\right)' \boldsymbol{\varepsilon} = \frac{1}{\sigma^2} \sum_{t=1+j}^{T} \sum_{s=1}^{t-j} \varepsilon_t \varepsilon_s^+ \psi_{t-s-j}. \end{aligned}$$

Hence the expectation of $s_{\alpha,i}s_{\beta,j}$ can be rewritten as

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \frac{1}{\sigma^4} \sum_{t=1+i}^{T} \sum_{s=1}^{t-i} \sum_{l=1+j}^{T} \sum_{k=1}^{l-j} \psi_{t-s-i}\psi_{l-k-j}\mathbb{E}\left[\varepsilon_t\varepsilon_s\varepsilon_l\varepsilon_k^+\right].$$

Note that the above expectations are nonzero only if the four indices of ε_t are pairwise equal. More precisely, the only possible case is when $t = l$ and $s = k$. We thus obtain the following expression

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \frac{\phi_2}{\sigma^2} \sum_{t=1+\max(i,j)}^{T} \sum_{s=1}^{t-\max(i,j)} \psi_{t-s-i}\psi_{t-s-j} = \frac{\phi_2}{\sigma^2} \operatorname{tr}\left[\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\right)\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j\right)'\right].$$

(ii) Proof of fact (2.13) goes along the same line. Rewrite the expectation of $s_{\beta,i}s_{\beta,j}$ as

$$\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] = \frac{1}{\sigma^4} \sum_{t=1+i}^{T} \sum_{s\le t-1} \sum_{l=1+j}^{T} \sum_{k\le l-1} \psi_{t-s-i}\psi_{t-s-j}\mathbb{E}\left[\varepsilon_t\varepsilon_s^+\varepsilon_l\varepsilon_k^+\right].$$

In this situations the expectations are nonzero only if the indices of ε satisfy conditions $t = l \neq s = k$ and $s \neq k \neq t = l$. Which in turn leads to (2.13) since

$$\begin{aligned} \mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \frac{\phi_2}{\sigma^2} \sum_{t=1+\max(i,j)}^{T} \sum_{s=1}^{t-\max(i,j)} \psi_{t-s-i}\psi_{t-s-j} \\ &\quad + \frac{\phi_1^2}{\sigma^2} \sum_{t=1+\max(i,j)}^{T} \sum\sum_{1\le s\neq k\le t-\max(i,j)} \psi_{t-s-i}\psi_{t-k-j}, \end{aligned}$$

where

$$\sum_{t=1+\max(i,j)}^{T} \sum_{s=1}^{t-\max(i,j)} \psi_{t-s-i}\psi_{t-s-j} = \operatorname{tr}\left[\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\right)\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j\right)'\right],$$

$$\sum_{t=1+\max(i,j)}^{T} \sum_{s=1}^{t-\max(i,j)} \psi_{t-s-i} \sum_{k=1}^{t-\max(i,j)} \psi_{t-k-j} = \boldsymbol{l}'(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{i})(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{j})'\boldsymbol{l},$$

with $\boldsymbol{l}$ being a $T \times 1$ vector of ones.

Proof of Lemma A.4

(i) Consider the following decomposition of the elements of $s_{\beta,i}$ into two terms

$$s_{\beta,i} = -\frac{1}{\sigma^2} \sum_{t=1+i}^{T} \sum_{s=1}^{t-i-1} \varepsilon_t \varepsilon_s 1\left(\varepsilon_{t-i} \geq 0\right) \psi_{t-s-i} - \frac{1}{\sigma^2} \sum_{t=1+i}^{T} \varepsilon_t \varepsilon_{t-i}^{+}, \tag{A.1}$$

for $i = 1, ..., p$. Hence the expectation of $s_{\beta,i}s_{\beta,j}$ can be expressed as

$$\begin{aligned}
\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \frac{1}{\sigma^4} \sum_{t=1+i}^{T} \sum_{s=1}^{t-i-1} \sum_{l=1+j}^{T} \sum_{k=1}^{l-j-1} \mathbb{E}\left[\varepsilon_t \varepsilon_s \varepsilon_l \varepsilon_k 1\left(\varepsilon_{t-i} \geq 0\right) 1\left(\varepsilon_{l-j} \geq 0\right)\right] \psi_{t-s-i} \psi_{l-k-j} \\
&+ \frac{1}{\sigma^4} \sum_{t=1+i}^{T} \sum_{s=1}^{t-i-1} \sum_{l=1+j}^{T} \mathbb{E}\left[\varepsilon_t \varepsilon_s \varepsilon_l \varepsilon_{l-j}^{+} 1\left(\varepsilon_{t-i} \geq 0\right)\right] \psi_{t-s-i} \qquad \text{(A.2)} \\
&+ \frac{1}{\sigma^4} \sum_{t=1+j}^{T} \sum_{s=1}^{t-j-1} \sum_{l=1+i}^{T} \mathbb{E}\left[\varepsilon_t \varepsilon_s \varepsilon_l \varepsilon_{l-i}^{+} 1\left(\varepsilon_{t-j} \geq 0\right)\right] \psi_{t-s-j} \\
&+ \frac{1}{\sigma^4} \sum_{t=1+i}^{T} \sum_{l=1+j}^{T} \mathbb{E}\left[\varepsilon_t \varepsilon_{t-i}^{+} \varepsilon_l \varepsilon_{l-j}^{+}\right].
\end{aligned}$$

Consider first $i = j$. Then the second and the third term in (A.2) are zero. The only relevant cases when expectation is non zero for the first term are when $t = l$; $s = k$ and for the fourth term when $t = l$. These facts together with the fact that

$$\mathcal{F}_0 := \mathbb{E}\left[1\left(\varepsilon_{t-i} \geq 0\right)\right] = \int_0^{\infty} dF_{\varepsilon}\left(x\right) = 1 - F_{\varepsilon}\left(0\right),$$

implies that

$$\begin{aligned}
\mathbb{E}\left[s_{\beta,i}s_{\beta,i}\right] &= \mathcal{F}_0 \sum_{t=1+i}^{T} \sum_{s=1}^{t-i-1} \psi_{t-s-i}^2 + \frac{\phi_2}{\sigma^2}(T-i) \\
&= \mathcal{F}_0 \sum_{t=1+i}^{T} \sum_{s=1}^{t-i} \psi_{t-s-i}^2 + \frac{\phi_2 - \sigma^2 \mathcal{F}_0}{\sigma^2}(T-i) \qquad \text{(A.3)} \\
&= \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{i})(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{i})'\right] + \frac{\phi_2 - \sigma^2 \mathcal{F}_0}{\sigma^2}(T-i). \qquad \text{(A.4)}
\end{aligned}$$

When $i > j$, the second term in (A.2) as well is zero and the only relevant case for the first term is when $t = l$; $s = k$ and for the fourth term when $t = l$. However, the third term in (A.2) when $t = l$ and $s = t - i$ has non zero expectation and can be expressed as

$\sigma^2\phi_2(1-\mathcal{F}_0)\sum_{t=1+i}^{T}\psi_{i-j}$. This results in the following outcome

$$\begin{aligned}\mathbb{E}\left[s_{\beta,i}, s_{\beta,j}\right] &= \mathcal{F}_0^2 \operatorname{tr}\left[(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i)(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j)'\right] \\ &\quad + \frac{(T-i)}{\sigma^2}\left((\phi_2-\sigma^2\mathcal{F}_0)\mathcal{F}_0\psi_{i-j}+\phi_1^2\right). \end{aligned} \tag{A.5}$$

Finally, for $i<j$ the results are identical to those obtained for $i>j$ due to the symmetry of the variance covariance matrix.

(ii) The same techniques are used to find the covariance between $s_{\alpha,i}$ and $s_{\beta,j}$. For the case when $j<i$ we have that

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \mathcal{F}_0\sum_{t=1+i}^{T}\sum_{s=1}^{t-i}\psi_{t-s-i}^2 \tag{A.6}$$

$$= \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{B}^i\mathbf{M}_{\boldsymbol{\alpha}}^{-1})(\mathbf{B}^j\mathbf{M}_{\boldsymbol{\alpha}}^{-1})'\right], \tag{A.7}$$

and for $j\geq i$ additional terms enter the expression, i.e.,

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{B}^i\mathbf{M}_{\boldsymbol{\alpha}}^{-1})(\mathbf{B}^j\mathbf{M}_{\boldsymbol{\alpha}}^{-1})'\right] + \frac{(\phi_2-\sigma^2\mathcal{F}_0)}{\sigma^2}\psi_{i-j}(T-j), \tag{A.8}$$

which completes the proof of the Lemma.

Proof of Theorem 8

To lighten the notational load in what follows we omit the argument $\boldsymbol{\theta}_0$. Then rewrite the score vector as $\mathbf{s}_{\boldsymbol{\beta}} = \frac{1}{\sigma^2}\sum_t \mathbf{Z}_{t,T}$, where $\mathbf{Z}_{t,T} = \left(Z_{t,T}^{(1)},...,Z_{t,T}^{(p)}\right)'$ with $Z_{t,T}^{(i)}$ defined as

$$Z_{t,T}^{(i)} = \sum_{s=1}^{t-i}\varepsilon_t\varepsilon_s^{+}\psi_{t-s-i} = \varepsilon_t\xi_{t-i},$$

and ξ_{t-i} denotes $\sum\limits_{s=1}^{t-i}\varepsilon_s^{+}\psi_{t-s-i}$. To investigate the limiting behavior the Cramer-Wold device will be applied which tells that it is sufficient to study the limiting distribution of a sequence of scalars $\eta_{t,T} = \boldsymbol{\lambda}'\mathbf{Z}_{t,T}$, where $\boldsymbol{\lambda}$ is a $p\times 1$ vector such that $\|\boldsymbol{\lambda}\|=1$ and $\|\cdot\|$ defines an L_2 vector norm.

The central limit theorem for martingale difference sequences (hereafter, mds) applies to the $\{\eta_{t,T}\}$ if the following holds: [4]

(i) $\{\eta_{t,T},\mathcal{F}_{t,T}\}$ is mds, where $\mathcal{F}_{t,T}$ is defined as an associated σ-algebra to the sequence $\eta_{t,T}$ such that $\eta_{t,T}$ is measurable with respect to $\mathcal{F}_{t,T}$;

(ii) $\mathbb{E}\left|\eta_{t,T}\right|^{2+r} < C < \infty$ for some $r>0$ and all t;

(iii) define $\overline{\sigma}_{\eta,T}^2 \equiv \frac{1}{T}\mathbb{E}\left[\left(\sum\limits_t \eta_{t,T}\right)^2\right]$, where $\overline{\sigma}_{\eta,T}^2 > r' > 0$ and

$$\frac{1}{T}\sum_t \eta_{t,T}^2 - \overline{\sigma}_{\eta,T}^2 \xrightarrow{p} 0.$$

[4] see, e.g., White (2001), Corollary 5.26

It is straightforward to see that condition (i) is satisfied since $\mathbb{E}\left[\eta_{t,T}|\mathcal{F}_{t-1,T}\right] = \lambda'\mathbb{E}\left[\boldsymbol{Z}_{t,T}|\mathcal{F}_{t-1,T}\right] = 0$ and the assumption on ε_t assures that $\mathbb{E}\left|\eta_{t,T}\right| < \infty$. To verify condition (ii) notice first that by Cauchy- Schwarz and Minkowski's inequalities

$$\mathbb{E}\left|\eta_{t,T}\right|^{2+r} \leq \|\boldsymbol{\lambda}\|^{2+r}\,\mathbb{E}\left\|\mathbf{Z}_{t,T}\right\|^{2+r} \leq \left(\sum_i \left(\mathbb{E}\left|Z_{t,T}^{(i)}\right|^{2+r}\right)^{\frac{1}{2+r}}\right)^{2+r}.$$

Hence, condition (ii) follows from uniform L_{4+r} boundedness of ε_t, uniform L_{4+r} boundedness of ε_t^+ (implied by assumption 3) and the following arguments

$$\begin{aligned}
\mathbb{E}\left|Z_{t,T}^{(i)}\right|^{2+r} &\leq \left(\mathbb{E}\left|\varepsilon_t\right|^{4+r}\mathbb{E}\left|\xi_{t-i}\right|^{4+r}\right)^{\frac{1}{2}} \\
&\leq C\left(\sum_{s=1}^{t-i}\left(\mathbb{E}\left|\varepsilon_s^+\psi_{t-s-i}\right|^{4+r}\right)^{\frac{1}{4+r}}\right)^{2+r} \\
&\leq C_1\left(\sum_{s=1}^{t-1}\left|\psi_{t-s-i}\right|\right)^{2+r} < \infty,
\end{aligned}$$

where the second inequality follows from the Minkowski's inequality and the last one from invertibility and stability of the process.

Regarding condition (iii), first it is clear that $\overline{\sigma}_{\eta,T}^2$ is bounded away from zero, i.e.,

$$\overline{\sigma}_{\eta,T}^2 = \frac{1}{T}\mathbb{E}\left[\left(\sum_t \boldsymbol{\lambda}'\mathbf{Z}_{t,T}\right)^2\right] = \frac{1}{T}\boldsymbol{\lambda}'\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\boldsymbol{\lambda} > 0.$$

Second, to show the convergence of $\frac{1}{T}\sum_t \eta_{t,T}^2 - \overline{\sigma}_{\eta,T}^2$ it is sufficient to show the convergence of

$$\frac{1}{T}\sum_t Z_{t,T}^{(i)}Z_{t,T}^{(j)} - \frac{1}{T}V_{\boldsymbol{\beta}}(i,j) = \frac{1}{T}\sum_t\left(\varepsilon_t^2 - \sigma^2\right)\xi_{t-i}\xi_{t-j} + \frac{1}{T}\sigma^2\sum_t X_{t-1}, \tag{A.9}$$

where $X_{t-1} \equiv \sum_t\left(\xi_{t-i}\xi_{t-j} - \gamma_2\left(\sum_{t=1+\max(i,j)}^T \overline{\psi}_t^{(i)}\overline{\psi}_t^{(j)} - J_{i,j}\right)\right)$. The first term on the r.h.s. of (A.9) satisfies the mds property and $\mathbb{E}\left|\left(\varepsilon_t^2 - \sigma^2\right)\xi_{t-1}^2\right|^{2+r} < \Delta < \infty$. Therefore the law of large numbers for mds gives that $\frac{1}{T}\sum_t\left(\varepsilon_t^2 - \sigma^2\right)\xi_{t-i}\xi_{t-j} \xrightarrow{p} 0$. Moreover, assumption 3 with standard arguments (see, e.g., Hamilton, 1994, Chapter 7, pp.192-193) implies that X_{t-1} is uniformly integrable L^1 mixingale which in turn gives that $\frac{1}{T}\sum_t X_{t-1} \xrightarrow{p} 0$.

Proofs of limiting results of AsAR model are similar to those given for AsMA and hence are omitted.

B Appendix: Tables

Table 2.1: Rejection frequencies (in%) under the null of no asymmetric effects for AsMA and AsAR processes

		$\varepsilon_t \sim N(0,1)$						$\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$						$\varepsilon_t \sim \text{GARCH}(1,1)$					
		MA(1)			AR(1)			MA(1)			AR(1)			MA(1)			AR(1)		
	α	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$	$\text{LM}_T^{(IM)}$	$\text{LM}_T^{(OPG)}$	$\text{LM}_T^{(Reg)}$
$T=50$	**0.0**	7.9	6.3	6.5	5.8	6.0	4.9	5.6	7.4	5.7	4.8	7.1	4.8	9.1	6.6	8.8	7.8	6.3	8.9
	0.1	6.1	7.2	5.6	4.2	6.5	6.1	6.6	7.7	5.9	4.6	7.3	5.9	9.1	6.8	8.9	6.5	5.0	7.5
	0.2	7.1	5.8	4.7	4.4	5.5	5.1	5.5	5.4	5.6	4.1	6.4	5.2	8.9	6.4	9.1	6.9	6.2	9.0
	0.3	6.6	7.7	6.3	4.2	5.3	4.8	6.4	8.0	6.1	3.4	7.3	5.2	8.5	6.0	8.2	6.1	5.3	8.3
	0.4	6.7	8.0	6.9	3.8	5.5	4.3	6.0	7.9	5.4	2.7	6.8	5.8	8.4	5.0	8.8	5.3	5.2	7.5
	0.5	7.3	6.5	5.1	3.0	5.0	4.7	6.8	7.1	4.9	2.5	6.1	4.8	8.4	6.1	8.9	4.5	4.8	7.0
	0.6	7.4	7.8	5.8	2.6	6.0	5.2	7.8	8.5	6.1	2.4	5.8	4.7	8.1	6.0	9.1	3.5	5.0	7.2
	0.7	8.3	8.4	6.4	3.2	5.1	4.4	8.1	9.9	6.7	2.1	7.0	5.1	8.5	7.1	10.9	3.4	5.1	7.4
	0.8	8.7	10.9	7.4	1.9	5.3	4.7	9.4	11.5	8.0	2.3	5.1	3.5	8.6	8.8	12.2	4.1	5.5	7.7
	0.9	7.9	11.9	9.6	3.0	5.7	4.6	8.9	13.9	9.0	2.6	5.6	4.5	7.1	11.8	12.2	5.0	5.8	8.7
$T=100$	**0.0**	5.9	5.2	4.7	5.1	5.8	5.3	4.8	6.3	5.3	4.9	5.8	5.4	8.0	6.1	8.9	7.9	5.2	8.2
	0.1	5.5	5.8	5.4	4.7	6.0	4.9	4.5	6.2	4.9	4.7	6.2	5.3	9.2	5.9	9.0	8.5	6.0	9.0
	0.2	5.0	5.5	4.5	4.1	5.7	4.6	4.9	6.7	5.8	5.2	5.2	4.3	8.7	6.1	8.7	7.3	5.8	8.1
	0.3	5.7	5.3	4.8	4.3	5.6	4.8	4.0	5.8	4.5	4.5	6.6	5.5	8.3	4.9	7.6	6.1	4.8	7.8
	0.4	5.0	5.3	4.7	3.8	5.9	4.9	5.1	5.8	5.5	3.9	6.3	5.4	8.5	4.8	8.7	6.4	5.5	7.8
	0.5	5.4	5.1	4.6	3.5	5.8	4.8	4.4	5.0	4.6	3.5	5.6	4.6	8.3	4.8	7.1	5.0	4.6	6.8
	0.6	4.5	6.4	5.7	3.6	5.3	5.3	4.3	6.1	6.2	2.6	4.7	4.2	7.6	5.4	8.5	5.4	5.7	7.4
	0.7	4.9	5.3	4.8	3.2	5.3	5.1	4.4	6.1	5.7	3.0	5.7	4.7	8.0	4.7	8.9	3.9	4.0	6.0
	0.8	5.2	6.2	6.4	3.3	5.7	4.8	5.2	6.2	5.8	3.4	5.3	4.1	7.2	5.5	8.9	5.4	5.0	7.2
	0.9	5.3	7.4	7.7	3.7	5.2	4.3	5.9	8.5	6.3	3.2	5.2	4.1	6.8	8.1	10.8	6.4	5.6	8.2
$T=200$	**0.0**	4.8	6.1	5.9	5.0	5.5	5.1	4.1	5.6	5.5	4.9	6.1	5.5	8.3	5.1	8.4	9.4	6.2	9.5
	0.1	4.2	5.4	4.8	4.6	5.6	4.6	4.7	6.3	5.8	5.5	5.5	5.3	9.5	6.0	9.5	7.8	5.2	8.2
	0.2	5.0	5.1	5.1	4.8	5.4	5.0	4.0	5.8	5.6	5.3	6.4	5.7	8.2	4.6	7.8	9.3	6.3	9.4
	0.3	4.2	5.4	5.2	5.2	6.5	5.9	3.8	4.6	4.1	4.9	6.0	5.9	9.6	6.0	9.4	7.6	4.9	7.9
	0.4	4.3	6.7	6.5	4.1	4.7	4.4	4.3	5.2	5.1	5.0	5.6	4.6	7.8	4.7	7.7	7.1	5.3	8.0
	0.5	4.7	5.0	5.2	4.9	5.7	5.5	4.4	5.7	5.2	3.8	5.2	4.2	9.1	5.3	8.7	7.0	5.5	8.2
	0.6	4.0	5.1	5.0	4.1	4.8	4.5	3.4	5.3	5.0	4.8	6.1	5.3	8.0	5.1	8.1	6.1	4.8	7.9
	0.7	4.5	5.7	5.9	3.8	5.0	4.6	3.8	6.4	6.8	3.5	5.3	4.6	7.5	4.5	7.7	7.0	5.7	8.3
	0.8	4.4	4.9	4.6	4.1	5.7	5.3	4.8	5.4	5.1	3.7	6.3	5.4	7.3	5.2	8.5	7.2	6.5	9.2
	0.9	4.5	6.3	7.2	3.9	4.4	4.5	5.4	5.6	5.5	3.5	4.4	4.1	7.7	6.5	10.2	6.4	5.3	7.7

Notes: The nominal size is 5%. The errors ε_t are drawn from $N(0,1)$ (left panel), $\mathcal{B}(0,1,0.8,3)$ (middle panel) and GARCH(1,1) (right panel).

Table 2.2: Lag selection frequencies (in%) under different AsMA and AsAR DGPs

		$y_t = \varepsilon_t + 0.5\varepsilon^-_{t-1} + 0.4\varepsilon^-_{t-2} + 0.3\varepsilon^+_{t-1} + 0.2\varepsilon^+_{t-2}$						$y_t = \varepsilon_t + 0.5y^-_{t-1} + 0.4y^-_{t-2} + 0.3y^+_{t-1} + 0.2y^+_{t-2}$					
$T \setminus p$		1	2	3	4	5	6	1	2	3	4	5	6
100	AIC	3.9	67	12.2	6.3	5.6	5	6.05	68.15	12	5.4	4.35	4.05
	BIC	20.05	75.95	3.2	0.5	0.15	0.15	20.75	75.85	2.65	0.5	0.25	0
	HQ	9.85	77.5	7.6	2.75	1.25	1.05	11.7	77.45	7.15	1.75	1.3	0.65
200	AIC	0.25	73.75	10.85	6.7	4.75	3.7	0.15	73.55	12.65	6	4.5	3.15
	BIC	2.1	95.45	2.05	0.4	0	0	2.35	94.95	2.6	0.05	0.05	0
	HQ	0.45	89.95	5.8	2.4	0.85	0.55	0.75	88.5	7.85	1.4	1.25	0.25
400	AIC	0	74.4	11.75	6.9	3.95	3	0	71.05	13.9	5.9	5.3	3.85
	BIC	0	97.85	1.9	0.25	0	0	0.05	97.1	2.6	0.2	0.05	0
	HQ	0	90.9	6.4	2	0.4	0.3	0	88.9	8.35	1.75	0.5	0.5
		$y_t = \varepsilon_t + 0.5\varepsilon^-_{t-1} + 0.3\varepsilon^-_{t-2} + 0.1\varepsilon^+_{t-1} + 0.1\varepsilon^+_{t-2}$						$y_t = \varepsilon_t + 0.5y^-_{t-1} + 0.3y^-_{t-2} + 0.1y^+_{t-1} + 0.1y^+_{t-2}$					
100	AIC	19.45	53.2	10.4	6.95	5.65	4.35	21.1	55.5	10.05	5.8	4.4	3.15
	BIC	49.95	46.7	2.35	0.8	0.15	0.05	50.2	47.85	1.45	0.35	0.15	0
	HQ	32.65	55.8	6.55	2.7	1.65	0.65	34.5	56.75	5.2	1.9	1.3	0.35
200	AIC	4.8	65.1	13.25	7.45	5.15	4.25	5.05	69.35	11.3	5.9	5.05	3.35
	BIC	23.45	71.4	2	0.25	0.2	0	22.05	76.15	1.3	0.45	0.05	0
	HQ	10.65	78	2.4	2.4	0.9	0.65	10.9	81.2	4.7	2.05	0.85	0.3
400	AIC	0.15	67.95	14.05	7.75	5.3	4.8	0.15	67.25	10.7	8.2	7.25	6.45
	BIC	2.65	94.3	2.7	0.35	0	0	2.3	95.45	1.7	0.45	0.1	0
	HQ	0.5	88.1	7.5	2.4	0.95	0.55	0.7	89.15	4.8	2.95	1.7	0.7

Table 2.3: Estimation results for the growth of U.S. unemployment rate

Regressor	Estimate	t-statistic
y_{t-1}	-0.0500	-0.9343
y_{t-2}	0.2124	5.7014*
y_{t-3}	0.1452	3.9078*
y_{t-4}	0.1265	3.3830*
y^+_{t-1}	0.1658	2.3093*

Notes: * denotes significance at the 1% level

C Appendix: Figures

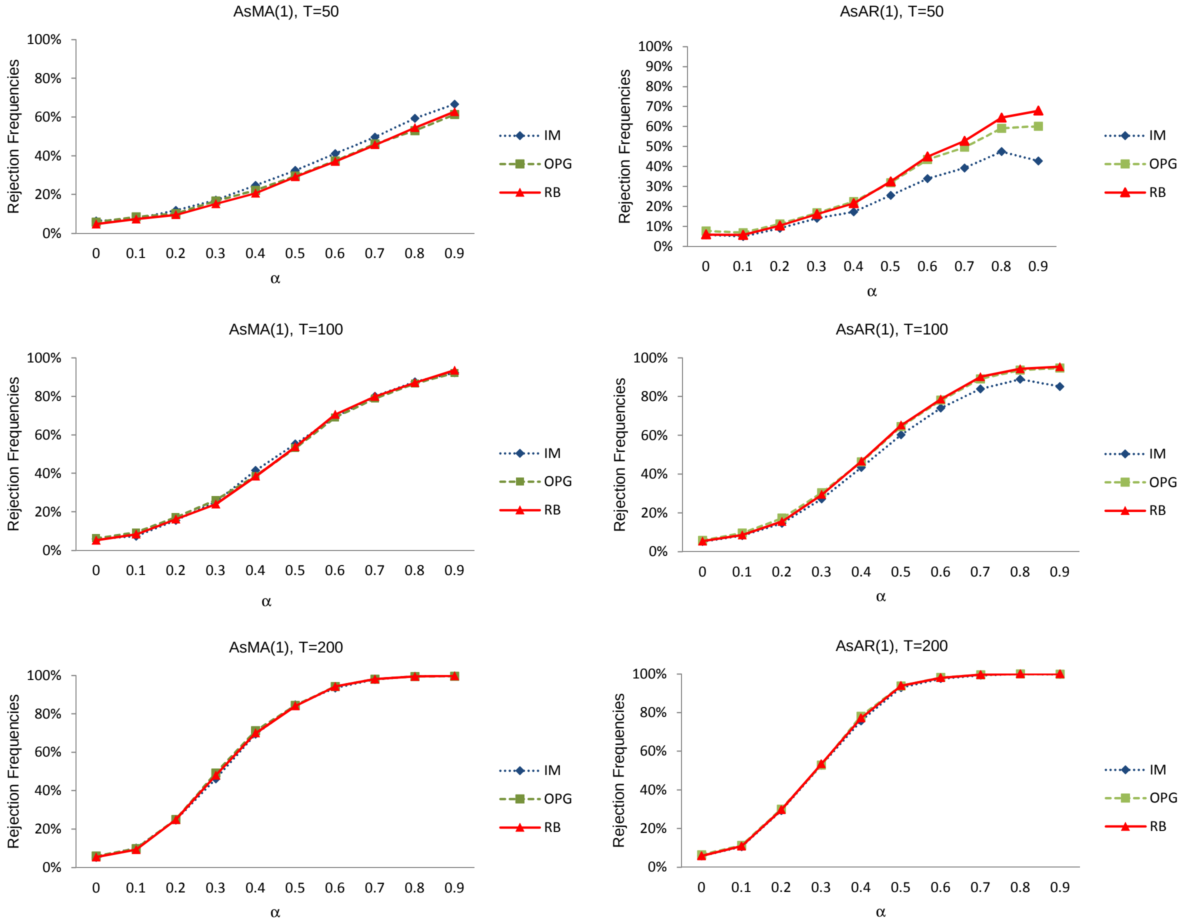


Figure 2.1: Power of the LM_T variants when $\varepsilon_t \sim N(0,1)$. Figures are generated for $\beta = 0$ and α runs from 0 to 0.9 with step 0.1. The left panel presents results for the AsMA DGP and the right panel for the AsAR one. Number of MC replications for each output is 2000

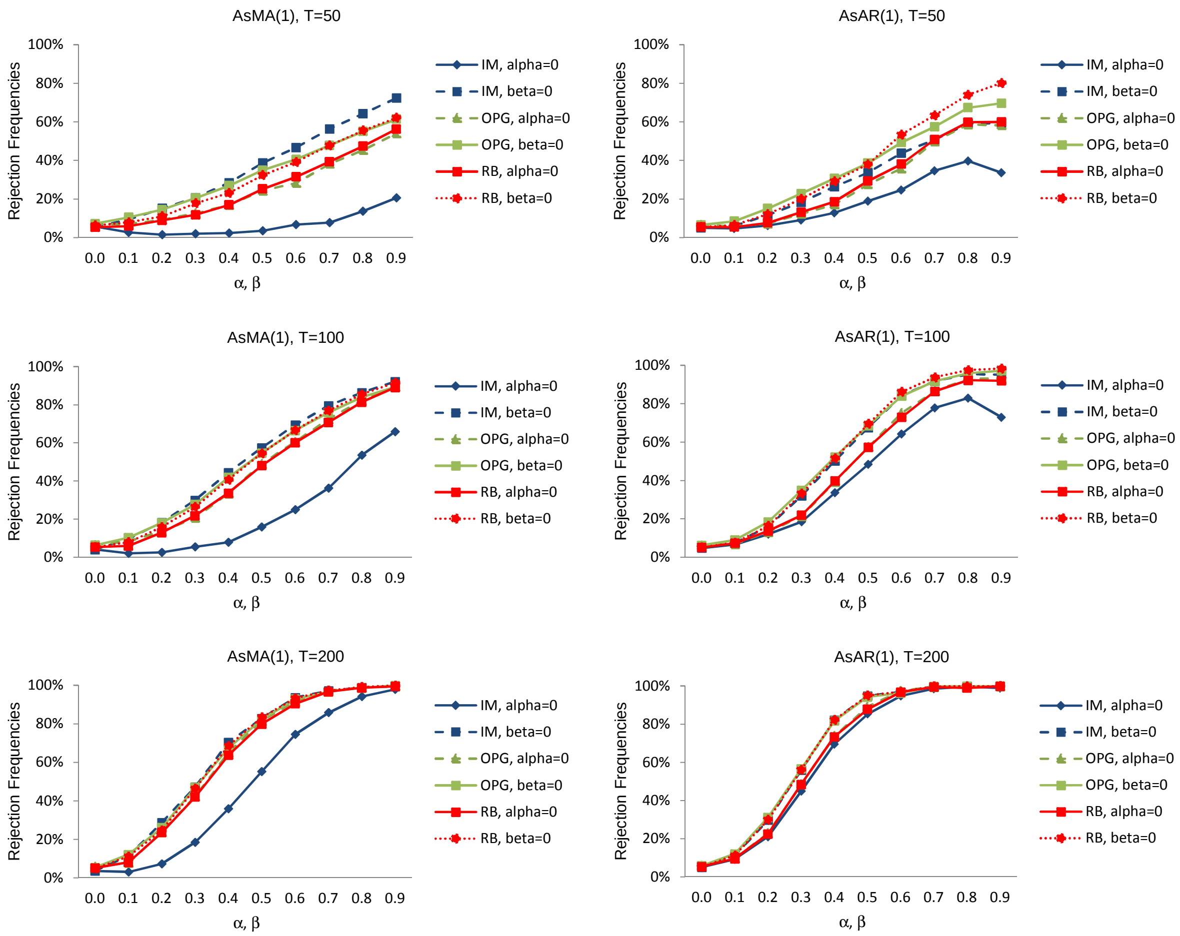


Figure 2.2: Power of the LM_T variants when $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$. All figures are generated for two scenarios: $\beta = 0$, α runs from 0 to 0.9 with step 0.1, and $\alpha = 0$, β runs from 0 to 0.9 with step 0.1. The left panel presents results for the AsMA DGP and the right panel for the AsAR one. Number of MC replications for each output is 2000

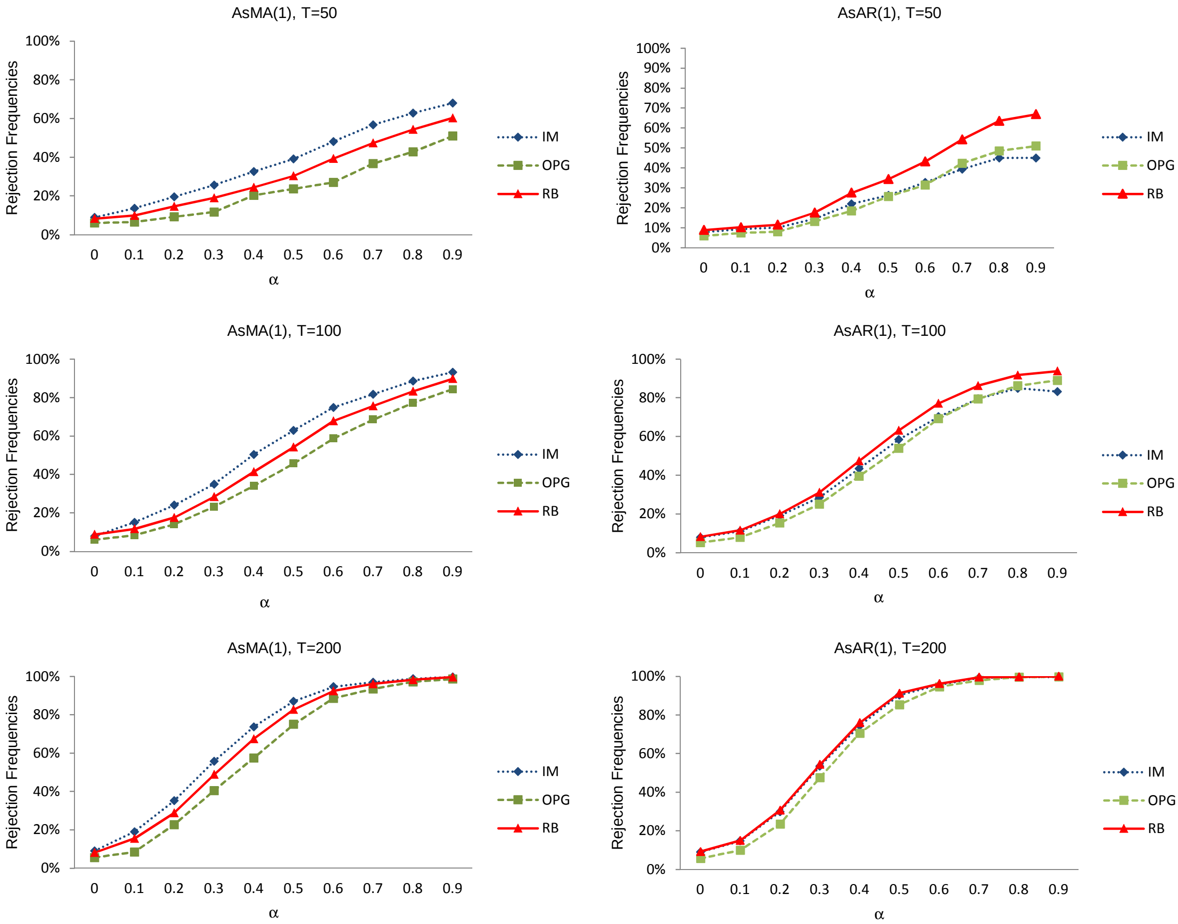


Figure 2.3: Power of the LM_T variants when $\varepsilon_t \sim \mathrm{GARCH}\,(1,1)$. Figures are generated for $\beta = 0$ and α runs from 0 to 0.9 with step 0.1. The left panel presents results for the AsMA DGP and the right panel for the AsAR one. Number of MC replications for each output is 2000

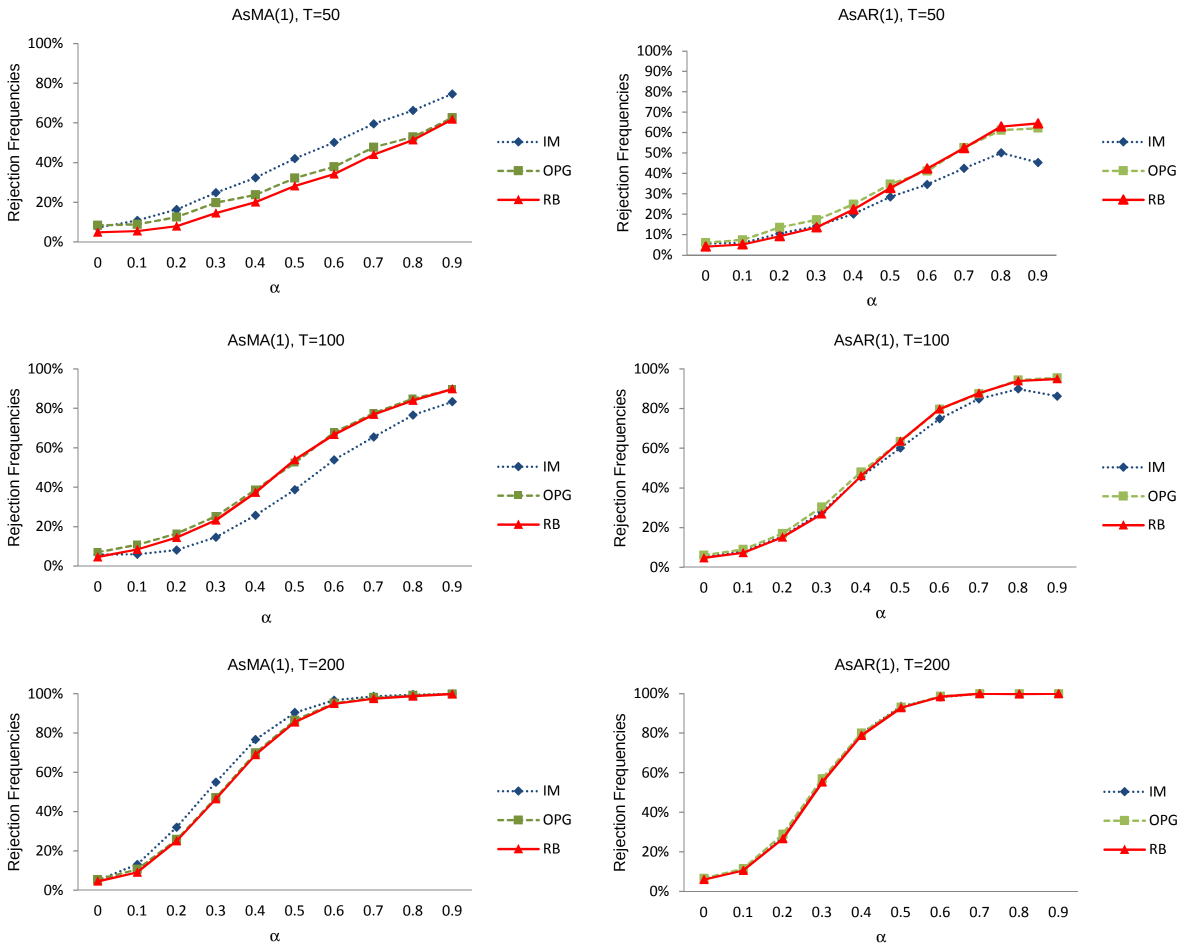


Figure 2.4: Power of the LM_T variants when the lag length is determined using the BIC. Errors are generated as $N\left(0, 1\right)$. Each figure is generated for $\beta = 0$ and α runs from 0 to 0.9 with step 0.1. The left panel presents results for the AsMA DGP and the right panel for the AsAR one. Number of MC replications for each output is 2000

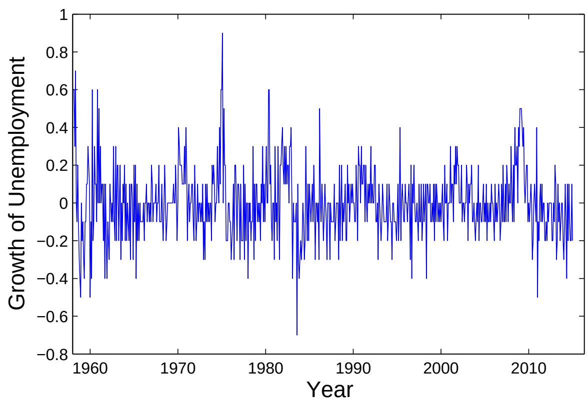


Figure 2.5: The growth of the U.S. unemployment rate from January 1958 to December 2014.

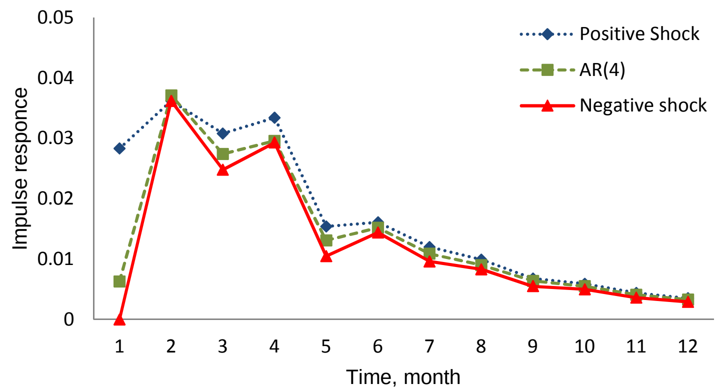


Figure 2.6: Impulse-response analysis of the growth of the U.S. unemployment rate based on the empirical AsAR(4) and AR(4) models. The effects of the positive shock $\varepsilon_0 = \widehat{\sigma}$ and the negative one $\varepsilon_0 = -\widehat{\sigma}$ are illustrated by the blue line with "diamonds" and the red line with "triangles", respectively. The impulse-response of the AR(4) model is given by green line with squares.

Bibliography

Brännäs, K. and J. G. De Gooijer (1994). Autoregressive-Asymmetric Moving Average Models for Business Cycle Data. *Journal of Forecasting 13*(6), 529–544.

Brännäs, K., J. G. De Gooijer, C. Lönnbark, and A. Soultanaeva (2012, January). Simultaneity and Asymmetry of Returns and Volatilities: The Emerging Baltic States' Stock Exchanges. *Studies in Nonlinear Dynamics & Econometrics 16*(1), 1–24.

Brännäs, K., J. G. De Gooijer, and T. Teräsvirta (1998). Testing Linearity against Nonlinear Moving Average Models. *Communications in Statistics, Theory and Methods 27*, 2025–2035.

Caporale, G. M. and L. A. Gil-Alana (2007). Nonlinearities and Fractional Integration in the US Unemployment Rate. *Oxford Bulletin of Economics and Statistics 69*(4), 521–544.

Davidson, R. and J. MacKinnon (1983). Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test. *Economics Letters 12*, 269–275.

Davidson, R. and J. G. MacKinnon (2001, January). Artificial Regressions. Working Papers 1038, Queen's University, Department of Economics.

Elwood, S. K. (1998). Is the Persistence of Shocks to Output Asymmetric? . *Journal of Monetary Economics 41*(2), 411 – 426.

Gelfand, I. M. and G. Shilov (1964). *Generalized Functions, Properties and Operations*, Volume 1. New York and London: Academic Press.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, B. (1997). Inference in TAR Models. *Studies in Nonlinear Dynamics & Econometrics 2*(1), 1–14.

Karras, G. and H. H. Stokes (1999). On the Asymmetric Effects of Money-Supply Shocks: International Evidence from a Panel of OECD Countries. *Applied Economics 31*(2), 227–235.

Kilian, L. and R. J. Vigfusson (2011, November). Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases? *Quantitative Economics 2*, 419–453.

Koutmos, G. (1999). Asymmetric Index Stock Returns: Evidence from the G7. *Applied Economics Letters 6*(12), 817–820.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Ltd.

Pelagatti, M. M. and F. Lisis (2009). *"Variance Initialisation in GARCH Estimation" Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction*.

Phillips, P. (1991, December). A Shortcut to LAD Estimator Asymptotics. *Econometric Theory 7*(04), 450–463.

Phillips, P. (1997). Towards a Unified Asymptotic Theory for Autoregression. *Biometrica 74*(3), 535–547.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.

Wecker, W. E. (1981). Asymmetric Time Series. *Journal of the American Statistical Association 76*(373), 16–21.

White, H. (2001). *Asymptotic Theory for Econometricians: Revised Edition*. Academic Press.

Yilanci, V. (2008). Are Unemployment Rates Nonstationary or Nonlinear? Evidence from 19 OECD Countries. *Economics Bulletin 3*(47), 1–5.

Chapter 3

Forecasting Methods for Functional Time Series

3.1 Introduction

In recent years advances in data collection and storage led to the possibility of recording many real life processes at increasingly high accuracy. Examples include high frequency data such as financial transactions, environmental data such as ozone or insolation maps and economic data such as income distributions or yield curves. The availability of large amounts of data offers manifold opportunities for researchers to obtain a better understanding of the underlying processes. However, to make use of this growing information and efficiently handle big data sets, suitable statistical tools are required to describe, model and forecast the relevant characteristics of this data. Functional data analysis (FDA) has emerged as a response to this request and has consequently been growing into an important field of statistical research.

In FDA, where large data sets are utilized in the form of functional observations (or curves), the focus has been mostly on independent and identically distributed observations. In many empirical applications data is collected sequentially over time. Consequently, we expect that the functional observations in a given time period are affected by past observations. Therefore, additional tools are required to analyze data that is given in the form of a functional time series (FTS). This paper studies the problem of describing and forecasting FTS and consists of two main parts. In a first step we provide a simple yet broad framework to quantify time dependencies in FTS. Second, we develop forecasting techniques for FTS under the given definition of time dependency.

Stochastic processes with time dependencies have been considered in the statistical literature. In the context of classical (i.e., finite dimensional) time series analysis, ergodicity and various mixing conditions are well established and frequently used (see, e.g. Hamilton (1994) and Davidson (1994) for a review). In the functional context, however, only few concepts are available when dealing with time-dependent observations. A key reference is Hörmann and Kokoszka (2010) who introduce a moment based notion of weak dependence using m-dependence. In this paper we complement the approach of Hörmann and Kokoszka (2010) by suggesting an alternative concept of time dependencies for FTS. Using the spectral Karhunen-Loeve representation functional observations can be represented by their functional principal component (FPC) scores. Therefore, the dependence between functional observations can be quantified through their respective FPC scores. This approach allows us to adapt various concepts of dependence available in the time series literature to the functional context. In particular, we consider dependence based on the autocovariances and cumulants of FPC scores. Further, since FPCs play a major role in explaining time dependencies it is necessary to establish the consistency of their estimates. We derive the convergence rates for the estimators of the FPCs under quite general serial dependence that allows for the long range dependence of the FPC scores. This in turn extends the result in Hörmann and Kokoszka (2010).

In the second part of the paper we discuss forecasting methods for FTS. Most work dedicated to the prediction of (FTS) has focused on the functional autoregressive model of order one (FAR(1)) suggested in Bosq (2000). In particular, Bosq (2000) derives the estimator and the predictor for the FAR model using the Yule Walker equation and shows their consistency. Besse et al. (2000) propose a local adaptation of the FAR(1) model by introducing a nonparametric weighted kernel estimator. The issue of weak convergence for estimates of the FAR(1) model is addressed in Mas (2007). Kargin and Onatski (2008) develop a predictive factor technique for the estimation of the autoregressive operator. Park and Qian (2012) apply the FAR(1) framework to model FTS of distributions. Didericksen et al. (2012) provide a small sample simulation study of the performance of the FAR(1) model and several competing prediction techniques. More recently, Kokoszka and Reimherr (2013) suggest a testing procedure to determine the lag order for more general FAR(p) processes. Aue et al. (2015) suggest a simple alternative procedure to transform the FAR model into a vector autoregressive model of functional principal scores, where standard multivariate techniques can be used to model and predict FTS.

In order to forecast FTS that follow our concept of time dependence we discuss two forecasting techniques. First, FTS processes that have a linear response to the past functional observations can be forecasted by the FAR model. We show that the autocovariance estimator given in Bosq (2000) is consistent under our notion of time dependence and derive its convergence rate. However, the concept of time dependence we introduce covers a broader class of processes than described by FAR. More precisely, the behavior of the autocovariances of the FPC scores is less restrictive (in particular we can allow for long range dependence) and non-linear responses are possible. For this reason we generalize the FAR model to the functional additive autoregressive model (FAAR). The idea of functional additive models was introduced by Müller and Yao (2008) in the context of functional linear regressions. This approach gives rise to a more flexible and essentially nonparametric model and allows us to consider the problem of prediction as a problem of nonlinear response of the FPC scores. To estimate the nonlinear responses we propose a k-nearest neighbors classification approach that is simple to implement and in the finite-dimensional setting well understood. As this approach has been successfully applied to classical time series analysis (see, e.g., Cover and Hart (1967), Stone (1977), Stute (1984) and Yakowitz (1987)), we can use the available theoretical results to derive the convergence rate of our predictor in the FAAR model.

To assess the performance of the proposed forecasting methods in small samples we provide a Monte Carlo simulation study. In particular, we compare the accuracy of the prediction of the FAAR model to the FAR model, the multivariate score model suggested by Aue et al. (2015) and benchmark models such as mean predictor, naive predictor and prediction of VAR for discrete observations. Further, we compare the performance of the above mentioned FTS models in forecasting electricity consumption in Denmark, Finland, Norway and Sweden. Our results show that FAAR models and multivariate score models provide the most accurate forecasts.

The remainder of this paper is organized as follows. Section 3.2 introduces the notion of dependence for functional time series. Section 3.3 discusses the impact of time dependence on the estimators of the functional principal components. In Section 3.4 we address the FAR model, while a generalization of the FAR model to FAAR, its estimation and asymptotic properties are presented in Section 3.5. A supporting small sample study is presented in Section 3.6. An empirical application to electricity consumption is described in Section 3.7 and concluding remarks are given in Section 3.8. All proofs, figures and tables are collected in the Appendix.

3.2 Methodology and Assumptions

We shall assume that we observe a series of functional observations $\{X_i(t)\}$ for $t \in [a, b]$ and $i = 1, ..., N$, where the interval $[a, b]$ is normalized to $[0, 1]$. For each i the observation X_i belongs to the Hilbert space $H = L^2([0, 1], \|\cdot\|)$ of square integrable functions which is equipped with a norm $\|\cdot\|$ induced by the inner product $\langle x, y\rangle \equiv \int_0^1 x(t)y(t)\mathrm{d}t$. The object $\{X_i(t)\}_{i=1}^N$ is referred to as functional time series (see e.g., Horváth and Kokoszka, 2012, Chapter 13-16 and Bosq, 2000 for a survey on FTS analysis) and we refer to i as the time index. In what follows the data $\{X_i\}$ are assumed to be given in a functional form since the problem of data representation in functional form has been extensively studied in the literature (see, e.g., Ramsay and Silverman, 2005 for a review of the available techniques and general description of FDA).

Our attention is restricted to weakly stationary processes allowing for the standard time series representation

$$X_i = G\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right), \tag{3.1}$$

where $\{\varepsilon_i\}$ denotes the series of errors or innovations which are i.i.d elements from Hilbert space H, and G is a measurable function $G : H^\infty \to H$. In this paper two cases of representation (3.1) are considered. The first is the functional autoregressive (FAR) model that models linear responses of a FTS to its lags (see Section 3.4). Second, To account for possible nonlinear responses we extend the FAR framework to more general settings using the functional additive approach suggested in Müller and Yao (2008) for functional regressions (see Section 3.5). Representation (3.1) can also be extended to non-stationary sequences $\{X_i\}$. We do not pursue this topic in our paper and refer the interested reader to Horváth et al. (2014) for additional insights. For future reference, $\mathcal{S}$ denotes the space of Hilbert-Schmidt operators from H to H and is equipped with the operator norm $\|\cdot\|_{\mathcal{S}}$ (i.e., for some $\Psi \in \mathcal{S}$, $\|\Psi\|_{\mathcal{S}} = \left(\sum_{h=1}^{\infty} \|\Psi(e_h)\|^2\right)^{1/2}$ for any orthonormal basis $\{e_h\}_{h\geq 1}$) and the space of bounded linear operators on H is denoted by $\mathcal{L}$ with the norm $\|\Psi\|_{\mathcal{L}} = \sup_{\|x\|\leq 1} \{\|\Psi(x)\|,\ x \in H\}$.

We begin by describing the concept of time dependency in functional time series. It is founded on the spectral decomposition of random functions as follows. All random functions are defined on a common probability space $(\Omega, \mathcal{A}, P)$. Let $L^p_H\left(\Omega, \mathcal{A}, P\right)$ denote the space of H valued random variables X such that for $p \geq 1$, $\mathbb{E}\|X\|^p < \infty$. Every function $X \in L^2_H$ possesses a mean function $\mu := \mathbb{E}\left(X\right)$ and a covariance operator $C(x) := \mathbb{E}\left[\langle X-\mu, x\rangle X-\mu\right]$, where $x \in L^2$ and C admits the spectral decomposition. That is,

$$C(x) = \sum_{\ell=1}^{\infty} \lambda_\ell \langle \psi_\ell, x\rangle \psi_\ell, \tag{3.2}$$

where $\{\lambda_\ell\}_{\ell\geq 1}$ is the strictly positive decreasing sequence of eigenvalues and $\{\psi_\ell\}_{\ell\geq 1}$ denotes the corresponding sequence of eigenfunctions (i.e., $C(\psi_\ell) = \lambda_\ell \psi_\ell$) which forms an orthonormal basis system of H. It follows that X admits the Karhunen-Loève representation

$$X(t) = \mu(t) + \sum_{\ell=1}^{\infty} \theta_\ell \psi_\ell(t), \tag{3.3}$$

where $\theta_\ell = \langle X, \psi_\ell\rangle$ denotes the ℓ-th functional principal component *score* of X. By construction, the sequence of functional principal component scores $\{\theta_\ell\}_{\ell\geq 1}$ is such that the elements θ_ℓ are uncorrelated across the spectral dimension ℓ, have mean zero and variance λ_ℓ. Then for a given weakly stationary FTS $\{X_i\}$ (such that for each $i = 1, ..., N$, $X_i \in L^2_H$) X_i admits a Karhunen-Loève decomposition which in turn yields a sequence of scores $\{\theta_{i,\ell}\}$, and the corresponding sequences of eigenvalues $\{\lambda_\ell\}$ and eigenfunctions $\{\psi_\ell\}_{\ell\geq 1}$.

The following assumption formalizes how time dependencies between functional observations $\{X_i\}$ are translated into their score series. Let $\kappa_{\ell_1,\ldots,\ell_q}(0,\tau_1,\ldots,\tau_{q-1})$ denote the q-th order cumulant of $(\theta_{i,\ell_1}, \theta_{i+\tau_1,\ell_2}, \ldots, \theta_{i+\tau_{q-1},\ell_q})$, where $\tau_1, \ldots, \tau_{q-1} \in \mathbb{N}$ are integers (see, e.g., Brillinger, 2001, p.19 for a more detailed description of cumulants). Then we shall assume:

Assumption 3″

(**i**) *For some* $\alpha > 2$ *and all* $\ell \geq 1$,

$$\lambda_\ell - \lambda_{\ell+1} \sim \ell^{-\alpha-1}.$$

(**ii**) *Define* $B_{\ell,s}^{(h)} := \sup_i |\mathbb{E}[\theta_{i,\ell}\theta_{i-h,s}]|$. *Then there exists a constant* $B > 0$ *and some* $\beta > 0$ *such that*

$$B_{\ell,s}^{(h)} \leq B\, h^{-\beta}\sqrt{\lambda_\ell \lambda_s}.$$

(**iii**) *For fixed* $q \geq 3$ *and some constant* $B > 0$, *the joint* q*-th order cumulants are absolutely summable*

$$\sum_{\tau_1,\ldots,\tau_{q-1}=-\infty}^{\infty} \cdots \sum \left|\kappa_{\ell_1,\ldots,\ell_q}(0,\tau_1,\ldots,\tau_{q-1})\right| \leq B \prod_{j=1}^{q} \lambda_{\ell_j}^{1/2}.$$

Part (**i**) of Assumption 3″ is the standard assumption that prevents the spacing between adjacent eigenvalues λ_ℓ from being too small. It also implies that $\lambda_\ell \sim \ell^{-\alpha}$. The importance of spacing property (**i**) will become particularly apparent from the results of Corollary 4, where the asymptotic properties of eigenfunction estimators are studied.

Part (**ii**) and (**iii**) of Assumption 3″ describe the form of time dependencies that we allow for the scores $\{\theta_{i,\ell}\}_{i,\ell \geq 1}$. The assumed behavior of $B_{\ell,s}^{(h)}$, which represents a measure of absolute covariances between score series $\{\theta_{i,\ell}\}$ and lagged series $\{\theta_{i-h,s}\}$, is only a mild restriction. In particular, part (**ii**) implies an intuitive restriction on the absolute summability of the h-th autocovariances of the score series $\{\theta_{i,\ell}\}_i$ across the spectral dimension ℓ, since $\sum_{\ell \geq 1} |\mathbb{E}[\theta_{i,\ell}\theta_{i-h,\ell}]| \leq \sum_{\ell \geq 1} B_{\ell,\ell}^{(h)} \leq Ch^{-\beta}$. However, absolute summability of the autocovariances of the score series is not required across the time dimension i and fixed spectral dimention ℓ. More precisely, for $0 < \beta < 1$ one can conclude that $\sum_{h=1}^{N} \mathbb{E}[\theta_{i,\ell}\theta_{i-h,\ell}] \leq \sum_{h=1}^{N} B_{\ell,\ell}^{(h)}$ is of order $N^{1-\beta}\lambda_\ell$ which diverges for fixed ℓ and large N. In what follows we refer to this as a long range dependence property. A similar restriction holds for the covariances of the score series across time dimension with fixed the spectral dimensions $\ell \neq s$, i.e.,

$$\sum_{h=1}^{N} |\mathbb{E}[\theta_{i,\ell}\theta_{i-h,s}]| = O\left(N^{1-\beta}\sqrt{\lambda_\ell \lambda_s}\right)$$

Finally, Assumption 3″ (**iii**) requires absolute summability of the joint cumulants of $\{\theta_{i,\ell}\}$ up to q-th order. This allows us to control the temporal dependencies in the q-th moments of the score series across spectral and time dimension. In particular, condition (**iii**) for one fixed spectral direction ℓ, $\sum_{\tau_1,\ldots,\tau_{q-1}=-\infty}^{\infty} |\kappa_{\ell,\ldots,\ell}(0,\tau_1,\ldots,\tau_{q-1})| \leq C\lambda_\ell^{q/2}$, implies the finiteness of the q-th moment, i.e., $\mathbb{E}\|X_i\|^q < \Delta < \infty$ for all i. For more details on how moments are related to cumulants see Appendix A equation (A.1). In general this cumulant condition is standard for the time series literature (see, e.g. Andrews, 1991, Brillinger, 2001, and Demetrescu et al., 2008) and provides us with a useful measure of the joint statistical dependence of higher order moments and a convenient tool for deriving the rates of convergence. It should be noted that the value of q is method-specific and as we shall see in the sequel relaxing linear structure of the model may require strengthening the restrictions on the moments.

Furthermore, note that the concept of α-mixing is closely related to the form of time dependencies assumed in (**ii**)-(**iii**). In fact, α-mixing together with finite sixth moments implies absolute summability of the joint cumulants up to sixth order (see, e.g. Andrews, 1991 or Gonçalves and Kilian, 2007). Hence, the main difference between the two approaches lies in the way autocovariances are handled. In general we find that conditions (**ii**) and (**iii**) have several advantages in a functional setting. First, they allow for a broader scope of time dependencies (in that absolutely summable autocovariances are not necessary which can be controlled through parameter β). Second, incorporating decay across the spectral dimension ℓ is straighforward, which is crucial for the analysis. Third, the stated conditions have an intuitive interpretation of the time dependence concept for functional data when compared to various mixing properties. Moreover, using standard time series techniques it can be easily verified in practice if there is time dependence between the scores of the FTS.

3.3 Properties of Functional Principal Components

The fundamental ingredients for describing time dependence in functional data are principal component scores. However, in practice scores and other FPC (C and its eigenvalues and eigenfunctions) are not known and must be estimated. Therefore, before developing forecasting methods that rely on Assumption 3″, it is crucial to verify the convergence of the estimated FPC to their population counterparts. Consistency results for the FPC are available for independent observations (see, e.g., Dauxois et al., 1982) and for L^4-m-dependent functional data (see e.g., Hörmann and Kokoszka, 2010). In this section we show that consistency of the corresponding estimators extends to our time dependency settings.

We start with the preliminaries. Suppose we observe $X_1, \ldots, X_N$. The standard estimators for the mean function, μ, and the covariance operator, $C(x)$, are given by the following sample averages

$$\hat{\mu}(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t), \tag{3.4}$$

$$\widehat{C}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \langle X_i - \hat{\mu}, x\rangle (X_i(t) - \hat{\mu}(t)), \quad x \in L^2. \tag{3.5}$$

Further, we denote the estimators of eigenvalues and eigenfunctions as $\{\hat{\lambda}_\ell\}_{\ell=1}^{L}$ and $\{\widehat{\psi}_\ell\}_{\ell=1}^{L}$, respectively. Using $\widehat{C}_N(t)$, they are computed from the eigenequation

$$\widehat{C}_N(\widehat{\psi}_\ell) = \widehat{\lambda}_\ell \widehat{\psi}_\ell.$$

Typically estimates of eigenelements ($\hat{\lambda}_\ell$ and $\widehat{\psi}_\ell$) can be obtained for an arbitrary fixed level L such that $L < N$. The asymptotic results in Section 3.4 and 3.5 provide a discussion of this issue, where L is set to be a function of N, such that $L \to \infty$ as $N \to \infty$. Ramsay and Silverman (2005, Section 6.4) discuss practical/computational methods for solving eigenequations.

Remark 13 *In what follows we shall assume without loss of generality that X_i have means equal to zero for all $i = 1, ..., N$. For any practical application the methodology introduced in this paper remains unchanged if data are centered prior to the forecasting exercise. For the completeness of the discussion we state the following result for the estimator of μ. For the weakly stationary FTS $\{X_i\}_{i=1}^{N}$ that fulfills Assumption 3″* (**i**)-(**ii**) *we have*

$$\mathbb{E}\left\|\hat{\mu}_N - \mu\right\|^2 = O\left(\max\left\{N^{-\beta}, N^{-1}\right\}\right).$$

The following result establishes the consistency of estimator (3.5).

Theorem 9 *If a weakly stationary FTS $\{X_i\}_{i=1}^{N}$ fulfills Assumption 3″ with joint cumulants up to order 4 then*

$$\mathbb{E}\left\|\widehat{C}_N - C\right\|_{\mathcal{S}}^2 = O\left(N^{-2\beta^*}\right),$$

where $\beta^* := \min\{\beta, 1/2\}$.

Theorem 9 implies that the fastest convergence speed that can be achieved for the empirical estimator of the covariance operator is N^{-1} when $\beta \geq 1/2$. This extends previously obtained results in Bosq (2000) and Hörmann and Kokoszka (2010) showing that the fastest convergence can also be achieved for processes that potentially posses long range dependencies. In other words, the absolute summability of the autocovariances of the functional principal component score series $\{\theta_{i,\ell}\}_{i\geq 1}$ across the time dimension i, is not necessary to get rate N^{-1}. If one is only interested in establishing the consistency of the covariance operator estimator, part (**ii**) of Assumption 3$''$ can be relaxed to $B_{\ell,s}^{(h)} \leq Bb_h\sqrt{\lambda_\ell, \lambda_s}$ with $\sum_{h=1}^{\infty} h^{-1}b_h < \infty$. This condition allows for a slow decay of the time dependencies represented by component b_h that can even be of logarithmic order $b_h = O\left(\ln(h)^{-1-\beta}\right)$ for $\beta > 0$ (see, e.g., Davidson, 1994, Theorem 2.31).

The autocovariance operator defined as

$$\Gamma_h = \mathbb{E}\left[\langle X_i, x\rangle X_{i-h}\right], \tag{3.6}$$

for $i = 1, ..., N$ and some h, can estimated similarly by the sample analogue

$$\widehat{\Gamma}_{h,N} = \frac{1}{N-1}\sum_{i=1}^{N-1} \langle X_i, x\rangle \left(X_i\left(t\right)\right). \tag{3.7}$$

Furthermore, the following holds for any autocovariance operator of order h.

Corollary 3 *If a weakly stationary FTS $\{X_i\}_{i=1}^N$ fulfills Assumption 3$''$ with joint cumulants up to order 4 then*

$$\mathbb{E}\left\|\widehat{\Gamma}_{h,N} - \Gamma_h\right\|_{\mathcal{S}}^2 = O\left(N^{-2\beta^*}\right).$$

Our next result gives explicit bounds for the mean squared error of the eigenelement estimators.

Corollary 4 *If a weakly stationary FTS $\{X_i\}_{i=1}^N$ fulfills Assumption 3$''$ with joint cumulants up to order 4 then*

$$\textbf{(i)} \qquad \mathbb{E}\left(\sup_{\ell\geq 1}\left|\hat{\lambda}_\ell - \lambda_\ell\right|^2\right) = O\left(N^{-2\beta^*}\right),$$

$$\textbf{(ii)} \qquad \mathbb{E}\left(\sup_{1\leq\ell\geq L}\left\|a_\ell\widehat{\psi}_\ell - \psi_\ell\right\|^2\right) = O\left(\delta_\ell^2 N^{-2\beta^*}\right),$$

where $a_\ell := sign(\langle\widehat{\psi}_\ell, \psi_\ell\rangle)$, $\delta_\ell := \max_{1\leq k\leq\ell}(\lambda_k - \lambda_{k+1})^{-1}$.

The results in Corollary 4 indicate that, as ℓ increases, it becomes more difficult to estimate the eigenfunctions ψ_ℓ associated with λ_ℓ since the expected L^2 error is proportional to δ_ℓ^2. As a consequence, the spacing between adjacent eigenvalues $\{\lambda_\ell\}_{\ell\geq 1}$ cannot decrease too fast. In particular, by Assumption 3$''$(**i**) $\mathbb{E}\left(\sup_{1\leq\ell\geq L}\left\|a_\ell\widehat{\psi}_\ell - \psi_\ell\right\|^2\right) = O\left(L^{2(1+\alpha)}N^{-2\beta^*}\right)$. Therefore, restriction $L = o\left(N^{\beta^*/(1+\alpha)}\right)$ has to hold for estimators $\{\widehat{\psi}_\ell\}_{\ell=1}^L$ to be consistent. Further, the estimator $\widehat{\psi}_l$ of ψ_l is only identified up to a change in sign. As is standard in the literature, we shall tacitly assume that the sign of $\widehat{\psi}_l$ is chosen such that $\int\widehat{\psi}_l\psi_l \geq 0$.

Note, recently Hörmann and Kidziński (2015) proofed that for the consistency of FPCs estimators the spacing property given in Assumption 3$''$(**i**) can be relaxed to more general settings. However, our subsequent analysis of the forecasting techniques in Sections 3.4 and 3.5 requires explicit rates of convergence for the estimators $\widehat{\lambda}_\ell$ and $\widehat{\psi}_\ell$ and consequently the spacing property.

3.4 Forecasting Linear FTS

In this section we discuss estimation and forecasting techniques for FAR models. As pointed out in the introduction the FAR(1) model is the model most commonly used in the FTS analysis and it is natural to use it as the main linear FTS benchmark model. The theory of FAR(1) processes in Hilbert and Banach spaces is studied in Bosq (2000) to which we refer the reader for a general overview. In this section we study the estimator suggested in Bosq (2000) and derive its convergence rate under the time dependency assumption stated in Section 3.2. For simplicity of exposition we consider the FAR model of order one.[1] The model takes the form

$$X_i = \rho\left(X_{i-1}\right) + \varepsilon_i, \tag{3.8}$$

where ε_i is a strong white noise in L^2_H, i.e., ε_i is a zero mean iid sequence in L^2_H with the covariance operator $C_\varepsilon(x) := \mathbb{E}\left[\langle \varepsilon_i, x\rangle \varepsilon_i\right]$ being a positive definite Hilbert-Schmidt operator. The autoregressive operator ρ is a assumed to be Hilbert-Schmidt operator satisfying

$$\|\rho^k\|_{\mathcal{L}} < 1 \text{ for some } k \geq 1. \tag{3.9}$$

This condition assures strict stationarity for process X_i (see, e.g., Bosq, 2000, Theorem 3.1). In other words, if (3.9) holds then function $G(\cdot)$ in FTS representation (3.1) takes an additive linear form

$$X_i = \sum_{h=1}^{\infty} \rho^h(\varepsilon_{i-h}).$$

To formulate the estimator of $\rho(\cdot)$ and derive its convergence rate we first address the well known issue often referred to as an ill-posed inverse problem. Recall that $C(x) = \mathbb{E}\left[\langle X_i, x\rangle X_i\right]$ and $\Gamma_h(x) = \mathbb{E}\left[\langle X_i, x\rangle X_{i-h}\right]$, and both operators allow for spectral representations

$$C(x) = \sum_{\ell=1}^{\infty} \lambda_\ell \langle \psi_\ell, x\rangle \psi_\ell, \tag{3.10}$$

$$\Gamma_h(x) = \sum_{\ell=1}^{\infty}\sum_{s=1}^{\infty} \mathbb{E}\left[\theta_{i,\ell}, \theta_{i-h,s}\right] \langle \psi_\ell, x\rangle \psi_s. \tag{3.11}$$

[1] See, e.g., Bosq (2000, Section 5) and Horváth and Kokoszka (2012, Chapter 15.1) for the review on how to estimate higher order FAR models

It follows from (3.8) that operator equation $\Gamma_1 = \rho C$ holds and formally gives the solution $\rho = \Gamma_1 C^{-1}$. However, the operator C does not have a bounded inverse on the entire space H. It follows from (3.10) that $C^{-1} = \sum_{\ell=1}^{\infty} \lambda_\ell^{-1} \langle \psi_\ell, x \rangle \psi_\ell$, where $\lambda_\ell^{-1} \to \infty$ as $\ell \to \infty$ and the domain of C^{-1} is restricted to $\mathcal{D}(C^{-1}) = \{y \in H \,|\sum_{\ell=1}^{\infty} \langle y, \psi_\ell \rangle^2 / \lambda^2 < \infty\}$. The standard method in the literature to circumvent this problem is to use only the first L functional components. That is, for $\lambda_1 > \lambda_2 > \ldots > 0$ we define H_L, a subspace of H spanned by the L-eigenvectors $\psi_1, ..., \psi_L$ associated with $\lambda_1 > \ldots > \lambda_L$, and consider

$$C_L^{-1} = \sum_{\ell=1}^{L} \lambda_\ell^{-1} \langle \psi_\ell, x \rangle \psi_\ell, \tag{3.12}$$

where C_L^{-1} is the inverse of C on H_L and L is the function of N such that $L \to \infty$ as $N \to \infty$. Then the estimator of ρ is based on (3.7), the sample analog of (3.12) and can be formulated as

$$\widehat{\rho}_N(x) = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{\ell,s=1}^{L} \widehat{\lambda}_\ell^{-1} \langle \widehat{\psi}_\ell, x \rangle \widehat{\theta}_{i,\ell} \widehat{\theta}_{i+1,s} \widehat{\psi}_s. \tag{3.13}$$

Remark 14 *Note that the FAR process (3.8)-(3.9) satisfies the time dependence notion discussed in Section 3.2, however it impose stricter conditions on the autocovariances of the FPC scores:*

1. *The FAR process (3.8)-(3.9) does not posses the long range dependence property (i.e., $\beta > 1$). Indeed, condition (3.9) implies $\sum_{h=1}^{\infty} \left\|\rho^h\right\|_{\mathcal{L}} < \infty$ which in turn implies $\sum_{h=1}^{\infty} \|\Gamma_h\|_{\mathcal{L}} < \infty$. Using expression (3.11) one can conclude that $\sum_{h=1}^{\infty} \|\Gamma_h\|_{\mathcal{L}} < \infty$ if $\beta > 1$.*

2. *The autocovariances of the FPC scores $\mathbb{E}[\theta_{i,\ell}\theta_{i-h,\ell}]$ decay faster then the variances $\mathbb{E}[\theta_{i,\ell}\theta_{i,\ell}]$ across spectral dimension ℓ. To see this note that the autoregressive operator ρ admits the representation*

$$\rho(x) = \sum_{\ell=1}^{\infty} \sum_{s=1}^{\infty} a_{\ell,s} \langle \psi_\ell, x \rangle \psi_s, \;\; \textit{with } x \in H, \tag{3.14}$$

where $a_{\ell,s} = \mathbb{E}\left[\theta_{i,\ell}, \theta_{i-1,s}\right] \lambda_\ell^{-1}$ denote the spectral coefficients. Further, we adopt the approach of Hall and Horowitz (2007) for functional linear regressions and substitute Assumption 3′ **(ii)** *with one, that allows us to control the decrease of the spectral coefficients $a_{\ell,s}$ with more flexibility (see Assumption 3.3 in Hall and Horowitz, 2007). That is, instead of Assumption 3′* **(ii)** *assume there exists a constant $B > 0$, some $\beta > 1$ and $\gamma > 1/2 + \alpha$ such that for all $\ell \geq 1$,*

$$B_{\ell,s}^{(h)} \leq B\; h^{-\beta} \ell^{-\gamma} s^{-\gamma}. \tag{3.15}$$

Then, since ρ is the Hilbert-Schmidt operator we have $\sum_{s=1}^{\infty}\sum_{\ell=1}^{\infty} a_{\ell,s}^2 < \infty$. The squared summability of $a_{\ell,s}$ is assured if and only if $\gamma > 1/2+\alpha$. In turn, the autocovariances of the FPC scores behave as $\mathbb{E}[\theta_{i,\ell}\theta_{i-h,\ell}] = O(\ell^{2\gamma})$ and decay faster then the variances $\mathbb{E}[\theta_{i,\ell}\theta_{i,\ell}] = O(\ell^{\alpha})$.

The following result shows the consistency of $\widehat{\rho}_N$ and its speed of convergence.

Theorem 10 *If a FAR process* (3.8)-(3.9) *satisfies Assumption 3″* (**i**) *and* (**iii**) *with joint cumulants up to order* 4*, and condition (3.15) then*

$$\|\widehat{\rho}_N - \rho\|_{\mathcal{L}} = O_p\left(\max\left\{\frac{L^{2\alpha+\frac{3}{2}}}{\sqrt{N}}, L^{1+2(\alpha-\gamma)}\right\}\right). \tag{3.16}$$

The rate of convergence for the estimator of the autoregressive operator consists of two parts. The first one, $\frac{L^{2\alpha+\frac{3}{2}}}{\sqrt{N}}$, characterizes the convergence of estimator $\widehat{\rho}_N$ to the truncated true operator $\rho_L = \Gamma_1 C_L^{-1}$. Moreover, it restricts L for the estimator $\widehat{\rho}_N$ to be consistent such that $L = o\left(N^{1/(4\alpha+3)}\right)$ and $L \to \infty$ as $N \to \infty$. The second part, $L^{1+2(\alpha-\gamma)}$, describes asymptotic behaviour of the reminder $\|\rho_L - \rho\|_{\mathcal{L}}$, which converge in probability to zero since $1+2(\alpha-\gamma) < 0$. Note that the fastest convergence rate $O_p\left(N^{-1/2}\right)$ can be achieved when space H is finite dimensional which is inline with the results for the OLS estimator of stationary multivariate autoregressive models (such as VAR for instance).

3.5 Forecasting Nonlinear FTS

As the correct model specification for FTS is not known in practice it might be too restrictive to assume a linear modeling framework, as for instance, FAR model. For this reason, in this section we propose a simple, yet robust and versatile approach to tackle potential nonlinearity in FTS. We use the functional additive approach of Müller and Yao (2008) to generalize FAR(1) model (3.8) and rewrite it as a functional additive autoregressive model. Using equation (3.14) the FAR model can be rewritten as standard linear regression model with infinitely many FPC score as predictors,

$$\mathbb{E}\left[X_{i+1}|X_i\right] = \sum_{s=1}^{\infty}\sum_{\ell=1}^{\infty} a_{\ell,s}\theta_{i,\ell}\psi_s,$$

In particular, the relationship between the response and predictor scores is modeled linearly as $\mathbb{E}\left[\theta_{i+1,s}|X_i\right] = \sum_{s=1}^{\infty} a_{\ell,s}\theta_{i,\ell}$. Furthermore, the linear framework of the FAR model and the uncorrelatedness of the FPS scores imply that $\mathbb{E}\left[\theta_{i+1,s}|\theta_{i,\ell}\right] = a_{\ell,s}\theta_{i,\ell}$. As suggested in Müller and Yao (2008), this model can be generalized by replacing the linear terms $a_{\ell,s}\theta_{i,\ell}$ by functional counterparts $m_{\ell,s}(\theta_\ell)$. This transforms the FAR model into a functional additive autoregressive model (FAAR)

$$\mathbb{E}\left[X_{i+1}|X_i\right] = \sum_{s=1}^{\infty}\sum_{\ell=1}^{\infty} m_{\ell,s}(\theta_{i,\ell})\psi_s, \tag{3.17}$$

where it is assumed that $\mathbb{E}[m_{\ell,s}(\theta_{i,\ell})] = 0$ for all $\ell, s \geq 1$ to assure identifiability. We impose a mild restriction on the model (3.17). Let the random principal component scores $\theta_{i,\ell}$ have unconditional probability density function $f_\ell(\theta_{i,\ell})$, and write $f_{\ell,s}(\theta_{i+1,s}|\theta_{i,\ell})$ for the conditional probability density of $\theta_{i+1,s}$ given $\theta_{i,\ell}$.

Assumption 2 *$m_{\ell,s}(\cdot)$, $f_\ell(\cdot)$ and $f_{\ell,s}(\cdot)$ are twice continuously differentiable and $f_\ell(\cdot)$, and $f_\ell(\cdot)$ are bounded. Furthermore, the functional principal component scores $\theta_{i,\ell}$ and $\theta_{i,s}$ are independent for $\ell \neq s$.*

That is, the only requirement for functions $m_{\ell,s}(\cdot)$ is smoothness. Further, Assumption 2 strengthens contemporaneous uncorrelatedness of the FPC scores to independency. This in turn implies that

$$\mathbb{E}\left[\theta_{i+1,s}|\theta_{i,\ell}\right] = \mathbb{E}\left[\mathbb{E}\left[\theta_{i+1,s}|X_i\right]|\theta_{i,\ell}\right] = \mathbb{E}\left[\sum_{q=1}^{\infty} m_{q,s}(\theta_{i,q})|\theta_{i,\ell}\right] = m_{\ell,s}\left(\theta_{i,\ell}\right).$$

The simple and flexible framework of model (3.17) provides us with a non-linear alternative to the FAR model. In particular, representation (3.17) motivates a straightforward forecasting scheme to predict the expected value of X_{N+1} through estimates of the conditional means $m_{\ell,s}(\theta_{N,\ell})$. Define the predictor $M(X_N) := \mathbb{E}\left[X_{N+1}|X_N\right]$. Then using the approximation $\widehat{X}_{i,L} = \sum_{\ell=1}^{L}\widehat{\theta}_{i,\ell}\widehat{\psi}_\ell$ instead of real functions X_i the estimator of $M(X_N)$ can be constructed as

$$\widehat{M}_{N,L}(\widehat{X}_{N,L}) = \sum_{\ell=1}^{L}\sum_{s=1}^{L} \widehat{m}_{\ell,s}(\widehat{\theta}_{N,\ell})\widehat{\psi}_s, \tag{3.18}$$

where L is set to be a function of N such that $L \to \infty$ as $N \to \infty$. While the estimation of the functional principal components ψ_ℓ and $\theta_{i,\ell}$ has already been discussed in Section 3.3, we propose in the following section an estimator for the conditional means $m_{\ell,s}(\theta_{i,\ell})$.

3.5.1 k-Nearest Neighbors Estimator

In this section a simple method based on the k-nearest neighbors approach (KNN) is suggested to estimate predictor $M(X_N)$. The main idea behind forecasting with KNN is to identify the past observations of the time series that are most similar (in terms of some distance) to the last onservation and use a combination of their future values to predict the next value of the series.

If FTS satisfies model (3.17) and Assumptions $3''$ and 2 then the KNN method can be adopted directly to the series of the FPC scores. The estimation procedure consists of three basic steps:

1. Use data $X_1, ..., X_N$ and the FPC analysis to compute estimates $\widehat{\psi}_\ell$, $\widehat{\lambda}_\ell$ and FPC scores $\{\widehat{\theta}_{i,\ell}\}_{i=1}^N$ for $\ell = 1, ..., L$ (as described in Section 3.3).

2. Compute the distance between the most recent FPC score $\widehat{\theta}_{N,\ell}$ and each element in the rest of the score series $\{\widehat{\theta}_{i,\ell}\}_{i=1}^{N-1}$. A typical choice for this task Minkowski distance. Denote the index set of the k_N closest neighbors to the feature score component $\widehat{\theta}_{N,\ell}$ by $\mathcal{I}(k_N; \widehat{\theta}_{N,\ell})$, where the number of neighbors depends on sample size N such that $k_N \to \infty$ as $N \to \infty$.

3. Once the k_N closest elements are identified their subsequent values are averaged to obtain the final estimator, i.e.,

$$\widehat{m}_{\ell,s}(\widehat{\theta}_{N,\ell}) := \frac{1}{k_N} \sum_{i \in \widehat{\mathcal{I}}\left(k_N; \hat{\theta}_{N,\ell}\right)} \hat{\theta}_{i+1,\ell}, \tag{3.19}$$

   for $\ell, s = 1, ..., L$.

Substituting estimates $\widehat{m}_{\ell,s}(\widehat{\theta}_{N,\ell})$ and $\widehat{\psi}_s$ where $\ell, s = 1, ..., L$ back to (3.18) gives the functional predictor. Note that KNN estimator (3.19) is presented with equal weights $1/k_N$. Alternative weighting schemes can be considered as well. For instance, weights can be set to be inversely proportional to the distance between the last observation $\widehat{\theta}_{N,\ell}$ and a neighbor from $\widehat{\mathcal{I}}\left(k_N; \hat{\theta}_{N,\ell}\right)$, i.e.,

$$w_i = \frac{\frac{1}{d_i}}{\sum\limits_{j=1}^{k_N} \frac{1}{d_j}},$$

where d_i is a distance between $\widehat{\theta}_{N,\ell}$ and a neighbor $i \in \widehat{\mathcal{I}}\left(k_N; \hat{\theta}_{N,\ell}\right)$.

3.5.2 Asymptotic properties of FKNN

We split the investigation of the asymptotic properties of predictor (3.18)-(3.19) for FAAR model into two parts as follows. Consider the *infeasible* estimator of $m_{\ell,s}(\theta_\ell)$ given by

$$\widetilde{m}_{\ell,s}(\theta_{N,\ell}) := \frac{1}{k_N} \sum_{i \in \mathcal{I}\left(k_N; \theta_{N,\ell}\right)} \theta_{i+1,\ell}.$$

where all quantities of spectral decomposition, λ_ℓ, ψ_ℓ and $\theta_{i,\ell}$ are assumed to be known. Consequently, the *infeasible* functional predictor $M_{N,L}(x_L)$ with the additional smoothing step based on a approximation $X_{i,L}(t) = \sum_{\ell=1}^{L} \theta_{i,\ell}\psi_\ell(t)$ is defined by

$$M_{N,L}\left(X_{N,L}\right) := \sum_{\ell=1}^{L}\sum_{s=1}^{L} \widetilde{m}_{\ell,s}(\theta_{N,\ell})\psi_s.$$

Then to obtain the convergence rate of the estimator (3.18)-(3.19) to the true predictor it suffices to obtain the convergence rate of infeasible estimator to the true predictor, $\mathbb{E}\|M_{N,L}(X_{N,L})-M(X_N)\|^2$, and convergence rate of the feasible estimator (3.18)-(3.19) to infeasible one, $\mathbb{E}\left\|\widehat{M}_{N,L}(\hat{x}_L) - M_{N,L}(x_L)\right\|^2$. The following theorems present the respective convergence rates.

Theorem 11 *Let a weakly stationary FTS $\{X_i\}_{i=1}^N$ fulfills Assumption 3′ with joint cumulants up to order* 4*, Assumption 2 and follows model* (3.17)*. Moreover, it is assumed that $L^{\alpha-1}\sum_{\ell=L}^{\infty}\mathbb{E}\left[m_{\ell,s}^2(\theta_{i,\ell})\right] = O\left(\lambda_s\right)$. Then we have*

$$\mathbb{E}\|M_{N,L}(X_{N,L}) - M(X_N)\|^2 = O\left(\max\left\{k_N^{-1}, L^{1-\alpha}\right\}\right),$$

where $k_N \sim N^{4/5}$.

Theorem 12 *If a weakly stationary FTS $\{X_i\}_{i=1}^N$ fulfills Assumption 3′ with joint cumulants up to order* 6*, Assumption 2 and follows model* (3.17) *then*

$$\mathbb{E}\left\|\widehat{M}_{N,L}(\hat{x}_L) - M_{N,L}(x_L)\right\|^2 = O\left(\frac{L^{3+2\alpha}\log(N)}{N^{2\beta^*}}\right), \tag{3.20}$$

where $\beta^* = \min\{\beta, 1/2\}$.

The result of Theorem 11 implies that the infeasible estimator is consistent and its convergence rate consists of two parts. The first part, k_N^{-1}, describes the convergence of the infeasible estimator to the truncated true predictor $M_L(X_{N,L}) = \sum_{s,\ell=1}^{L} m_{\ell,s}(\theta_{N,\ell})\psi_s$. It also shows that the consistency result requires the number of neighbors to be the function of the sample size such that $k_N \sim N^{4/5}$. The second one characterizes the convergence of the remainder $\mathbb{E}\|M_L(X_{N,L}) - M(X_N)\|^2$ which is of order $O\left(L^{1-\alpha}\right)$.

Theorem 12 delivers the convergence between feasible and infeasible estimators. One benefit of this result is that it allows us to state the restrictions on the principal component cutoff L. It is required that $L = o\left(N^{2\beta^*/(2\alpha+3)}/\log(N)^{1/(2\alpha+3)}\right)$ and $L \to \infty$ as $N \to \infty$ to obtain the consistent FAAR predictor.

3.6 Small Sample Performance

We now turn to study the small-sample properties of the proposed models. The objective of this section is twofold. The first objective is to evaluate the forecasting performance of the FAR and the FAAR frameworks in different setups, relating to the asymptotic results obtained in Sections 3.4 and 3.5. The second one is to conduct a comparison of the proposed models with other alternatives available in the related forecasting/functional literature. The last aspect is covered by examining the comparative forecast performance of the FAR model and FAAR approach with that of the

1. *VAR model.* It is natural to investigate when functional settings provide an advantage compared to standard multivariate techniques. For this reason we include the VAR method, where functional observations X_i are treated as $T \times 1$ vectors $\mathbf{X}_i = [X_i(t_1), ..., X_i(t_K)]'$. These vectors are obtained by evaluating the original functions at T equidistant points $t_s = \frac{s-1}{T-1}$, $s = 1, ..., T$ and $i = 1, ..., N$;
2. *Improved FAR* [iFAR]. This approach is suggested by Kokoszka and Zhang (2010) to control for possibly small values of $\widehat{\lambda}_\ell$ that potentially can be translated into large errors in $\widehat{\lambda}_\ell^{-1}$. It is suggested to add a positive baseline to $\widehat{\lambda}_\ell$ in (3.13) for $\ell \geq 2$;
3. *Multivariate score model.* This model is recently suggested by Aue et al. (2015) and is based on the standard multivariate techniques applied to the vector of scores. Here we employ the VAR model for the score series which provides a simplified and elegant alternative for the FAR model.. In what follows this method will be referred to as MSM method.

We also supplement our comparative analysis with two standard benchmarks commonly employed in functional data analysis (see, e.g., Didericksen et al., 2012). The first is *Mean prediction* [MP], where predictors are obtained as the mean of the sample $\widehat{X}_{N+1} = \frac{1}{N}\sum_{i=1}^{N} X_i$, and the second is *Naive Prediction* [NP] given as $\widehat{X}_{N+1} = X_N$.

We use the FAR(1) model as the *main benchmark* design for FTS processes

$$X_i(t) = \int_0^1 \rho(t, s) X_{i-1}(s) \mathrm{d}s + \varepsilon_i(t), \tag{3.21}$$

for $i = 1, ..., N$. The error terms are generated as Brownian bridges

$$\varepsilon_i(t) = W(t) - tW(1), \tag{3.22}$$

where $W(\cdot)$ is the standard Wiener process generated as $W(\frac{k}{K}) = \frac{1}{\sqrt{K}}\sum_{j=1}^{k} Z_j$ for $k = 1, ..., K$ and Z_j are independent standard normals.

Three different forms of the kernel $\rho(t, s)$ are used: $\rho(t, s) = Ce^{\frac{-(t^2+s^2)}{2}}$, $\rho(t, s) = C$ and $\rho(t, s) = Ct$. In all cases the constant C is chosen such that $\|\rho\|_{\mathcal{S}} = 0.5$. Samples of size $N = 50, 100$ and 200 have been generated with a burn-in period of 100 functional observations. In all cases $N - 1$ observations where used for the estimation and on the last observation a one-step ahead forecast was computed. All results were repeated

$N_r = 1000$ times. For the FAAR model, the number of nearest neighbors k_N was set to $N^{4/5}$ as suggested by Theorem 11. To estimate and forecast with the VAR model the size of the grid has to be specified and the following rule was applied $T = 0.1N$. Finally, to measure the forecasts performance, the mean squared error (MSE) and the mean median error (MME) were computed, i.e.,

$$MSE \equiv \frac{1}{N_r}\sum_{j=1}^{N_r} \|X_{N+1}^j - \widehat{X}_{N+1}^j\|^2, \tag{3.23}$$

$$MME \equiv \frac{1}{N_r}\sum_{j=1}^{N_r} \int_0^1 \left| X_{N+1}^j(s) - \widehat{X}_{N+1}^j(s)\mathrm{d}s \right|, \tag{3.24}$$

where X_{N+1}^j and $\widehat{X}_{N+1}^j$ represent real observations and obtained forecasts, respectively, for j's replication. It should be mentioned that we used two approaches to estimat the number of FPC L. First, L is selected such that FPCs explain at least 99% or 95% of the variability in the sample. Second, we apply the selection criteria suggested in Aue et al. (2015). We report that the second approach provides forecasts with smaller MSE and MME errors. Therefore, the results based on the first approach are omitted here and are available upon request.

We report our results in the form of boxplots of the errors MSE and MME for different sample sizes and kernels. Figures 3.1, 3.2 and 3.3 present the results for the case when the kernel is given as $\rho(t,s) = Ce^{\frac{-(t^2+s^2)}{2}}$, $\rho(t,s) = C$ and $\rho(t,s) = Ct$, respectively. All models based on functional observations (e.g., FAAR, FAR, iFAR and MSM) perform significantly better than the benchmark predictors and the VAR model, except for the special case when $\rho(t,s) = C$. In this setup, the mean predictor provides the best forecasting results due to the structure of the DGP. In general, none of the FAR, iFAR and MSM dominates the others, while the FAAR model has marginally higher median and variance of the forecast errors. This stems from the fact that the aim of the FAAR model is to forecast general autoregressive processes while FAR, iFAR and MSM are explicitly tailored for the considered FAR DGP.

3.7 Forecasting electric load demand in the Nordic countries

In this section we are considering the prediction of daily electric load demand curves in the Nordic countries from a functional perspective. This problem has been of high interest to decision makers in the energy sector and has seen numerous contributions in the statistical literature. Traditionally, parametric time series models have been applied to this problem - both classical time series methods and machine learning type methods such as artificial neural networks and support vector machines (see, e.g., Kyriakides and Polycarpou, 2007, Feinberg and Genthliou, 2005, Hippert et al. (2001) and Chen et al. (2004) among others). This section describes the implementation and comparison of the FTS models discussed in Section 3.6.

The data that is used in this application has been provided by Nord Pool Spot AS, the energy exchange of the Nordic and Baltic countries in Oslo, Norway [2]. Hourly demand data is made available for Denmark, Finland, Norway and Sweden since 2013. The time stamps of the raw data are converted to UTC such that every day has always 24 hours. That is, our sample for each country consist of $N = 987$ daily observations from January 1, 2013 till September 15, 2015, where each one is observed at 24 equidistant time points (e.g., hourly). Figure 3.4 plots a typical daily observation in a summer period. Further, a visual inspection of the data reveals that the level of the electricity demand significantly changes between different seasons of the year. Therefore, the data was centered and adjusted for monthly seasonality by subtracting from each observation the corresponding monthly average. Figure 3.5 plots the seasonal monthly components for each country.

Since we treat discrete observations as realizations of continuous functions, a preliminary smoothing step is required to reconstruct the underlying functional observations. For reconstruction of the deseasonalized load demand functions we consider a basis representation in terms of fourth-order B-splines with knots placed at each observed hour. Thus, the number of employed basis functions is 24 per curve. This amount of basis functions leads inevitably to overfitting the data and we thus penalize the sum of squared errors for roughness (as measured through the squared second derivative). The optimal choice of the smoothing parameter λ can be determined through minimizing a generalized cross-validation criterion (GCV). The FDA package offered by Ramsay et al. (2009) for the Matlab was applied here.[3]

We start with the report on the estimation of the functional principal components. For each country the first three principle components combined account for more than 90% of the total variation in the sample. Figure 3.6 plots the eigenfunctions and their respective percentages. Further, an analysis of the estimated score series provides evidence of the time dependencies for each sample. In particular, we verify the presence of the dependencies by looking at autocovariances and partial autocovariances of the score series. Figure 3.7 illustrates our findings for the first FPC score series.

We apply FAAR, FAR, iFAR, MSM, VAR models and benchmark models such as the naive prediction and the mean prediction to obtain forecasts for the deseasonalized electric load demand functions. The original sample is split into two parts. The first one from January 1, 2013 till December 31, 2014 is reserved for the estimation and learning purposes and the second for the evaluation of the one step ahead forecast performance. Finally,

[2] http://www.nordpoolspot.com/historical-market-data/
[3] http://www.psych.mcgill.ca/misc/fda/downloads/FDAfuns/

MSE and MME given in (3.24) and (3.24), respectively, are used for the comparison of the quality of the competing procedures. The number of principal components and lags is selected according to the selection criteria suggested in Aue et al. (2015). Further, more attention is paid to choosing the number of neighbors for the predictor in the FAAR model. More precisely, we forecast the last observation in the estimating part of the sample using (3.18)-3.19 with different values of $k_N = 1, ..., N^{4/5}$. Then the number k_N is selected to minimize the MSE between the obtained predictors and the last observation.

The results are reported in Figure 3.7 in the form of boxplots of the MSE and the MME errors. In general the MSM model is the best framework for forecasting electricity demand in Nordic countries except Denmark. In the case of Denmark the FAAR model provides forecasts with smaller errors when compared to MSM and for other cases is a runner-up. This finding indicates that there is a nonlinear response of the FPC score series to the past observations. This statement is also supported by the evidence from scatter plots illustrated in Figure 3.9. The bold lines show the best polynomial fit of order 3. In all countries but Denmark we can see that the relationship between the current first FPC score value and its lag is linear. Finally, FAR, iFAR and VAR models deliver equally good results and in general are able to outperform the naive predictors.

3.8 Conclusion

In this paper a time dependence concept for functional observations is proposed. It is based on the idea of the Karhunen-Loève decomposition of functional observations which gives us the vector valued time series of FPC scores. In particular, time dependence in FTS is quantified through the autocovariances and cumulants of its FPC scores series. To operate with this concept in practice we show that the estimates of the FPCs are consistent under the described dependencies. Further, two forecasting techniques for functional time series are discussed. The first one is the FAR model for processes that have a linear relation with the past observations. We then extend this linear framework using the functional additive approach suggested in Müller and Yao (2008) and offer a simple forecasting technique based on the kNN approach. Asymptotic consistency is derived. Further our simulations indicate that the loss of efficiency against the FAR model when the true underlying DGP is linear is only marginal.

A Appendix: Auxiliary results

To economize notations we use $\sum_{i,j=1}^{N}$ and $\sum_{i\neq j=1}^{N}$ instead of full expressions $\sum_{i=1}^{N}\sum_{j=1}^{N}$ and $\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$ throughout this appendix. Further, the following combinatorial representation of p-th order moments in terms of joint cumulants is often used for proofs and is stated here for future reference. For a set of random variables $x_1,\dots,x_p$ one has

$$\mathbb{E}\left[x_1\cdot\ldots\cdot x_p\right]=\sum_{\pi}\prod_{B\in\pi}\kappa(x_i{:}i\in B), \tag{A.1}$$

were π cycles through all possible partitions of the set $\{1,2,\dots,p\}$ and B cycles through all blocks of partition π. For instance, zero mean random variables satisfies the following expressions: $\kappa(x_1,x_2)=\mathbb{E}\left[x_1,x_2\right]$ for $p=2$, $\kappa(x_1,x_2,x_3)=\mathbb{E}\left[x_1,x_2,x_3\right]$ for $p=3$ and

$$\begin{aligned}\kappa(x_1,x_2,x_3,x_4) &= \mathbb{E}\left[x_1,x_2,x_3,x_4\right]-\mathbb{E}\left[x_1,x_2\right]\mathbb{E}[x_3,x_4]\\ &- \mathbb{E}\left[x_1,x_3\right]\mathbb{E}[x_2,x_4]-\mathbb{E}\left[x_1,x_4\right]\mathbb{E}[x_2,x_3]\,.\end{aligned}$$

To facilitate understanding of the following proofs we collect intermediate steps into auxiliary Lemmas.

Lemma A.1 *Let a weakly stationary FTS $\{X_i\}_{i=1}^{N}$ satisfies Assumption 3′ with $q=4$ then*

$$\sup_{\ell\geq 1}\left|\hat{\lambda}_\ell-\lambda_\ell\right| \leq \left\|\widehat{C}_N-C\right\|_{\mathcal{L}}, \tag{A.2}$$

$$\left\|c_\ell\widehat{\psi}_\ell-\psi_\ell\right\| \leq C\delta_\ell\left\|\widehat{C}_N-C\right\|_{\mathcal{L}},\ \textit{for } 2\leq\ell\leq L \tag{A.3}$$

where $c_\ell=sign\left(\langle\widehat{\psi}_\ell,\psi_\ell\rangle\right)$, $\delta_\ell=\max_{1\leq k\leq\ell}\left(\lambda_k-\lambda_{k+1}\right)^{-1}$, and C is some positive constant.

Proof. Both results (A.2) and (A.3) follow from Bosq (2000, Lemma 4.2 and 4.3), respectively. ■

Lemma A.2 *A FAR process (3.8)-(3.9) satisfies Assumption 3″* (**i**) *and* (**iii**) *with joint cumulants up to order 4, and condition (3.15) then:*

- (**i**) $\frac{1}{N}\sum_{i=1}^{N}\|X_i\|^2=\sum_{\ell=1}^{\infty}\lambda_\ell+O_p\left(N^{-1/2}\right)$;
- (**ii**) $\widehat{\lambda}_L^{-1}=O_p\left(L^{\alpha}\right)$ *as* $N\to\infty$, $L\to\infty$ *and* $\frac{L^{\alpha}}{N^{1/2}}\to 0$;
- (**iii**) $\left\|\widehat{\Gamma}_{1,N}\right\|_{\mathcal{L}}=O_p(1)$;
- (**iv**) $\left\|\widehat{\Gamma}_{1,N}\left(\widehat{\psi}_\ell\right)\right\|\leq 2\widehat{\lambda}_\ell^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\|X_i\|^2\right)^{1/2}$;
- (**v**) $\sum_{\ell=L}^{\infty}\left\|\rho\left(\widehat{\psi}_\ell\right)\right\|^2=O_p\left(\max\left\{\frac{L^{2+\alpha}}{N^{1/2}},L^{1+2(\alpha-\gamma)}\right\}\right)$;

Proof.
Proof of item (i): To establish item (**i**) we show that $\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}\|X_i\|^2-\sum_{\ell=1}^{\infty}\lambda_\ell\right|^2=$

$O\left(N^{-1}\right)$ and then by Chebyshev inequality (**i**) will follow. First, notice that $\frac{1}{N}\sum_{i=1}^{N}\|X_i\| = \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell=1}^{\infty}\theta_{i,\ell}^2$, and denote $Z_i = \sum_{\ell=1}^{\infty}\theta_{i,\ell}^2$, $\overline{Z}_N = \frac{1}{N}\sum_{i=1}^{N} Z_i$ and $m = \sum_{\ell=1}^{\infty}\lambda_\ell$. Then

$$
\begin{aligned}
Var\left(\overline{Z}_N\right) &= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}^2\theta_{j,s}^2\right] - m^2 \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell,s=1}^{\infty}\left(\kappa_{\ell,\ell,s,s(0,0,|i-j|,|i-j|)} + 2\mathbb{E}\left[\theta_{i,\ell}\theta_{j,s}\right]^2\right),
\end{aligned}
$$

where the last equality comes from relation (A.1). For the first term by Assumption 3″(**iii**) we have

$$
\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell,s=1}^{\infty}\kappa_{\ell,\ell,s,s(0,0,|i-j|,|i-j|)} \leq \frac{B}{N^2}\sum_{i=1}^{N}\sum_{\ell,s=1}^{\infty}\lambda_\ell\lambda_s = O\left(N^{-1}\right),
$$

and for the second

$$
\begin{aligned}
\frac{2}{N^2}\sum_{i,j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,s}\right]^2 &= \frac{2}{N^2}\sum_{i\neq j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,s}\right]^2 + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 \\
&\leq \frac{4}{N^2}\sum_{h=1}^{N-1}\sum_{i=h+1}^{N}\sum_{\ell,s=1}^{\infty}\left(B_{\ell,s}^{(h)}\right)^2 + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 \\
&\leq \frac{B}{N}\sum_{h=1}^{N-1}h^{-2\beta}\sum_{\ell,s=1}^{\infty}\ell^{-\gamma}s^{-\gamma} + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 = O\left(N^{-1}\right),
\end{aligned}
$$

where the last result comes from Assumption 3″ (**i**) and (**iii**) and condition 3.15.
Proof of item (**ii**): It follows immediately from Corollary 4 and Chebyshev inequality $\widehat{\lambda}_\ell = O_p\left(\max\left\{L^{-\alpha}, N^{-1/2}\right\}\right)$ and $\widehat{\lambda}_\ell^{-1} = O_p\left(\frac{1}{\max\left\{L^{-\alpha}, N^{-1/2}\right\}}\right)$. The item (**ii**) will follow from the fact $N^{-1/2}$ will go to zero faster then $L^{-\alpha}$ since $L^\alpha/N^{1/2} \to 0$.
Proof of item (**iii**): Follows from Corollary 3 and Chebyshev inequality.
Proof of item (**iv**): Follows from Lemma 8.3 in Bosq (2000).
Proof of item (**v**): Item (**v**) is obtained by using the proof from Lemma 8.2 in Bosq (2000) and the facts that $\left\|\widehat{C}_N - C\right\|_{\mathcal{L}} = O_p(N^{-1/2})$, $\sum_{\ell=1}^{L}\delta_\ell = O(L^{2+\alpha})$ and $\sum_{\ell=L}^{\infty}\|\rho\left(\psi_\ell\right)\|^2 = O\left(L^{1+2(\alpha-\gamma)}\right)$ ■

B Appendix: Proofs

Proof of Remark 13

We have

$$\begin{aligned}
\mathbb{E}\left\|\hat{\mu}-\mu\right\|^{2} &= \frac{1}{N^{2}}\sum_{i,j=1}^{N}\mathbb{E}\left\langle X_{i}-\mu,X_{j}-\mu\right\rangle=\frac{1}{N^{2}}\sum_{i,j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell},\theta_{j,s}\right] \\
&= \frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{\ell=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell},\theta_{i,\ell}\right]+\frac{1}{N^{2}}\sum_{i\neq j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell},\theta_{j,s}\right].
\end{aligned}$$

As a consequence of Assumption 3″ part (**i**) $\sum_{\ell=1}^{\infty}\lambda_{\ell}<\infty$ such that the first term in the last equation above behaves as $O\left(N^{-1}\right)$. Rearranging the second term and invoking Assumption 3″ (**ii**) gives

$$\begin{aligned}
\frac{1}{N^{2}}\sum_{i\neq j=1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell},\theta_{j,s}\right] &= \frac{2}{N^{2}}\sum_{h=1}^{N-1}\sum_{i=h+1}^{N}\sum_{\ell,s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell},\theta_{j,s}\right] \\
&\leq \frac{2}{N^{2}}\sum_{h=1}^{N-1}\sum_{i=h+1}^{N}\sum_{\ell,s=1}^{\infty}B_{\ell,s}^{(h)} \\
&\leq \frac{C}{N^{2}}\sum_{h=1}^{N-1}(N-h)h^{-\beta}\sum_{\ell,s=1}^{\infty}\sqrt{\lambda_{\ell}\lambda_{s}}=O\left(\max\left\{N^{-\beta},N^{-1}\right\}\right).
\end{aligned}$$

The last equality uses Davidson (1994, Theorem 2.27) and the fact that $\sum_{\ell=1}^{\infty}\sqrt{\lambda_{\ell}}<\infty$ which follows from Assumption 3″.

Proof of Theorem 9

We have,

$$\begin{aligned}
\mathbb{E}\left\|\widehat{C}_{N}-C\right\|_{\mathcal{S}}^{2} &= \sum_{\ell=1}^{\infty}\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\left\langle X_{i},\psi_{\ell}\right\rangle X_{i}-\mathbb{E}\left[\left\langle X_{i},\psi_{\ell}\right\rangle X_{i}\right]\right)\right\|^{2} \\
&= \frac{1}{N^{2}}\sum_{i,j=1}^{N}\sum_{\ell=1}^{\infty}\left(\sum_{s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\theta_{i,s}\theta_{j,s}\right]-\lambda_{\ell}^{2}\right) \qquad \text{(A.4)} \\
&= \frac{1}{N^{2}}\sum_{i,j=1}^{N}\sum_{\ell=1}^{\infty}\left(\mathbb{E}\left[\theta_{i,\ell}^{2}\theta_{j,\ell}^{2}\right]-\lambda_{\ell}^{2}\right) \\
&+ \frac{1}{N^{2}}\sum_{i,j=1}^{N}\sum_{\ell\neq s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\theta_{i,s}\theta_{j,s}\right]:=a+b. \qquad \text{(A.5)}
\end{aligned}$$

It follows from relation (A.1) that

$$a=\frac{1}{N^{2}}\sum_{i,j=1}^{N}\sum_{\ell=1}^{\infty}\left(\kappa_{\ell,\ell,\ell,\ell(0,0,|i-j|,|i-j|)}+2\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\right]^{2}\right),$$

where $\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell=1}^{\infty}\kappa_{\ell,\ell,\ell,\ell}(0,0,|i-j|,|i-j|) = O\left(N^{-1}\right)$ by Assumption 3″(**iii**) and

$$\begin{aligned}
\frac{2}{N^2}\sum_{i,j=1}^{N}\sum_{\ell=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\right]^2 &= \frac{2}{N^2}\sum_{i\neq j=1}^{N}\sum_{\ell=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\right]^2 + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 \\
&\leq \frac{4}{N^2}\sum_{h=1}^{N-1}\sum_{i=h+1}^{N}\sum_{\ell=1}^{\infty}\left(B_{\ell,\ell}^{(h)}\right)^2 + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 \\
&\leq \frac{B}{N}\sum_{h=1}^{N-1}h^{-2\beta}\sum_{\ell=1}^{\infty}\lambda_\ell^2 + \frac{2}{N}\sum_{\ell=1}^{\infty}\lambda_\ell^2 \\
&= O\left(\max\left\{N^{-2\beta}, N^{-1}\right\}\right),
\end{aligned}$$

where the last equality comes from Assumption 3″(**i**) and (**ii**).

Similar arguments apply to term b, i.e.,

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell\neq s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\theta_{i,s}\theta_{j,s}\right] = \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell\neq s=1}^{\infty}\left(\kappa_{\ell,\ell,s,s}(0,0,|i-j|,|i-j|)+\right. \tag{A.6}$$

$$\left. + \mathbb{E}\left[\theta_{i,\ell}\theta_{j,\ell}\right]\mathbb{E}\left[\theta_{i,s}\theta_{j,s}\right] + \mathbb{E}\left[\theta_{i,\ell}\theta_{j,s}\right]\mathbb{E}\left[\theta_{i,s}\theta_{j,\ell}\right]\right) \tag{A.7}$$

by relation (A.1). The first terms on the r.h.s of (A.6) is $O\left(N^{-1}\right)$ by Assumption 3″(**iii**). The second and the third terms on the r.h.s of (A.6) are $O\left(\max\{N^{-2\beta}, N^{-1}\}\right)$ by the same arguments as above. In particular, for the third term we have

$$\begin{aligned}
\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell\neq s=1}^{\infty}\mathbb{E}\left[\theta_{i,\ell}\theta_{j,s}\right]\mathbb{E}\left[\theta_{i,s}\theta_{j,\ell}\right] &\leq \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{\ell\neq s=1}^{\infty}\left(B_{\ell,s}^{(i-j)}\right)^2 = \frac{2}{N^2}\sum_{h=1}^{N-1}\sum_{i=h+1}^{N}\sum_{\ell\neq s=1}^{\infty}\left(B_{\ell,s}^{(h)}\right)^2 \\
&\leq \frac{B}{N}\sum_{h=1}^{N-1}h^{-2\beta}\sum_{\ell\neq s=1}^{\infty}\lambda_\ell\lambda_s = O\left(\max\left\{N^{-2\beta}, N^{-1}\right\}\right).
\end{aligned}$$

Putting together rates for a and b yields the statement of the theorem.

Proof of Theorem 10

Recall that $H_L = span\{\psi_1, ..., \psi_L\}$ and let $\widehat{H}_L = span\{\widehat{\psi}_1, ..., \widehat{\psi}_L\}$ and denote π_L and $\widehat{\pi}_L$ projections on H_L and $\widehat{H}_L$, respectively. Then we can consider the following decomposition

$$\begin{aligned}
\left(\widehat{\rho}_N - \rho\right)(x) &= \left(\widehat{\rho}_N - \rho\pi_L(x)\right) + \left(\rho\pi_L(x) - \rho\widehat{\pi}_L(x)\right) + \left(\rho\widehat{\pi}_L(x) - \rho(x)\right) \\
&:= a_N(x) + b_N(x) + c_N(x).
\end{aligned}$$

Further, denote $a_N(x) = \sum_{k=1}^4 a_{k,N}(x)$, where

$$\begin{aligned}
a_{1,N}(x) &= \widehat{\Gamma}_{1,N}\left(\sum_{\ell=1}^{L}\left(\widehat{\lambda}_\ell^{-1} - \lambda_\ell^{-1}\right)\langle x, \widehat{\psi}_\ell\rangle \widehat{\psi}_\ell\right), \\
a_{2,N}(x) &= \widehat{\Gamma}_{1,N}\left(\sum_{\ell=1}^{L}\lambda_\ell^{-1}\left(\langle x, \widehat{\psi}_\ell\rangle - \langle x, \psi'_\ell\rangle\right)\widehat{\psi}_\ell\right), \\
a_{3,N}(x) &= \widehat{\Gamma}_{1,N}\left(\sum_{\ell=1}^{L}\lambda_\ell^{-1}\langle x, \psi'_\ell\rangle\left(\widehat{\psi}_\ell - \psi'_\ell\right)\right), \\
a_{4,N}(x) &= \left(\widehat{\Gamma}_{1,N} - \Gamma\right)\left(\sum_{\ell=1}^{L}\lambda_\ell^{-1}\langle x, \psi'_\ell\rangle\psi'_\ell\right).
\end{aligned}$$

For the first term we have

$$\|a_{N,1}(x)\| \leq \sum_{\ell=1}^{L}\frac{|\widehat{\lambda}_\ell - \lambda_\ell|}{\widehat{\lambda}_\ell\lambda_\ell}|\langle x, \widehat{\psi}_\ell\rangle|\left\|\widehat{\Gamma}_{1,N}(\widehat{\psi}_\ell)\right\|.$$

Using (A.2), Cauchy-Schwartz inequality and item (**iv**) of Lemma A.2 we obtain

$$\|a_{N,1}\|_{\mathcal{L}} \leq 2\left(\frac{1}{N}\sum_{i=1}^{N}\|X_i\|^2\right)^{1/2}\|C_N - C\|_{\mathcal{L}}\left(\sum_{\ell=1}^{L}\widehat{\lambda}_\ell^{-1/2}\lambda_\ell^{-1}\right).$$

From Theorem 9 and Chebyshev inequality $\|C_N - C\|_{\mathcal{L}} = O_p(N^{-1/2})$. Assume for now that $L^\alpha/N^{1/2} \to 0$, then by using item (**i**) and (**ii**) of Lemma A.2 one gets

$$\|a_{N,1}\|_{\mathcal{L}} = O_p\left(\frac{L^{\frac{3}{2}\alpha+1}}{N^{1/2}}\right). \tag{A.8}$$

Finally, to archive the consistency it is required that $L^{\frac{3}{2}\alpha+1}/N^{1/2} \to 0$ which in turn implies the condition $L^\alpha/N^{1/2} \to 0$ has to hold. That is, $L^\alpha/N^{1/2} \to 0$ is necessary but not sufficient to obtain the statement of the theorem.

Turning to $a_{N,2}(x)$, from item (**iv**) of Lemma A.2 and Cauchy-Schwartz inequality we have

$$\|a_{N,2}\|_{\mathcal{L}} \leq 2\left(\frac{1}{N}\sum_{i=1}^{N}\|X_i\|^2\right)^{1/2}\sum_{\ell=1}^{L}\widehat{\lambda}_\ell^{1/2}\lambda_\ell^{-1}\left\|\widehat{\psi}_\ell - \psi_\ell\right\|,$$

where (A.3) together with and the fact that $\sum_{\ell=1}^{L}\delta_\ell = O(L^{\alpha+2})$ yield

$$\|a_{N,2}\|_{\mathcal{L}} = O_p\left(\frac{L^{\frac{3}{2}\alpha+2}}{N^{1/2}}\right). \tag{A.9}$$

Concerning $a_{N,3}(x)$, Cauchy-Schwartz inequality and orthogonality of $\widehat{\psi}_\ell$ and ψ_ℓ yield the bound

$$\|a_{N,3}\|_{\mathcal{L}} \leq \left\|\widehat{\Gamma}_{1,N}\right\|_{\mathcal{L}}\left(\sum_{\ell=1}^{L}\lambda_\ell^{-2}\langle x, \widehat{\psi}_\ell\rangle^2\left\|\widehat{\psi}_\ell - \psi_\ell\right\|^2\right)^{1/2}.$$

Then using item (**iii**) of Lemma A.2 and the fact that $\left(\sum_{\ell=1}^{L} \sigma_\ell^2\right)^{1/2} = O(L^{\alpha+3/2})$ yield

$$\|a_{N,3}\|_{\mathcal{L}} = O_p\left(\frac{L^{2\alpha+\frac{3}{2}}}{N^{1/2}}\right). \tag{A.10}$$

Finally,

$$\|a_{N,4}\|_{\mathcal{L}} = \left\|\widehat{\Gamma}_{1,N} - \Gamma\right\|_{\mathcal{L}} \left(\sum_{\ell=1}^{L} \lambda_\ell^{-2}\langle x, \psi_\ell\rangle^2\right)^{1/2}.$$

Then Corollary 3 entail

$$\|a_{N,4}\|_{\mathcal{L}} = O_p\left(\frac{L^{\alpha+\frac{1}{2}}}{N^{1/2}}\right). \tag{A.11}$$

Next we turn to $b_N(x)$ and $c_N(x)$. First observe that

$$\|b_N\|_{\mathcal{L}} \leq C\left(\sum_{\ell=L}^{\infty}\left\|\rho\left(\widehat{\psi}_\ell\right)\right\|^2 + \sum_{\ell=L}^{\infty}\|\rho\left(\psi_\ell\right)\|^2\right). \tag{A.12}$$

which behave as $O_p\left(\max\left\{\frac{L^{2+\alpha}}{N^{1/2}}, L^{1+2(\alpha-\gamma)}\right\}\right)$ by item (**v**) of Lemma A.2. For $c_N(x)$ we have $\|c_N\|_{\mathcal{L}} = \sum_{\ell=L}^{\infty}\|\rho\left(\psi_\ell\right)\|^2 = O_p\left(L^{1+2(\alpha-\gamma)}\right)$ and statement of the theorem is proofed.

Proof of Theorem 11

First, define $M_L(X_{N,L}) := \sum_{s,\ell=1}^{L} \mathbb{E}\left[\theta_{N+1,s}|\theta_{N,\ell}\right]\psi_s = \sum_{s,\ell=1}^{L} m_{\ell,s}(\theta_{N,\ell})\psi_s$, where in comparison to $M_{N,L}(x_L)$ the k_N-NN estimators of the scores have been replaced by the corresponding conditional population means. Since our interest is in analyzing $\mathbb{E}\|M_{N,L}(X_{N,L}) - M(X_N)\|^2$, it suffices, upon adding and subtracting $M_L(X_{N,L})$ in the argument of our object of interest, to consider the two terms

$$\mathbb{E}\|M_L(X_{N,L}) - M(X_N)\|^2 \text{ and } \mathbb{E}\|M_{N,L}(X_{N,L}) - M_L(X_{N,L})\|^2 \tag{A.13}$$

For simplicity of notation let θ_ℓ denote $\theta_{N,\ell}$. Then for the first term in (A.13) by using the orthonormality of the $\{\psi_\ell\}$ we have

$$\mathbb{E}\|M_L(X_{N,L}) - M(X_N)\|^2 = \mathbb{E}\left\|\sum_{s,\ell=1}^{L} m_{\ell,s}(\theta_\ell)\psi_\ell - \sum_{s,\ell=1}^{\infty} m_{\ell,s}(\theta_\ell)\psi_\ell\right\|^2$$

$$= \sum_{s,\ell=L+1}^{\infty} \mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right] + \sum_{s=L+1}^{\infty}\sum_{\ell=1}^{L} \mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right] + \sum_{s=1}^{L}\sum_{\ell=L+1}^{\infty} \mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right]. \tag{A.14}$$

Now observe that from $L^{\alpha-1}\sum_{\ell=L}^{\infty}\mathbb{E}\left[m_{\ell,s}^2(\theta_{i,\ell})\right] = O(\lambda_s)$ it follows immediately that $\sum_{s,\ell=L+1}^{\infty}\mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right] = O(L^{2(1-\alpha)})$, $\sum_{s=L+1}^{\infty}\sum_{\ell=1}^{L}\mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right] = O(L^{1-\alpha})$ and $\sum_{s=1}^{L}\sum_{\ell=L+1}^{\infty}\mathbb{E}\left[m_{\ell,s}(\theta_\ell)^2\right] = O(L^{1-\alpha})$

Now we consider the second term in (A.13) which can be written as

$$\mathbb{E}\|M_{N,L}(X_{N,L}) - M_L(X_{N,L})\|^2 = \mathbb{E}\left\|\sum_{s,\ell=1}^{L}\left(\widetilde{m}_{\ell,s}(\theta_l) - m_{\ell,s}(\theta_l)\right)\psi_l\right\|^2$$

$$= \sum_{s,\ell=1}^{L}\mathbb{E}\left[\left(\widetilde{m}_{\ell,s}(\theta_\ell) - m_{\ell,s}(\theta_\ell)\right)^2\right], \tag{A.15}$$

where the second equality follows from the orthonormality of the sequence of eigenfunctions $(\psi_\ell)_{\ell=1}^{L}$. For fixed $\ell = 1,\ldots,L$, rates of convergence of the mean squared error in (A.15) can be derived by following results in Yakowitz (1987). A careful inspection of the proofs in Yakowitz (1987) reveals that analyzing the second moment of the distance between (the given) θ_l and its farthest (of the k_N) neighbor is of key importance. Denote this farthest neighbor to θ_l by $\theta_{N(k_N),l}$ and write $R_{i,l}(\theta_l) := |\theta_{i,l} - \theta_l|$ such that $R_{(k_N),l}(\theta_l) := |\theta_{N(k_N),l} - \theta_l|$ denotes the k_N-th order statistic of the $R_{i,l}(\theta_l)$. Results in Yakowitz (1987) indicate that $\mathbb{E}[R_{(k_N),l}(\theta_l)^2] \leq C_1(l)k_N^{-1/2}$, where $C_1(l)$ is some constant that depends only on l. While this holds true for fixed l, we have to consider asymptotics where L goes to infinity. Now observe that

$$\mathbb{E}\left[R_{(k_N),l}(\theta_l)^2\right] = \mathbb{E}\left[|\theta_{N(k_N),l} - \theta_l|^2\right] \leq C_2(N)\lambda_l$$

for fixed N, where $C_2(N)$ is some constant only depending on N. Combining these results gives us $\mathbb{E}[R_{(k_N),l}(\theta_l)^2] \leq C_3 k_N^{-1/2}\lambda_l$, where now C_3 is a constant that is independent of both l and N. Moreover, Yakowitz (1987) shows that the number of neighbors k_N has to grow with the sample size where $k_N \sim \lfloor N^{4/5}\rfloor$.

The desired result now follows from the Theorem 2.1 Yakowitz (1987) and the arguments presented above.

Proof of Theorem 12

Denote, for $i = 1, \dots, k_N$, by $N(i) \in \mathcal{I}(k_N; \theta_\ell)$ the index of the i-th nearest neighbor to θ_ℓ. Then upon adding and subtracting $\sum_{\ell,s=1}^{L} \widetilde{m}_{\ell,s}(\theta_\ell)\widehat{\psi}_s$ to the argument of $\mathbb{E}\left\|\widehat{M}_{N,L}(\hat{x}_L) - M_{N,L}(x_L)\right\|^2$ it suffices to analyze the quantities

$$\mathbb{E}\left\|\sum_{\ell,s=1}^{L} \widetilde{m}_{\ell,s}(\theta_\ell)\left(\widehat{\psi}_s - \psi_s\right)\right\|^2 \quad \text{and} \quad \mathbb{E}\left\|\sum_{\ell,s=1}^{L} \left(\widehat{m}_{\ell,s}(\hat{\theta}_\ell) - \widetilde{m}_{\ell,s}(\theta_\ell)\right)\widehat{\psi}_s\right\|^2.$$

For the first term we have

$$\begin{aligned}
&\mathbb{E}\quad\left\|\sum_{\ell,s=1}^{L} \widetilde{m}_{\ell,s}(\theta_\ell)\left(\widehat{\psi}_s - \psi_s\right)\right\|^2 = \mathbb{E}\left[\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L} \widetilde{m}_{\ell,s}(\theta_\ell)\widetilde{m}_{k,\tau}(\theta_k)\left\langle \widehat{\psi}_s - \psi_s, \widehat{\psi}_\tau - \psi_\tau\right\rangle\right] \\
&\leq \mathbb{E}\left[\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L} \widetilde{m}_{\ell,s}(\theta_\ell)\widetilde{m}_{k,\tau}(\theta_k)\left\|\widehat{\psi}_s - \psi_s\right\|\left\|\widehat{\psi}_\tau - \psi_\tau\right\|\right] \\
&\leq \frac{1}{k_N^2}\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L}\sum_{i,j=1}^{k_N} \mathbb{E}\left[\theta_{N(i)+1,\ell}\theta_{N(j)+1,k}\delta_s\delta_\tau\left\|\widehat{C}_N - C\right\|_{\mathcal{S}}^2\right], \qquad \text{(A.16)}
\end{aligned}$$

where the last inequality follows from Lemma A.1. As already discussed in the proof of Theorem 9 we have

$$\begin{aligned}
\left\|\widehat{C}_N - C\right\|_{\mathcal{S}}^2 &= \frac{1}{N^2}\sum_{n,m=1}^{N}\left(\sum_{h_1,h_2=1}^{\infty} \theta_{n,h_1}\theta_{n,h_2}\theta_{m,h_1}\theta_{m,h_2}\right. \\
&\quad \left. + \sum_{h_1=1}^{\infty}\lambda_{h_1}^2 - \sum_{h_1=1}^{\infty}\lambda_{h_1}\theta_{n,h_1}^2 - \sum_{h_1=1}^{\infty}\lambda_{h_1}\theta_{m,h_1}^2\right).
\end{aligned}$$

Thus the expression in (A.16) can be rewritten as

$$\frac{1}{k_N^2}\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L}\sum_{i,j=1}^{k_N} \mathbb{E}\left[\theta_{N(i)+1,\ell}\theta_{N(j)+1,k}\delta_s\delta_\tau\left\|\widehat{C}_N - C\right\|_{\mathcal{S}}^2\right] = A_1 + A_2 - 2A_3,$$

where

$$\begin{aligned}
A_1 &:= \frac{1}{k_N^2 N^2}\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\sum_{h_1,h_2=1}^{\infty} \delta_s\delta_\tau\mathbb{E}\left[\theta_{N(i)+1,\ell}\theta_{N(j)+1,k}\theta_{n,h_1}\theta_{n,h_2}\theta_{m,h_1}\theta_{m,h_2}\right], \\
A_2 &:= \frac{1}{k_N^2 N^2}\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\sum_{h_1=1}^{\infty} \delta_s\delta_\tau\lambda_{h_1}^2\mathbb{E}\left[\theta_{N(i)+1,\ell}\theta_{N(j)+1,k}\right], \\
A_3 &:= \frac{1}{k_N^2 N^2}\sum_{\ell,s=1}^{L}\sum_{k,\tau=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\sum_{h_1=1}^{\infty} \delta_s\delta_\tau\lambda_{h_1}\mathbb{E}\left[\theta_{N(i)+1,\ell}\theta_{N(j)+1,k}\theta_{n,h_1}^2\right].
\end{aligned}$$

The analysis of the terms above now proceeds by considering the relationship between higher order moments and joint cumulants as defined in (A.1) and noting that the random variables $\theta_{\cdot,h} = \langle X_\cdot, \psi_h\rangle$ have zero mean by construction and are independent across h by assumption.

We start with term A_2. The relevant case for us to consider is $\ell = k$ as otherwise $A_2 = 0$ by the above arguments. Distinguishing the cases where $\ell \neq h_1$ and $\ell = h_1$ then yields

$$\begin{aligned} A_2 &= \frac{1}{k_N^2 N^2} \sum_{\ell,s=1}^{L} \sum_{\tau=1}^{L} \sum_{i,j=1}^{k_N} \sum_{n,m=1}^{N} \sum_{h_1\neq\ell=1}^{\infty} \delta_s\delta_\tau\lambda_{h_1}^2 \kappa_{\ell,\ell(0,|N(i)-N(j)|)} \\ &+ \frac{1}{k_N^2 N^2} \sum_{\ell,s=1}^{L} \sum_{\tau=1}^{L} \sum_{i,j=1}^{k_N} \sum_{n,m=1}^{N} \delta_s\delta_\tau\lambda_\ell^2 \kappa_{\ell,\ell(0,|N(i)-N(j)|)} \\ &=: A_{2,1} + A_{2,2}. \end{aligned} \tag{A.17}$$

Now consider the term A_3 and again note that it suffices to consider only the case $\ell = k$. Again distinguishing the cases where $\ell \neq h_1$ and $\ell = h_1$ we have by (A.1) that

$$\begin{aligned} A_3 &= \tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \lambda_{h_1}^2 \kappa_{l(0,|N(i)-N(j)|)} \\ &+\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \delta_l^2 \lambda_l \kappa_{\ell,\ell(0,|N(i)-N(j)|,|N(i)+1-n|,|N(i)+1-n|)} \\ &+\tfrac{2}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \delta_l^2 \lambda_l \kappa_{l(0,|N(i)+1-n|)}\kappa_{l(0,|N(j)+1-n|)} \\ &+\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \delta_l^2 \lambda_l \kappa_{l(0,|N(i)-N(j)|)}\kappa_{l(0,0)} \\ &=: A_{3,1} + A_{3,2} + A_{3,3} + A_{3,4}. \end{aligned} \tag{A.18}$$

Note that the term A_3 enters the object of interest twice with a negative sign, such that all terms of which A_2 is comprised are canceled in view of $A_{2,1} = A_{3,1}$ and $A_{2,2} = A_{3,4}$ and since $\kappa_{l(0,0)=\lambda_l}$.

We now tun to term A_1 and first decompose into the cases where $h_1 \neq h_2$ and $h_1 = h_2$. The second case is furthermore decomposed into cases where $l = k$ and $l \neq k$. This yields

$$\begin{aligned} A_1 &= \tfrac{1}{k_N^2 N^2} \textstyle\sum\sum_{l,k=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum\sum_{h_1\neq h_2}^{\infty} \delta_l\delta_k \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,k}\theta_{n,h_1}\theta_{n,h_2}\theta_{m,h_1}\theta_{m,h_2}\right] \\ &+\tfrac{1}{k_N^2 N^2} \textstyle\sum\sum_{l\neq k}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=1}^{\infty} \delta_l\delta_k \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,k}\theta_{n,h_1}^2\theta_{m,h_1}^2\right] \\ &+\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,k}\theta_{n,h_1}^2\theta_{m,h_1}^2\right] \\ &=: A_{1,1} + A_{1,2} + A_{1,3}. \end{aligned} \tag{A.19}$$

Now note that $A_{1,2} = 0$ by the same arguments as above. For term $A_{1,3}$, we decompose into the cases where $l \neq h_1$ and $l = h_1$ which yields

$$\begin{aligned} A_{1,3} &= \tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\theta_{n,l}^2\theta_{m,l}^2\right] \\ &+\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \mathbb{E}\left[\theta_{n,h_1}^2\theta_{m,h_1}^2\right] \end{aligned} \tag{A.20}$$

We consider first the first term of (A.20). By (A.1) and writing, with some abuse of notation, $\kappa^{(p)}$ for the p-th order cumulant, we have

$$\begin{aligned}&\mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\theta^2_{n,l}\theta^2_{m,l}\right]\\&=\kappa_l^{(6)}+15\kappa_l^{(4)}\kappa_l^{(2)}+10\kappa_l^{(3)}\kappa_l^{(3)}+15\kappa_l^{(2)}\kappa_l^{(2)}\kappa_l^{(2)}.\end{aligned}$$

There are 15 instances of $\kappa_l^{(2)}$ which are of the form

$$\begin{aligned}&1\times\kappa_{l(0,|N(i)-N(j)|)}\\&2\times\kappa_{l(0,|N(i)+1-n|)}\\&2\times\kappa_{l(0,|N(i)+1-m|)}\\&2\times\kappa_{l(|N(i)-N(j)|,|N(i)+1-n|)}\\&2\times\kappa_{l(|N(i)-N(j)|,|N(i)+1-m|)}\\&4\times\kappa_{l(|N(i)+1-n|,|N(i)+1-m|)}\\&1\times\kappa_{l(|N(i)+1-n|,|N(i)+1-n|)}\\&1\times\kappa_{l(|N(i)+1-m|,|N(i)+1-m|)}\end{aligned}$$

Now note that there are precisely four instances where $\kappa_l^{(2)}$ is such that the first term in (A.20) takes the form

$$\frac{1}{k_N^2N^2}\sum_{l=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\delta_l^2\lambda_l\kappa_{l(0,|N(i)+1-n|)}\kappa_{l(0,|N(j)+1-n|)}$$

and precisely one instance where $\kappa_l^{(2)}$ is such that the first term in (A.20) takes the form

$$\frac{1}{k_N^2N^2}\sum_{l=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\delta_l^2\lambda_l^2\kappa_{l(0,|N(i)-N(j)|)}$$

which are canceled by $A_{3,3}$ and $A_{3,4}$, respectively, since these terms enters twice with a negative sign. By similar arguments, we have two instances in which $\kappa_l^{(4)}$ is such that the first term in (A.20) takes the form

$$\frac{1}{k_N^2N^2}\sum_{l=1}^{L}\sum_{i,j=1}^{k_N}\sum_{n,m=1}^{N}\delta_l^2\lambda_l\kappa_{l(0,|N(i)-N(j)|,|N(i)+1-n|,|N(i)+1-n|)}$$

which are canceled by $A_{3,2}$, again since that term enters twice with a negative sign. The remaining terms of the first term in (A.20) do not provide the dominant rate of convergence such that we skip the further analysis and consider next the second term in (A.20). By (A.1) we have

$$\begin{aligned}&\mathbb{E}\left[\theta^2_{n,h_1}\theta^2_{m,h_1}\right]\\&=\kappa_{h_1(0,0,|n-m|,|n-m|)}+\kappa_{h_1(0,0)}\ \kappa_{h_1(|n-m|,|n-m|)}+2\kappa_{h_1(0,|n-m|)}\ \kappa_{h_1(0,|n-m|)}\end{aligned}$$

such that we obtain for the second term of (A.20)

$$\begin{aligned}
&\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \mathbb{E}\left[\theta_{n,h_1}^2 \theta_{m,h_1}^2\right] \\
&= \tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \lambda_{h_1}^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \\
&+ \tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \kappa_{h_1}(0,0,|n-m|,|n-m|) \\
&+ 2\tfrac{1}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \kappa_{h_1}(0,|n-m|)^2.
\end{aligned}$$

Observe now that the first term in the above display is canceled by $A_{3,1}$ as it enters twice with a negative sign. As a consequence, the terms A_2, A_3 and parts of A_1 cancel each other out. The dominant rate of convergence is now obtained by considering the third term in the above display for which we have

$$\begin{aligned}
&\tfrac{2}{k_N^2 N^2} \textstyle\sum_{l=1}^{L} \sum\sum_{i,j=1}^{k_N} \sum\sum_{n,m=1}^{N} \sum_{h_1=L+1}^{\infty} \delta_l^2 \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right] \kappa_{h_1}(0,|n-m|)^2 \\
&= 2\left(\tfrac{1}{k_N N} \textstyle\sum_{l=1}^{L} \delta_l^2 \sum\sum_{i,j=1}^{k_N} \mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right]\right) \times \\
&\qquad \left(\tfrac{1}{k_N N} \textstyle\sum_{h_1=L+1}^{\infty} \sum\sum_{n,m=1}^{N} \kappa_{h_1}(0,|n-m|)^2\right). \qquad \text{(A.21)}
\end{aligned}$$

For the first term in brackets in (A.21) we have, for some constant $C > 0$,

$$\begin{aligned}
(\dots) &\le \tfrac{1}{k_N N} \textstyle\sum_{l=1}^{L} \delta_l^2 \sum_{i=1}^{k_N} \mathbb{E}\left[\theta_{N(i)+1,l}^2\right] + \tfrac{1}{k_N N} \sum_{l=1}^{L} \delta_l^2 \sum\sum_{i\neq j}^{k_N} \left|\mathbb{E}\left[\theta_{N(i)+1,l}\theta_{N(j)+1,l}\right]\right| \\
&\le \tfrac{1}{k_N N} \textstyle\sum_{l=1}^{L} \delta_l^2 \sum_{i=1}^{k_N} \lambda_l + \tfrac{2}{k_N N} \sum_{m=1}^{k_N-1} \sum_{i=m+1}^{k_N} \sum_{l=1}^{L} \delta_l^2 B_{m,l} \\
&\le \tfrac{1}{N} \textstyle\sum_{l=1}^{L} \delta_l^2 \lambda_l + \tfrac{C}{k_N N} \sum_{m=1}^{k_N-1} (k_N - m) m^{-\beta} \sum_{l=1}^{L} \delta_l^2 \lambda_l \\
&= O\left(\frac{k_N^{1-\tilde{\beta}} L^{3+\alpha}}{N}\right),
\end{aligned}$$

where the last equality follows from Assumption 3″. For the second term in brackets in (A.21) we have by similar arguments for some constants $C, C^* > 0$,

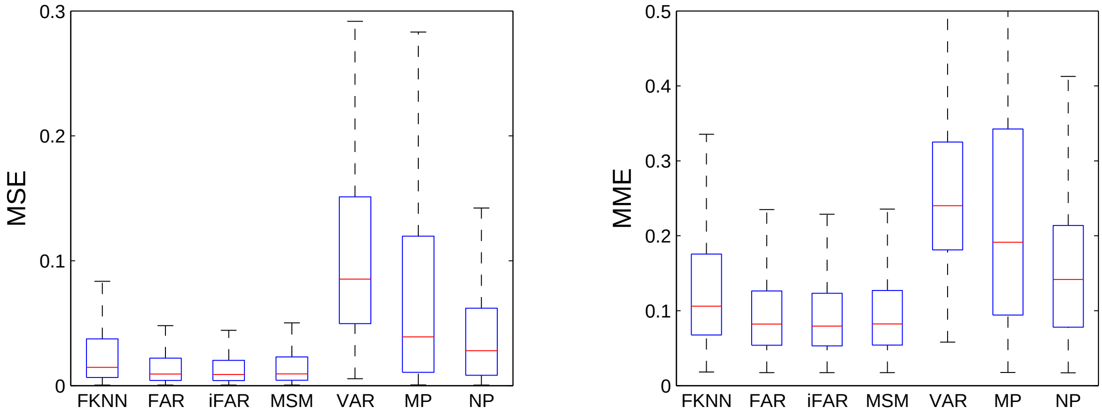
$$\begin{aligned}
(\dots) &\le \frac{1}{k_N N} \sum_{h_1=1}^{\infty} \sum\sum_{n,m=1}^{N} \mathbb{E}\left[\theta_{n,h_1}\theta_{m,h_1}\right]^2 \\
&\le \frac{1}{k_N N} \sum_{h_1=1}^{\infty} \sum_{n=1}^{N} \mathbb{E}\left[\theta_{n,h_1}^2\right]^2 + \frac{1}{k_N N} \sum_{h_1=1}^{\infty} \sum\sum_{n\neq m}^{N} \left|\mathbb{E}\left[\theta_{n,h_1}\theta_{m,h_1}\right]\right|^2 \\
&\le \frac{1}{k_N} \sum_{h_1=1}^{\infty} \lambda_{h_1}^2 + \frac{2}{k_N N} \sum_{m=1}^{N-1} \sum_{i=1}^{N} \sum_{h_1=1}^{\infty} B_{m,h_1}^2 \\
&\le \frac{C}{k_N} + \frac{C^*}{k_N N} \sum_{m=1}^{N-1} \sum_{i=1}^{N} m^{-2\beta} \sum_{h_1=1}^{\infty} \lambda_{h_1}^2 = O\left(\frac{N^{1-2\beta^*}}{k_N}\right).
\end{aligned}$$

where $\beta^* = \min\{\beta, 1/2\}$. Combining these results we obtain the following rate of convergence

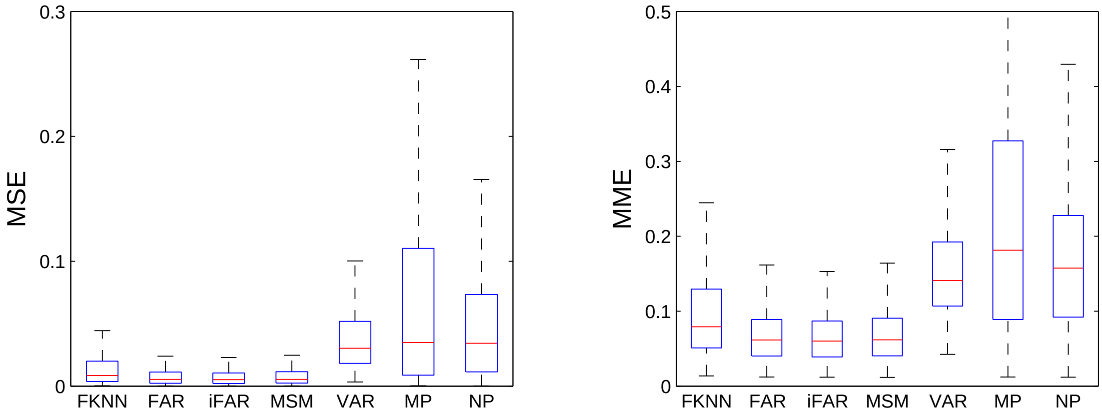
$$O\left(\frac{L^{3+\alpha}}{k_N^{\beta^*} N^{2\beta^{**}}}\right).$$

Note that we omit the analysis of term $A_{1,1}$ for brevity as it follows by the same arguments presented above and yields the same rate of convergence.

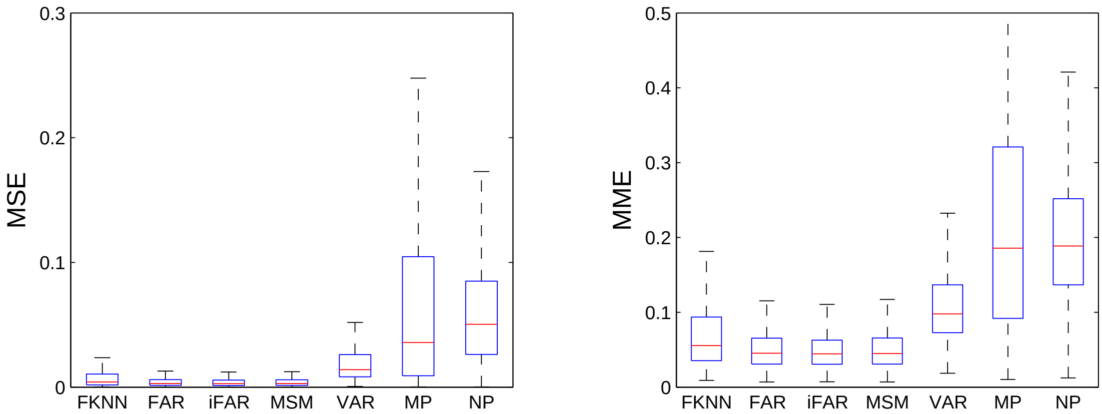
C Appendix: Figures



(a) $T = 50$

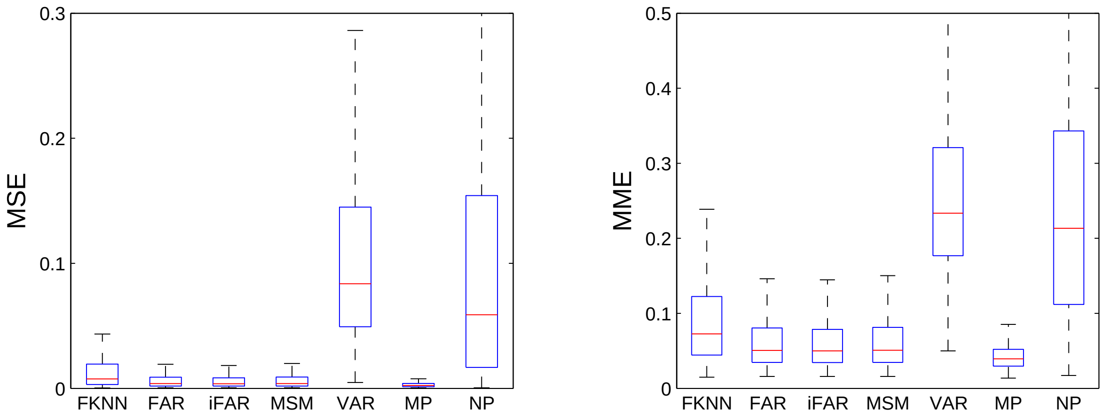


(b) $T = 100$

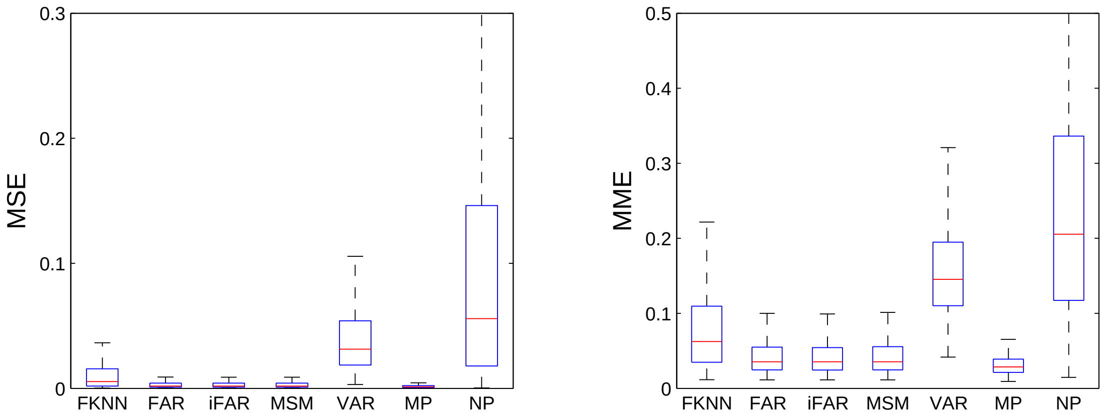


(c) $T = 200$

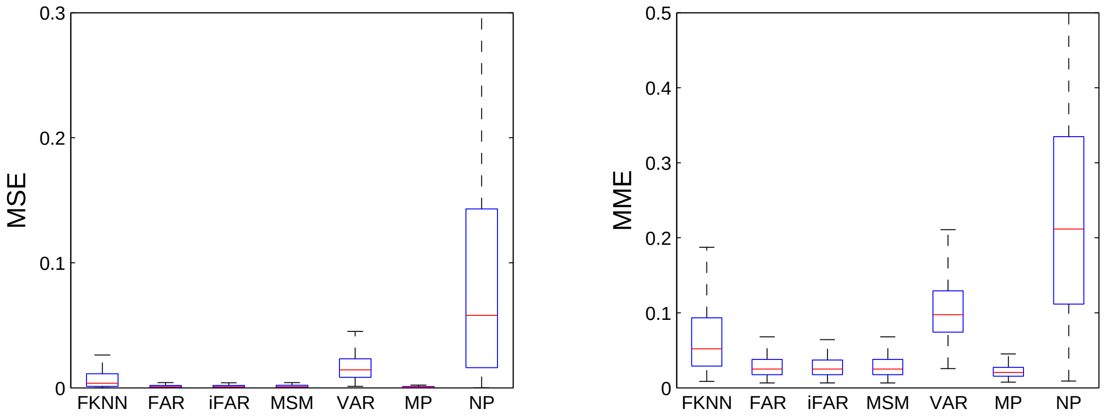
Figure 3.1: Boxplots of the prediction errors MSE (left panel) and MME (right panel) when DGP has kernel $\rho(t, s) = Ce^{\frac{-(t^2+s^2)}{2}}$.



(a) $T = 50$

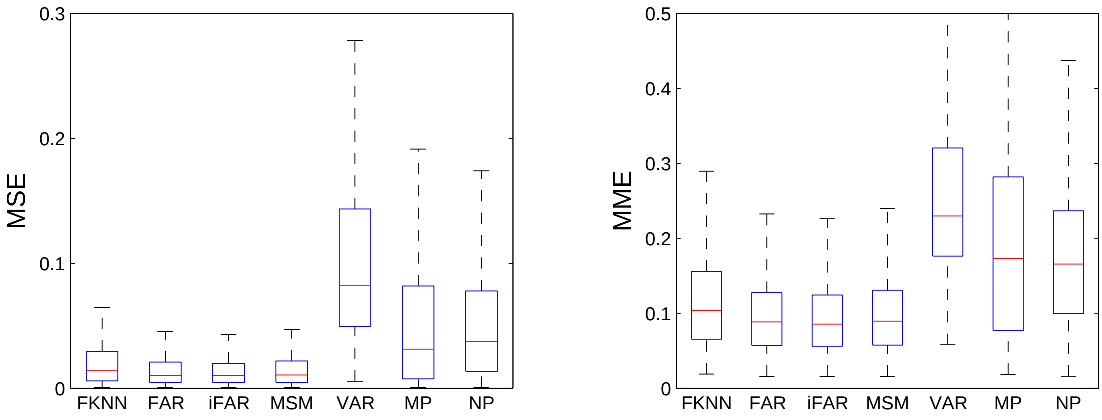


(b) $T = 100$

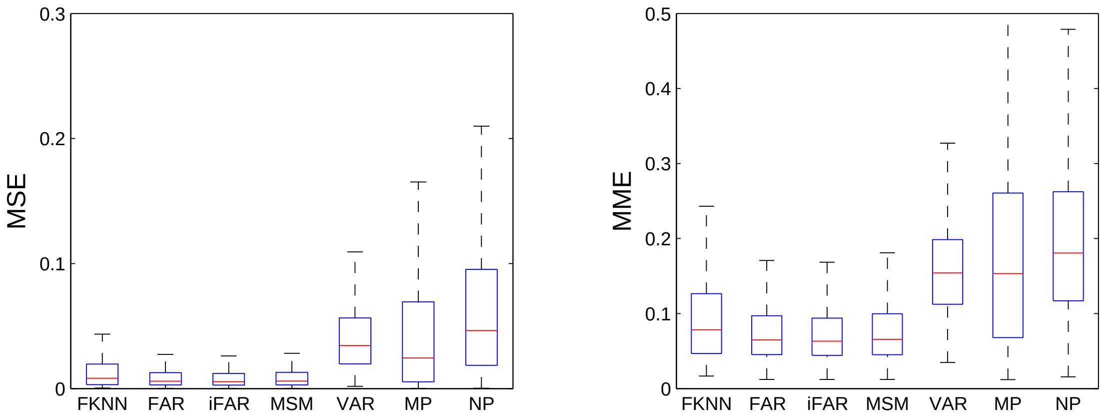


(c) $T = 200$

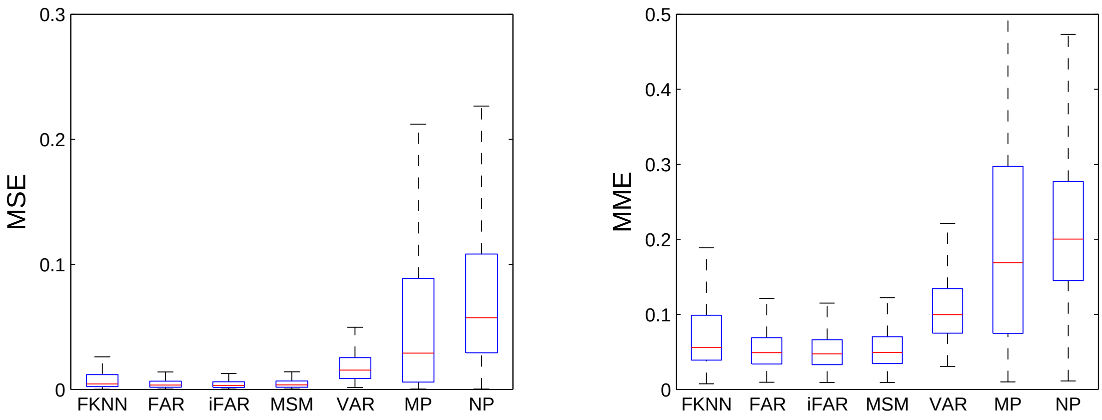
Figure 3.2: Boxplots of the prediction errors MSE (left panel) and MME (right panel) when DGP has kernel $\rho(t, s) = C$.



(a) $T = 50$



(b) $T = 100$



(c) $T = 200$

Figure 3.3: Boxplots of the prediction errors MSE (left panel) and MME (right panel) when DGP has kernel $\rho(t, s) = Ct$.

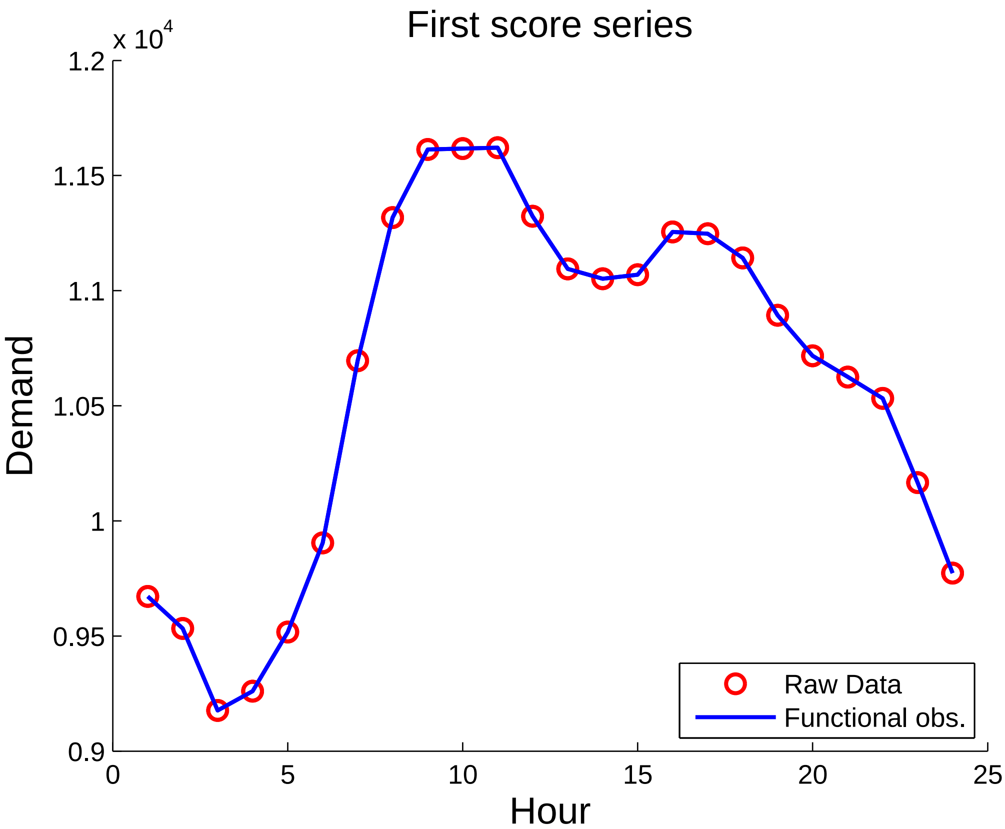
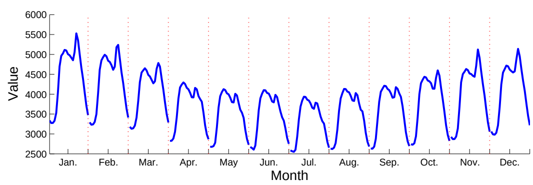
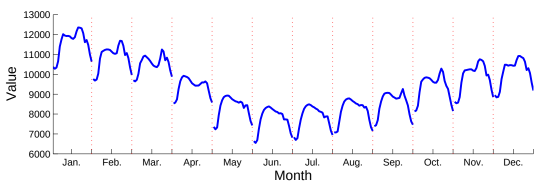


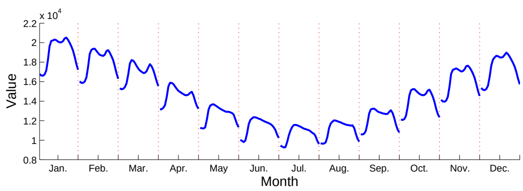
Figure 3.4: Typical daily discrete observation and reconstructed functional observation for electricity demand in Norway (June 1, 2013).



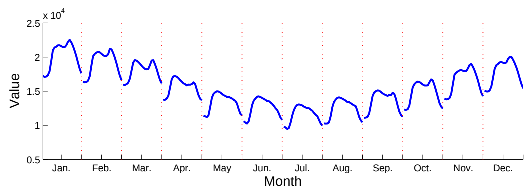
(a) Denmark



(b) Finland

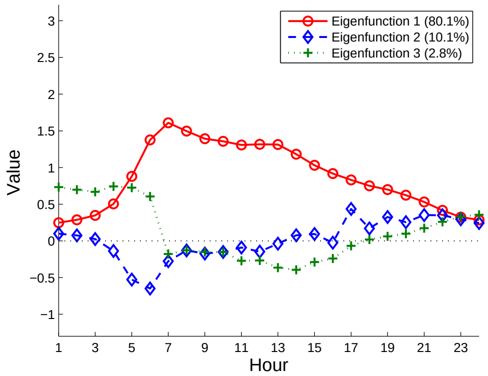


(c) Norway

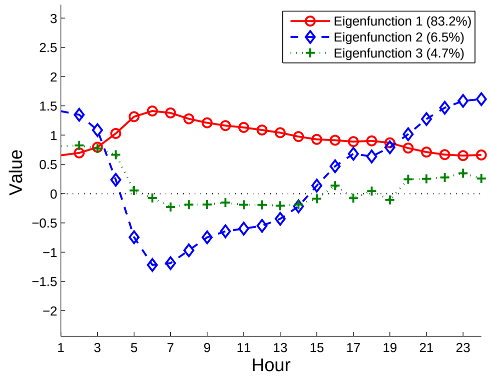


(d) Sweden

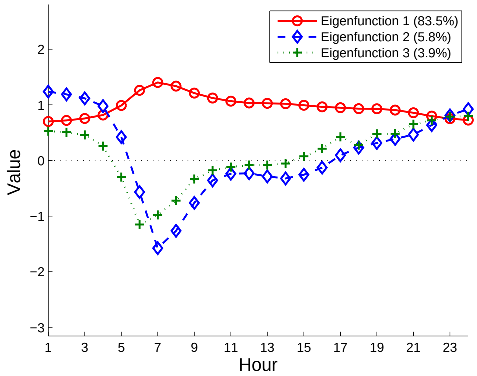
Figure 3.5: Seasonal monthly averages of the electricity demand in the Nordic countries.



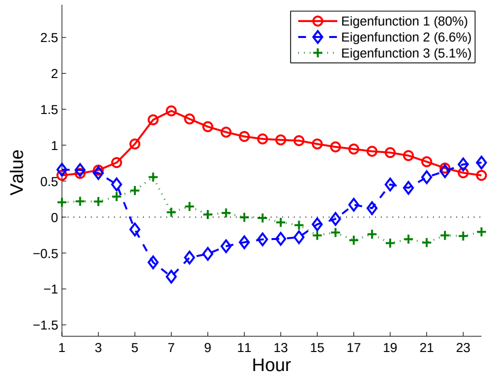
(a) Denmark



(b) Finland

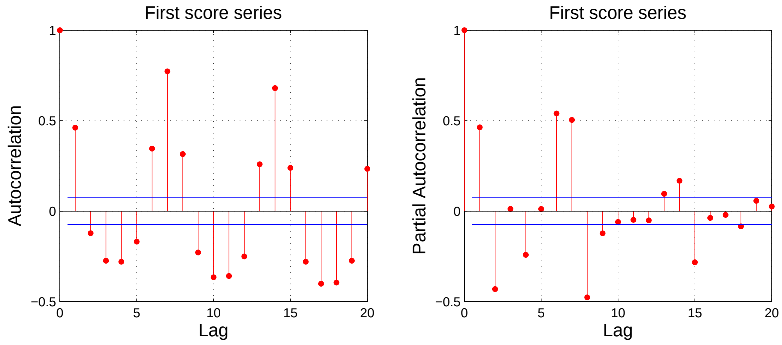


(c) Norway

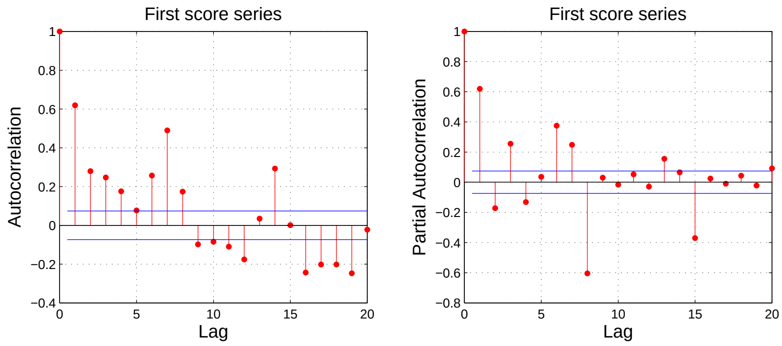


(d) Sweden

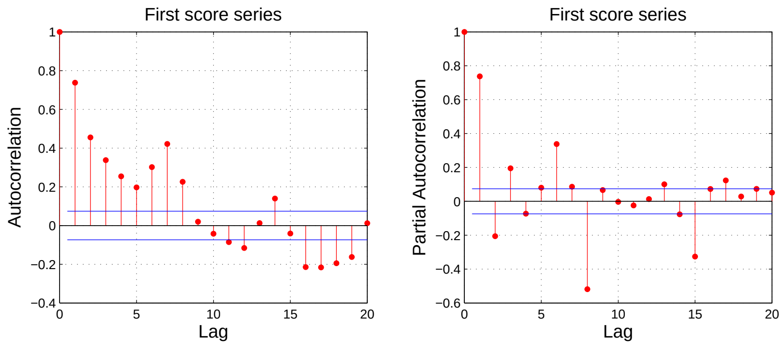
Figure 3.6: The first three estimated eigenfunctions of the electricity demand in the Nordic countries. The percentages indicate the amount of total variation accounted for by each eigenfunction.



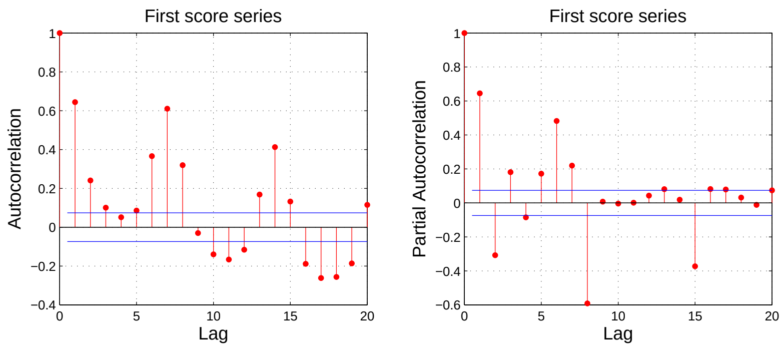
(a) Denmark



(b) Finland

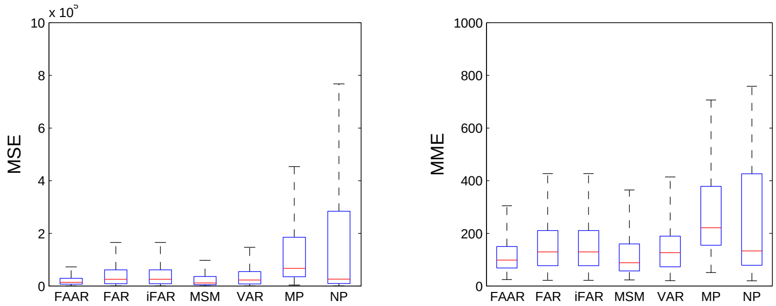


(c) Norway

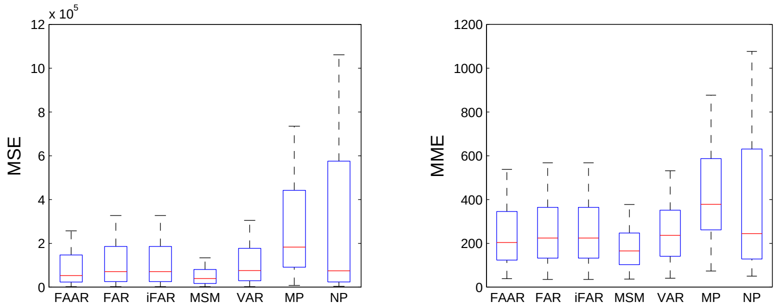


(d) Sweden

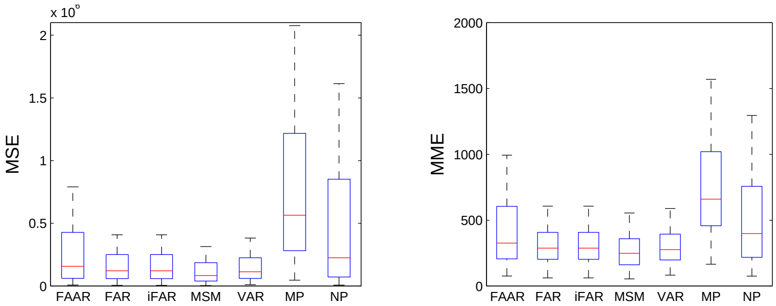
Figure 3.7: Time dependencies in score series. Left panel: sample autocorrelation of the first empirical FPC score series. Right panel: sample partial autocorrelation function of the first empirical FPC score series.



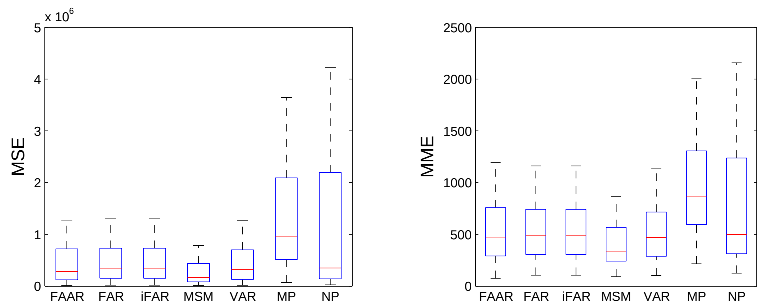
(a) Denmark



(b) Finland

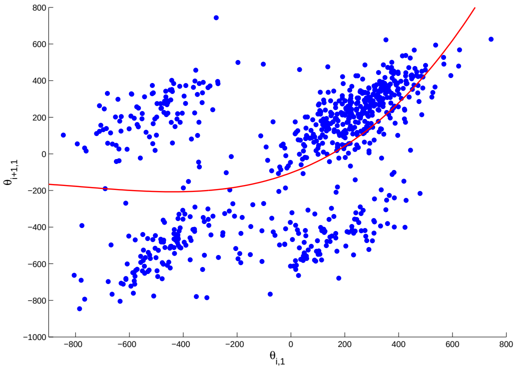


(c) Norway

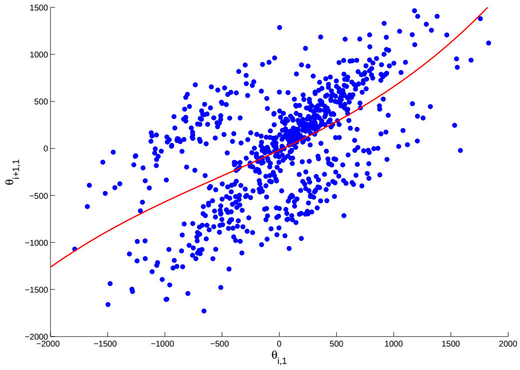


(d) Sweden

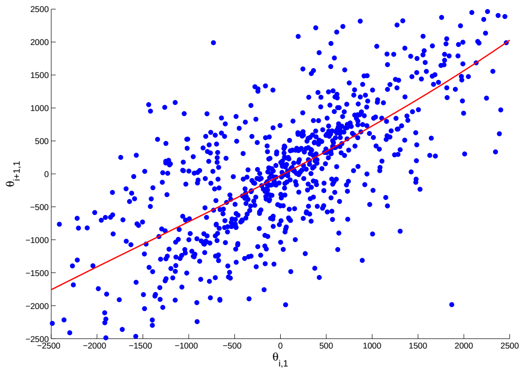
Figure 3.8: Boxplots of the prediction errors MSE (left panel) and MME (right panel).



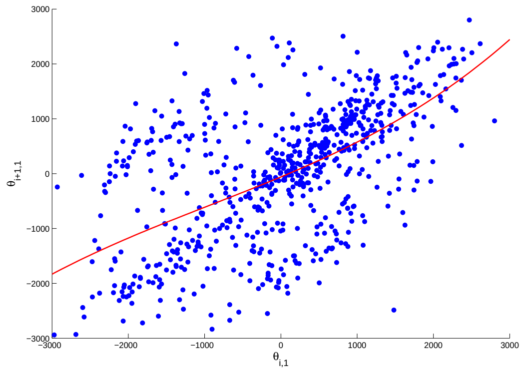
(a) Denmark



(b) Finland



(c) Norway



(d) Sweden

Figure 3.9: Scatter plots of the relationship between for the first FPC score and it lag.

Bibliography

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.

Besse, P. C., H. Cardot, and D. B. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics 27*(4), 673–687.

Bosq, D. (2000). *Linear Processes in Function Spaces*. New York: Springer.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Philadelphia: Society for Industrial and Applied Mathematics.

Chen, B. J., M. W. Chang, and C. J. Lin (2004). Load forecasting using support vector machines: A study on eunite competition 2001. *IEEE Transactions on Power Systems 19*, 1821–1830.

Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory 13*, 21–27.

Dauxois, J., A. Pousse, and Y. Romain (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis 12*(1), 136 – 154.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Demetrescu, M., V. Kuzin, and U. Hassler (2008). Long memory testing in the time domain. *Econometric Theory 24*(1), 176–215.

Didericksen, D., P. Kokoszka, and X. Zhang (2012). Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics 27*(2), 285–298.

Feinberg, E. A. and D. Genthliou (2005). Load forecasting. In J. H. Chow, F. F. Wu, and J. J. Momoh (Eds.), *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control and Computational Intelligence, Power Electronics and Power Systems*, pp. 269–285. New York: Springer.

Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews 26*(6), 609–641.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hippert, H. S., C. E. Pedreira, and R. C. Souza (2001). Neural netowrks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems 16*, 44–55.

Hörmann, S. and Ł. Kidziński (2015). A note on estimation in hilbertian linear models. *Scandinavian Journal of Statistics 42*(1), 43–62.

Hörmann, S. and P. Kokoszka (2010). Weakly dependend functional data. *The Annals of Statistics 38*, 1845–1884.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. New York: Springer.

Horváth, L., P. Kokoszka, and G. Rice (2014). Testing stationarity of functional time series. *Journal of Econometrics 179*(1), 66 – 82.

Kokoszka, P. and M. Reimherr (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis 34*(1), 116–129.

Kokoszka, P. and X. Zhang (2010). Improved estimation of the kernel of the functional autoregressive process. *Technical Report. University of Chicago*.

Kyriakides, E. and M. Polycarpou (2007). Short term electric load forecasting: A tutorial. In C. K. and L. Wang (Eds.), *Trends in Neural Computation, Studies in Computational Intelligence, vol. 35*, pp. 391–418. New York: Springer.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis 98*(6), 1231 – 1261.

Müller, H.-G. and F. Yao (2008). Functional additive models. *Journal of the American Statistical Association 103*, 1534–1544.

Park, J. Y. and J. Qian (2012). Functional regression of continuous state distributions. *Journal of Econometrics 167*(2), 397 – 412. Fourth Symposium on Econometric Theory and Applications (SETA).

Ramsay, J., G. Hooker, and S. Graves (2009). *Functional Data Analysis with R and MATLAB*. Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). New York: Springer.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics 5*, 595–620.

Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *The Annals of Statistics 12*(3), 917–926.

Yakowitz, S. (1987). Nearest-neighbour methods for time series analysis. *Journal of Time Series Analysis 8*, 235–247.

Chapter 4

On Estimation of Heterogeneous Panels with Systematic Slope Variations

4.1 Introduction

It becomes common in the panel data analysis to allow unobserved heterogeneity not only enter the model through the individual specific constant but also through the slope of the model. One of the standard and common approaches to handle the slope heterogeneity is to consider a random coefficient model where the slope coefficients are randomly distributed across individuals with a common mean parameter (See, e.g., Hsiao and Pesaran, 2008). This topic gained considerable attention in the resent literature, where number of testing procedures have been developed to test for slope homogeneity (see, e.g., Pesaran and Yamagata, 2008, Juhl and Lugovskyy, 2014 and Breitung et al., 2016). It is also widely recognized that such a modeling framework can have important consequences for the estimation and inference in the panel models (see, e.g., Pesaran et al., 1996 and Breitung, 2014 for a review of this topic).

The main aim of this paper is to analyze estimation procedures in heterogeneous panels with a particular focus on systematic slope variations - dependence of any form between covariates and their respective coefficients. The properties of the random coefficient panel model when coefficients are assumed to be independent of the covariates are well studied. However, this setup provides a restrictive modeling framework for many economic applications. (See, e.g., Wooldridge, 2005, who stress the importance of this issue) Therefore, estimators robust to (potentially) systematically varying slopes have to be developed.

There are two general concepts to construct an estimator of a slope or a common parameter in heterogeneous panel models.[1] The first one uses pooled data across indi-

[1] See for instance Pesaran and Smith (1995) for a detailed review of different estimation concepts in

viduals and time for estimation (pooled or within-group estimator), while the second one estimates the parameter for each individual/group which later are pooled to obtain a single estimator (mean-group estimator). In the presence of the systematic slope variations the within-group estimator (and also the GLS estimator) provides inconsistent results, whereas the mean-group estimator is robust in this situation. On the other hand the robustness of the mean group estimator comes at the price of a higher variance in comparison with pooled type estimator. For this reason a Hausman test can be used to choose an appropriate estimator as suggested in Pesaran et al. (1996).

In this work we develop an alternative solution to the estimation problem of a heterogeneous panel with (potentially) systematically varying slopes. We propose an estimation procedure that is based on the pooled estimator with Mundlak type correction (see, Mundlak, 1978 for more details). This solution is appealing due to its simplicity of implementation since it only requires to add well define addition regressor to the panel model and then perform the pooled estimation procedure. Further, it is asymptotically equivalent to the mean-group estimator in terms of bias and efficiency when N and T are large. This in turn allows to concentrate on one estimation technique and to avoid the additional testing as suggested in Pesaran et al. (1996). Finally, when N is large and T is fixed the new estimation procedure can provide an attractive alternative in terms of efficiency when compared to the mean-group estimator. This findings are supported with Monte Carlo experiments in small samples.

The reminder of the paper is structured as follows. Section 2 discusses the modeling framework, available estimation procedures and suggests a robust pooled estimator. In Section 3 asymptotic properties of the estimator are derived and discussed. The finite sample properties are studied Section 4. Section 5 concludes.

4.2 Model, Assumptions and Estimators

We assume that data are generated by the random coefficient model for panels, where the slope coefficients are constant over time but differ randomly across individuals, i.e.,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \tag{4.1}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i, \tag{4.2}$$

for $i = 1, 2, ..., N$ and $t = 1, ..., T$, where $\mathbf{x}_{it}$ is a $K \times 1$ vector of exogenous regressors. The vector of (random) coefficients consist of a common non-stochastic vector $\boldsymbol{\beta}$ and a vector of a individually specific disturbances $\mathbf{v}_i$. Inserting (4.2) into (4.1) and stacking

panels. This work also considers the other types of estimators based on the between-group regression and the time series regression. However, these estimators are found to be less efficient than the pooled and mean-group ones and are not considered in this paper over the time dimetion yields

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i\,, \tag{4.3}$$

where $\mathbf{u}_i = [u_{i1}, ..., u_{iT}]'$ with $u_{it} = \mathbf{x}'_{it}\mathbf{v}_i + \varepsilon_{it}$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$ and $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$.

Assumption 3 *(i) The error vector $\boldsymbol{\varepsilon}_i$ is iid$(0, \sigma_\varepsilon^2\mathbf{I}_T)$, where $\mathbf{I}_T$ is $T \times T$ identity matrix. Moreover, $\boldsymbol{\varepsilon}_i$ is independent of $\mathbf{X}_i$ and $\mathbf{v}_i$ for all i.*

Assumption 4 *(i) The $K \times 1$ strictly exogenous vector of regressors x_{it} is weakly stationary and $\mathbb{E}\,|x_{it,k}|^{4+\delta} < C < \infty$ for some $\delta > 0$, $C > 0$ and all $i = 1, ..., N$ and $t = 1, ..., T$. (ii) Further, matrices $\mathbf{S}_{i,T} \equiv \mathbf{X}'_i\mathbf{X}_i/T$ and $\mathbf{S}_{N,T} \equiv \sum_{i=1}^N \mathbf{X}'_i\mathbf{X}_i/NT$ are positive definite for all N and T and have non-stochastic positive definite limits, i.e.,*

$$\begin{aligned}
\mathbf{S}_i &\equiv \operatorname*{plim}_{T\to\infty} \mathbf{S}_{i,T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathbf{x}_{it}\mathbf{x}'_{it}], \\
\mathbf{S} &\equiv \operatorname*{plim}_{N,T\to\infty} \mathbf{S}_{N,T} = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathbf{S}_i.
\end{aligned}$$

Assumption 5 *(i) The error vectors $\mathbf{v}_i|\mathbf{X} \sim iid(0, \boldsymbol{\Theta}_{i,NT})$, where is $\boldsymbol{\Theta}_{i,NT}$ diagonal. (ii) For each $i, j = 1, ..., N$ and $t = 1, ..., T$ $\mathbf{v}_i$ is independent of ε_{jt}*

Estimators:

There are two well established approaches to estimate the common parameter $\boldsymbol{\beta}$ that represents the central tendency among heterogeneous responses.

First, we may just ignore parameter heterogeneity and pool the data which will yield the pooled OLS estimator

$$\widehat{\boldsymbol{\beta}}_p = \left(\sum_{i=1}^N \mathbf{X}'_i\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^N \mathbf{X}'_i\mathbf{y}_i\right). \tag{4.4}$$

Furthermore, it is customary to use generalized LS version of the pooled estimator

$$\widehat{\boldsymbol{\beta}}_{gls} = \left(\sum_{i=1}^N \mathbf{X}'_i\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^N \mathbf{X}'_i\boldsymbol{\Omega}_i^{-1}\mathbf{y}_i\right). \tag{4.5}$$

where $\boldsymbol{\Omega}_i = E(\mathbf{u}_i\mathbf{u}'_i|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\Theta}_{i,NT}\mathbf{X}'_i + \sigma_\epsilon^2\mathbf{I}_T$, and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iT})'$.

Second, the parameter $\boldsymbol{\beta}$ may be estimated separately for each group and then the individual specific estimators are pooled to obtain an estimator of $\boldsymbol{\beta}$. This approach was advocated by Pesaran and Smith (1995) and it is referred to as mean-group estimator, i.e.,

$$\widehat{\boldsymbol{\beta}}_{mg} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i \ , \tag{4.6}$$

where $\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\mathbf{X}_i'\mathbf{y}_i\right)$.

It can be easily seen that the consistency of (4.4) and (4.5) depends on the relation between the $\mathbf{x}_{it}$ and u_{it}:

$$\mathbb{E}\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) = T\mathbb{E}\left(\sum_{i=1}^{N}\mathbf{S}_{i,T}\mathbf{v}_i\right) + \mathbb{E}\left(\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right). \tag{4.7}$$

Under assumption 3 we have $\mathbb{E}\left(\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right) = 0$ for all i. Hence, it follows that for the consistency of these estimators we require that $\mathbf{S}_{i,T}$ and $\mathbf{v}_i$ are uncorrelated. For this reason Wooldridge (2005) advocated a sufficient condition, $\mathbb{E}\left(\mathbf{v}_i|x_{it}\right) = 0$ for all t, to make the pooled estimator unbiased. In this work we propose to consider more general settings by following the Mundlak (1978) and introduce the auxiliary regression

Assumption 3′

$$\mathbf{v}_i = \left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)\boldsymbol{\gamma} + \boldsymbol{\xi}_i, \tag{4.8}$$

where $\boldsymbol{\xi}_i$ is iid$(0,\Delta\mathbf{I}_K)$, $\boldsymbol{\xi}_i$ is uncorrelated with $\mathbf{S}_{i,T}$ and ε_{jt} for all $i, j = 1, ..., N$.

The demeaning $\mathbf{S}_{i,T} - \mathbf{S}_{N,T}$ in (4.8) is used to ensure that $\mathbb{E}\left(\mathbf{v}_i\right) = 0$. Clearly, $\boldsymbol{\gamma} = 0$ if and only if the $\mathbf{S}_{i,T}$ are uncorrelated with the effects $\mathbf{v}_i$. It also follows from assumption 5 and 3′ that $\boldsymbol{\Theta}_{i,NT} = \Delta\mathbf{I}_K + \left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'\left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)$. Further, model (4.1) under assumption 3′ takes the form

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\boldsymbol{\gamma} + \eta_{it},$$

where $\mathbf{z}_{it} = \left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)\mathbf{x}_{it}$ and $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\xi}_i + \varepsilon_{it}$. Accordingly, a consistent estimator of $\boldsymbol{\beta}$ can be obtained as

$$\widetilde{\boldsymbol{\beta}}_p = \left(\mathbf{X}'\mathbf{M}_z\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{M}_z\mathbf{y}\right), \tag{4.9}$$

where $\mathbf{M}_z = \mathbf{I}_{NT} - \mathbf{Z}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}'$, $\mathbf{Z} = \left(\mathbf{Z}_1', ..., \mathbf{Z}_N'\right)'$ and $\mathbf{Z}_i = \left(\mathbf{z}_{i,1}', ..., \mathbf{z}_{i,T}'\right)$ for $i = 1, ..., N$.

4.3 Asymptotic Properties

In this section we investigate the asymptotic properties of the estimators considered in the previous section: $\widetilde{\boldsymbol{\beta}}_p$, $\widehat{\boldsymbol{\beta}}_p$, $\widehat{\boldsymbol{\beta}}_{gls}$ and $\widehat{\boldsymbol{\beta}}_{MG}$. Next two propositions present first order asymptotics of the considered estimators.

Proposition 1 (Bias) *Given model (4.1)-(4.2) satisfies the assumptions 3, 4 and 3′ then for a fixed T and $N \to \infty$ the following holds*

(i) *For the pooled estimator:*

$$\operatorname*{plim}_{N\to\infty} \widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta} = \mathbf{S}^{-1}\mathbf{S}^{(2)}\boldsymbol{\gamma}, \tag{4.10}$$

where $\mathbf{S}^{(2)} \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} Var\left[\mathbf{S}_{i,T}\right]$.

(ii) *For the GLS estimator:*

$$\operatorname*{plim}_{N\to\infty} \widehat{\boldsymbol{\beta}}_{gls} - \boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\boldsymbol{\gamma} + O_p(T^{-1/2}), \tag{4.11}$$

where $\boldsymbol{\Omega} \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Theta}_i^{-1}$, $\boldsymbol{\Lambda} \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Theta}_i^{-1}\left(\mathbf{S}_i - \mathbf{S}\right)$ *and* $\boldsymbol{\Theta}_i \equiv \lim_{N,T\to\infty} \boldsymbol{\Theta}_{i,NT}$.

(iii) *For the pooled-Mundlak estimator:*

$$\operatorname*{plim}_{N\to\infty} \widetilde{\boldsymbol{\beta}}_p - \boldsymbol{\beta} = 0, \tag{4.12}$$

(iv) *For the mean-group estimator:*

$$\operatorname*{plim}_{N\to\infty} \widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta} = 0. \tag{4.13}$$

Proof. See Appendix. ■

Proposition 1 illustrates several key facts and findings about the consistency of estimators of heterogeneous panels with systematically varying slopes. First, as discussed above the standard pooled OLS estimator $\widehat{\boldsymbol{\beta}}_p$ has bias which will vanish only if $\boldsymbol{\gamma} = 0$, which in turn is associated with no correlation between disturbances v_i and second empirical moment of the covariates. Second, an interesting result is obtained for $\widehat{\boldsymbol{\beta}}_{gls}$ estimator that is known to be asymptotically equivalent to $\widehat{\boldsymbol{\beta}}_{mg}$ and consistent for heterogeneous panels when $N, T \to \infty$. Item (ii) of Proposition 1 shows that in fact under systematic slope variations ($\boldsymbol{\gamma} \neq 0$) the GLS estimator will be consistent only if $\boldsymbol{\Lambda} = 0$. That is, the mean of scaled variances of covariates has to be equal zero. Further, it is shown in Proposition 1 item (iii) that inclusion of the additional regressor z_i in the model can fix the problem of the bias of the pooled estimator. Finally, item (iv) confirms the consistency of the mean-group estimator.

The next question of interest is the efficiency of relevant (consistent) estimators. For simplicity of exposition (and without loss of generality) we analyze the case where the model contains only one regressor ($K = 1$), generated independently across i, t and identically across t, i.e.,

$$x_{it} \sim \text{id}(0, S_i). \tag{4.14}$$

Our next result presents the asymptotic variance of the pooled estimator with Mundlak correction.

Proposition 2 (Efficiency) *If model (4.1)-(4.2) satisfies the assumptions 3, 4, 3′ and additionally covariates behaves as in* (4.14) *with* $\mathbb{E}\left|x_{it,k}\right|^8 < \infty$ *for all* $i = 1, ..., N$ *and* $t = 1, ..., T$*, then*

$$\lim_{N\to\infty} N \operatorname{Var}\left(\widetilde{\beta}_p\right) = \frac{\sigma_\varepsilon^2}{T} S^{-1} + \Delta \Sigma_S S^{-2} + O(T^{-1}), \tag{4.15}$$

where $S \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N S_i$ *and* $\Sigma_S \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N S_i^2$.

Recall that for the mean group estimator a similar result is obtained (see, e.g., Hsiao and Pesaran, 2008)

$$N \operatorname{Var}\left(\widehat{\beta}_{MG}\right) \underset{N\to\infty}{\rightarrow} \frac{\sigma_\varepsilon^2}{T} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{S_i} + \Delta. \tag{4.16}$$

Further, from Cauchy-Schwarz inequality and Jensen's inequality it follows that $[S]^{-1} \leq \frac{1}{N} \sum_{i=1}^N \frac{1}{S_i}$ and $\Sigma_S \geq S^2$, respectively. Therefore, it becomes clear from (4.15) and (4.16) that in the settings with large N and fixed T both estimators can have gains in terms of the efficiency, when compared to each other. In particular, if the variance of idiosyncratic errors dominates the variance of the slope coefficients (i.e., $\sigma_\varepsilon^2 >> \Delta$) then $\widetilde{\beta}_p$ will provide more efficient estimates, otherwise (i.e., $\Delta >> \sigma_\varepsilon^2$) $\widehat{\beta}_{MG}$ will be preferred option in terms of efficiency.

4.4 Monte Carlo Experiments

In this section we investigate the finite sample properties of the estimation procedures for heterogeneous panels discussed in this paper, $\widehat{\beta}_p$, $\widetilde{\beta}_p$, $\widehat{\beta}_{gls}$ and $\widehat{\beta}_{mg}$. The aim of this section is to evaluate and compare the performance of the estimators in terms of their bias and efficiency for several different setups, relating to the theoretical discussion of Section 4.3.

The following data-generating process is used to conduct experiments

$$\begin{aligned} y_{it} &= x_{it}\beta_i + \varepsilon_{it}, \\ x_{it} &\sim iid\, N(0, S_i), \end{aligned}$$

where $\varepsilon_{it} \sim iid\, N(0,1)$. Variances of the regressors we generate as $S_i = 1$, $S_i \sim \chi_1^2$ and $S_i \sim U[0.5, 3.5]$. The dependencies between v_i and s_i are modeled as in the assumption 3′, i.e., $v_i = \gamma\left(s_i - \overline{s}\right) + \xi_i$. Therefore, in our benchmark specification we generate the

	Est.	Bias	Var	Ratio	Bias	Var	Ratio	Bias	Var	Ratio	Bias	Var	Ratio
$S_i = 1$,		$\gamma = 0$						$\gamma = 0.5$					
		$T = 20$			$T = 100$			$T = 20$			$T = 100$		
$N = 20$	$\widehat{\beta}_p$	0.002	0.058	1.08	0.000	0.052	1.02	0.048	0.057	1.09	0.001	0.051	1.02
	$\widehat{\beta}_{gls}$	0.003	0.054	1.00	-0.000	0.051	1.00	0.003	0.052	1.00	-0.009	0.050	1.00
	$\widetilde{\beta}_p$	0.003	0.054	1.01	-0.000	0.051	1.00	0.002	0.053	1.01	-0.009	0.050	1.00
	$\widehat{\beta}_{MG}$	0.003	0.054	·	-0.000	0.051	·	0.002	0.052	·	-0.009	0.050	·
$N = 100$	$\widehat{\beta}_p$	-0.002	0.011	1.08	0.001	0.010	1.02	0.011	0.053	1.02	0.007	0.011	1.02
	$\widehat{\beta}_{gls}$	-0.001	0.011	1.00	0.001	0.010	1.00	0.001	0.052	1.00	-0.003	0.010	1.00
	$\widetilde{\beta}_p$	-0.001	0.011	1.00	0.001	0.010	1.00	0.001	0.052	1.00	-0.003	0.010	1.00
	$\widehat{\beta}_{MG}$	-0.001	0.011	·	0.001	0.010	·	0.001	0.052	·	-0.003	0.010	·
$S_i \sim \chi^2(1)$,													
$N = 20$	$\widehat{\beta}_p$	0.002	0.144	0.07	-0.008	0.141	0.26	0.968	0.588	0.24	0.888	0.394	0.75
	$\widehat{\beta}_{gls}$	0.003	0.066	0.03	-0.001	0.058	0.11	-0.054	0.077	0.03	-0.078	0.066	0.13
	$\widetilde{\beta}_p$	0.004	0.126	0.06	-0.006	0.120	0.22	-0.005	0.124	0.05	-0.002	0.118	0.22
	$\widehat{\beta}_{MG}$	-0.007	2.084	·	0.008	0.545	·	0.020	2.443	·	0.010	0.527	·
$N = 100$	$\widehat{\beta}_p$	-0.001	0.032	0.07	-0.000	0.030	0.33	1.106	0.183	0.39	0.995	0.112	1.28
	$\widehat{\beta}_{gls}$	-0.001	0.013	0.03	-0.001	0.012	0.13	-0.051	0.016	0.03	-0.089	0.014	0.16
	$\widetilde{\beta}_p$	0.001	0.030	0.07	-0.001	0.028	0.31	0.002	0.030	0.06	0.000	0.029	0.33
	$\widehat{\beta}_{MG}$	-0.003	0.453	·	-0.001	0.091	·	0.017	0.467	·	0.005	0.087	·
$S_i \sim U[0.5, 3.5]$,													
$N = 20$	$\widehat{\beta}_p$	-0.004	0.082	1.65	-0.001	0.017	1.62	1.580	0.301	5.82	1.652	0.067	6.51
	$\widehat{\beta}_{gls}$	-0.004	0.050	1.00	-0.000	0.011	1.00	-0.218	0.100	1.92	-0.231	0.020	1.96
	$\widetilde{\beta}_p$	-0.005	0.072	1.44	-0.002	0.016	1.47	0.000	0.074	1.43	-0.002	0.015	1.48
	$\widehat{\beta}_{MG}$	-0.004	0.050	·	-0.000	0.011	·	0.001	0.052	·	-0.002	0.010	·
$N = 100$	$\widehat{\beta}_p$	-0.007	0.080	1.54	-0.002	0.016	1.60	1.656	0.068	6.54	1.369	0.035	3.52
	$\widehat{\beta}_{gls}$	-0.002	0.052	1.00	-0.001	0.010	1.00	-0.228	0.020	1.91	-0.179	0.020	1.97
	$\widetilde{\beta}_p$	-0.001	0.069	1.33	0.000	0.014	1.42	-0.002	0.015	1.49	-0.000	0.014	1.44
	$\widehat{\beta}_{MG}$	-0.002	0.052	·	-0.001	0.010	·	-0.000	0.010	·	0.000	0.010	·

Table 4.1: Bias and efficiency of the estimators for heterogeneous panels with systematic slope variations.

slopes as

$$\beta_i \sim N(1, \Delta) + \gamma(s_i - \bar{s}),$$

where $\Delta = 1$, $\gamma = \{0, 0.5\}$. All results are based on 5000 relications. We examine four combinations of $(N, T) = \{(20, 20), (20, 100), (100, 20), (100, 100)\}$.

Results of the simulations are presented Table 1. In particular, the bias, the MSE of the estimators and ratio of the estimator's MSE with respect to the MSE of the mean-group estimator are reported. Finally, the left panel represents the case when parameter $\gamma = 0$ indicating no correlation between v_i and s_i, while $\Delta = 0$ presents the case when three is no heterogeneity in slopes. The main results of the experiments confirm the theoretical findings of Proposition 1 and 2.

A Appendix: Proofs

First, an auxiliary Lemma is provided.

Lemma A.3 *Given that $x_{it} \sim id(0, S_i)$ and $\mathbb{E}\left|x_{it}\right|^8 < \infty$ for all i and t then the first four moments of $S_{i,T} = \mathbf{x}_i' \mathbf{I}_T \mathbf{x}_i / T$ are*

(i) $\mathbb{E}[S_{i,T}] = S_i$,

(ii) $\mathbb{E}[S_{i,T}^2] = S_i^2 \left(1 + \frac{\lambda_{1,i}}{T}\right)$,

(iii) $\mathbb{E}[S_{i,T}^3] = S_i^3 \left(1 + 3\frac{\lambda_{1,i}}{T} + \frac{\lambda_{2,i}}{T^2}\right)$,

(iv) $\mathbb{E}[S_{i,T}^4] = S_i^4 \left(1 + 6\frac{\lambda_{1,i}}{T} + \frac{\lambda_{3,i}}{T^2} + \frac{\lambda_{4,i}}{T^3}\right)$,

where $\lambda_{1,i} = \left(2 + p_i^{(2)}\right)$, $\lambda_{2,i} = \left(8 + p_i^{(4)} + 12p_i^{(2)} + 10\left(p_i^{(1)}\right)^2\right)$, $\lambda_{3,i} = \left(44 + 60p_i^{(2)} + 4p_i^{(4)} + 40\left(p_i^{(1)}\right)^2 + 3\left(p_i^{(2)}\right)^2\right)$, $\lambda_{4,i} = \left(48 + 144p_i^{(2)} + 24p_i^{(4)} + p_i^{(6)} + 240\left(p_i^{(1)}\right)^2 + 32\left(p_i^{(2)}\right)^2 + 56p_i^{(1)}p_i^{(3)}\right)$ *and* $p_i^{(1)}$ *and* $p_i^{(2)}$ *are Persons measure of skewness and kurtosis of the* x_{it} *distribution and* $p_i^{(3)}, ..., p_i^{(6)}$ *are regarded as measure for deviation from normality as in Ullah (2004).*

Proof. To obtain (**i**), (**ii**) and (**iii**) we make use of results derived in Appendix A.5 of Ullah (2004). Item (**iv**) follows from Theorem 2 of Bao and Ullah (2010). ■

Proof of Proposition 1

Item (i)

For the pooled estimator it holds

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta} &= \left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\mathbf{v}_i\right) + O_p(N^{-1/2}) \\ &= \left[\mathbf{S}_{N,T}\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{i,T}^2 - \mathbf{S}_{N,T}^2\right]\boldsymbol{\gamma} + O_p(N^{-1/2}),\end{aligned}$$

where in turn by Assumption 4 $\mathbf{S}_{N,T} \xrightarrow{p} \mathbf{S}$ and LLN for independent heterogeneous distributed random variables (see, e.g., White, 2001, Corrolary 3.9) we have

$$\left[\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{i,T}^2 - \mathbf{S}_{N,T}^2\right] - \left[\frac{1}{N}\sum_{i=1}^N \left(\mathbb{E}\left[\mathbf{S}_{i,T}^2\right] - \mathbb{E}\left[\mathbf{S}_{i,T}\right]^2\right)\right] \quad \xrightarrow{a.s.} \quad 0,$$

as $N \to \infty$ and $\xrightarrow{a.s.}$ denotes almost sure convergence. 1.

Item (ii)

The GLS estimator can be written as a matrix weighted average of the least squares estimator for each cross-sectional unit i.e.,

$$\widehat{\boldsymbol{\beta}}_{gls} = \frac{1}{N}\sum_i \mathbf{R}_i \widehat{\boldsymbol{\beta}}_i, \tag{A.1}$$

where

$$\begin{aligned}
\mathbf{R}_i &= \left[\frac{1}{N}\sum_j \left(\sigma_\varepsilon^2\left(\mathbf{X}_j'\mathbf{X}_j\right) + \boldsymbol{\Theta}_{i,NT}\right)^{-1}\right]^{-1}\left(\sigma_\varepsilon^2\left(\mathbf{X}_i'\mathbf{X}_i\right) + \boldsymbol{\Theta}_{i,NT}\right)^{-1}, \\
\boldsymbol{\Theta}_{i,NT} &\equiv \mathbb{E}[v_i v_i'|\mathbf{X}] = \Delta\mathbf{I}_K + \left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'\left(\mathbf{S}_{i,T} - \mathbf{S}_{N,T}\right)'.
\end{aligned}$$

Further rewrite weights $\mathbf{R}_i$ as $\mathbf{R}_i = \left[\mathbf{Q}_{N,T}\right]^{-1}\mathbf{Q}_{i,T}$ where $\mathbf{Q}_{i,T} = \left(\sigma_\varepsilon^2\left(\mathbf{X}_i'\mathbf{X}_i\right) + \boldsymbol{\Theta}_{i,NT}\right)^{-1}$ and $\mathbf{Q}_{N,T} = \sum_i \mathbf{Q}_{i,T}/N$.

Remark 1: The following holds

(i) $\mathbf{S}_{i,T} = \mathbf{S}_i + O_p\left(T^{-1/2}\right)$;

(ii) $\mathbf{S}_{i,T}^{-1} = \mathbf{S}_i^{-1} + O_p\left(T^{-1/2}\right)$;

(iii) $\boldsymbol{\Theta}_{i,NT} = \boldsymbol{\Theta}_{i,N} + O_p\left(T^{-1/2}\right)$, where $\boldsymbol{\Theta}_{i,N} = \left(\mathbf{S}_i - \mathbf{S}_N\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'\left(\mathbf{S}_i - \mathbf{S}_N\right)' + \Delta\mathbf{I}_K$ and $\mathbf{S}_N = \frac{1}{N}\sum_i \mathbf{S}_i$.

(iv) $\boldsymbol{\Theta}_{i,NT} = \boldsymbol{\Theta}_i + O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\right)$, where $\boldsymbol{\Theta}_i = \left(\mathbf{S}_i - \mathbf{S}\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'\left(\mathbf{S}_i - \mathbf{S}\right)' + \Delta\mathbf{I}_K$

Proof of Remark 1: (i) follows from Lindeberg-Levy CLT; (ii) comes from the fact that $\mathbf{S}_i$ is positive definite and first order Taylor expansion of the inverse function $g\left(\mathbf{S}_{i,T}\right) = \mathbf{S}_{i,T}^{-1}$ in the local neighborhood of $\mathbf{S}_i$. For item (iii) and (iv) notice that $\mathbf{S}_{NT} = \mathbf{S}_N + O_p\left(T^{-1/2}\right) = \mathbf{S} + O_p\left(T^{-1/2}\right) + O_p\left(N^{-1/2}\right)$. Then results will follow from (i) and uniform $L_{4+\delta}$ boundedness of regressors (i.e., Assumption 4). ■

Remark 2:

(i) $\mathbf{Q}_{i,T} = \boldsymbol{\Theta}_{i,N}^{-1} - \frac{1}{T}\mathbf{W}_{i,N} + O_p\left(T^{-1/2}\right)$, where $\mathbf{W}_{i,N} = \sigma_\varepsilon^2\boldsymbol{\Theta}_{i,N}^{-1}\mathbf{S}_i^{-1}\boldsymbol{\Theta}_{i,N}^{-1}$.

(ii) $\left[\mathbf{Q}_{N,T}\right]^{-1} = \boldsymbol{\Omega}_N^{-1} + \frac{1}{T}\boldsymbol{\Omega}_N^{-1}\mathbf{W}_N\boldsymbol{\Omega}_N^{-1} + O_p\left(T^{-1/2}\right)$, where $\boldsymbol{\Omega}_N = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Theta}_{i,N}^{-1}$ and $\mathbf{W}_N = \frac{1}{N}\sum_{i=1}^N \mathbf{W}_{i,N}$.

Proof of Remark 2: By the Remark 1 we have $\mathbf{Q}_{i,T}^{-1} = \boldsymbol{\Theta}_{i,N} + \frac{1}{T}\sigma_\varepsilon^2\mathbf{S}_i^{-1} + O_p\left(T^{-1/2}\right)$. Then by using the first order Taylor expansion of the inverse of matrix sum (i.e., $(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$) it follows

$$\begin{aligned}
\mathbf{Q}_{i,T} &= \left[\boldsymbol{\Theta}_{i,N} + \frac{1}{T}\sigma_\varepsilon^2\mathbf{S}_i^{-1}\right]^{-1} + O_p\left(T^{-1/2}\right) \\
&= \boldsymbol{\Theta}_{i,N}^{-1} - \frac{1}{T}\sigma_\varepsilon^2\boldsymbol{\Theta}_{i,N}^{-1}\mathbf{S}_i^{-1}\boldsymbol{\Theta}_{i,N}^{-1} + O_p\left(T^{-1/2}\right).
\end{aligned}$$

Summing $\mathbf{Q}_{i,T}$ over i and using again Taylor expansion for the inverse will yield item (ii) of Remark 2. ■

Notice that $\mathbf{\Omega}_N^{-1} = O(1)$ and $\mathbf{W}_N = O(1)$ hence $[\mathbf{Q}_{N,T}]^{-1} = \mathbf{\Omega}_N^{-1} + O_p\left(T^{-1/2}\right)$. Then by putting together expression for GLS estimator (A.1) and Remark 2 we have

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{gls} &= \left(\mathbf{\Omega}_N^{-1} + O_p\left(T^{-1/2}\right)\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Theta}_{i,N}^{-1}\widehat{\boldsymbol{\beta}}_i - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_{i,N}^{-1}\widehat{\boldsymbol{\beta}}_i + O_p\left(T^{-1/2}\right)\right)\\ &= \mathbf{\Omega}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Theta}_{i,N}^{-1}\widehat{\boldsymbol{\beta}}_i + O_p\left(T^{-1/2}\right).\end{aligned}$$

where the last equality comes from the fact that $\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_{i,N}^{-1}\widehat{\beta}_i = O_p\left(T^{-1}\right)$. Further from Remark 1 and since $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta} = \mathbf{v}_i + O_p\left(T^{-1/2}\right)$ the result for item (**ii**) will follow, i.e.,

$$\mathbf{\Omega}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Theta}_{i,N}^{-1}\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta} + \mathbf{\Omega}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Theta}_i^{-1}\left(\mathbf{S}_i - \mathbf{S}\right)\boldsymbol{\gamma} + O_p\left(T^{-1/2}\right) + O_p\left(N^{-1/2}\right).$$

Item (iii) and (iv)

(**iii**) Mundlak-type pooled estimator:

$$\widetilde{\boldsymbol{\beta}}_p - \boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{M}_Z\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{M}_Z(\mathbf{X}\odot\boldsymbol{\eta})\right) = O_p(N^{-1/2}),$$

where $\odot$ denotes Hadamard product, $\boldsymbol{\eta}$ is $NT\times 1$ vector with typical element $\eta_{it} = x_{it}\xi_i + \varepsilon_{it}$ and the last equality comes from the Kolmogorov LLN.
(**iv**) Mean-group estimator:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta} &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\mathbf{X}_i'\mathbf{X}_i\mathbf{v}_i\right) + O_p(N^{-1/2})\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_{i,T} - \mathbf{S}_{NT}\right)\boldsymbol{\gamma} + O_p(N^{-1/2}) = O_p(N^{-1/2}).\end{aligned}$$

Proof of Proposition 2

First note that

$$\begin{aligned}N\,\mathrm{Var}\left(\widetilde{\beta}_p\right) &= \sigma^2\mathbb{E}\left(\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{X}}{N}\right)^{-1}\\ &+ \Delta\mathbb{E}\left(\left(\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{D}_X\mathbf{D}_X'\mathbf{M}_Z\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{X}}{N}\right)^{-1}\right),\end{aligned}$$

where $D_X = diag\{\mathbf{X}_1, ..., \mathbf{X}_N\}$. Then by LLN for independent heterogeneously distributed observations and Lemma A.3 it follows,

$$\frac{\mathbf{X}'\mathbf{X}}{N} - \frac{T}{N}\sum_{i=1}^{N}\mathbb{E}\left[S_{i,T}\right] \overset{a.s.}{\to} 0,$$

where $\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[S_{i,T}\right] = \frac{1}{N}\sum_{i=1}^{N} S_i \to S$ as $N \to \infty$. Same techniques will provide the rest of the results

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{Z}}{N} &\overset{a.s.}{\to} \lim_{N\to\infty}\frac{T}{N}\sum_{i=1}^{N}\left(\mathbb{E}(S_{i,T}^2) - \mathbb{E}\left(S_{i,T}\right)^2\right) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(S_i^2\left(2+p_i^{(2)}\right)\right),\\
\frac{\mathbf{Z}'\mathbf{Z}}{N} &\overset{a.s.}{\to} \lim_{N\to\infty}\frac{T}{N}\sum_{i=1}^{N}\left(\mathbb{E}[S_{i,T}^3] - 2\mathbb{E}[S_{i,T}^2]\mathbb{E}[S_{i,T}] + \mathbb{E}[S_{i,T}]^3\right)\\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(S_i^3\left(2+p_i^{(2)}\right)\right) + O(T^{-1});\\
\frac{\mathbf{X}'\mathbf{D}_X\mathbf{D}_X'\mathbf{X}}{N} &\overset{a.s.}{\to} \lim_{N\to\infty}\frac{T^2}{N}\sum_{i=1}^{N}\mathbb{E}[S_{i,T}^2] = T^2\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}S_i^2 + O(T);
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{D}_X\mathbf{D}_X'\mathbf{Z}}{N} &\overset{a.s.}{\to} \lim_{N\to\infty}\frac{T^2}{N}\sum_{i=1}^{N}\mathbb{E}[S_{i,T}^3] - \mathbb{E}[S_{i,T}^2]\mathbb{E}[S_{i,T}]\\
&= T\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}S_i^3\left(2(2+p_i^{(2)})\right) + O(1);\\
\frac{\mathbf{Z}'\mathbf{D}_X\mathbf{D}_X'\mathbf{Z}}{N} &\overset{a.s.}{\to} \lim_{N\to\infty}\frac{T^2}{N}\sum_{i=1}^{N}\mathbb{E}[S_{i,T}^4] - 2\mathbb{E}[S_{i,T}^3]\mathbb{E}[S_{i,T}] + \mathbb{E}[S_{i,T}^2]\mathbb{E}[S_{i,T}]^2\\
&= T\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}S_i^4\left((2+p_i^{(2)})\right) + O(1).
\end{aligned}
$$

Putting together all results from above will yield,

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{X}}{N} &\overset{p}{\to} ST + O(1),\\
\frac{\mathbf{X}'\mathbf{M}_Z\mathbf{D}_X\mathbf{D}_X'\mathbf{M}_Z\mathbf{X}}{N} &\overset{p}{\to} \Sigma_S T^2 + O(T),
\end{aligned}
$$

where $S \equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}S_i$ and $\Sigma_S \equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}S_i^2$, which in turn yield the statement of the proposition.

Bibliography

Bao, Y. and A. Ullah (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference 140*(5), 1193–1205.

Breitung, J. (2014). The Analysis of Macroeconomic Panel Data. In B. Baltagi (Ed.), *Panel Data, Oxford Handbooks*, Chapter 15, pp. forthcoming.

Breitung, J., C. Roling, and N. Salish (2016). Lm-type tests for slope homogeneity in panel data models. *The Econometrics Journal*, forthcoming.

Hsiao, C. and M. H. Pesaran (2008). Random coefficient models. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Chapter 6. Springer.

Juhl, T. and O. Lugovskyy (2014). A test for slope homogeneity in fixed effects models. *Econometric Reviews 33*, 906–935.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Pesaran, M. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics 68*, 79–113.

Pesaran, M. H., R. Smith, and K. Im (1996). Dynamic linear models for heterogeneous panels. In L. Matys and P. Serestre (Eds.), *The econometrics of panel data*, Chapter 8, pp. 145–195.

Pesaran, M. H. and T. Yamagata (2008). Testing slope homogeneity in large panels. *Journal of Econometrics 142*, 50–93.

Ullah, A. (2004). *Finite-sample econometrics*. Oxford University Press.

White, H. (2001). *Asymptotic Theory for Econometricians* (Revised Edition ed.). ACADEMIC PRESS.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *The Review of Economics and Statistics 87*(2), 385–390.