Essays on diagnostic testing in time series model

Essays On Diagnostic Testing In Time Series Models

Charisios Grivas

Thesis submitted for the degree of Doctor of Philosophy.

Department of Economics, Mathematics, and Statistics

Birkbeck College

University of London

August 2023

Declaration

I declare that the work presented in this Ph.D. thesis is my own original work. Where information has been derived from other sources, I confirm that this has been clearly and fully identified and acknowledged. No part of the thesis contains material previously submitted to this University or to any other institution for a degree.

Charisios Grivas, August 2023

Statement of Published Work

A version of the second chapter, Chapter 2 of my thesis has been published as:

Grivas, Charisios. "An Automatic Portmanteau Test For Nonlinear Dependence." Econometrics and Statistics (2023).

Statement of Conjoint Work

One out of my three primary chapters that form this thesis involves conjoint work, as specified below.

Chapter III "Bandwidth Selection for Estimators of Time-Varying Stochastic Coefficient Models" is conjoint work with Zacharias Psaradakis. Overall, my contribution amounts to two thirds of the total paper.

Acknowledgements

Throughout my PhD I have been helped, encouraged and supported by multiple individuals, and I apologize for not being able to name and thank each person individually.

First and foremost, I would like to thank my principal supervisor Zacharias Psaradakis, who has provided me with invaluable guidance and insights, given me many opportunities, and has taught me a lot about research. His contribution and kindness has been deeply appreciated and will always be remembered. I would also like to thank my secondary supervisor, Walter Beckert, for his guidance, advice and support, that have been instrumental in helping me shape my work and my academic thinking.

I have been fortunate to interact with many during my PhD, each of whom have taught me something valuable and are owed my thanks. To name a few: Weining Wang, Liudas Giraitis, Marian Vávra, Ron Smith, Yunus Aksoy, Raffaella Giacomini, Karolos Arapakis, Andreas Drichoutis, Alejandro Riaño, Nikos Paltalidis, Rubens Morita, Spyros Symeonidis, Konstantinos Adamopoulos and Jose Camarena Brenes.

The Economics Department at Birkbeck University has been a great environment to learn and conduct research. I have greatly benefited from the comments and questions of many faculty members, as well as fellow PhD students.

Funding Acknowledgements

I gratefully acknowledge the funding that made my PhD possible. I was funded by the Birkbeck School of Business, Economics and Informatics through a fee waiver and a teaching assistant scholarship.

Abstract

The Ph.D thesis, titled *Essays On Diagnostic Testing In Time Series Models*, investigates several issues related to inference in time series models. The aim is to develop a deeper understanding of issues involving hypothesis testing and inference in models that exhibit some non-linear dependence or time-varying endogeneity. This thesis is made up of five main chapters,

In the first chapter (Chapter 1) we provide a motivation for the thesis.

In the second chapter (Chapter 2), we develop a data-driven version of a portmanteau test for detecting nonlinear types of statistical dependence. The test properly controls the type I error without being sensitive with respect to the number of autocorrelations used. In addition, the automatic test is found to have higher power in simulations when compared to the standard portmanteau test, for both raw data and residuals.

In the third chapter (Chapter 3), we propose a bootstrap version of a time-varying Hausman test statistic, which compares kernel based time-varying OLS and IV estimators of regression coefficients, allowing for possible changes in the endogeneity status of the regressors over time. In this chapter, we examine the finite-sample performance of the asymptotic and the bootstrap version of the test by means of Monte Carlo simulations and we establish the asymptotic validity of a simple, easy to use bootstrap procedure. The bootstrap test has more accurate size and higher power than its asymptotic counterpart. What is more, it is demonstrated that the size and power of the bootstrap test are insensitive with respect to the choice of the bandwidth parameters. This is of particular importance since in current practice researchers use a variety of ad hoc approaches to bandwidth selection which are typically based on objective functions that address estimation concerns rather than test accuracy.

In the fourth chapter (Chapter 4), we study the problem of bandwidth choice for non-parametric instrumental variable and least square estimation for econometric models whose coefficients can vary over time either deterministically or stochastically, under both endogeneity and exogeneity. In this chapter, we compare different data-driven selectors for the smoothing parameter. We find that data-driven methods perform well for both the estimators. Quite interestingly, we find that selecting the bandwidth parameter in a data-driven way for the time-varying least square estimation under endogeneity provides a way to reduce the finite small sample bias of the estimator.

In the last chapter we summarize the results of the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Non-linearity, endogeneity, structural change and parameter instability are prominent features of many relationships among economic and financial variables. It is not surprising, therefore, that problems relating to effective detection of such features and to statistical inference in their presence have attracted a phenomenal amount of interest in econometrics and statistics. A wide variety of methods have been developed that offer ways to test for neglected nonlinearity in econometric models, for potential endogeneity of explanatory variables in regression models, and for parameter instability and structural change of various forms. At the same time, a large array of models with built-in non-linear features have been proposed, as well as parametric and non-parametric methods for optimal inference in the presence of instrumental variables and/or parameter instability of specific or general forms.

The work in this thesis contributes to several strands of this vast literature. Focusing primarily on models for time series data, we consider problems relating to diagnostic tests for non-linear dependencies, to tests for endogeneity in the presence of parameter instability of general unspecified forms, and to the selection of smoothing parameters for non-parametric inference in models with stochastically or deterministically time-varying parameters.

The problem of testing for non-linear dependence in raw data or in residuals from a parametric time series model is taken up in the second chapter of the thesis. A common limitation of most portmanteau tests for serial correlation and non-linear dependence based on sample autocorrelations is that the choice of the number of autocorrelations to be used in the construction of the relevant test statistics is arbitrary. Building on a recently proposed idea for an automatic version of a portmanteau test for autocorrelation in raw data, we develop a data-driven version of a widely used portmanteau test for detecting non-linear types of statistical dependence. The proposed test has good size properties, without being sensitive with respect to the number of autocorrelations used, the latter being chosen in an automated (i.e., data-driven) way. In addition, the automatic test has superior power when compared to the standard portmanteau test based on a pre-specified number of autocorrelations. Most importantly, the automatic test procedure is extended to deal with the case of autocorrelations based on squared residuals from a fitted model, thus providing a general diagnostic test for (second-order) non-linear dependence.

In the third chapter of the thesis, we develop a bootstrap version of a time-varying Hausman test for endogeneity in linear regression models with time-varying parameters. The test is based on non-parametric kernel least squares and instrumental variables estimators of regression coefficients, when the latter vary smoothly, either deterministically or stochastically, subject only to appropriate boundedness and smoothness conditions. What is more, the test allows for possible changes in the endogeneity status of the explanatory variables over time. This is a departure from the traditional setting in which the exogeneity status of an independent variable is treated as binary throughout the sample period. In view of the substantial empirical evidence in favour of structural change in macroeconomic and finance models, such an assumption about the time invariance of the exogeneity status of a variable appears restrictive, and it is doubtful that a conventional Hausman test is the most appropriate way of detecting failures of orthogonality between explanatory variables and unobservable disturbances. Taking as a point of departure a recently proposed time-varying version of the Hausman test, we consider its accuracy in samples of sizes that are relevant for applications. Since a test based on critical values obtained from the large-sample asymptotic null distribution of the test statistic tends to suffer from size distortions and low power, we propose using a version of the test that relies on critical values obtained from a suitable bootstrap approximation to the sampling distribution of the test statistic. Such a bootstrap-based test is shown to have more accurate size and higher power than its asymptotic counterpart. What is more, the size and power of the bootstrap test are found to insensitive with respect to the choice of the smoothing, or bandwidth, parameters that are required to construct kernel estimators of the regression parameters. The asymptotic validity of an easy-to-use bootstrap version of the time-varying Hausman test is also established.

Linear regression models with time-varying parameters and potentially endogenous explanatory variables are also at the centre of the fourth chapter of the thesis. Here, we focus on the problem of choosing the bandwidth parameter for non-parametric kernel-based least squares and instrumental variables estimators of regression coefficients. As is well known from the literature on non-parametric kernel estimation, the choice of bandwidth significantly impacts the performance of kernel estimators. Small bandwidth values can yield under-smoothed estimates which have high variance, while large values may result in over-smoothing and large bias. In our context, the potential presence of endogenous explanatory variables in the model further complicates the already difficult problem of effective bandwidth choice. To address this problem, we consider a variety of automated (i.e., data-driven) bandwidth selection methods. These include selection methods based on cross-validation, wild bootstrap and dependent wild bootstrap resampling, and a non-parametric variant of Akaike's information criterion. Unlike earlier work, which considered non-parametric regressions with fixed or random (but exogenous) explanatory variables, time-invariant coefficients, and white-noise errors, the models considered in this chapter have stochastically varying coefficients, explanatory variables that may be exogenous or endogenous, and errors which may be serially uncorrelated and homoskedastic, heterogeneously distributed, or serially correlated. It is shown that cross-validation and bootstrap-based procedures provide effective data-driven choices of the bandwidth under a variety of conditions that are relevant in econometrics.

The final chapter of the thesis summarises the main findings and concludes.

Chapter 2

An Automatic Portmanteau Test for Nonlinear Dependence

1 Introduction

One of the most popular approaches for detecting serial correlation is based on the Box–Pierce portmanteau test (see Box and Pierce (1970)), or its finite-sample correction developed by Ljung and Box (1978). The test statistic is simply the sample size times the sum of squares of the first p sample autocorrelations. A problem that practitioners often face is how to correctly specify ad hoc the order of lags p for the autocorrelation in a way that will properly control the probability of type I error while having a high power.

An attractive application of portmanteau tests is on the residuals of fitted autoregressive moving average (ARMA) models. Following the methodology of Box and Jenkins (1990), once a model has been fitted, the econometrician performs a number of tests on the residuals to check the adequacy of the model. This process of checking the residuals for any remaining serial correlation continues until the resulting residuals contain no detectable additional structure. In practice, portmanteau tests are more useful for disqualifying unsatisfactory models from consideration than for selecting the best-fitting model among closely competing candidates, as pointed out by Brockwell and Davis (2009) (page 312).

However, as Anderson (1979) observed, the autocorrelation function of the squared series can be useful in identifying non-linearity in time series. In particular, even when the series appears not to be autocorrelated, the squared series could be autocorrelated, hence revealing some form of nonlinear dependence. Motivated by this observation, McLeod and Li (1983) proposed a modification of the Ljung–Box test based on the squared series or squared residuals from a fitted model. As Luukkonen et al. (1988) demonstrate, the McLeod-Li($\mathcal{ML}$) test is particularly successful in the presence of autoregressive conditional heteroskedasticity (ARCH) and stochastic volatility, the test being asymptotically equivalent to a Lagrange Multiplier test for ARCH (see also Psaradakis and Vávra (2019)). In fact, Luukkonen et al. (1988) call it an ARCH test to "remind the reader that it is an ARCH test". Although the test generally has good control of the probability type I error, this is dependent on the number of autocorrelations p used to construct the test being carefully chosen.

The aim of this chapter is to propose a data-driven version of the McLeod–Li test that deals with the problem of selecting the number of autocorrelations p in an effective manner. Building upon the seminal work by Escanciano and Lobato (2009), we let the data determine the order p, by automatically adapting to the order of the serial correlation present. Under the null hypothesis of independent, identically distributed data (i.i.d) data, the proposed test statistic follows asymptotically a chi-square distribution with one degree of freedom. Under the alternative hypothesis of non-i.i.d. data, the test chooses p depending on the serial correlation present and is consistent. An attractive feature of the automatic version of the McLeod–Li test, is that its size properties are not dependent on the choice of the p value.

Adapting the methodology of Escanciano and Lobato (2009), our chapter contributes to the literature in two distinct ways. First, by considering squared data instead of raw data, we provide a means of identifying nonlinear dependence in a time series. Second, and more importantly, we extend the procedure to squared residuals from a fitted model, thus providing a general diagnostic test for (second-order) nonlinear dependence.

The chapter is organized as follows. In Section 2, we briefly present an overview of the McLeod-Li test, introduce the proposed test for both raw data and residuals, and establish their asymptotic properties. In Section 3, we explore the finite-sample behaviour of the test by means of Monte Carlo simulations. Finally, some concluding remarks and suggestions for possible future research are given in Section 4.

2 The McLeod and Li Portmanteau Test

In this section, we first revisit the McLeod–Li test for raw data and propose an automatic version of the test. We then consider the case of residuals from a fitted model.

2.1 Preliminaries

Consider a strictly stationary time series $\{y_t\}_{t=1}^{T}$ with $\mathbb{E}[y_t^4] < \infty$. For any $0 \leq j \leq T-1$, let $\gamma(j) = \operatorname{cov}\left(y_t^2, y_{t-j}^2\right)$ and its sample analogue $\hat{\gamma}(j) = \frac{1}{T-j}\sum_{t=1+j}^{T}\left(y_t^2 - \bar{\delta}\right)\left(y_{t-j}^2 - \bar{\delta}\right)$, with $\bar{\delta} = \frac{1}{T}\sum_{t=1}^{T} y_t^2$. The portmanteau test of McLeod and Li (1983) is based on the statistic:

$$\mathcal{ML}(p) = T\left(T+2\right)\sum_{j=1}^{p}\frac{\hat{\rho}^2\left(j\right)}{T-j}, \tag{2.1}$$

where $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$ is the lag-j sample autocorrelation of the squared series and p is the maximum lag order (number of autocorrelations) specified ad hoc by the econometrician. Under the null hypothesis that $\{y_t\}_{t=1}^T$ is i.i.d., $\rho(j) = \gamma(j)/\gamma(0) = 0$ for all $j \geq 1$ and $\mathcal{ML}(p)$ has a $\chi^2(p)$ asymptotic distribution as $T \to \infty$.

2.2 Automatic Portmanteau Test

In their novel paper, Escanciano and Lobato (2009) proposed a modification of the Ljung–Box portmanteau test that allows the data to automatically determine the number of autocorrelations on the basis of an information criterion. The proposed test statistic is the maximum value of the portmanteau test statistics penalized by a term that is an increasing function of the number of autocorrelations.

Analogously to Inglot and Ledwina (2006) and Escanciano and Lobato (2009), our automatic version of the McLeod–Li test is based on the statistic:

$$\mathcal{AML} = \mathcal{ML}(\tilde{p}),$$

where

$$\tilde{p} = \min\{p : 1 \leq p \leq d, \mathcal{L}_p = \max_{1 \leq u \leq d} \mathcal{L}_u\},$$

and

$$\mathcal{L}_u = \mathcal{ML}(u) - \pi(u, T, q).$$

Here, d serves as a fixed upper bound, q is a fixed positive number, to be defined later, and $\pi(u, T, q)$ is a penalty function that takes the form

$$\pi(u, T, q) = \begin{cases} u \log T, & \text{if } \max_{1 \leq j \leq d} \sqrt{T}|\hat{\rho}(j)| \leq \sqrt{q \log T}, \\ 2u, & \text{if } \max_{1 \leq j \leq d} \sqrt{T}|\hat{\rho}(j)| > \sqrt{q \log T}. \end{cases} \tag{2.2}$$

Note that in (2.2) the penalty function involves a switching rule between the Bayesian information criterion (BIC) criterion (Schwarz (1978)) and the Akaike information criterion (AIC) (Akaike (1974)). This combination of the two criteria is desirable since BIC is able to properly control type I error and is more powerful when there is first-order serial correlation in the data. When higher-order serial correlation is present, AIC yields more powerful tests. The threshold $\sqrt{qlogT}$ employed here is an adaptation of the standard solution for a white noise sequence of $i.i.d$ $N(0,1)$ random variables; see Inglot and Ledwina (2006) (page 126-127) for a motivation of such a data-driven rule.

Our first theorem establishes the asymptotic null distribution of the automatic McLeod–Li test.

Theorem 1. *If $\{y_t\}_{t=1}^T$ is i.i.d with $\mathbb{E}[y_t^4] < \infty$, then $\mathcal{AML}$ has a $\chi^2(1)$ asymptotic distribution as $T \to \infty$.*

The next theorem shows that the test is consistent against a fixed alternative under which $\rho(j) \neq 0$ for some

$j \geq 1$.

Theorem 2. *If $\{y_t\}_{t=1}^{T}$ is strictly stationary and ergodic with $\mathbb{E}[y_t^4] < \infty$, then the test based on $\mathcal{AML}$ is pointwise consistent, as $T \to \infty$, against the alternative $H_a^K : \rho(1) = \cdots = \rho(K-1) = 0$, $\rho(K) \neq 0$ for $1 \leq K \leq d$.*

The proofs of the theorems are in the Appendix 2.A.2.

2.3 Residual Portmanteau Test

We now consider the case where the adequacy of a fitted model is checked by using a portmanteau test based on the autocorrelations of the squared residuals. The econometrician observes a finite stretch of data $\{y_t\}_{t=1}^{T}$ from a stochastic process with mean μ satisfying:

$$y_t = \mu + \sum_{j=0}^{\infty} c(\beta, j)\epsilon_{t-j}, \tag{2.3}$$

where $c(\beta, j)$ are real weights, assumed to be known functions of an unknown finite-dimensional vector of parameters β and satisfying $\sum_{j=0}^{\infty} |c(\beta, j)| < \infty$ and $c(\beta, 0) = 1$, and $\{\epsilon_t\}_{t=1}^{\infty}$ are strictly stationary white noise errors with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[|\epsilon_t|^s] < \infty$ for some $s \geq 8$. A well-known special case of (2.3), as considered in McLeod and Li (1983), are ARMA(r_1, r_2) processes, for which the weights $c(\beta, j)$ satisfy $\sum_{j=0}^{\infty} c(\beta, j)z^j = \vartheta(z)/\varphi(z)$ for all complex $|z| \leq 1$, where $\vartheta(\cdot)$ and $\varphi(\cdot)$ are polynomials of degree r_1 and r_2, respectively, having no common roots and no roots inside or on the unit circle.

Following McLeod and Li (1983), the autocorrelation function of the squared errors $\{\epsilon_t^2\}$ can be useful in identifying nonlinear dependence. The presence of such nonlinear dependence is of importance since it can be an indication of misspecification of the fitted model, or can, if taken into consideration, lead to improved forecast accuracy (Granger and Ap (1978)). The errors $\{\epsilon_t\}$ are, of course, unobservable in practice and hence one must use residuals $\{\hat{\epsilon}_t\}_{t=1}^{T}$ in their place.

Given a consistent estimator $\hat{\theta} = (\hat{\beta}, \hat{\mu})$ of $\theta = (\beta, \mu)$, and assuming (2.3) is invertible, residuals can be computed as (e.g., Kreiss (1991)):

$$\hat{\epsilon}_t = y_t - \hat{\mu} - \sum_{j=1}^{t-1} c\left(\hat{\beta}, j\right)(y_{t-j} - \hat{\mu}). \tag{2.4}$$

Suitable estimators of θ can be obtained by quasi-maximum likelihood, instrumental variables, or least-squares methods (see, e.g., Kuersteiner (2001) and the references therein). Then, given a prespecified fixed $p \geq 1$, a test for nonlinear dependence in $\{\epsilon_t\}$ may be based on the McLeod–Li portmanteau statistic:

$$\mathcal{ML}^*(p) = T(T+2)\sum_{j=1}^{p} \frac{\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}^2(j)}{T-j}, \tag{2.5}$$

where $\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(j) = \hat{\gamma}_{\hat{\epsilon}\hat{\epsilon}}(j) / \hat{\gamma}_{\hat{\epsilon}\hat{\epsilon}}(0)$, $\hat{\gamma}_{\hat{\epsilon}\hat{\epsilon}}(j) = \frac{1}{T-j}\sum_{t=1+j}^{T}\left(\hat{\epsilon}_t^2 - \bar{\delta}_{\hat{\epsilon}}\right)\left(\hat{\epsilon}_{t-j}^2 - \bar{\delta}_{\hat{\epsilon}}\right)$ for $0 \leq j \leq T-1$, and $\bar{\delta}_{\hat{\epsilon}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2$. Under suitable regularity conditions (see Assumption 1 below), $\mathcal{ML}^*(p)$ has a $\chi^2(p)$ asymptotic distribution as $T \to \infty$ when $\{\epsilon_t\}$ are i.i.d. with $\mathbb{E}[\epsilon_t^8] < \infty$ (see Psaradakis and Vávra (2019)).

Instead of relying on an ad hoc choice for the number of estimated autocorrelations p in (2.5), we suggest employing an automatic version of the test, analogous to that introduced in the previous subsection. The test statistic is $\mathcal{AML}^* = \mathcal{ML}^*(\tilde{p})$, with $\tilde{p}$ determined in the same manner as before but now using a penalty function $\pi(u, T, q)$ that takes the form:

$$\pi(u, T, q) = \begin{cases} u \log T, & \text{if } \max \sqrt{T}|\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(j)| \leq \sqrt{q \log T}, \\ 2u, & \text{if } \max_{1 \leq j \leq d} \sqrt{T}|\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(j)| > \sqrt{q \log T}. \end{cases} \tag{2.6}$$

To establish the asymptotic properties of the automatic test, we will need the following assumption.

Assumption 1. *(a) $C(z) = \sum_{j=0}^{\infty} c(\beta, j) z^j$ is analytic and without zeros inside and on the unit circle, and differentiable with respect to β; (b) $\hat{\theta}$ is $\sqrt{T}$-consistent for θ; (c) $\sqrt{T}\,|\partial\tilde{\gamma}_{\epsilon\epsilon}(j)/\partial\theta| = O_p(1)$ for $0 \leq j \leq T-1$, where $\tilde{\gamma}_{\varepsilon\varepsilon}(j) = \frac{1}{T-j}\sum_{t=1+j}^{T}\left(\epsilon_t^2 - \bar{\delta}_{\epsilon}\right)\left(\epsilon_{t-j}^2 - \bar{\delta}_{\epsilon}\right)$ and $\bar{\delta}_{\epsilon} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$.*

These conditions, which are similar to those of Psaradakis and Vávra (2019), ensure that the residuals in (2.4) are well defined and the estimated autocorrelations $\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(\cdot)$ are consistent and asymptotically normal under the null hypothesis of i.i.d. errors. In the ARMA case mentioned earlier, with parameters estimated by conventional methods (see, e.g., Brockwell and Davis (2009)[Ch. 8]), all the requirements of the assumption are satisfied.

The asymptotic properties of the automatic test are given in the next two theorems. Here, $\rho_{\epsilon\epsilon}(j) = \gamma_{\epsilon\epsilon}(j)/\gamma_{\epsilon\epsilon}(0)$ for $j \geq 0$, where $\gamma_{\epsilon\epsilon}(j) = \operatorname{cov}\left(\epsilon_t^2, \epsilon_{t-j}^2\right)$.

Theorem 3. *If $\{\epsilon_t\}_{t=1}^{\infty}$ is i.i.d. with $\mathbb{E}[\epsilon_t^8] < \infty$ and Assumption 1 holds, then $\mathcal{AML}^*$ has a $\chi^2(1)$ asymptotic distribution as $T \to \infty$.*

Theorem 4. *If Assumption 1 holds and $\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(j) = \rho_{\epsilon\epsilon}(j) + o_p(1)$ for all $0 \leq j \leq T-1$, then the test based on $\mathcal{AML}^*$ is consistent, as $T \to \infty$, against the fixed alternative $H_a^K : \rho_{\epsilon\epsilon}(1) = \cdots = \rho_{\epsilon\epsilon}(K-1) = 0$, $\rho_{\epsilon\epsilon}(K) \neq 0$ for $1 \leq K \leq d$.*

The proofs of the theorems are in the Appendix 2.A.2.

Note that the assumption that $\mathbb{E}[\epsilon_t^8] < \infty$ is a standard requirement for $\mathcal{ML}$-type tests (e.g., McLeod and Li (1983), Psaradakis and Vávra (2019)) and ensures that sample autocovariances of squared errors have a well-defined asymptotic distribution when the errors are not autocorrelated. It may be possible to relax the moment condition somewhat along the lines of Anderson (1991), although this would typically require imposing stronger restrictions on the parameters of the data-generating process. A different approach that has been considered in the literature relies on using the absolute value of a time series or log-squared time series in the construction of $\mathcal{ML}$ test statistics as these only require the existence of the fourth moment. These approaches, however, are beyond the scope of the chapter and the interested reader is referred to Pérez and Ruiz (2003), and the references therein, for some discussion.

It is also worth remarking that methods based on automatic model selection typically entail biases, as shown by Leeb and Pötscher (2005), and the sampling distributions of estimators and test statistics may be affected as a result. Ideally, any such effects should not be ignored post estimation. When the McLeod–Li test, or its automatic version, are applied to the residuals of a particular model to test for possible nonlinear dependence, the diagnostic testing may affect subsequent inference. However, a number of strategies have been proposed to deal with this issue, as pointed out by Fenga and Politis (2011). For example, Broersen and de Waele (2004) suggest a model selection method based on hierarchical models. They develope a model selection strategy limiting the selection to ARMA$(r, r-1)$ candidate models and show that this does not lead to a lower-quality model being selected even in finite samples. On the other hand, Buckland et al. (1997) consider bootstrap as a way of incorporating model uncertainty into inference. The bootstrap, with model selection applied independently to each resample, allows inference without conditioning on a single selected model.

2.4 Linearity Test

A time series $\{y_t\}$ is sometimes considered to be linear if it admits a representation such as (2.3) with respect to an i.i.d. sequence $\{\epsilon_t\}$. This is the notion of linearity considered in McLeod and Li (1983), Lawrance and Lewis (1987), Berg et al. (2010) and Giannerini et al. (2015), among others. Within this framework, deviations from the i.i.d. assumption about the errors $\{\epsilon_t\}$ are viewed as evidence of nonlinear behaviour of $\{y_t\}$. Hence, one may employ tests based on the $\mathcal{ML}^*$ or $\mathcal{AML}^*$ statistics constructed from the residuals of a suitable approximation to (2.3) as general portmanteau tests for detecting deviations from linearity. Lee et al. (1993), for example, use a first-order autoregressive model to obtain the required residuals, while McLeod and Li (1983) rely on autoregressive models the order of which is estimated by means of information criteria. Note however, that this is not the only notion of non-linearity in the literature. For example, Hannan (1973) considers a different characterization based on the properties of one-step-ahead linear predictors.

3 Simulation Results

In this section, Monte Carlo experiments are carried out to investigate the finite-sample performance of the $\mathcal{AML}$ test. The main objective is to compare the properties of the automatic $\mathcal{ML}$ to the standard $\mathcal{ML}$, in the presence and absence of nonlinear serial dependence in the errors of a model. To conserve space, we only consider the case of residuals with a slight abuse of notation, denoting as $\mathcal{AML}$ and $\mathcal{ML}$ the tests applied to the residuals instead of $\mathcal{AML}^{\star}$ and $\mathcal{ML}^{\star}$. The case of raw data is provided in the supplement.

In all simulations in this section, 10,000 independent artificial time series y_t of length $100+T$ with $T \in \{50, 100, 200, 500\}$ are generated , but only the last T observations for each series are used. We consider $p/T = 2.5\%,\ 5\%,\ 7\%$ and 10% for the original test. The nominal level considered, for comparing the empirical significance levels of $\mathcal{AML}$ and $\mathcal{ML}$, is $\alpha = 5\%$. To conserve space, simulation results for nominal levels 0.01 and 0.10 are not reported here but provided similar results. We set $d = 75$, as in Escanciano and Lobato (2009), except when $T = 50$ where we use $d = 25$,and $q = 3.6$ following our simulations in the next section. Note that in both Escanciano and Lobato (2009) and Inglot and Ledwina (2006) q is set to 2.4.

3.1 Choice of q

As discussed in section 2 above, expression (2.2) involves a fixed parameter q which provides a switching rule between the BIC and AIC criteria. In this subsection, we provide evidence to support our choice of $q = 3.6$ used in some of the Monte Carlo simulations, in a similar manner to Escanciano and Lobato (2009) and Inglot and Ledwina (2006). Specifically, we estimate the empirical size of the test by Monte Carlo simulation for a sample of $T = 500$. For all the simulations in this subsection, $\mathcal{AML}$ is applied to the residuals obtained from an AR(1) model with a coefficient $\beta \in \{\pm 0.5, \pm 0.9\}$, for eight different values of q, ($q = 1.2,\ 1.5,\ 1.8,\ 2.4,\ 2.7,\ 3,\ \infty$). Similar results are obtained for other values of the β parameter but are not provided due to space constraints.

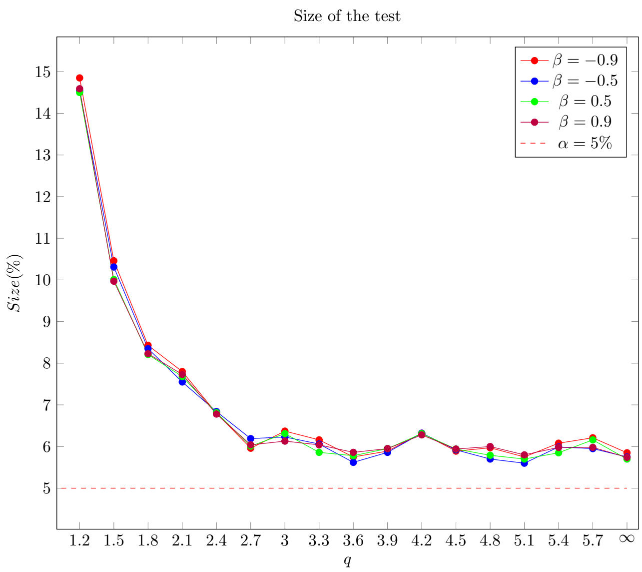


Figure 2.1: Rejection frequencies at 5% level for the $\mathcal{AML}$ test on the residuals of an $AR(1)$ model, where the error follows a standard normal distribution, and for different values of constant q. Simulations are based on $T = 500$.

Figure 2.1 above, presents graphically the rejection probabilities under the null, where the errors are generated from a standard normal distribution. Figure 2.1 shows that for $q > 3$ the rejection probability becomes relative flat near 5% and hence the value of $q = 3.6$ suffices to properly control for type I error.

We also provide some further justification for our choice for q, under the alternative hypothesis. Specifically, in Figure 2.2 below, we report the rejection frequencies for the automatic test under the alternative for $T = 500$.

We employ an EGARCH(1,1) model, which we will revisit in subsection 3.2, as model 5. Simulations based on a stochastic volatility model provided similar results and are hence omitted. Again, Figure 2.2 shows that $q = 3.6$ is a reasonable choice with high power, that does not vary much.

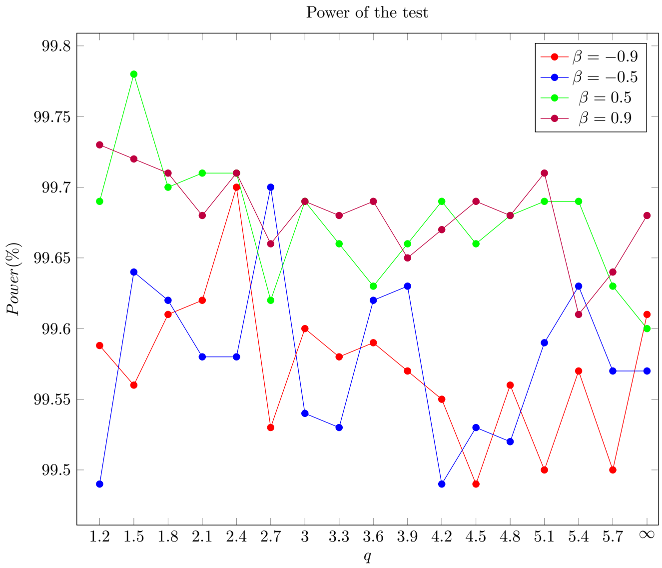


Figure 2.2: Rejection frequencies at 5% level for the $\mathcal{AML}$ test on the residuals of an $AR(1)$ model, where the error follows a EGARCH, and for different values of constant q. Simulations are based on $T = 500$.

3.2 Level and Power of Tests based on Residuals

In this subsection, we examine the application of the test on residuals. Following McLeod and Li (1983), we consider the residuals of an AR(1) model, that is:

$$y_t = \beta y_{t-1} + \epsilon_t \tag{2.7}$$

where $\beta \in \{0, \pm 0.6, \pm 0.9\}$. We estimate that model by least squares and allow for a constant. Under the null, the errors $\{\epsilon_t\}$ are an i.i.d. sequence, following:

1. a standard normal distribution,
2. a t-distribution with 9 degrees of freedom.

The latter choice is made so that the error term satisfies the finite eighth moment condition. A log-normal distribution was also considered as an alternative to the t-distribution but provided similar results. Under the alternative, we allow the errors to have one of the structures defined below.

Under the alternative, the following data-generating processes(DGPs) are used in the simulations for the error term:

1. $\epsilon_t = v_t \sigma_t$ where $\sigma_t^2 = 0.001 + 0.05\epsilon_{t-1}^2 + 0.90\sigma_{t-1}^2$,

2. $\epsilon_t = v_t \exp(\sigma_t^2)$, where $\sigma_t^2 = 0.936\sigma_{t-1}^2 + 0.32u_t$,

3. $\epsilon_t = v_t v_{t-1}$,

4. $\epsilon_t = v_{t-2}v_{t-1}(v_{t-2} + v_t + 1)$,

5. $\epsilon_t = v_t\sigma_t$ where $\log\sigma_t^2 = 0.001 + 0.5|v_{t-1}| - 0.2v_{t-1} + 0.95\log\sigma_{t-1}^2$,

6. $\epsilon_t = -0.5\epsilon_{t-1}I(\epsilon_{t-1} \leq 1) + 0.4\epsilon_{t-1}I(\epsilon_{t-1} > 1) + u_t$,

7. $\epsilon_t = -0.5\epsilon_{t-1}\{1 - G(\epsilon_{t-1})\} + 0.4\epsilon_{t-1}G(\epsilon_{t-1}) + u_t$,

8. $\epsilon_t = 0.8u_{t-2}^3 + u_t$.

In all the DGPs above, $\{u_t\}$ and $\{v_t\}$ are i.i.d. standard normal random variables independent of each other, $I(A)$ is the indicator of event A, and $G(x) = 1/(1+e^{-x})$ is the logistic distribution function. The DGPs cover a wide variety of nonlinear processes often encountered in economics and finance. Models 1-4 are taken from Romano and Thombs (1996) and represent a GARCH process, a stochastic volatility process, an 1-dependent process, and an uncorrelated non-martingale-difference process, respectively. 5 is an EGARCH process taken from Escanciano and Lobato (2009). The remaining three DGPs are taken from Psaradakis and Vávra (2019); they represent a threshold AR (TAR) [6], a smooth-transition AR [7], and a nonlinear MA (NLMA) [8]. Bilinear models were also considered but provided similar results. The Monte Carlo rejection frequencies of the conventional $\mathcal{ML}$ test and its automatic version, $\mathcal{AML}$ (at nominal level 5%) are shown in Figure 2.3-Figure 2.7. Under the null, the $\mathcal{AML}$ has empirical size close to the nominal level while the size of $\mathcal{ML}$ varies substantially. Figure 2.3 presents the empirical size of the tests under the null of i.i.d standard normal errors and t-distributed errors. The empirical size of the $\mathcal{ML}$ test seems to be quite close to the nominal level only for $T = 100$, and varies a lot depending on the choice of p, while the bigger the sample the more it deviates from 5%. The empirical size of the $\mathcal{AML}$, on the other hand, does not differ significantly from the nominal level regardless of the sample size T. It is worth mentioning that the $\mathcal{AML}$ experiences small size distortions for a 1% significance level for small samples. On the other hand, the automatic test seems to have proper size for significance levels of 10% independently of the sample size. Nevertheless, we argue that for sample sizes typically encountered in finance, the automatic test has size close to nominal at all significance levels.

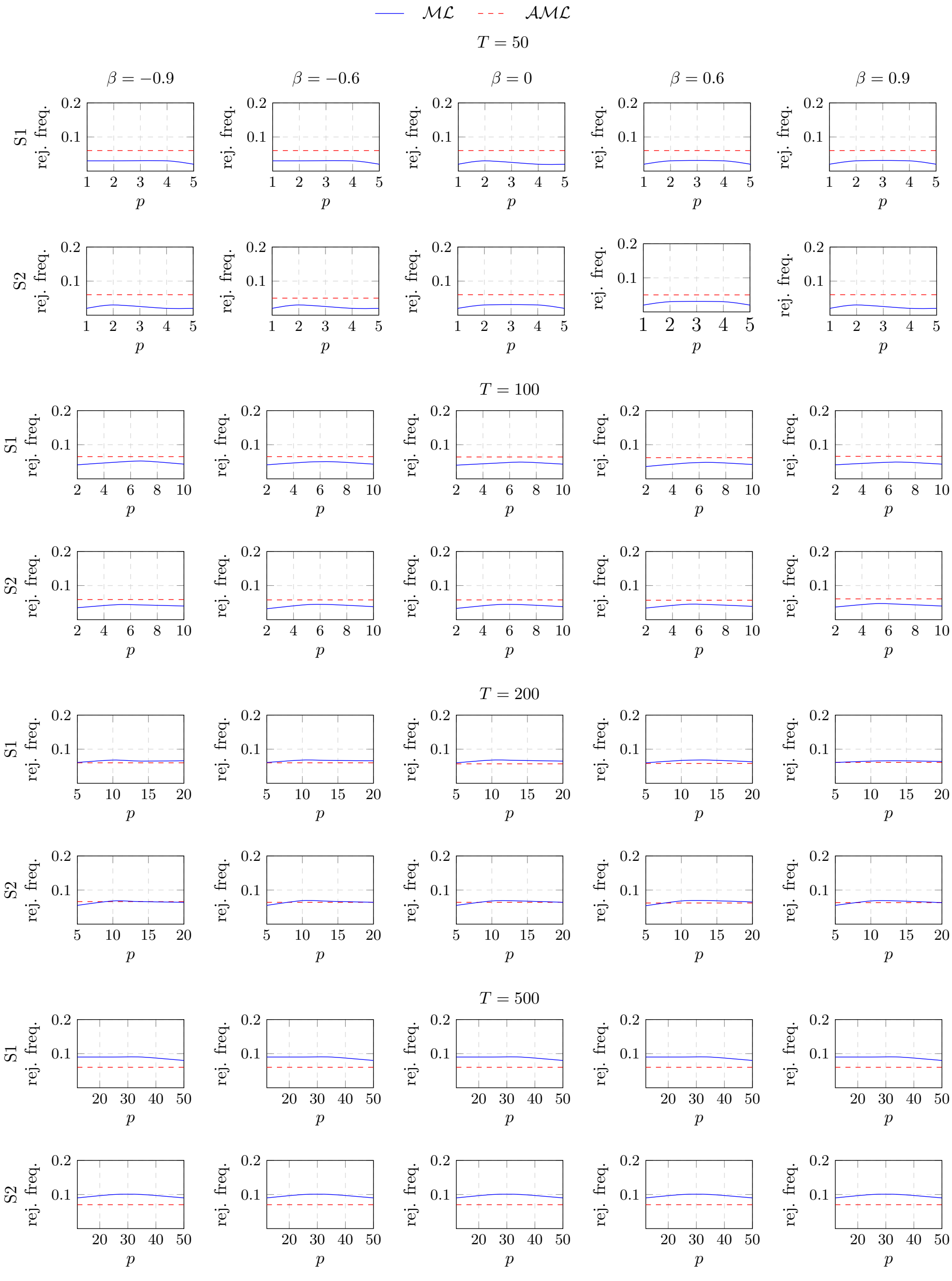


Figure 2.3: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ for the residuals under the null, at 5% significance level and for different sample sizes.

In Figure 2.4-Figure 2.7 we show the rejection frequencies under the alternative. For virtually all DGP, and across all sample sizes, the $\mathcal{AML}$ outperforms the $\mathcal{ML}$. The automatic test has higher rejection frequencies than the $\mathcal{ML}$ not only for ARCH models but also for a variety of different nonlinear models. Also, the value of the β does not seem to affect the size and the power of the tests.

Figure 2.4 below depicts the rejection frequencies for a sample size of $T = 50$. The $\mathcal{AML}$ test statistic provides higher power for all DGP, with the most obvious case for 5, while the power of the $\mathcal{ML}$ depends somewhat on the choice of p.

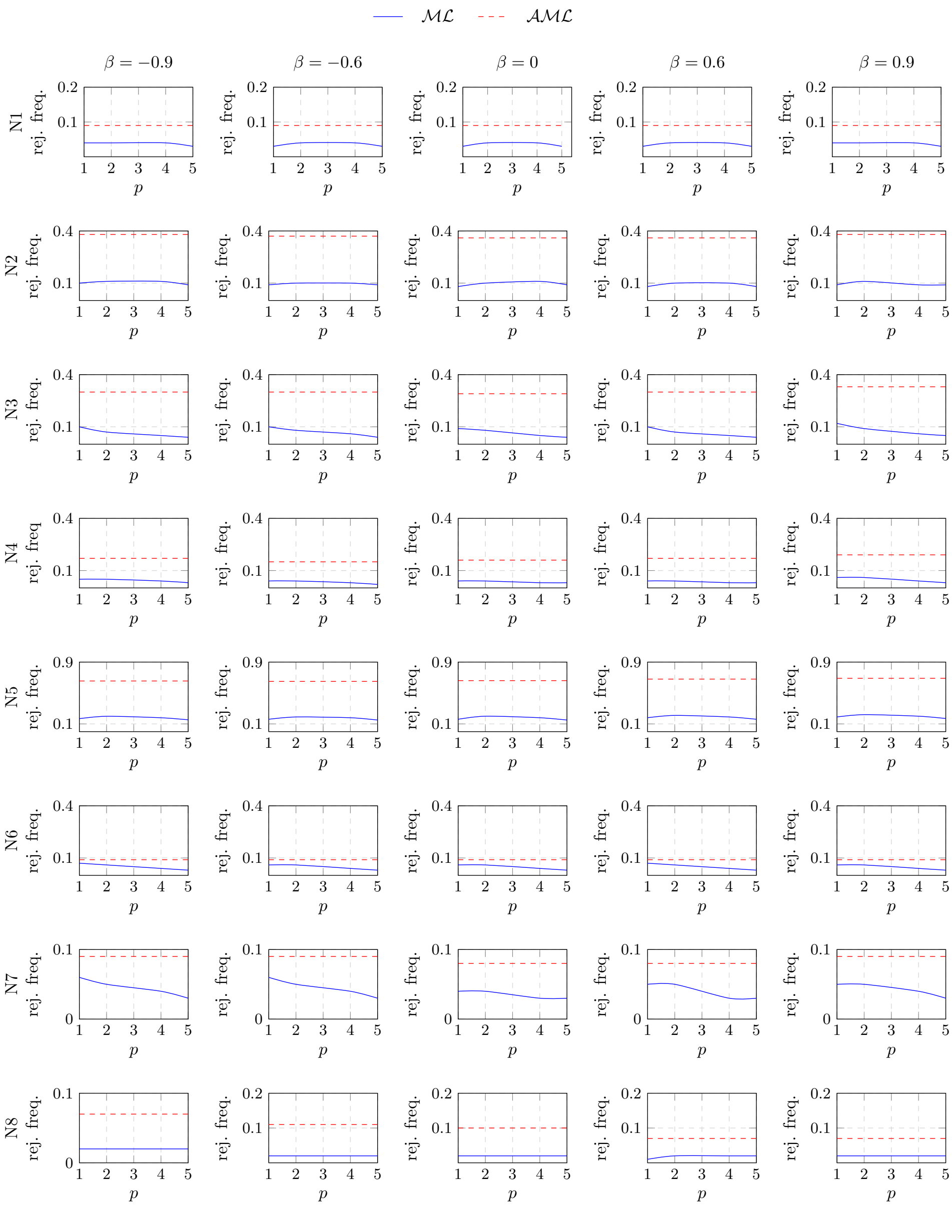


Figure 2.4: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ for the residuals under the alternative for $T = 50$ at 5% singificance level.

Figure 2.5 shows the rejection frequencies for $T = 100$. Again the $\mathcal{AML}$ provides higher power compared to $\mathcal{ML}$. The sensitivity of the $\mathcal{ML}$ to the choice of p is more apparent here. Note the increase in power of the automatic test for 5 and 4. Similarly as above, the value of the β does not seem to affect the power of the tests.

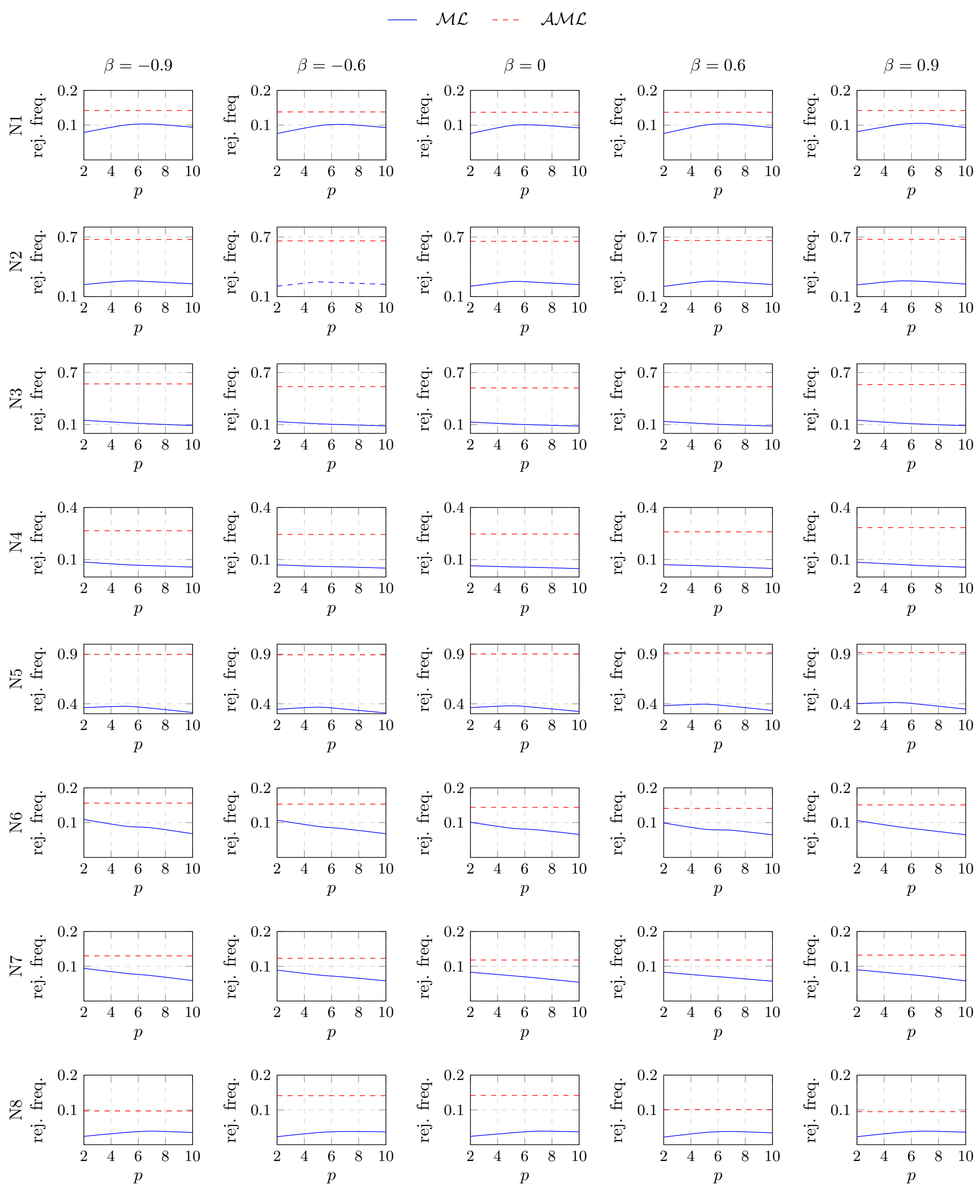


Figure 2.5: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ for the residuals under the alternative, for $T = 100$ and at 5% singificance level.

Moving on to sample sizes that are most common in finance, we see that our main conclusions remain unaffected as seen in Figure 2.6 and Figure 2.7. Starting from Figure 2.6 and a sample size of $T = 200$ we notice that the $\mathcal{AML}$ has higher power than the $\mathcal{ML}$, while the latter seems to have a large sensitivity to the parameter p.

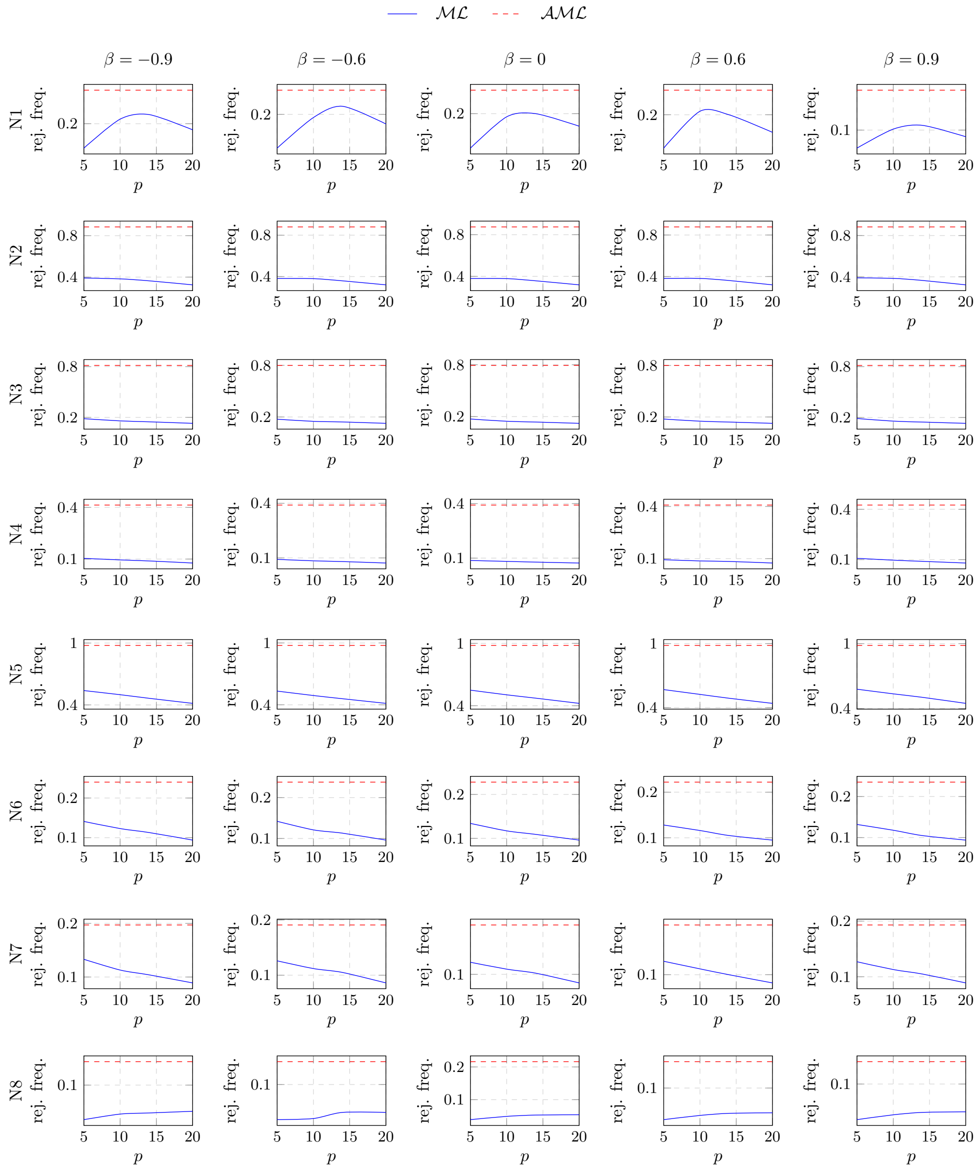


Figure 2.6: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ for the residuals under the alternative, for $T = 200$ and at 5% singificance level.

For example, for 1, the power of the $\mathcal{ML}$ is stable and around 0.2 for values of p close to 10 but it goes below 0.2 for values less or greater than 10. Similarly, for 5-7 the $\mathcal{ML}$ has a power around 0.15 when $p = 5$ while its power diminishes for other values of p. Finally, similar results are obtained for $T = 500$ as depicted in Figure 2.7.

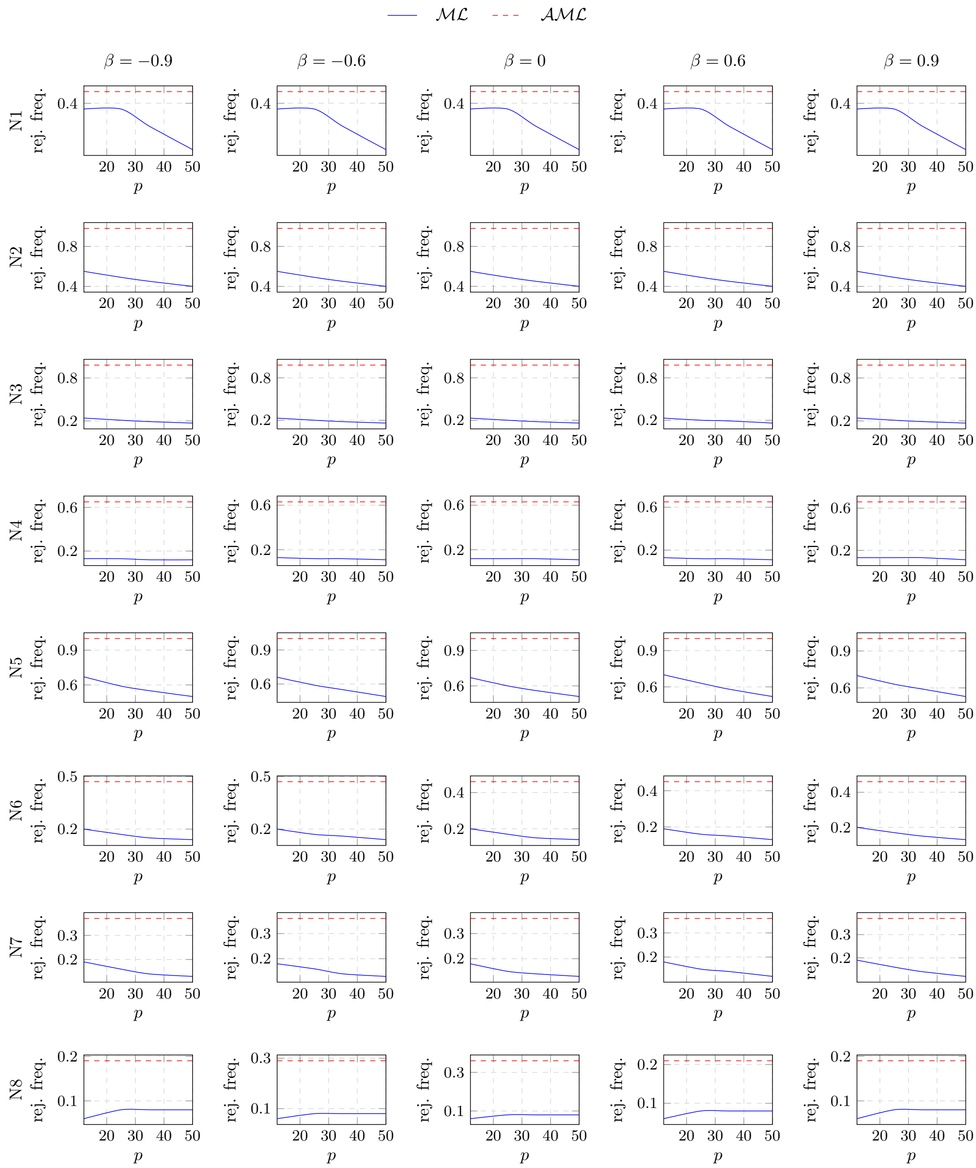


Figure 2.7: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ for the residuals under the alternative, for $T = 500$ and at 5% singificance level.

4 Summary and Conclusion

A data-driven version of the McLeod-Li portmanteau test for detecting nonlinear dependence is proposed. The proposed test is easy to implement, has a chi-square asymptotic distribution and, most importantly, it properly controls the probability of type I error for sample sizes that are common in applications. The simulation results for both raw data and residuals indicate good size and power properties in finite samples.

Further research could focus on choosing the parameter q in a data-driven way too. This could be done for example by means of bootstrapping or subsampling. Other interesting extensions of the automatic procedure would be to tests based on cross-correlations of the type considered in Psaradakis and Vávra (2019), and tests for multivariate time series and comparisons with the procedures developed in Escanciano et al. (2013).

2.A Appendix

2.A.1 Additional Simulation Results For Raw Data

In this section, we compare the $\mathcal{AML}$ with $\mathcal{ML}$ for raw data by means of Monte Carlo simulations. The raw data are assumed to follow the data generating processes(DGP) described in Section 3.2 of the main text.

Figure 2.8-Figure 2.13 present the rejection frequencies from the simulation studies conducted in section 3.2 of the main text. Figure 2.8 presents the results under the null of a i.i.d $\mathbf{N}(0,1)$ (S1), and t-distribution with 9 degrees of freedom (S2) for four different sample sizes, namely $T = 50$, $T = 100$, $T = 200$, $T = 500$ at 5% significance level. The size of the $\mathcal{AML}$ is close to nominal for all different DGP and across all sample sizes while the size of $\mathcal{ML}$ varies with p and the sample size. For example, the size of $\mathcal{ML}$ for $T = 500$ is close to 10% while the size of $\mathcal{AML}$ remains close to 5%.

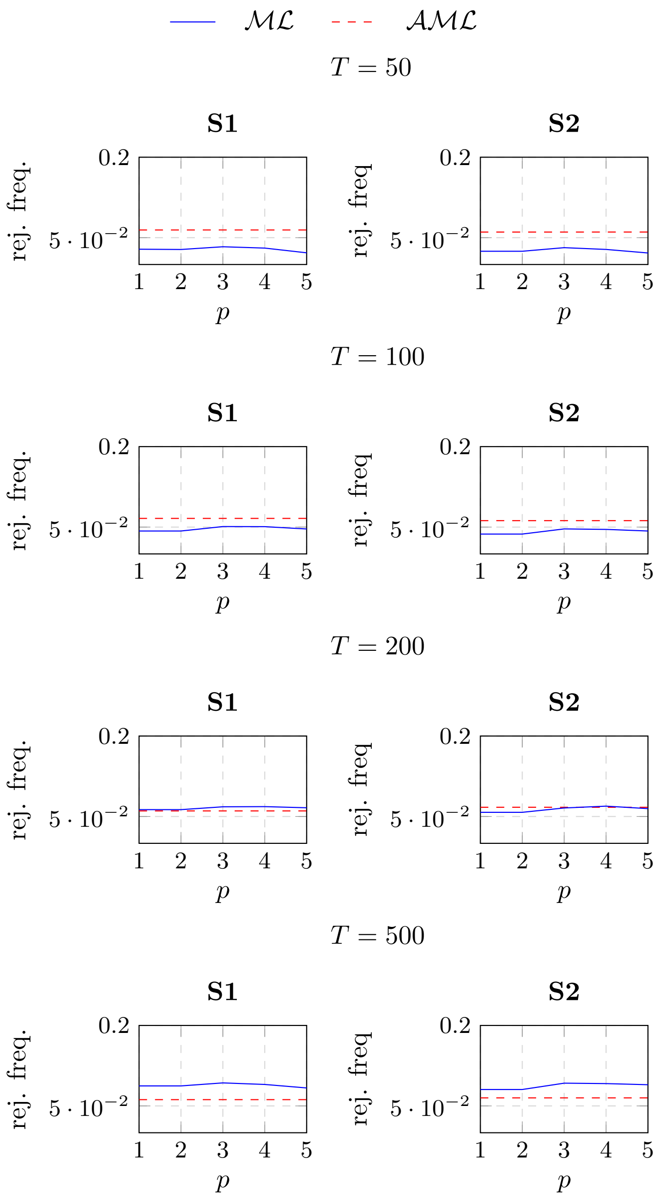


Figure 2.8: Rejection frequencies under the null at 5% significance level.

The rest of the graphs, namely Figure 2.9-Figure 2.13 depict the rejection frequencies under the alternatives as presented in section 3.2 of the main text. The DGPs cover a wide variety of nonlinear processes often encountered in economics and finance. For all DGPs and all sample sizes, the $\mathcal{AML}$ outperforms the $\mathcal{ML}$.

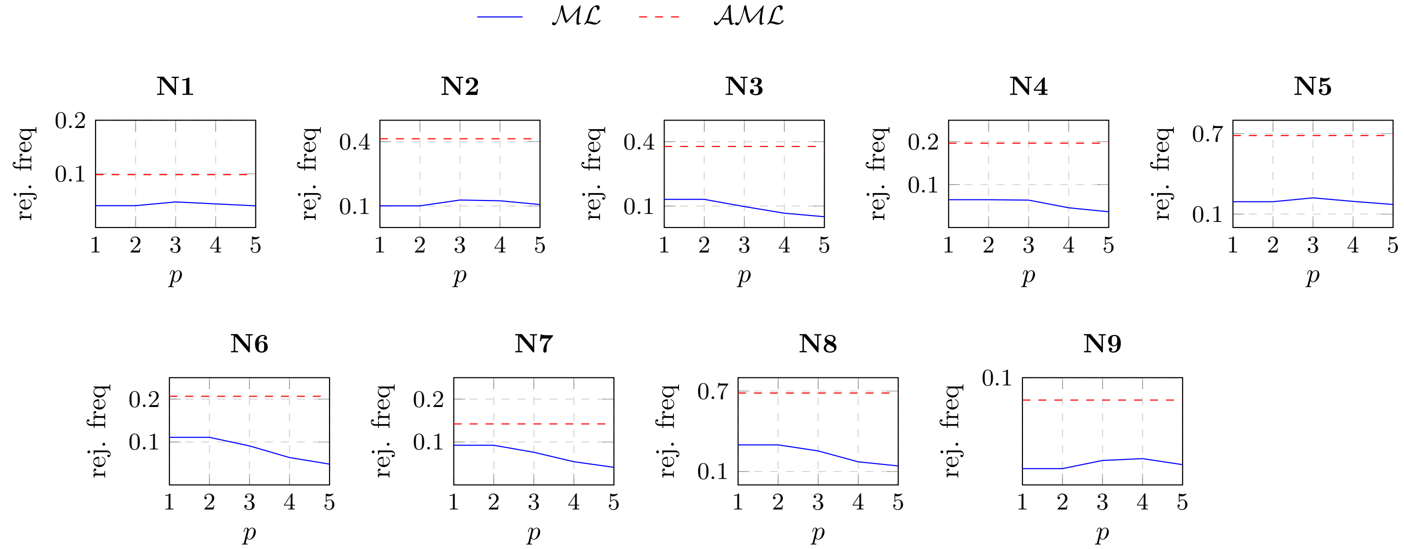


Figure 2.9: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ under the alternative, for T=50, at 5% significance level.

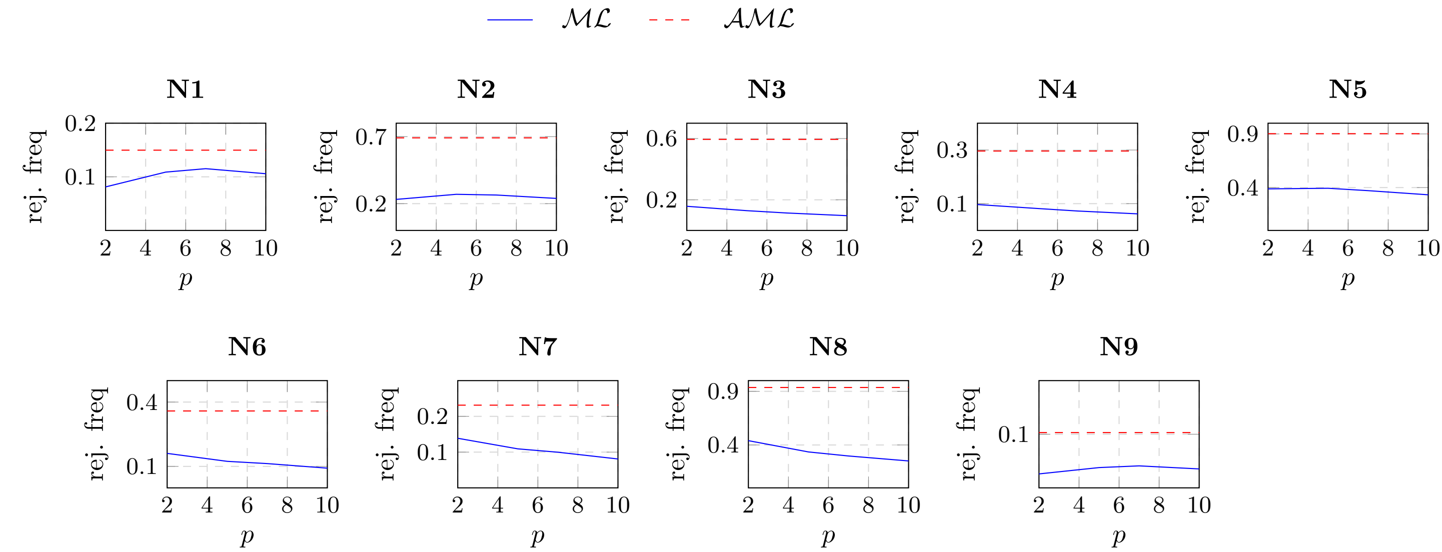


Figure 2.10: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ under the alternative, for $T = 100$, at 5% significance level.

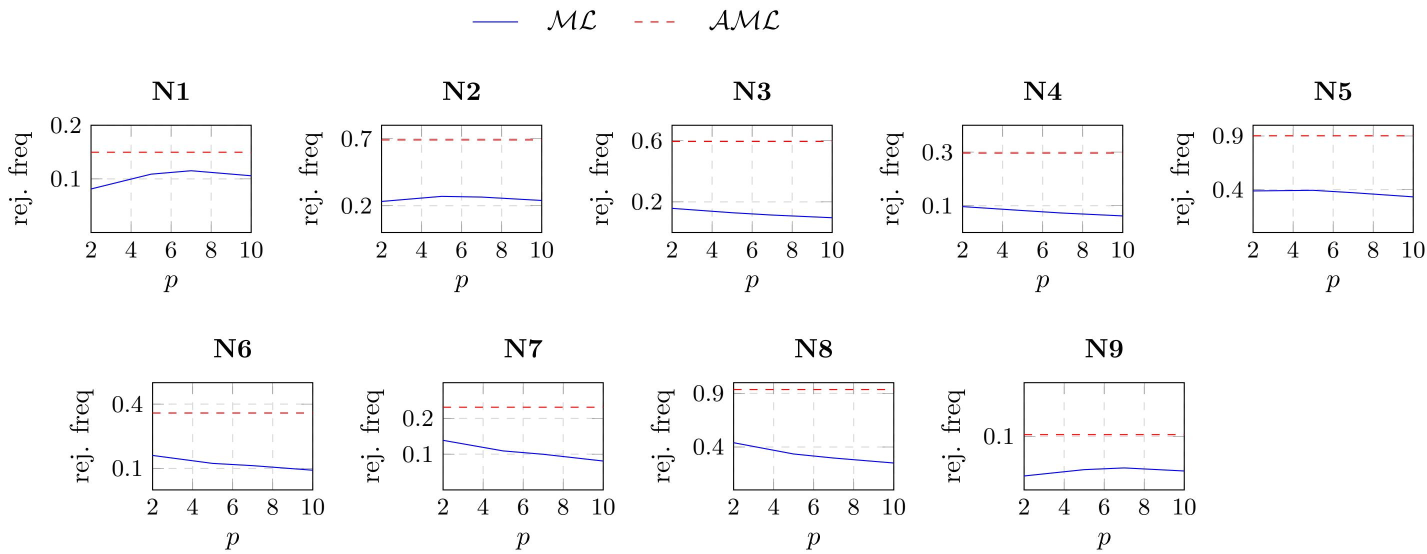


Figure 2.11: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ under the alternative, for $T = 100$, at 5% significance level.

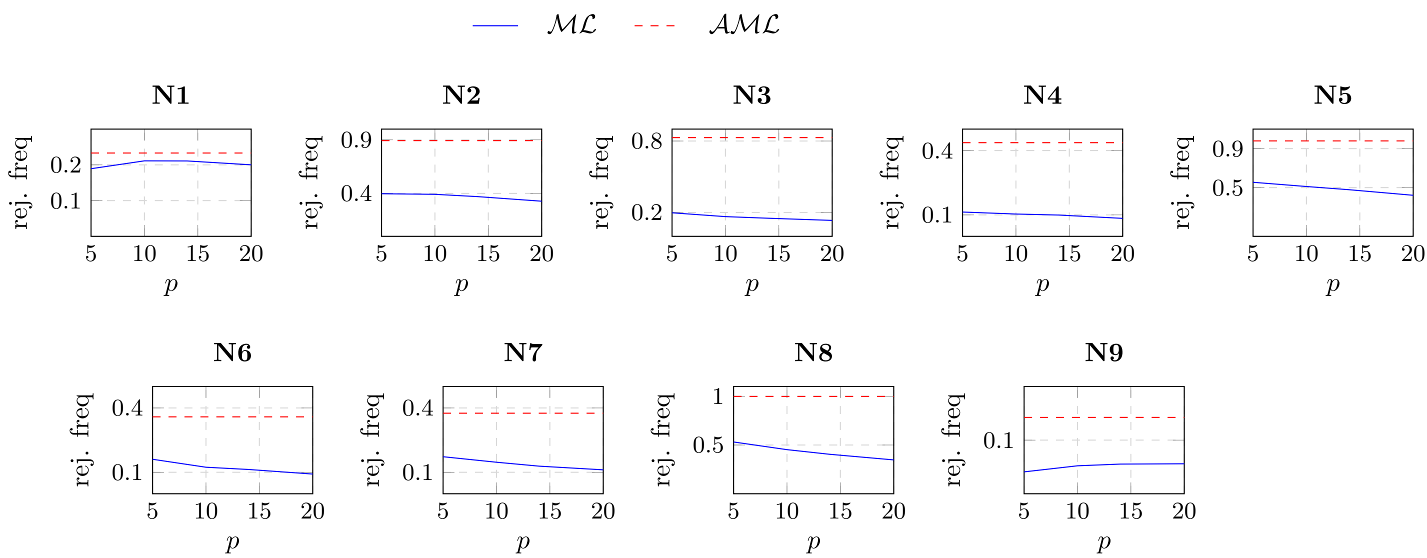


Figure 2.12: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ under the alternative, for $T = 200$, at 5% significance level.

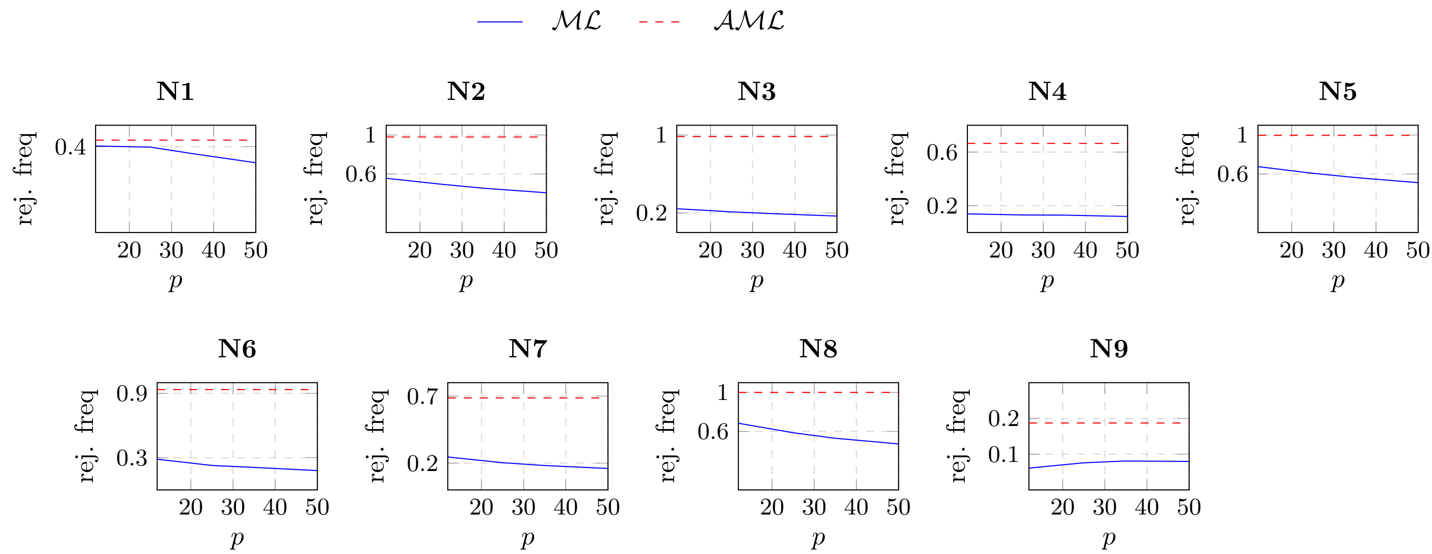


Figure 2.13: Rejection frequencies of $\mathcal{ML}$ and $\mathcal{AML}$ under the alternative, for $T = 500$, at 5% significance level.

2.A.2 Proofs

Proof of Theorem 1: Define

$$p_{BIC} = min\{m : 1 \leq m \leq d; L_{BIC}(m) \geq L_{BIC}(h), h = 1, 2, ..., d\},$$

where $L_{BIC}(p) = ML(p) - p\ logT$. We need to prove that:

$$lim_{T\to\infty} P(\tilde{p} = p_{BIC}) = 1, \tag{2.8}$$

and

$$lim_{T\to\infty} P(p_{BIC} = 1) = 1, \tag{2.9}$$

Under the null of $i.i.d$ data, $\sqrt{T}\hat{\rho}$ has a $N(0, I)$ asymptotic distribution, where $\hat{\rho} = (\hat{\rho}(1), ..., \hat{\rho}(p))$ and I is the identity matrix.

Now, consider the event

$$A_T(q) = \{\ max_{1\leq j\leq d}\sqrt{T}|\hat{\rho}(j)| > \sqrt{q\ log\ T}\}$$

and assume $q \geq 2$. We have $\max_{1\leq j\leq d}\sqrt{T}|\hat{\rho}(j)| = O_p(1)$ under the null, so that $P(A_T(q)) \to 0$, which implies that (2.8) holds. We also have

$$P(p_{BIC} = 1) = 1 - \sum_{i=2}^{d} P(p_{BIC} = i) \geq 1 - \sum_{i=2}^{d} P(L_{BIC}(i) \geq L_{BIC}(1)) \tag{2.10}$$

and,

$$P(L_{BIC}(i) \geq L_{BIC}(1)) \leq P(ML \geq (i-1)\ log\ T)$$

and since under the null ML is $O_p(1)$ we conclude that (2.9) holds. Therefore, Theorem 1 follows from an application of Lindeberg-Lévy CLT for $i.i.d$ random variables(see Billingsley (2013)27.2, p.359). $\square$

Proof of Theorem 2: Define

$$p_{AIC} = min\{m : 1 \leq m \leq d; L_{AIC}(m) \geq L_{AIC}(h), h = 1, 2, ..., d\}$$

where $L_{AIC}(p) = ML(p) - 2\,p$. We need to prove that:

$$\lim_{T\to\infty} P(\tilde{p} = p_{AIC}) = 1 \tag{2.11}$$

and

$$\lim_{T\to\infty} P(p_{AIC} \geq K) \to 1 \tag{2.12}$$

Consider the event

$$A_T(q) = \{\max_{1\leq j\leq d} \sqrt{T}|\hat{\rho}(j)| < \sqrt{q\ log\ T}\}$$

Then for $K \leq d$ we have $\hat{\rho}(K) \to \rho(K) \neq 0$ by the Ergodic Theorem and

$$P(A_T(q)) \leq P\left(\sqrt{T}|\hat{\rho}(K)| < \sqrt{q\ log\ T}\right) \to 0$$

and so (2.11) holds. Now for (2.12), we have:

$$\begin{aligned} P(p_{AIC} = k) &\leq P(L_{AIC}(k) \geq L_{AIC}(K)) \\ &\leq P(|ML(k)| \geq\ 2(k-K) + |ML(K)|) \to 0 \end{aligned}$$

which follows directly from an application of the Ergodic Theorem and so (2.12) also holds. Hence, $\forall\ M > 0$

$$\begin{aligned} P(ML(\tilde{p}) \leq M) =& P(ML(\tilde{p}) \leq M \cap \tilde{p} \geq K) + o(1) \\ &\leq P\left(|\hat{\rho}(K)|^2 \leq M\right) + o(1) \\ &= o(1) \end{aligned}$$

and so $ML(\tilde{p}) \to \infty$, asymptotically and hence the test is consistent against H_a^K, $\forall K \leq d$ $\square$

Proof of Theorem 3: By a similar argument as in the proof of Theorem 1, under the null, and and by defining

$$p_{BIC} = min\{m : 1 \leq m \leq d; L_{BIC}(m) \geq L_{BIC}(h), h = 1, 2, ..., d\}$$

where $L_{BIC}(p) = ML(p) - p\ logT$ and $ML(p)$ refers now to the statistic based on residuals we obtain the following:

$$lim_{T\to\infty}P(\tilde{p} = p_{BIC}) = 1 \tag{2.13}$$

$$lim_{T\to\infty}P(p_{BIC} = 1) = 1 \tag{2.14}$$

. □

Proof of Theorem 4: Define

$$p_{AIC} = min\{m : 1 \leq m \leq d; L_{AIC}(m) \geq L_{AIC}(h), h = 1, 2, ..., d\}$$

where $L_{AIC}(p) = ML(p) - 2p$. Under the alternative, we prove that:

$$\lim_{T\to\infty} P(\tilde{p} = p_{AIC}) = 1 \tag{2.15}$$

and,

$$\lim_{T\to\infty} P(p_{AIC} \geq K) \to 1 \tag{2.16}$$

Now consider the event

$$A_T(q) = \{\max_{1\leq j\leq d} \sqrt{T}|\hat{\rho}_{\hat{\epsilon}\hat{\epsilon}}(j)| < \sqrt{q\ log(T)}\} \tag{2.17}$$

By a similar argument as in Theorem 2 we can show that

$$P(A_T(q)) \to 0$$

and so (2.15) hold and the rest of the proof follows for (2.16). □

Chapter 3

Testing For Time-Varying Exogeneity: A Bootstrap Approach

1 Introduction

Endogeneity status of a variable is usually treated as binary so that a variable is allowed to be either exogenous or endogenous throughout the whole sample. This might seem restrictive in the presence of changes in regime resulting from, among others, significant economic and financial events or changes in government policies. For example, the exogeneity status of the short-term interest rate in a model for inflation may change over a period in which the policy of the central bank switched from inflation targeting to exchange-rate targeting. Consequently, specification tests of the type proposed by Hausman (1978) that usually guide the choice of the IV estimator over the LS need to allow for time-variation. Recently Giraitis et al. (2021)(GKM) proposed a non-parametric kernel-based, estimation for time-varying IV regression and derived time-varying versions of a Hausman exogeneity test, allowing for changes in the endogeneity status of the regressors over time. Although the finite-sample bias of the estimators is found by GKM to be relatively small, time-varying Hausman tests appear to suffer from size distortions and low power. Furthermore, these distortions are quite sensitive to the choice of the bandwidth parameters used to construct the relevant non-parametric estimators.

This paper proposes to use suitable bootstrap procedures to approximate the distribution of the time-varying Hausman test statistic under the null hypothesis of exogeneity, and thus obtain critical values and p-values for the exogeneity test. In the present context, bootstrap-based versions of the Hausman test are straightforward to implement, and as our Monte Carlo simulations reveal, control the test size better than the asymptotic test, that is the test that uses critical values from the asymptotic null distribution of the test statistic. More importantly perhaps for the time variant case studied here, the bootstrap test is quite insensitive to the bandwidth values used. In addition, we provide a consistency result for the time-varying LS estimator in the subsection 3.A.2.

This might be of independent interest as it is obtained under more general conditions than Giraitis et al. (2014).

In much of the literature, IV inferential procedures and related Hausman-type tests typically rely on the assumption that both the functional form and parameters of the model under consideration remain constant over time. Similarly, the relationship between the endogenous variables and the instruments is normally assumed to remain unchanged. These assumptions, although crucial, are obviously open to criticism. Zhang et al. (2008) for instance, argue that the conflicting conclusions about the importance of key variables in the determination of inflation based on the NKPC may be due to neglected parameter variation.

In view of these difficulties, attempts have been made to allow for structural breaks in models the parameters of which are estimated by IV. Prominent among these is Hall et al. (2012) who extended the framework of Bai and Perron (1998) to linear models with endogenous regressors. A limitation of this approach, however, is that the structural breaks are considered to be deterministic in nature. More recently, Giraitis et al. (2021) relaxed considerably this requirement. Building upon the work of Chen (2015a) and Giraitis et al. (2014), GKM proposed a non-parametric, kernel-based, estimation and inferential theory for time-varying IV regressions, allowing for both deterministic and random coefficients. What is more, their framework allows for possible changes in the endogeneity status of the explanatory variables. Tests for such changes in endogeneity may be performed using time-varying Hausman-type tests which, in this set up, rely on time-varying LS and IV coefficient estimates. Based on the simulations provided by GKM the true and nominal levels of the test differ when asymptotic critical values are used whilst these differences vary considerably with the bandwidth.

The ability of bootstrap procedures to provide accurate approximations to the sampling distributions of estimators and test statistics is well documented in the literature(see, e.g. Beran and Ducharme (1991)) and Hall (1992) among others). The basic idea of bootstrap testing amounts to treating the sample as the population and drawing a large number of bootstrap samples from a distribution that obeys the null. These samples are then used to compute the bootstrap critical values and p-values of the test. In many cases, the use of bootstrap critical values over asymptotic ones can provide significant reduction in errors in the level of the test statistic(c.f Horowitz (2001)). In the more familiar setup of regression models with time-invariant parameters fu Wong (1996) employed a bootstrap-based version of the conventional full-sample Hausman test and showed that it can provide substantial improvements over the corresponding asymptotic test.

The qualitative and quantitative differences between the bootstrap and asymptotic versions of the time-varying Hausman test are illustrated empirically by considering an emprical model for the US Phillips curve with time-varying parameters. Models for the NKPC have attacted a lot of attention in the literature as they are used to identify the forward-looking components of inflation and the trade-off between inflation and unemployment over the cycle. Following GKM, unemployment is used as a forcing variable for inflation. The bootstrap version of the time-varying Hausman test seems to be less sensitive to the choice of the bandwidth parameter than the asymptotic test, while also suggesting time endogeneity of unemployment for a longer period of time around 2000 when compared to the asymptotic test.

The paper is organized as follows: In Section 2, we briefly present an overview of the problem addressed by GKM and review the time-varying Hausman test proposed. In Section 3, we discuss the bootstrap approach in this context. Section 4 examines the behaviour of the test statistic by means of Monte Carlo simulations. Section 5 presents the empirical application. Finally, some concluding remarks and some directions on further research are given in Section 6.

2 Theory

Giraitis et al. (2014) introduced a non-parametric time-varying OLS estimation method that is based on a kernel generalisation of a rolling window. GKM expanded the results in the IV context with either deterministic or random coefficients, and derived a time-varying version of the Hausman exogeneity test comparing the time-varying OLS and IV estimators, allowing for a shift in endogeneity status over time.

To fix ideas, we consider the following regression model for a univariate series, y_t:

$$y_t = x_t^{'}\beta_t + u_t \tag{3.1}$$

$$x_t = \Psi_t^{'} z_t + v_t \tag{3.2}$$

where $x_t = (x_{1,t}, \dots, x_{p,t})^{'}$ is a $p \times 1$ vector of random variables, $\beta_t = (\beta_{1,t}, \dots, \beta_{p,t})^{'}$ is a $p \times 1$ parameter vector and u_t is random noise. In (3.2), $z_t = (z_{1,t}, \dots, z_{n,t})^{'}$ is a $n \times 1$ vector of random variables, $\Psi_t^{'} = (\psi_{lk,t})$ is a $p \times n$ parameter matrix and $v_t = (v_{1,t}, \dots, v_{p,t})^{'}$ is a $p \times 1$ noise vector.

Under the assumption of exogeneity of the regressors, the OLS estimator of β_t is

$$\hat{\beta}_t = (\sum_{j=1}^{T} b_{H,|j-t|} x_j x_j^{'})^{-1} (\sum_{j=1}^{T} b_{H,|j-t|} x_j y_j), \tag{3.3}$$

computed with kernel weights $b_{H,|j-t|}$ and bandwidth parameter H to be defined below in (3.7).

Assume now that endogenous variables x_t are correlated with u_t but there exist some exogenous instruments z_t such that:

$$E[z_t u_t] = 0, \quad E[z_t v_t^{'}] = 0, \quad t \geq 1. \tag{3.4}$$

GKM introduced[1] a kernel type estimator for β_t

$$\tilde{\beta}_t = (\sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j^{'} z_j x_j^{'})^{-1} (\sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j^{'} z_j y_j) \tag{3.5}$$

[1] In GKM's notation $\tilde{\beta}_t$ is $\tilde{\beta}_{1,t}$.

where $b_{H,|j-t|}$ are again the kernel weights with bandwidth parameter H and $\hat{\Psi}_j$ is the kernel OLS estimator

$$\hat{\Psi}_t = (\sum_{j=1}^{T} b_{L,|j-t|} z_j z_j^{'})^{-1} (\sum_{j=1}^{T} b_{L,|j-t|} z_j x_j^{'}) \tag{3.6}$$

which is a consistent estimate of Ψ_j.

Note the different bandwidth parameters L and H used in (3.6) and (3.5) respectively. The kernel weights are of the form:

$$b_{H,|j-t|} = K(\frac{|j-t|}{H}) \tag{3.7}$$

where $H \to \infty$, $H = o(T)$ is the bandwidth parameter and $K(x)$, $x \in (0, \alpha)$ is a non-negative continuous function with a finite or infinite support such that[2] for some $C > 0$ and $\nu > 3$,

$$K(x) \leq C(1 + x^{\nu})^{-1}, \quad |(d/dx)K(x)| \leq C(1 + x^{\nu})^{-1}, \quad x \in (0, \alpha). \tag{3.8}$$

Under assumptions[3] (1)-(5) in Giraitis et al. (2021), GKM proposed a time-varying version of the Hausman exogeneity test, which compares the time-varying IV and OLS estimators defined above, allowing for changes in the endogeneity status over time. Denote $K_t = \sum_{j=1}^{T} b_{H,|j-t|}$ and $K_{2,t} = \sum_{j=1}^{T} b_{H,|j-t|}^2$. Set $V_{T,t} = (S_{\hat{x}\hat{x},t})^{1/2}(S_{xx,t})^{1/2}(\hat{\beta}_t - \tilde{\beta}_t)$ where $S_{xx,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t| x_j x_j^{'}}$, $S_{\hat{x}\hat{x},t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{x}_j \hat{x}_j^{'}$, $\hat{x}_j = \Psi_j z_j$. The test statistic takes the form of

$$S_t = \frac{K_t^2}{K_{2,t}} V_{T,t}^{'} \hat{\Sigma}_{\hat{v}\hat{v},t}^{-1} V_{T,t} \hat{\sigma}_{\hat{u},t}^{-2} \tag{3.9}$$

where $\hat{\Sigma}_{\hat{v}\hat{v},t} := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{v}_j \hat{v}_j^{'}$, $\hat{\sigma}_{\hat{u},t}^2 := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{u}_j^2$ based on residuals $\hat{u}_j = y_j - x_j^{'} \tilde{\beta}_j$ and $\hat{v}_j = x_j - \hat{\Psi}_j^{'} z_j$ which can be used to test the null hypothesis $H_0 : E[v_t u_t] = 0$ that x_j is exogenous at time t. Under the null, the test statistic is asymptotically distributed as χ_p^2.

As shown in GKM, size distortions of the test statistic in (3.9) can be substantial even for large sample sizes. These size distortions seem to be quite sensitive to the choice of the bandwidth parameters H and L and also on the number of instruments. In the next section, we present an overview of the bootstrap approaches used and discuss a number of issues regarding their implementation.

3 The Bootstrap

Originally proposed by Efron (1979), bootstrap constitutes a major tool in the hands of statisticians for approximating the sampling distribution and variance of complicated statistics. Also, as pointed out by Politis

[2]These bounds are used to obtain Bernstein type inequalities for sums of α-mixing variables with thin or heavy tailed distributions, similarly to Dendramis et al. (2021).

[3]These assumptions will be referred as GKM assumptions.

(2003), Beran and Ducharme (1991) among many, bootstrap tests can often provide significant refinements to asymptotic tests. However, bootstrap often involves a number of choices to be made ex-ante by the researcher. This is especially important in the case of endogeneity.

The first issue that needs to be addressed is that any bootstrap procedure to be employed needs to retain any possible dependence between x and z. Here, we do not make any assumption regarding the parametric model for x and z and hence we use nonparametric bootstrap. Depending on the presence of dependence between these variables one might want to employ a fixed design bootstrap, as proposed by Kreiss (1997),where x and z are held fixed or resample blocks of rows of data from (x_t, z_t) using block bootstrap[4] as proposed by Kunsch (1989). A well known difficulty however, of the block bootstrap is that the block size b needs to be determined in advance by the practitioner.

The choice of the block size is critical for the performance of the bootstrap. If the block size chosen is too small then the dependency among the blocks is broken and hence it cannot be expected that the bootstrap samples will mimic closely the original data. If, on the other hand, the block size is chosen to be too large, the bootstrap samples are no longer random enough. However, only in few cases literature provides guidance regarding the selection of the block size (c.f Bühlmann and Künsch (1999), Lahiri (1999) and Hall et al. (1995a)). In most of the cases, simple rules specifying the rate at which b should increase with sample size are provided for specific applications. Usually those include $b = O(T^{1/3})$ or $b = O(T^{1/2})$. In this case, there is no simple rule guiding the choice of the block size and so different values for b need to be examined.

Note here, that the asymptotic test of GKM is obtained under the independence assumption of the error term. These errors however are allowed to be heterogeneously distributed and hence, it is important that any selected bootstrap procedure is able to capture such heterogeneity in the error term.

Another important issue, in testing for exogeneity, is the mode of resampling so that the null distribution is imposed, as discussed in Kapetanios (2010). This is especially important if, for example, the whole distribution is bootstrapped and used instead of the asymptotic χ^2 approximation. In particular, resampling from a distribution obeying the constraints of the null, even though the data may not satisfy the null, guarantees that the bootstrap test has power that is approaching one under the alternative(see e.g Lehmann et al. (1986), Sec 15.6).

Then, the practitioner needs to generate y_t using parametric bootstrap. This, of course requires to choose whether to use TV-OLS estimates or TV-IV estimates. Note here that under the null, both of these estimators are consistent although TV-OLS is more efficient. In this paper,[5] TV-OLS estimates are employed.

In view of the above, we suggest two alternative bootstrap procedures: The first alternative is the fixed-desing wild bootstrap. The algorithm is as follows:

[4] Alternatively, one could use variants of the MBB as in Politis and Romano (1991) who propose the Circular Bootstrap to deal with the end effects. This path was also explored. See the robustness analysis in section 3.A.

[5] A preliminary Monte Carlo analysis suggested that similar results are obtained using either OLS or IV estimator. Results are available upon request from the authors.

1. Estimate (3.1) by TV-OLS and obtain the residuals and the estimates of $\hat{u}$ and $\hat{\beta}_t$ respectively.

2. The wild bootstrap is applied to OLS residuals $\hat{u}_t$ by premultiplying them with η_t so that now the bootstrap errors are given by $u^\star = \eta_t \hat{u}_t$, where $\{\eta\}_{t=1}^T$ is an i.i.d standard normal distribution, independent of all other random processes.

3. Use $\hat{\beta}_t$, (z_t, x_t) and $\{u^\star\}_{t=1}^T$ in (3.1) to obtain bootstrap samples of size T for y_t and denote them $y_t^\star$.

4. Use $(z_t, x_t, y_t^\star)$ to obtain TV-OLS and TV-IV estimates given by (3.3) and (3.5) respectively, and calculate the Hausman test $S^\star$ for the bootstrap sample.

(iv) Repeat steps 2-3 B times to produce B bootstrap test statistics.

The second alternative is a combination of the moving block bootstrap for pairs of x and z and wild bootstrap, as proposed by Kapetanios (2010) (see also Goncalves and Kilian (2004) for an application in a different framework.) and the algorithm is the following:

1. Estimate (3.1) by TV-OLS and obtain the residuals and the estimates of $\hat{u}_t$ and $\hat{\beta}_t$ respectively.

2. Resample blocks of rows of (z_t, x_t) using a block bootstrap approach to produce bootstrap samples of size T, for the regressors, as suggested by Kunsch (1989).

3. The wild bootstrap is applied to the TV-OLS residuals $\hat{u}_t$ by premultiplying them with η_t so that I have $u_t^\star = \eta_t \hat{u}_t$, where $\{\eta\}_{t=1}^T$ is an i.i.d standard normal distribution, independent of all other random processes.

4. Use $\hat{\beta}$, $(z_t^\star, x_t^\star)$ and $u_t^\star$ in (3.1) to obtain bootstrap samples of size T for y_t and denote them $y_t^\star$.

5. Use $(z_t^\star, x_t^\star, y_t^\star)$ to obtain TV-OLS and TV-IV estimates given by (3.3) and (3.5) respectively, and calculate the Hausman test $S_b^\star$ for the bootstrap sample[6].

6. Repeat 2-5 B times to obtain B bootstrap tests.

The wild bootstrap uses a transformation of the residuals to construct the bootstrap error term $u_t^\star = \eta_t \hat{u}_t$ where η_t is a random variable with mean 0 and variance 1. Different distributions have been employed for η_t including those of Rademacher as suggested by Davidson and Flachaire (2008) and Mammen as proposed by Mammen (1993). Using wild bootstrap guarantees that the null is imposed as we now have $E^\star(x_t^\star u_t^\star) = E^\star(x_t^\star \hat{u}_t \eta_t) = 0$. The Wild bootstrap algorithm used, stems from Kapetanios (2010).

Consequently, an issue that needs to be addressed, is the number of bootstrap replications. Hall (1992) proposed to choose B, the number of bootstrap repetitions such that $\nu/(B+1) = 1 - \alpha$ for a positive integer ν. This implies in turn that α is a rational such that $\alpha = \alpha_1/\alpha_2$ for positive integers α_1 and α_2 with no common integer

[6] The notation $S_b^\star$ reads a Hausman test statistic for the bootstrap sample with block size b.

divisors. Then $B = \alpha_2 h - 1$ and $\nu = (\alpha_2 - \alpha_1)/h$ for positive integer h[7]. As Davidson and MacKinnon (2000) showed, using $B = 399$ is about the minimum for a test that guarantees a loss of power less than 1% at 0.05 level.

Finally, the consistency of the bootstrap approximation to the sampling distribution of the local Hausman test statistic can be concluded by showing convergence in probability of the conditional law of $S^\star$ given the sample y, x denoted as $L(S^\star)$ to the same limit as the law of S, under the null. This result is summarized in Theorem 5 below which is formally proven in subsection 3.A.2. The following conditions included as an assumption below, are needed for the bootstrap procedure.

Assumption 2. *Assume that*

1. $E(||S_{xx,t}^{-1}||_{sp}^4) = O(1)$,
2. $E(||x_t||^{8p}) = O(1)$ *where* $p = 1 + \epsilon$ *for* $\epsilon > 0$.

Theorem 5. *Under Assumptions 1-4 in GKM and Assumption 2 above and as* $T \to \infty$

$$L(S^\star) \Rightarrow L(S) \text{ in probability.}$$

The condition (1) from Assumption 2 guarantees the existence of the finite fourth moment of the matrix $S_{xx,t}^{-1}$ while (2) ensures that (3.A.58) is bounded and imposes a further restriction of magnitude $p > 0$ on Assumption 1(i) in GKM.

4 Simulation Results

In this Section, Monte Carlo experiments are carried out to investigate the finite-sample performance of the time-varying local Hausman test proposed by GKM and the bootstrap version of it. In all simulations[8], in this section, we use 5000 Monte Carlo replications and 999 bootstrap replications. The experiments found in this section are the same as those employed in GKM.

As data generating process (DGP) under the exactly identified case, we consider the following model:

$$y_t = \beta_t x_t + u_t, \quad x_t = \psi_t z_t + v_t \tag{3.10}$$

for $t = 1, \ldots, T$. Following GKM, correlation between u_t and v_t is introduced by specifying them as

$$u_t = se_{1,t} + (1 - s)e_{2,t} \quad v_t = se_{1,t} + (1 - s)e_{3,t} \tag{3.11}$$

[7]For instance, for $\alpha = 0.05$ I have $\alpha_1 = 1$, $\alpha_2 = 20$,$B = 20h - 1$ and $\nu = 19h$ for a positive integer h. So $B = 19, 39, 59, ...$ etc.

[8]All simulations were performed using an Apple M1 with a 8 GB unified memory. The code was written and executed in JuliaPro-1.5.3.

where $s = 0, 0.2, 0.5, 0.8, 0.9$ and $\{e_{1,t}\}$,$\{e_{2,t}\}$ and $\{e_{3,t}\}$ are mutually independent $NIID(0, 1)$ sequences.

The parameters $\beta_t = T^{-1/2}\xi_{1,t}$, $\psi_t = T^{-1/2}\xi_{2,t}$,$t = 1, \ldots, T$ are generated as two independent rescaled random walks, such that $\xi_{l,t} - \xi_{l,t-1} \sim N(0, 1)$ for $l = 1, 2$ that are also independent of $\{\psi_t\}$,$\{u_t\}$ and $\{v_t\}$. This implies that both the structural and the reduced form regressions have time-varying coefficients. Exogeneity of x_t is implied by $s = 0$, while for $s = 0.2, 0.5, 0.8, 0.9$, x_t is endogenous. The magnitude of s hence, provides a means for controlling the extent of endogeneity.

We examine two estimators of β_t namely: the TV-OLS $\hat{\beta}_t$ and the TV-IV$\tilde{\beta}_t$. These are computed using the Gaussian kernel[9] $K(x) = exp(-x^2/2)$ with a variety of bandwidth values H for estimation of β_t and L for ψ_t. Specifically, we set $H = T^{h_1}$ and $L = T^{h_2}$ with $h_1, h_2 = 0.4$ and 0.5 as in GKM. Results for values of 0.7 are also reported in the section 3.A. Lower values for the bandwidth increase robustness of estimates to parameter changes but decrease efficiency. Further, we consider three sample sizes of length $100 + T$ with $T \in \{100, 200, 400\}$. The first 100 observations are then discarded in order to eliminate initial value effects and only the remaining T observations are used. We now discuss the bootstrap implementation that is employed in the rest of the section.

Two bootstrap procedures are considered, as discussed in section 3. The first bootstrap procedure is the fixed-design wild bootstrap. where no resampling of (x_t, z_t) is taking place. The bootstrap version of the time-varying Hausman test under this bootstrap framework is denoted as $S^\star$. The second bootstrap method is a combination of block resampling of rows of (x_t, z_t) with block size b and wild bootstrap. The bootstrap version of the time-varying Hausman test is denoted as $S_b^\star$. Multiple values for the block size b,that are proportional or multiples to the sample size T are examined. In both of the aforementioned procedures, the bootstrap error term is constructed as $u_t^\star = \eta_t\hat{u}_t$ where η_t is an i.i.d standard Normal[10] distributed sequence.

To evaluate the performance of the test statistic, we examine the rejection frequencies of the local time varying Hausman test at 5% significance level and for $t = T/2$ which is representative as there is no change of the endogeneity status in the DGP. These frequencies are reported in Table 3.1-Table 3.3.

[9]Similar results were obtained using both the Epanechnikov kernel $K(x) = 0.75(1 - x^2)$ for $|x| < 1$ and the exponential kernel $K(x) = exp(-cx^\alpha)$ where $c > 0$ and $\alpha > 0$.

[10]The Rademacher latice distribution was also used as an alternative to the standard Normal but provided similar results.

Table 3.1: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (3.10)-(3.11). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

			T=100					
s	h_1	h_2	S	$S^\star$	$S^\star_{b=2}$	$S^\star_{b=4}$	$S^\star_{b=6}$	$S^\star_{b=8}$
0	0.4	0.4	0.024	0.045	0.056	0.056	0.053	0.052
	0.4	0.5	0.296	0.046	0.055	0.053	0.052	0.049
	0.5	0.4	0.030	0.045	0.063	0.060	0.059	0.058
	0.5	0.5	0.030	0.050	0.061	0.056	0.057	0.055
0.2	0.4	0.4	0.031	0.047	0.062	0.058	0.058	0.056
	0.4	0.5	0.036	0.048	0.057	0.052	0.053	0.052
	0.5	0.4	0.039	0.055	0.066	0.064	0.061	0.060
	0.5	0.5	0.038	0.057	0.065	0.065	0.065	0.062
0.5	0.4	0.4	0.306	0.333	0.377	0,3744	0.369	0.369
	0.4	0.5	0.308	0.328	0.351	0.345	0.342	0.337
	0.5	0.4	0.447	0.477	0.511	0.511	0.513	0.513
	0.5	0.5	0.445	0.472	0.499	0.498	0.496	0.495
0.8	0.4	0.4	0.712	0.790	0.801	0.798	0.795	0.795
	0.4	0.5	0.689	0.758	0.771	0.765	0.763	0.758
	0.5	0.4	0.447	0.477	0.511	0.511	0.513	0.513
	0.5	0.5	0.783	0.852	0.862	0.860	0.859	0.858
0.9	0.4	0.4	0.713	0.798	0.807	0.802	0.801	0.800
	0.4	0.5	0.690	0.751	0.776	0.773	0.769	0.763
	0.5	0.4	0.806	0.873	0.871	0.871	0.868	0.869
	0.5	0.5	0.783	0.848	0.864	0.860	0.857	0.859

Table 3.1 shows the rejection frequencies for the local Hausman test for $T = 100$. The asymptotic test appears to exhibit considerable size distortions for all bandwidth parameters and also low power. On the other hand, bootstrap procedures seem to outperform the asymptotic test both in terms of size and power irrespectively of the bandwidth parameter chosen. The size is close to the nominal 5% while the power is at least 38.5% larger than the power of the asymptotic test statistic.

Table 3.2: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (3.10)-(3.11). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

				T=200				
s	h_1	h_2	S	$S^\star$	$S^\star_{b=4}$	$S^\star_{b=6}$	$S^\star_{b=8}$	$S^\star_{b=16}$
0	0.4	0.4	0.023	0.048	0.060	0.061	0.058	0.056
	0.4	0.5	0.028	0.049	0.057	0.054	0.051	0.048
	0.5	0.4	0.031	0.050	0.058	0.060	0.056	0.055
	0.5	0.5	0.040	0.053	0.063	0.061	0.063	0.061
0.2	0.4	0.4	0.029	0.048	0.061	0.057	0.057	0.054
	0.4	0.5	0.035	0.050	0.057	0.052	0.052	0.047
	0.5	0.4	0.039	0.056	0.068	0.066	0.066	0.061
	0.5	0.5	0.040	0.053	0.063	0.061	0.063	0.061
0.5	0.4	0.4	0.367	0.404	0.445	0.444	0.443	0.435
	0.4	0.5	0.362	0.397	0.413	0.409	0.407	0.399
	0.5	0.4	0.533	0.569	0.609	0.607	0.605	0.598
	0.5	0.5	0.528	0.570	0.604	0.597	0.595	0.591
0.8	0.4	0.4	0.714	0.812	0.811	0.807	0.805	0.798
	0.4	0.5	0.694	0.770	0.768	0.766	0.762	0.759
	0.5	0.4	0.818	0.897	0.879	0.877	0.875	0.869
	0.5	0.5	0.799	0.879	0.871	0.868	0.868	0.861
0.9	0.4	0.4	0.713	0.806	0.813	0.812	0.810	0.801
	0.4	0.5	0.691	0.756	0.771	0.768	0.765	0.760
	0.5	0.4	0.824	0.893	0.885	0.882	0.881	0.870
	0.5	0.5	0.040	0.053	0.063	0.061	0.063	0.061

Table 3.2 presents the rejection frequencies for the local Hausman test for $T = 200$. The size of the bootstrap procedures are close to the nominal level irrespective of the bandwidth parameters. In terms of power, there is at least a 10.5% increase through the use of bootstrap. This is especially the case for small values of s where the power refinements through the use of bootstrap account for at least a 41% gain. The hausman test obtained under the fixed-design wild bootstrap seems to outperform both the asymptotic test and the test obtained under the pairwise block bootstrap in terms of size while only having a slightly lower power.

Similar results are obtained from Table 3.3 for $T = 400$. The power is at least 9.8% larger than the power of the asymptotic test. Overall, the fixed-design wild bootstrap version of the Hausman test seems to outperform the asymptotic test statistic in terms of power while retaining size close to nominal.

Table 3.3: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (3.10)-(3.11). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

			T=400					
s	h_1	h_2	S	$S^\star$	$S^\star_{b=6}$	$S^\star_{b=8}$	$S^\star_{b=16}$	$S^\star_{b=32}$
0	0.4	0.4	0.023	0.046	0.058	0.059	0.054	0.052
	0.4	0.5	0.028	0.047	0.053	0.052	0.048	0.050
	0.5	0.4	0.028	0.050	0.062	0.064	0.060	0.055
	0.5	0.5	0.028	0.051	0.063	0.063	0.059	0.056
0.2	0.4	0.4	0.029	0.055	0.071	0.071	0.063	0.061
	0.4	0.5	0.035	0.056	0.062	0.060	0.051	0.053
	0.5	0.4	0.041	0.060	0.078	0.076	0.075	0.067
	0.5	0.5	0.041	0.062	0.078	0.078	0.073	0.069
0.5	0.4	0.4	0.433	0.500	0.522	0.521	0.515	0.509
	0.4	0.5	0.425	0.494	0.478	0.476	0.460	0.465
	0.5	0.4	0.622	0.678	0.693	0.689	0.687	0.676
	0.5	0.5	0.615	0.672	0.688	0.684	0.681	0.679
0.8	0.4	0.4	0.729	0.857	0.812	0.808	0.804	0.796
	0.4	0.5	0.701	0.815	0.771	0.766	0.762	0.760
	0.5	0.4	0.832	0.929	0.893	0.891	0.885	0.876
	0.5	0.5	0.810	0.918	0.876	0.877	0.870	0.867
0.9	0.4	0.4	0.725	0.879	0.818	0.814	0.807	0.800
	0.4	0.5	0.694	0.814	0.769	0.766	0.757	0.758
	0.5	0.4	0.835	0.941	0.896	0.896	0.885	0.878
	0.5	0.5	0.809	0.929	0.878	0.877	0.871	0.864

Next, the overidentified case is also examined. Following GKM, we study the following model

$$y_t = \beta_t x_t + u_t, \quad x_t = \psi_{1,t} z_{1,t} + \psi_{2,t} z_{2,t} + v_t \tag{3.12}$$

for $t = 1, \ldots, T$ where $(\psi_{1,t})$ and $(z_{1,t})$ have the same specification as (ψ_t) and (z_t) from above, $(\psi_{2,t}) = T^{-1/2}\xi_{3,t}$ for $t = 1, \ldots, T$ is generated such that $\xi_{3,t} - \xi_{3,t-1} \sim NIID(0,1)$ and $(z_{2,t})$ is a sequence of standard normal i.i.d random variables.

Using (3.12) we now report the rejection frequencies in Table 3.4-Table 3.6 below.

Table 3.4: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$ for model (3.12). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

				T=100				
s	h_1	h_2	S	$S^\star$	$S^\star_{b=2}$	$S^\star_{b=4}$	$S^\star_{b=6}$	$S^\star_{b=8}$
0	0.4	0.4	0.055	0.074	0.072	0.073	0.075	0.074
	0.4	0.5	0.051	0.064	0.071	0.070	0.067	0.068
	0.5	0.4	0.069	0.079	0.085	0.084	0.086	0.084
	0.5	0.5	0.056	0.062	0.068	0.069	0.070	0.068
0.2	0.4	0.4	0.065	0.078	0.079	0.081	0.080	0.080
	0.4	0.5	0.061	0.070	0.073	0.073	0.072	0.071
	0.5	0.4	0.084	0.092	0.094	0.094	0.094	0.094
	0.5	0.5	0.069	0.075	0.075	0.077	0.076	0.077
0.5	0.4	0.4	0.379	0.373	0.410	0.413	0.409	0.410
	0.4	0.5	0.386	0.380	0.398	0.398	0.396	0.394
	0.5	0.4	0.535	0.511	0.555	0.555	0.557	0.554
	0.5	0.5	0.542	0.533	0.562	0.558	0.562	0.564
0.8	0.4	0.4	0.883	0.863	0.912	0.911	0.910	0.909
	0.4	0.5	0.892	0.886	0.921	0.918	0.917	0.916
	0.5	0.4	0.945	0.932	0.955	0.956	0.956	0.954
	0.5	0.5	0.952	0.948	0.969	0.968	0.968	0.966
0.9	0.4	0.4	0.907	0.951	0.935	0.934	0.932	0.934
	0.4	0.5	0.906	0.958	0.936	0.935	0.933	0.933
	0.5	0.4	0.956	0.968	0.968	0.968	0.967	0.968
	0.5	0.5	0.958	0.976	0.972	0.972	0.971	0.972

Table 3.4 reports the rejection frequencies for the local Hausman test for $T = 100$. The size of the asymptotic test is close to the nominal value although it varies quite a bit with respect to the bandwidth parameter. Note for $h_1 = 0.5$ and $h_2 = 0.4$ the size of the asymptotic test is 0.069. The size of the test based on the Fixed-Design WB seems to be higher than the both the nominal and the asymptotic. In terms of power, the power of the asymptotic test is always lower than the power of the test based on the Fixed-Design WB in the case of weak endogeneity while for strong endogeneity,the two have similar power.

Turning next to Table 3.5, we show the rejection frequencies for $T = 200$. Again similar results to $T = 100$ are obtained. The asymptotic test has size close to nominal while the size of the Fixed-design WB based test approaches the nominal size. There are severe size distortions for both the asymptotic test and the bootstrap based tests for $h_1 = 0.5$ and $h_2 = 0.4$. The power of the asymptotic test is close to the power of the Fixed-design WB based test although the power of the latter increases with s. In all cases, pairwise block bootstrap test performs worse than both the asymptotic test and the Fixed-design WB based test.

Table 3.5: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$ for model (3.12). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

				T=200				
s	h_1	h_2	S	$S^\star$	$S^\star_{b=4}$	$S^\star_{b=6}$	$S^\star_{b=8}$	$S^\star_{b=16}$
0	0.4	0.4	0.048	0.063	0.069	0.063	0.067	0.067
	0.4	0.5	0.047	0.058	0.064	0.06	0.063	0.067
	0.5	0.4	0.067	0.0083	0.086	0.087	0.084	0.087
	0.5	0.5	0.054	0.057	0.074	0.074	0.074	0.071
0.2	0.4	0.4	0.053	0.055	0.078	0.077	0.075	0.074
	0.4	0.5	0.052	0.056	0.064	0.061	0.062	0.061
	0.5	0.4	0.078	0.065	0.095	0.095	0.096	0.094
	0.5	0.5	0.064	0.063	0.079	0.079	0.079	0.078
0.5	0.4	0.4	0.474	0.500	0.522	0.521	0.518	0.517
	0.4	0.5	0.476	0.517	0.503	0.499	0.498	0.493
	0.5	0.4	0.666	0.672	0.696	0.694	0.698	0.693
	0.5	0.5	0.675	0.700	0.704	0.700	0.700	0.699
0.8	0.4	0.4	0.927	0.963	0.952	0.951	0.950	0.949
	0.4	0.5	0.920	0.960	0.943	0.942	0.940	0.938
	0.5	0.4	0.970	0.986	0.982	0.982	0.981	0.980
	0.5	0.5	0.967	0.988	0.982	0.981	0.981	0.981
0.9	0.4	0.4	0.930	0.975	0.958	0.959	0.956	0.955
	0.4	0.5	0.920	0.966	0.947	0.946	0.944	0.944
	0.5	0.4	0.975	0.992	0.985	0.984	0.984	0.982
	0.5	0.5	0.970	0.990	0.984	0.983	0.982	0.981

Finally, in Table 3.6 we present the results for $T = 400$. The asymptotic test has small size distortions while the size of the Fixed-design WB based test is close to the nominal value. On the other hand, the block bootstrap's size experiences some size distortions. In terms of power, asymptotic bootstrap and Fixed-design WB based test's performance is quite close for low values of s while the difference increases for larger values of s in favour of the Fixed-design WB based test. As expected, the size distortions of the asymptotic test decrease with sample size. Overall, Fixed-design WB based test retains the proper size across all sample sizes, bandwidth parameter values and number of instruments whilst also outperforms the asymptotic test in terms of power.

Table 3.6: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$ for model (3.12). Bandwidths are set as $H = T^{h_1}$ and $L = T^{h_2}$. Parameter s controls exogeneity($s = 0$) or endogeneity status($s \neq 0$).

			T=400					
s	h_1	h_2	S	$S^\star$	$S^\star_{b=6}$	$S^\star_{b=8}$	$S^\star_{b=16}$	$S^\star_{b=32}$
0	0.4	0.4	0.042	0.048	0.066	0.066	0.065	0.060
	0.4	0.5	0.045	0.051	0.061	0.061	0.059	0.057
	0.5	0.4	0.049	0.052	0.075	0.072	0.073	0.069
	0.5	0.5	0.046	0.052	0.061	0.062	0.062	0.064
0.2	0.4	0.4	0.049	0.056	0.077	0.075	0.074	0.071
	0.4	0.5	0.053	0.059	0.064	0.064	0.061	0.062
	0.5	0.4	0.065	0.059	0.085	0.084	0.084	0.080
	0.5	0.5	0.060	0.062	0.073	0.075	0.075	0.075
0.5	0.4	0.4	0.601	0.629	0.661	0.661	0.659	0.649
	0.4	0.5	0.607	0.644	0.636	0.633	0.626	0.631
	0.5	0.4	0.815	0.825	0.847	0.847	0.843	0.837
	0.5	0.5	0.819	0.839	0.848	0.848	0.845	0.843
0.8	0.4	0.4	0.935	0.973	0.957	0.957	0.956	0.954
	0.4	0.5	0.925	0.965	0.945	0.944	0.942	0.944
	0.5	0.4	0.981	0.993	0.990	0.989	0.988	0.987
	0.5	0.5	0.972	0.990	0.982	0.982	0.981	0.980
0.9	0.4	0.4	0.936	0.979	0.960	0.960	0.960	0.957
	0.4	0.5	0.920	0.969	0.945	0.944	0.942	0.942
	0.5	0.4	0.978	0.992	0.985	0.986	0.984	0.982
	0.5	0.5	0.972	0.991	0.982	0.981	0.981	0.979

5 Empirical Application

In this section, we follow GKM and employ the local Hausman exogeneity test in a time varying version of the traditional Phillips curve. The goal here is to compare the asymptotic test with its bootstrap counterpart based on the Fixed-design wild bootstrap and see whether different results are obtained.

The original article by GKM does not mention the dataset used and hence we consider monthly data obtained from Louis FRED[11]. Inflation π_t is computed as 100 times the seasonal log difference of the *CPIAUCSL* variable and the variable *UNRATE* is used for unemployment u_t. The sample period ranges from 1959:1 to 2021:12 to include COVID-19 pandemic. The model used is

$$\Delta\pi_t = c_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + e_t \tag{3.13}$$

where change in inflation is the dependent variable and change in unemployment together with one lag of the change in inflation are the independent variables. Differences are used throughout, due to the high persistence of the series. A Gaussian kernel is employed with bandwidth parameters $H = L = T^{0.7}$ since for these values the asymptotic Hausman test is shown to suffer the most in terms of size and power compared to its bootstrap version. It is worth noting that for our sample, there appears to be significant serial correlation and hence results should be viewed with caution because neither asymptotic test nor the bootstrap test allow for serial

[11] Following the suggestion by Lucchetti and Valentini (2021).

correlation.

Figure 3.1 shows the time varying OLS and IV coefficient estimates of α_t and γ_t with their associated 90% confidence intervals, respectively. The time-varying IV[12] estimator is quite different from the time-varying OLS for the parameter α until 2000 while for the remaining, the two overlap. The average values over time of $\hat{\alpha}$ and $\tilde{\alpha}$ are about -0.157 and -0.649 which are comparable to the full sample constant parameters OLS and 2SLS values, -0.109 and 0.127 respectively. The latter, however is not statistical significant.

The lower panel of Figure 3.1 graphs $\hat{\gamma}_t$ and $\tilde{\gamma}_t$ with the associated 90% confidence intervals. The two estimators provide similar results and the two lines seem almost indistinguishable. The average value over time for $\hat{\gamma}_t$ is 0.356 while the full sample constant parameter OLS value amounts to 0.376.

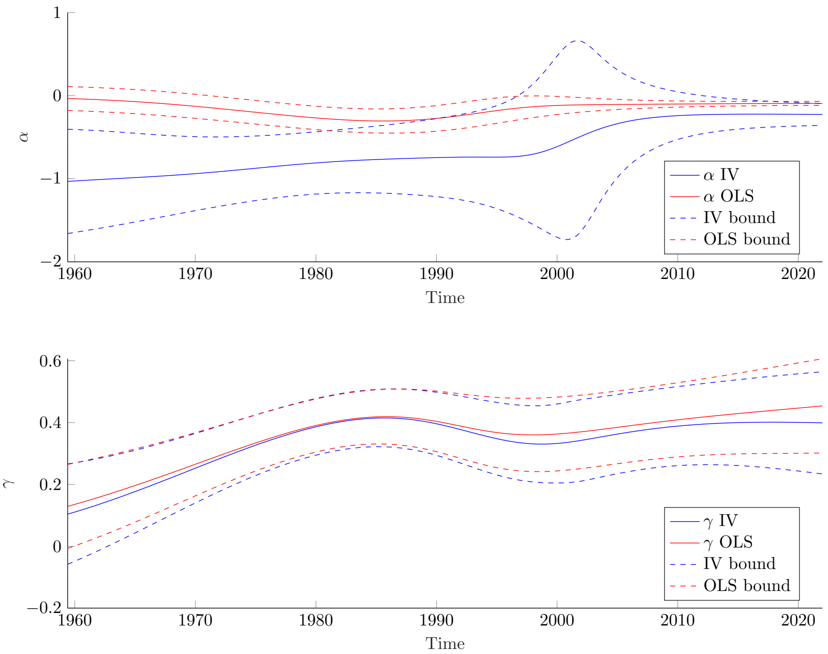


Figure 3.1: Empirical results for model (3.13).The two panels graph the OLS and IV coefficient estimates for α_t and γ_t respectively using $H = L = T^{0.7}$.

Figure 3.2 presents the p-values of the asymptotic time-varying Hausman test and its bootstrap version based on the fixed-design wild bootstrap. The upper panel shows the empirical p-values of the two tests for $H = L = T^{0.7}$. For most of the sample, the two tests seem to provide the same results while for the period around 2000 the two tests show conflicting results. Specifically, the bootstrap version of the Hausman test rejects the null of exogeneity at 10% significance level while the asymptotic does not reject.

Turning now to the bandwidth parameter values of $H = L = T^{0.5}$ the lower panel of Figure 3.2 shows the p-values of the asymptotic test and its bootstrap counterpart. Given the Monte Carlo simulations in section 4, these values were the values for which the asymptotic Hausman test experienced the smallest distortions in terms of size and power and hence can serve as a benchmark to compare the differences found in the two tests.

[12] Henceforth, OLS estimates will be denoted with a hat and IV estimates with a tilde.

The two versions of the test for the lower panel in Figure 3.2 seem to be quite similar for most parts, except for the periods around 1970s and the period from 1990 to 1999 where the bootstrap version of the test points out to rejecting the null while the asymptotic test shows exogeneity of the regressor.

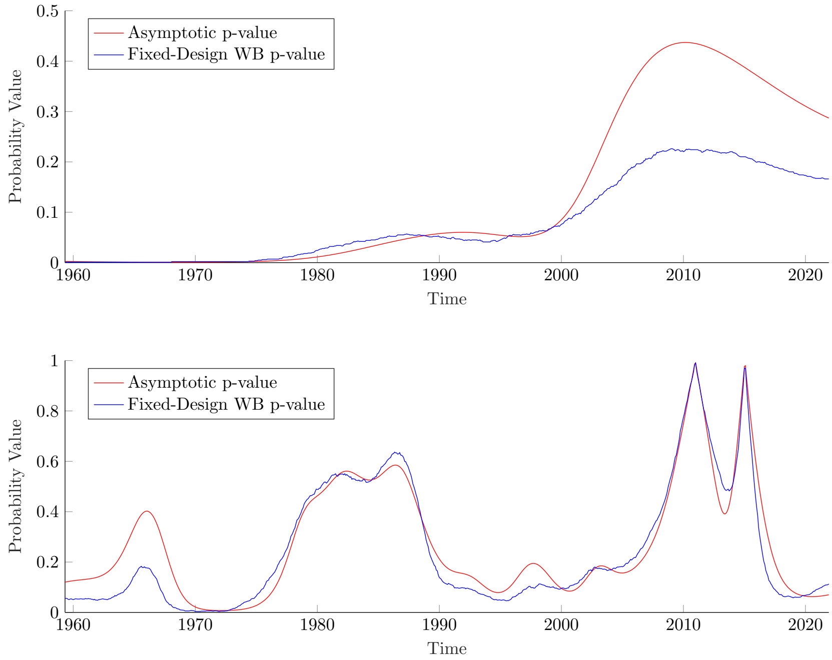


Figure 3.2: Empirical results for model (3.13).The two panels graph the empirical p-values of the aymptotic and bootstrap Hausman tests using $H = L = T^{0.7}$ and $H = L = T^{0.5}$ respectively.

Next, a forward looking (New-Keynesian) Phillips curve is also considered, as found in GKM and along the lines of Galí and Gertler, 1999. The New-Keynesian Phillips curve arises from the Calvo model and is given by:

$$\Delta\pi_t = c_t + \rho_t \Delta\pi^e_{t+1} + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + v_t \tag{3.14}$$

or written differently

$$\Delta\pi_t = c_t + \rho_t \Delta\pi_{t+1} + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + \epsilon_t \tag{3.15}$$

where $\epsilon_t = \rho_t(\Delta\pi^e_{t+1} - \Delta\pi_{t+1}) + v_t$, $\Delta\pi^e_{t+1}$ is the optimal one-step ahead forecast of $\Delta\pi_{t+1}$ made in period t, and v_t is an i.i.d error which is uncorrelated with all leads and lags with the forecast error $(\Delta\pi^e_{t+1} - \Delta\pi_{t+1})$. Obviously $\Delta\pi_{t+1}$ is correlated with the error term ϵ_t and hence a time-varying IV estimator is employed. Note that Galí and Gertler (1999) only consider the case of constant parameters. We repeat the same experiment as above, using four lags of the change in unemployment and inflation as instruments, a Gaussian kernel and $H = L = 0.7$.

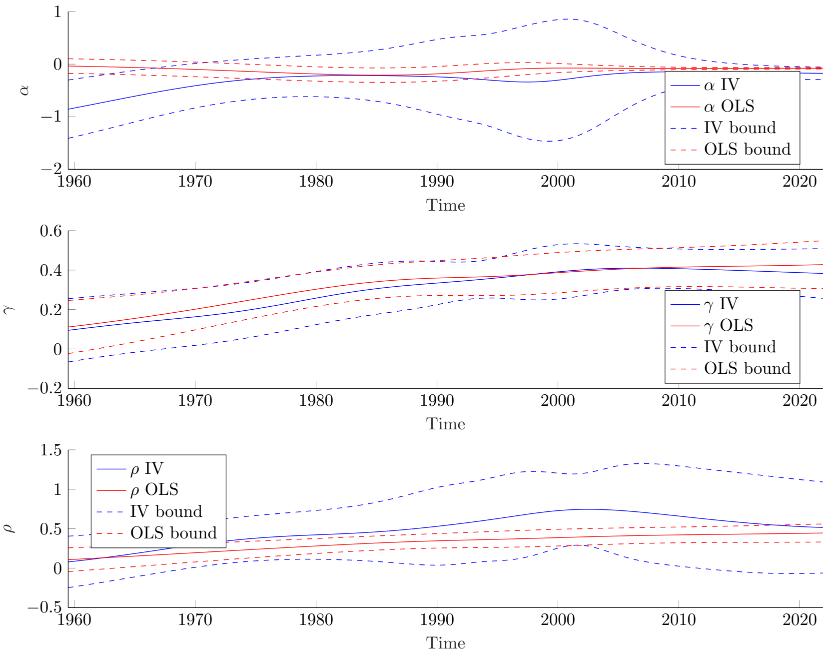


Figure 3.3: Empirical results for model (3.15).The three panels graph the OLS and IV coefficient estimates for α_t, γ_t and ρ_t respectively using $H = L = T^{0.7}$.

Figure 3.3 reports the coefficient estimates for this model. The upper panel of Figure 3.3 shows the coefficient estimates of $\hat{\alpha}_t$ and $\tilde{\alpha}_t$. The coefficient is close to 0 when estimated by time-varying OLS and never significant. Similarly, the time-varying IV estimator provides estimates close to 0 except for a small period around '60s. The confidence intervals seem to be quite wide around 2000 which might have an impact on the test statistic[13].

The middle panel of Figure 3.3 depicts the estimates of $\hat{\gamma}_t$ and $\tilde{\gamma}_t$. They both seem to perform quite similarly as the two lines seem almost identical. The lower panel graphs the results for $\hat{\rho}_t$ and $\tilde{\rho}_t$. The two lines appear to be deviating only for a short period around the '2000s.

Finally, the upper panel of Figure 3.4 graphs the p-values of the time-varying Hausman test and its bootstrap version for $H = L = T^{0.7}$. The asymptotic test provides lower p-values for the first half of the sample than the bootstrap test. Interestingly, the bootstrap test rejects the null of exogeneity around the '2000s while the asymptotic does not.

The lower panel of Figure 3.4 now depicts the p-values for the two tests for $H = L = T^{0.5}$. The two lines seem almost exact for most of the period except for the 1960s and '2000s where again the bootstrap rejects the exogeneity for a small period while the asymptotic does not.

[13] We further discuss this issue on section 3.A where we perform a robustness check.

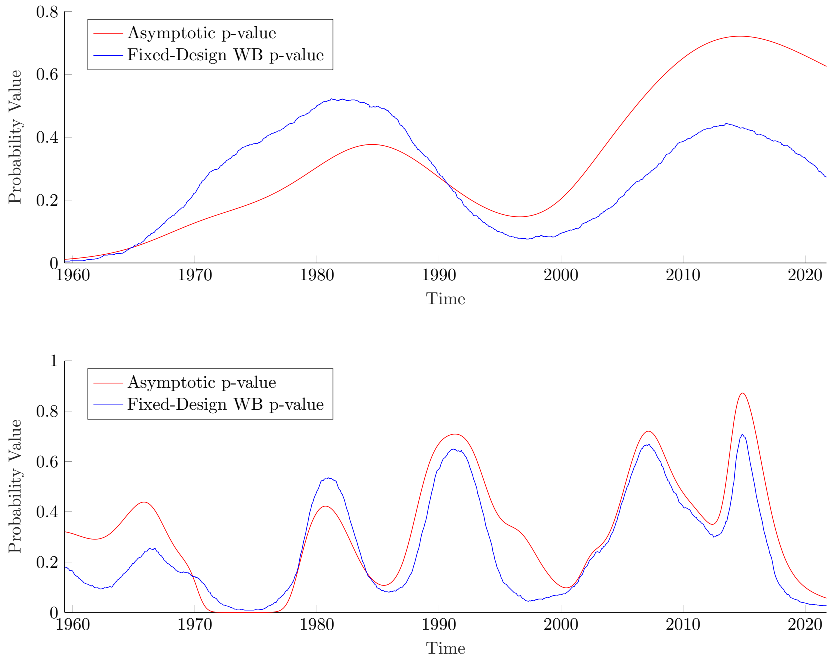


Figure 3.4: Empirical p-values of the time-varying Hausman test for model (3.15) and its bootstrap version for $H = L = T^{0.7}$ and $H = L = T^{0.5}$ respectively.

In summary, both the asymptotic and the time-varying boostrap version of the Hausman exogeneity test appear to perform quite similarly for most parts of the sample. However, contrary to the asymptotic test, its bootstrap version points out to endogeneity of unemployment for a short period around '2000s. In light of the results obtained in Section 4, this is consistent with our findings that the bootstrap test rejects the null more often than the asymptotic.

6 Conclusion

A usual assumption made when carrying out IV estimation is that the model, and hence the parameter vector, does not change through time. This in turn, also implies that the endogeneity status of the variable remains constant through time. This assumption, although crucial is often highly susceptible. A new strand of literature lead primarily by Giraitis et al. (2021) has proposed a non-parametric IV estimation based on kernels, and allowing for both deterministic or random coefficients. Consequently, a time-varying Hausman exogeneity test has been developed to test for a possible switching endogeneity status at a specific point. However, this test appears to experience size distortions and have low power.

In this paper, we propose an easy-to-use bootstrap version of the test. The performance of the bootstrapped test is evaluated through Monte Carlo simulations. As it is shown, the bootstrap test obtained under the fixed-design wild bootstrap has considerable refinements over the asymptotic test statistic with higher power, and size close to nominal for the exactly identified model. More importantly, the size and power of the bootstraped test are invariant to the choice of the bandwidth parameters.

On the other hand, the asymptotic test seems to perform relatively well in the case of the over- identified model both in terms of size and power. The Fixed-design Wild Bootstrap based test has size close to nominal for samples sizes typically encountered in macroeconomics while it experiences some small distortions in small samples.

Revisiting the empirical application by GKM, we estimate a Phillips curve for the USA, using unemployment as the forcing variable for inflation, and examine whether similar results are obtained using the asymptotic test and its bootstrap counterpart. The two tests seem to perform quite similarly for most parts except for a period around 2000 where the bootstrap test points out to endogeneity of unemployment. These results seem consistent with the Monte Carlo simulations, since the bootstrap test rejects the null more often that the asymptotic counterpart.

We finish with some open questions for further research. First, this article has focused only on the case of the local Hausman exogeneity test, but the seminal paper by Giraitis et al. (2021) also proposes a Global Hausman test for testing for possible endogeneity in a specified interval. Preliminary analysis suggests that bootstrap refinements could be obtained also in the case of the Global Hausman test. Second, in this article we have addressed the issue of bootstrapping the local Hausman test so that the test has high power while also having size close to the nominal, noticing that for the bootstrap test, these do not vary with bandwidth choice as much as for the asymptotic test. However, one could also select the bandwidth parameter using some form of calibration as in Shao and Politis (2013) and examine whether similar results are obtained. Finally, our Monte Carlo simulations reveal some small size distortions of the bootstrap procedures in the case of the over-identified models which requires further study.

3.A Appendix

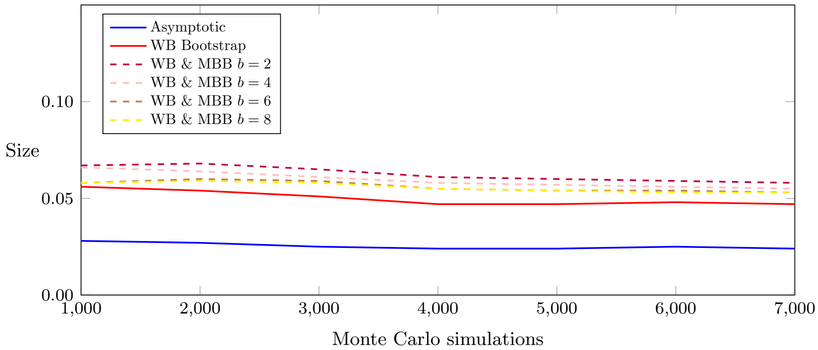
In section 3.A, we assess the robustness of the results with respect to a number of different attributes of the experiments employed in the paper.

3.A.1 Robustness Analysis

1. **Number of Monte Carlo Simulations**

   Figure 3.5 shows the results for the size of the asymptotic and bootstrap versions of the test for (3.10)-(3.11), $T = 100$ and $h_1 = h_2 = 0.4$ across different number of Monte Carlo simulations. Convergence is attained at 5000 simulations.

Figure 3.5: Coverage-size convergence for bootstrap and asymptotic versions of the Hausman test for model (3.10)-(3.11), $T = 100$ and $h_1 = h_2 = 0.4$.



2. **Bootstrap Alternatives**

Although in this article we employ the MBB and the Fixed-design WB, we also consider another alternative. To begin, by employing the MBB we implicitly assume that the size of each block is equal and fixed to b and that the sample size of each replication is equal to T which requires some sort of trimming in case T/b is not an integer and hence we have ignored any end effects. For example, because there is no data after $\{y_T, x_T\}$ the moving blocks method does not define a block of length b beginning at the end points. Politis and Romano (1991) proposed the Circular Bootstrap to "wrap" the data around in a circle so that $\{y_1, x_1\}$ follows directly after $\{y_T, x_T\}$. This path has also been explored in a preliminary analysis but provided similar results to MBB.

3. **Kernel Robustness**

As mentioned in Section 5, the confidence intervals for the parameter α_t appear to be quite wide especially around 2000 and hence it might well be the case that this is generating the difference of the p-values between the asymptotic and the bootstrap test statistic. As pointed out by Lucchetti and Valentini (2021), these wide confidence intervals are caused because of the first stage regression which includes the huge spike of the unemployment during the first lockdown, which seems to propagate throughout the whole sample when an infinite support kernel is employed. For this reason, we perform a sensitivity analysis of the empirical application with respect to different kernels. I consider two kernels with finite support, namely the Epanechnikov and the triangular. Figure 3.6 depicts the differences between the two

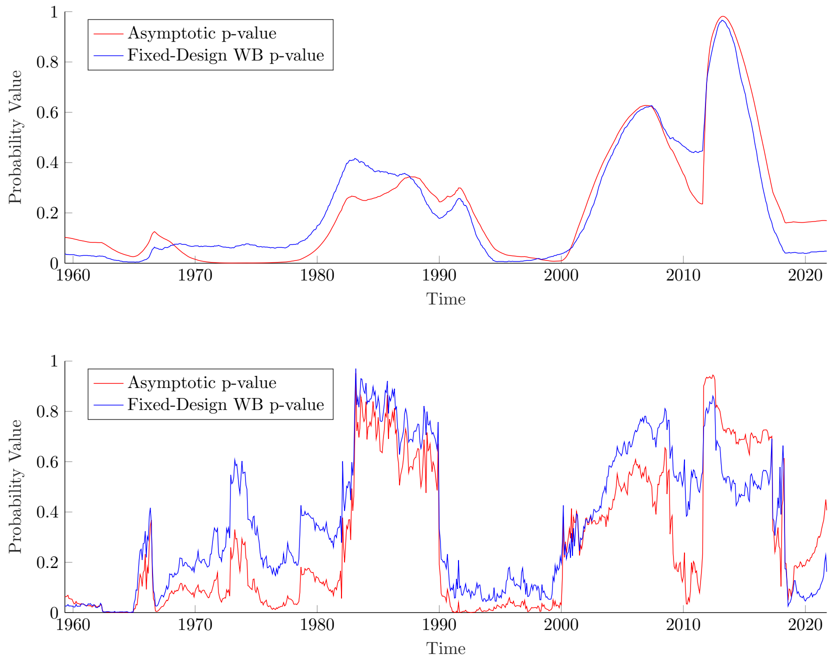


Figure 3.6: Empirical results for model (3.15).The two panels graph the empirical p-values of the asymptotic and bootstrap Hausman tests using $H = L = T^{0.7}$. The upper panel uses the Epanechnikov kernel while the lower uses the triangular.

3.A.2 Proofs

The following notation is used in the sequel: $L[\cdot]$ is used to denote "the law of" while '$\Rightarrow$' signifies weak convergence. O_p and o_p signify being bounded in probability and converging in probability to zero respectively; $P^\star$ and $E^\star$ denote bootstrap probability and expectation respectively, conditional on the data y, x; for a sequence of bootstrap quantities $\{Z_T^\star\}$, $Z_T^\star = O_p^\star(1)$ signifies that for any $c_T \to \infty$, $P^\star(|Z_T^\star| > c_T) = o_p(1)$ and $V_T^\star = o_p^\star(1)$ signifies that $P^\star(|V_T^\star| > \epsilon) = o_p(1)$ for any $\epsilon > 0$; similarly, $L^\star[\cdot]$ denotes bootstrap law conditional on the data y, x; $\hat{\sigma}^2_{\hat{u},t} := K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \hat{u}_j^2$ based on residuals $\hat{u}_j = y_j - x_j^{'} \tilde{\beta}_j$; Likewise, $\hat{\sigma}^{\star 2}_{\hat{u}^\star,t} := K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \hat{u}_j^{\star 2}$ is the bootstrap counterpart of $\sigma^2_{\hat{u},t}$. In what follows, the following relationship between spectral and Frobenius norms will be used frequently: $||AB|| \leq ||A||_{sp} ||B||$.

We first prove the consistency of TV-OLS under more general conditions than Giraitis et al. (2014).

Throughout the rest of the paper the following assumptions are made:

Assumption 3. *1. For $\theta > 8$, uniformly over l and t,*

$$E|x_{l,t}|^\theta, E|u_t|^\theta \leq C < \infty. \tag{3.A.16}$$

2. The processes $(x_{l,k}x_{k,t} - Ex_{l,t}x_{k,t})$, $(x_{l,t}u_t)$, (u_t) are α-mixing with mixing coefficients α_k such that for some $0 < \phi < 1$ and $c > 0$

$$\alpha_k \leq c\phi^k, \ \ k \geq 1. \tag{3.A.17}$$

3. The matrix $\Sigma_{xx,t} = E[x_t x_t^{'}]$ *and* $\sigma_{u,t} = E[u_t^2]$ *satisfy* $\max_{t\geq 1} ||\Sigma_{xx,t}^{-1}|| < \infty$, $\max_{t\geq 1} ||\sigma_{u,t}^{-1}||_{sp} < \infty$.

Assume also that $\beta_t = \beta_{T,t}$ are triangular arrays of matrices whose elements satisfy either Assumption 4 or Assumption 5 below.

Assumption 4. $(\beta_{k,t})$ *is a non-random sequence of real numbers that satisfies the following smoothness condition*

$$|\beta_{k,t} - \beta_{k,s}| \leq C\frac{|t-s|}{T},\ \ t, s = 1, \cdots, T, \tag{3.A.18}$$

and is uniformly bounded in t.

Assumption 5. $(\beta_{k,t})$ *is a random process that satisfies the smoothness condition*

$$|\beta_{k,t} - \beta_{k,s}| \leq (\frac{|t-s|}{T})^{1/2} q_{k,s},\ \ t, s = 1, \cdots, T, \tag{3.A.19}$$

where the distribution of variables $X{=}\beta_{k,t}q_{k,ts}$ *has a thin tail:*

$$P(|X| \geq \omega) \leq exp(-c_0|\omega|^{\lambda}),\ \ \omega > 0, \tag{3.A.20}$$

for $c_0, \lambda > 0$ *that do not depend on* k, t, s *and* T.

We also assume that the bandwidth parameter H satisfies the following regularity condition:

$$c_1 T^{1/(\theta/2-1)+\delta} \geq H, \tag{3.A.21}$$

where $c_1 > 0$ and $\delta > 0$ is arbitrary small.

Assumptions 3-5 above, are analogous to assumptions 1-3 in GKM. The assumption on the bandwidth parameter is analogous to equation (17) in GKM.

Finally, assume that

$$Ex_t u_t = 0 \text{ for } t \geq 1. \tag{3.A.22}$$

The next theorem establishes consistency of the TV-OLS estimator.

Theorem 6. *Suppose* x_t *and* u_t *satisfy Assumption 3 and* β_t *satisfy either Assumption 4 or Assumption 5. Then, as* $T \to \infty$, *the estimator* $\hat{\beta}_t$, *computed with bandwidth parameter* H *satisfying (3.A.21), has the property that:*

$$\max_{j=1,\cdots,T} ||\hat{\beta}_t - \beta_t|| = O_p(r_{T,H}) \tag{3.A.23}$$

where

$$r_{T,H} = \begin{cases} H^{-1/2}log^{1/2}T + HT^{-1}, & \textit{if } \beta_t \textit{ satisfies Assumption 4} \\ H^{-1/2}log^{1/2}T + (H/T)^{1/2}log^{1/\alpha}T, & \textit{if } \beta_t \textit{ satisfies Assumption 5} \end{cases}$$

Proof *of Theorem 6:* Write

$$\hat{\beta}_t = \beta_t + S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu,t}) \tag{3.A.24}$$

where $S_{xx,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_jx_j^{'}$, $\Delta_t^{(2)} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_jx_j^{'}(\beta_j - \beta_t)$ and $S_{xu,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_ju_j$. By (3.A.24) we now have,

$$\max_{t=1,\dots,T}||\hat{\beta}_t - \beta_t|| \leq \max_{t=1,\dots,T}||S_{xx,t}^{-1}||_{sp}(\max_{t=1,\dots,T}||\Delta_t^{(2)}|| + \max_{t=1,\dots,T}||S_{xu,t}||). \tag{3.A.25}$$

Hence, the desired result follows from showing that:

$$\max_{t=1,\dots,T}||S_{xx,t}^{-1}||_{sp} = O_p(1), \tag{3.A.26}$$

$$\max_{t=1,\dots,T}||\Delta_t^{(2)}|| = O_p(r_{T,H}), \tag{3.A.27}$$

$$\max_{t=1,\dots,T}||S_{xu,t}|| = O_p(H^{-1/2}log^{1/2}T). \tag{3.A.28}$$

Proof of (3.A.26): Expression (3.A.26) can be splitted in two parts as follows :

$$S_{xx,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}E[x_jx_j^{'}] + K_t^{-1}\sum_{j=1}^T (x_jx_j^{'} - E[x_jx_j^{'}]) =: S_{xx,t}^{(1)} + S_{xx,t}^{(2)}.$$

Now, write

$$S_{xx,t} = S_{xx,t}^{(1)}(1 + \tilde{\Delta}_t),\ \tilde{\Delta}_t = S_{xx,t}^{(1)\,-1}(S_{xx,t} - S_{xx,t}^{(1)}), \tag{3.A.29}$$

where if $||\tilde{\Delta}_t||_{sp} < 1$ then,

$$||S_{xx,t}^{-1}||_{sp} \leq ||(S_{xx,t}^{(1)})^{-1}||_{sp}||(1 + \tilde{\Delta}_t)^{-1}||_{sp} \leq ||(S_{xx,t}^{(1)})^{-1}||_{sp}(1 - ||\tilde{\Delta}_t||_{sp})^{-1} \tag{3.A.30}$$

$$\leq \max_{t=1,\dots,T}||(S_{xx,t}^{(1)})^{-1}||_{sp}(1 - \max_{t=1,\dots,T}||\tilde{\Delta}||_{sp})^{-1} \tag{3.A.31}$$

and will show that,

$$\max_{t=1,\dots,T}||(S_{xx,t}^{(1)})^{-1}||_{sp} = O_p(1), \tag{3.A.32}$$

$$\max_{t=1,\dots,T}||\tilde{\Delta}_t||_{sp} = o_p(1) \tag{3.A.33}$$

which together with (3.A.31) imply (3.A.26): $\max_{t=1,\dots,T}||S_{xx,t}^{-1}||_{sp} = O_p(1)$. Under Assumption 3(3) there exists $\nu > 0$, such that for all $t \geq 1$

$$\alpha^{'}\Sigma_{xx,t}\alpha \geq 1/\nu > 0,$$

thus for any $n \times 1$ vector $\alpha = (\alpha_1, \cdots, \alpha_p)^{'}$ such that $||\alpha||^2 = 1$,

$$\min_{||\alpha||=1} \alpha^{'} S^{(1)}_{xx,t} \alpha = \min_{||\alpha||=1} (K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \alpha^{'} \Sigma_{xx,t} \alpha) \geq \nu^{-1} (K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|}) = 1/\nu > 0$$

and hence the smallest eigenvalue of $S^{(1)}_{xx,t}$ is not smaller than $1/\nu > 0$ which proves (3.A.32). To show (3.A.33), bound

$$||\tilde{\Delta}_t||_{sp} \leq ||(S^{(1)}_{xx,t})^{-1}||_{sp} ||S_{xx,t} - S^{(1)}_{xx,t}||$$

and hence, in light of (3.A.32), it suffices to show that

$$\max_{t=1,\ldots,T} ||S_{xx,t} - S^{(1)}_{xx,t}|| = o_p(1).$$

By Assumption 3(2), the (l,k)th component $\omega_{l,k,j} = x_{l,j} x_{k,j} - E[x_{l,j} x_{k,j}]$ of $xx^{'}$ is α-mixing. Combining this with (3.A.21),Assumption 3(1) for $\theta^{'} = \theta/2$ and Lemma 5(i) of GKM, we obtain

$$\max_{t=1,\ldots,T} ||S_{xx,t} - S^{(1)}_{xx,t}|| = O_p(H^{-1/2} log^{1/2} T) = o_p(1), \tag{3.A.34}$$

which proves (3.A.33) and concludes the proof of (3.A.26).

Proof of (3.A.27): A typical element of $\Delta^{(2)}_t$ consists of a linear combination of sums

$$s_t := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \omega_{lk,j} (\beta_{k,j} - \beta_{k,t})$$

where $\omega_{lk,j} - E[\omega_{lk,j}]$ is an α-mixing sequence and $E|\omega_{lk,j}|^{\theta/2} \leq C < \infty$ for all j. Suppose β_t satisfies Assumption 2 in GKM. Then $|\beta_{l,j} - \beta_{l,t}| \leq C(\frac{|t-s|}{T})$ and (88) of lemma 5(ii) in GKM implies the bound (3.A.27) $\max_{t=1,\ldots,T} |s_t| = O_p(r_{T,H})$ with $r_{T,H}$ as defined for Assumption 4. Now suppose β_t satisfies Assumption 5 instead. Then $|\beta_{l,t} - \beta_{l,s}| \leq (\frac{|t-s|}{T})^{1/2} r_{l,ts}$ and Lemma 5(iii) of GKM implies (3.A.27) with $r_{T,H}$ as defined for Assumption 5. Proof of (3.A.28):

The $l-th$ element of the $p-vector$, S_{xu} is

$$\tilde{s}_t = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} x_{l,j} u_j$$

which under Assumption 3 (2) is an α-mixing sequence. Now, since (3.A.22) holds, Lemma 5 of GKM can be employed and hence $\max_{t=1,\ldots,T} ||S_{xu,t}|| = O_p(H^{-1/2} log^{1/2} T)$ which shows (3.A.28). This concludes the proof of the theorem. □

Proof *of Theorem 5:* Denote as $u^{\star}_t = \eta_t \hat{u}_t$ the bootstrap errors and $y^{\star}_t$ the dependent variable built from residuals as described at section 3 for Fixed-design Wild Bootstrap. Then $y^{\star}_j = x^{'}_j \hat{\beta}_j + u^{\star}_j = x^{'}_j \hat{\beta}_t + x^{'}_j (\hat{\beta}_j - \hat{\beta}_t) + u^{\star}_j$

and hence the OLS estimator can be rewritten as

$$\hat{\beta}_t^\star = \hat{\beta}_t + S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu^\star,t}^\star), \tag{3.A.35}$$

where $S_{xx,t}^{-1} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_j x_j^{'}$, $\Delta_t^{(2)} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_j^{'}(\beta_j - \beta_t)$ and $S_{xu^\star,t}^\star = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_j u_j^{\star'}$. Similarly, the IV estimator can be rewritten as

$$\tilde{\beta}_t^\star = \hat{\beta}_t + S_{\hat{x}x,t}^{-1}(\hat{\Delta}_t^{(1)} + S_{\hat{x}u^\star,t}^\star). \tag{3.A.36}$$

Now using the above expressions, the difference between the two estimators $\hat{\beta}_t^\star - \tilde{\beta}_t^\star$, can be rewritten as follows:

$$\hat{\beta}_t^\star - \tilde{\beta}_t^\star = (\hat{\beta}_t^\star - \hat{\beta}_t) - (\tilde{\beta}_t^\star - \hat{\beta}_t) \tag{3.A.37}$$

$$= S_{xx,t}^{-1}S_{xu^\star,t}^\star - S_{\hat{x}x,t}^{-1}S_{\hat{x}u^\star,t}^\star + \hat{R}_t \tag{3.A.38}$$

where $\hat{R}_t = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}\hat{\omega}_{t,j}$ and

$$\hat{\omega}_{t,j} = S_{xx,t}^{-1}x_j x_j(\hat{\beta}_j - \hat{\beta}_t) - S_{\hat{x}\hat{x},t}^{-1}\hat{x}_j x_j^{'}(\hat{\beta}_j - \hat{\beta}_t). \tag{3.A.39}$$

. Next, denote, $V_{T,t}^\star = V_t^\star = S_{\hat{x}\hat{x},t}^{1/2}S_{xx,t}^{1/2}(\hat{\beta}_t^\star - \tilde{\beta}_t^\star)$ to obtain,

$$V_t^\star = V_{1,t}^\star + \hat{V}_{2,t}, \tag{3.A.40}$$

$$V_{1,t}^\star = S_{\hat{x}\hat{x},t}^{1/2}S_{xx,t}^{1/2}\{S_{xx,t}^{-1}S_{xu^\star,t}^\star - S_{\hat{x}x,t}^{-1}S_{\hat{x}u^\star,t}^\star\}, \tag{3.A.41}$$

$$\hat{V}_{2,t} = S_{\hat{x}\hat{x},t}^{1/2}S_{xx,t}^{1/2}\hat{R}_t. \tag{3.A.42}$$

Now, the fixed-design wild bootstrap analogue of the Hausman test statistic takes the following form:

$$S^\star = \frac{K_t^2}{K_{2,t}}V_t^{'\star}\hat{\Sigma}_{\hat{v}\hat{v},t}^{-1}V_t^\star\sigma_{u^\star,t}^{-2\star} \tag{3.A.43}$$

which can be splitted into the main term and the remainder, that is

$$\sigma_{u^\star,t}^{-1\star}\Sigma_{vv,t}^{-1/2}K_t K_{2,t}^{-1/2}V_t^\star = U_t^\star + r_t^\star, \tag{3.A.44}$$

$$U_t^\star = \sigma_{u^\star,t}^{-1\star}\Sigma_{vv,t}^{-1/2}K_t K_{2,t}^{-1/2}V_{1,t}^\star, \tag{3.A.45}$$

$$r_t^\star = \sigma_{u^\star,t}^{-1\star}\Sigma_{vv,t}^{-1/2}K_t K_{2,t}^{-1/2}\hat{V}_{2,t}. \tag{3.A.46}$$

Hence, we need to prove

$$||r_t^\star|| = o_p^\star(1), \tag{3.A.47}$$

$$L^\star[U_t^\star] \overset{p}{\Rightarrow} \mathbb{N}(0, I). \tag{3.A.48}$$

Proof of (3.A.47): From Lemma 1, we have that:

$$||\hat{V}_{2,t}|| \leq (H/T)^\gamma A_T q_t = O_p((H/T)^\gamma). \tag{3.A.49}$$

Hence, from (3.A.49) combined with Assumption 1(iii) in GKM, $K_t K_{2,t}^{-1/2} = O(H^{1/2})$ we have,

$$||r_t^\star|| \leq K_t K_{2,t}^{-1/2} ||\Sigma_{\nu\nu,t}^{-1}|| \; ||\hat{V}_{2,t}|| \tag{3.A.50}$$

$$= O(H^{1/2}(H/T)^\gamma) = o_p^\star(1). \tag{3.A.51}$$

Proof of (3.A.48): It now remains to show that $L^\star(U_t^\star) \Rightarrow \mathbb{N}(0, I)$ which proves the desired result. For that, we rewrite $U_t^\star$ as

$$\dot{U}_t^\star = B_t^\star K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star, \; p_{t,j} = L_t x_j - L_t^{-1} \Psi_j^{'} z_j \tag{3.A.52}$$

where

$$\begin{aligned} L_{1,t} &= S_{\hat{x}\hat{x},t}^{1/2} S_{xx,t}^{-1/2}, \; L_{2,t} = S_{\hat{x}\hat{x},t}^{1/2} S_{xx,t}^{1/2} S_{\hat{x}\hat{x},t}^{-1} \\ L_t &= \Sigma_{x-\nu,x-\nu,t}^{1/2} \Sigma_{xx,t}^{-1/2}, \; B_t^\star = \sigma_{u^\star,t}^{\star -1} \Sigma_{\nu\nu,t}^{-1}. \end{aligned} \tag{3.A.53}$$

Now,

$$b^{'} \dot{U}_t^\star = b^{'} U_t + b^{'} (\dot{U}_t^\star - U_t) \tag{3.A.54}$$

where the first term on the RHS follows a $N(0, 1)$ as shown in GKM. Hence, to conclude the proof, it remains to show that

$$b^{'} (\dot{U}_t^\star - U_t) = o_{p^\star}(1). \tag{3.A.55}$$

Next, by substituting in the expression above:

$$b^{'} (B_t^\star K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star - B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j) \tag{3.A.56}$$

and by adding and subtracting in (3.A.56) the term $B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star$ the expression now becomes:

$$
\begin{aligned}
& b^{'}(B_t^\star K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star - B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star \\
& + B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star - B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j) = \\
& = b^{'}[\underbrace{(B_t^\star - B_t)}_{C_1} \underbrace{K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} u_j^\star}_{C_2} + \underbrace{B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} (u_j^\star - u_j)]}_{C_3}
\end{aligned}
$$

Hence it suffices to show that each of the three terms above are $o_{p^\star}(1)$. Before proceeding, note $||B_t|| = O_p(1)$ by the assumptions of the theorem in GKM.

For C_1, we have:

$$
\begin{aligned}
||B_t^\star - B_t|| &= ||\sigma_{u^\star,t}^{\star -1} \Sigma_{\nu\nu,t}^{-1} - \sigma_{u,t}^{-1} \Sigma_{\nu\nu,t}^{-1}|| \\
&\leq ||\Sigma_{\nu\nu,t}^{-1}|| \; ||\sigma_{u^\star,t}^{\star -1} - \sigma_{u,t}^{-1}|| \\
&\leq C||\sigma_{u^\star,t}^{\star -1} - \sigma_{u,t}^{-1}||
\end{aligned}
$$

by Assumption 1(iii) in GKM. Hence,

$$
\begin{aligned}
||\sigma_{u^\star,t}^{\star -1} - \sigma_{u,t}^{-1}|| &= ||\sigma_{u^\star,t}^{\star -1} - \hat{\sigma}_{\hat{u},t}^{-1} + \hat{\sigma}_{\hat{u},t}^{-1} - \sigma_{u,t}^{-1}|| \\
&\leq ||\sigma_{u^\star,t}^{\star -1} - \hat{\sigma}_{\hat{u},t}^{-1}|| + ||\hat{\sigma}_{\hat{u},t}^{-1} - \sigma_{u,t}^{-1}|| \\
&\leq ||\sigma_{u^\star,t}^{\star -1} - \hat{\sigma}_{\hat{u},t}^{-1}|| + o_p(1)
\end{aligned}
$$

by consistency of the estimator for $\sigma_{u,t}$. Hence, it now remains to show that $||\sigma_{u^\star,t}^{\star -1} - \hat{\sigma}_{\hat{u},t}^{-1}|| = o_{p^\star}(1)$, which is easily shown by Lemma 2 below.

Next, for C_2, using Lemma 3 and noting that the conditions of Lemma are satisfied since $u_t^\star$ is a mean zero random variable such that i) $E^\star(u_t^\star) = E^\star(\hat{u}_t \eta_t) = \hat{u}_t E^\star(\eta_t) = 0$ and ii) $\max_{k=1,\dots,T} \sum_{j=1}^{T} |E^\star[u_k^\star u_j^\star]| = 0$, we have

$$
\begin{aligned}
||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|}^2 p_{t,j} u_j^\star||^2 &= O_p(A_T^{02}) O(K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2) o_p^\star(1) \\
&= O_p(A_T^{02}) o_p^\star(1) \\
&= o_p^\star(1),
\end{aligned}
$$

and using (A.50), Lemma 9 in GKM, we have $||p_{t,j}|| \leq A_T r_{t,j}$ where $A_T = O_p(1)$ does not depend on t, j and $\max_{t,j=1,\dots,T} E r_{t,j}^2 = O(1)$.

Finally, for C_3, we have:

$$||B_t K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} (u_j^\star - u_j)|| \leq O(1) ||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} (u_j^\star - u_j)||$$

and since $p_{t,j}$ is conditionally independent of $(u_j^\star - u_j)$, we can add and subtract $\hat{u}_j$ in the above, and by noting that

$$\begin{aligned}
||\hat{u}_j - u_j|| &\leq ||(\tilde{\beta}_j - \beta_j|| \; ||x_j|| \\
&\leq (\max_{j=1,\dots,T} ||\tilde{\beta}_j - \beta_j||) \; ||x_j|| \\
&= o_p(1) \text{ (By theorem 1 in GKM and Assumption 1 about } x_j)
\end{aligned}$$

we have

$$\begin{aligned}
||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} (\hat{u}_j - u_j)|| &\leq K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} ||p_{t,j}|| \; ||(\hat{u}_j - u_j)|| \text{ (Note that } K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} = O(H^{1/2})) \\
&\leq O(H^{1/2}) O_p(1) r_{t,j} \; o_p(1) \\
&= o_p(1).
\end{aligned}$$

Next, for

$$||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} \underbrace{(u_j^\star - \hat{u}_j)}_{m_j^\star} ||,$$

since

$$\begin{aligned}
\max_{j=1,\dots,T} \sum_{j=1}^{T} |E^\star(m_k^\star m_j^\star)| &= \max_{j=1,\dots,T} \sum_{j=1}^{T} |\hat{u}_j^2 E^\star(\eta_j - 1)^2| \\
&= \max_{j=1,\dots,T} \sum_{j=1}^{T} |\hat{u}_j^2 E^\star(\eta_j^2 - 2\eta_j + 1)| \\
&= 2 \max_{j=1,\dots,T} \sum_{j=1}^{T} |\hat{u}_j^2| \\
&= O_p(1),
\end{aligned}$$

we can use Lemma 3, and hence we have:

$$||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} p_{t,j} m_j^\star||^2 \leq O_p^\star(A_T^{0\;2}) O(K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2 r_{t,j}^2) = O_p^\star(A_T^{0\;2}) = o_p^\star(1),$$

thus completing the proof. □

Next, define the following quantities: $A_{1,T} = \max_{t=1,\dots,T} ||S_{xx,t}^{-1}|| +$
$(\max_{t=1,\dots,T} ||S_{xx,t}^{-1}||)(\max_{t=1,\dots,T} ||\hat{\Psi}_j - \Psi_j|| + 1)$, $A_{2,T} = \max_{j=1,\dots,T} (||\hat{\Psi}_j - \Psi_j|| + 1)^2$, $\hat{q}_{1,T} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (\frac{|j-t|}{H})^\gamma \hat{\nu}_{t,j}$

where $\hat{\nu}_{t,j} = \{||x_j||^2 + (1+||\Psi_j||)||z_j x_j^{'}||\}||(\hat{\beta}_j - \hat{\beta}_t)(\frac{T}{|j-t|})^{\gamma}||$and $q_{2,T} = K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}(1+||\Psi_j||)^2||z_j||^2$ such that $A_T = A_{1,T}A_{2,T}$ and $q_t = q_{2,t}^{1/2}||S_{xx,t}||^{1/2}\hat{q}_{1,t}$.

Lemma 1. *Let the assumptions of Theorem 4 in GKM be satisfied. Let $\gamma = 1$ and $\gamma = 1/2$ if β_t satisfies Assumption 2 and Assumption 3, respectively. Then $\hat{V}_{2,t}$ has property:*

$$||\hat{V}_{2,t}|| \le (H/T)^{\gamma} A_T q_t = O_p((H/T)^{\gamma}) \tag{3.A.57}$$

where $A_T = O_p(1)$ and does not depend on t and also $\max_{t=1,\dots,T} Eq_t = O(1)$.

Proof *of Lemma 1.* : From (3.A.42) we have:

$$\begin{aligned} ||\hat{V}_{2,t}|| &\le ||S_{\hat{x}\hat{x},t}^{1/2}||_{sp}||S_{xx,t}^{1/2}||_{sp}||\hat{R}_t|| = ||S_{\hat{x}\hat{x},t}||_{sp}^{1/2}||S_{xx,t}||_{sp}^{1/2}||\hat{R}_t|| \\ &\le ||S_{\hat{x}\hat{x},t}||^{1/2}||S_{xx,t}||^{1/2}||\hat{R}_t||. \end{aligned}$$

Hence, we will show

$$||\hat{R}_t|| \le (H/T)^{\gamma} A_{1,T}\hat{q}_{1,t}, \tag{3.A.58}$$

$$||S_{\hat{x}\hat{x},t}||^{1/2} \le A_{2,T} q_{2,t}, \tag{3.A.59}$$

$$\max_{t=1,\dots,T} E||S_{xx,t}||^2 = O(1), \tag{3.A.60}$$

where $A_{i,T} = O_p(1)$ for $i = 1, 2$, $\max_{t=1,\dots,T} E|q_{2,t}|^2 = O(1)$, and $\max_{t=1,\dots,T} E|\hat{q}_{1,t}|^2 = O(1)$. Then

$$||\hat{V}_{2,t}|| \le ||S_{\hat{x}\hat{x},t}||^{1/2}||S_{xx,t}||^{1/2}||\hat{R}_t|| \tag{3.A.61}$$

$$\le (H/T)A_T q_t, \text{ where } A_T = A_{1,T}A_{2,T},\ q_t = q_{2,t}||S_{xx,t}||^{1/2}\hat{q}_{1,t}. \tag{3.A.62}$$

Clearly, $A_T = O_p(1)$ while

$$\max_{t=1,\dots,T} Eq_t^2 \le \max_{t=1,\dots,T} (E\hat{q}_{1,t}^2 + ||S_{xx,t}||^2 + Eq_{2,t}^2) = O_p(1),$$

which proves (3.A.57). Now for (3.A.58)-(3.A.60) recall $\hat{x}_j = \hat{\Psi}_j z_j$ and use the bound

$$||\hat{\Psi}_j|| \le ||\hat{\Psi}_j - \Psi_j|| + ||\Psi_j|| \le (||\hat{\Psi}_j - \Psi_j|| + 1)(1 + ||\Psi_j||). \tag{3.A.63}$$

Then $\hat{\omega}_{t,j}$ in (3.A.39) can now be bounded as follows:

$$|\hat{\omega}_{t,j}| \le (\frac{|j-t|}{T})^{\gamma} A_{1,t}\hat{\nu}_{t,j},$$

Then $\hat{R}_t \leq (H/T)^{\gamma} A_{1,T} \hat{q}_{1,t}$, $\hat{q}_{1,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (\frac{|j-t|}{H})^{\gamma} \hat{\nu}_{t,j}$. From (A.45) in GKM where

$$\max_{t=1,\ldots,T} ||S_{xx,t}^{-1}|| = O_p(1),$$

$$\max_{t=1,\ldots,T} ||S_{\hat{x}\hat{x},t}^{-1}|| = O_p(1)$$

and Theorem 1 in GKM we have that $A_{1,T} = O_p(1)$. Likewise for $\hat{q}_{1,t}$ we have:

$$\begin{aligned} E\hat{q}_{1,t}^2 \leq & E(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (\frac{|j-t|}{H})^{\gamma} \hat{\nu}_{t,j})^2 \\ \leq & (\max_{j=1,\ldots,T} E\hat{\nu}_{t,j}^2)(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (\frac{|j-t|}{H})^{\gamma})^2 \\ \leq & C \max_{j=1,\ldots,T} E\hat{\nu}_{t,j}^2 = O(1), \end{aligned} \tag{3.A.64}$$

because under Assumptions 1-2 in GKM and the fact that we have uniformly in j, t

$$\begin{aligned} E\hat{\nu}_{t,j}^2 \leq & E[(||x_j||^2 + ||z_j x_j^{'}||)^2 \{(1 + ||\Psi_j||)^2 ||(\hat{\beta}_j - \hat{\beta}_t)(\frac{T}{|j-t|})^{\gamma}||^2\}] \\ \leq & E[(||x_j||^2 + ||z_j x_j^{'}||)^4] + E[(1 + ||\Psi_j||)^4 ||(\hat{\beta}_j - \hat{\beta}_t)](\frac{T}{|j-t|})^{\gamma}||^4] \\ \leq & C, \end{aligned} \tag{3.A.65}$$

where $E||\hat{\beta}_j - \hat{\beta}_t||^4 = O(1)$. To see this, we argue as follows: by the c_r inequality (see White (2014) page 35)

$$E(||\hat{\beta}_j - \hat{\beta}_t||^4) \leq 2^3 (E(||\hat{\beta}_j||^4) + E(||\hat{\beta}_t||^4))$$

and so it suffices to prove that

$$E(||\hat{\beta}_t||^4) = O(1).$$

From the c_r inequality the above expression now becomes

$$E(||\hat{\beta}_t||^4) = E(||\beta_t + S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu,t})||^4) \tag{3.A.66}$$

$$\leq 2^3 [E(||\beta_t)||^4 + E(||S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu,t})||^4)], \tag{3.A.67}$$

hence I have to show that

$$E(||S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu,t})||^4) = O(1) \tag{3.A.68}$$

Now again, by applying the c_r inequality we obtain:

$$E(||S_{xx,t}^{-1}(\Delta_t^{(2)} + S_{xu,t})||^4) \leq 2^3 [E(||S_{xx,t}^{-1}\Delta_t^{(2)}||)^4 + E(||S_{xx,t}^{-1}S_{xu,t}||)^4].$$

Next, by using the property $||AB|| \leq ||A||_{sp}||B|| \leq ||A|| \; ||B||$ and Assumption 2.1 we only need to show that

$E(||\Delta_t^{(2)}||) = O(1)$.

So, for $\Delta_t^{(2)}$ we have

$$||\Delta_t^{(2)}||^4 \leq (K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \underbrace{||x_j x_j^{'}(\beta_j - \beta_t)}_{\text{Z}}||)^4. \tag{3.A.69}$$

Hence,

$$E(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} z_{jt})^4 = K_t^{-4} \sum_{j1,j2,j3,j4}^{T} b_{H,|j1-t|} \ldots b_{H,|j4-t|} E(z_{j1,t} \ldots z_{j4,t}) \tag{3.A.70}$$

Now by Hölder's inequality:

$$E(z_{j1,t} \ldots z_{j4,t}) \leq (E(z_{j1,t}^4)^{1/4} \ldots E(z_{j4,t}^4)^{1/4},$$

and so focusing on z_{j1}:

$$(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (E(z_{j1,t}^4)^{1/4}))^4 \leq C(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|})^4.$$

The last part is derived by

$$\begin{aligned} E(z_{j1,t}^4) &= E[(||x_j x_j^{'}(\beta_j - \beta_t))||)^4] \\ &\leq E(||x_j||^4 ||x_j||^4 ||\beta_j - \beta_t||^4) \\ &= E(||x_j||^8 ||\beta_j - \beta_t||^4), \end{aligned}$$

where now the last line becomes, by Hölder's inequality:

$$(E||x_j||^{8p})^{1/p} (E||\beta_j - \beta_t||^{4q})^{1/q}$$

for $p = 1 + \epsilon$ which follows from Assumption 2.2. Establishing $E(||x_t u_t||^4) = O(1)$ is straightforward from Assumption 1 in GKM. □

Lemma 2. *Under the assumptions of Theorem 1 of GKM, we have:*

$$||\sigma_{u^\star,t}^\star - \hat{\sigma}_{\hat{u},t}|| = o_p^\star(1).$$

Proof *of Lemma 2:* Let $\hat{u}_t^\star = y_t^\star - x_t^{'} \tilde{\beta}_t^\star$ and $y_t^\star = x_t^{'} \hat{\beta}_t + u_t^\star$, where $u_t^\star = \hat{u}_t \eta_t$ and $\hat{u}_t = y_t - x_t^{'} \hat{\beta}_t$. Then $\hat{u}_t^\star = u^\star + x_t^{'}(\tilde{\beta}_t^\star - \hat{\beta}_t)$ and hence $|\sigma_{u^\star,t}^{\star-1} - \hat{\sigma}_{\hat{u}^\star,t}^{\star-1}| = |K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} x_j^{'}(\tilde{\beta}_j^\star - \hat{\beta}_j)| \leq K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} |x_j^{'}(\tilde{\beta}_j^\star - \hat{\beta}_j)|$. Now, $K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} |x_j^{'}(\tilde{\beta}_j^\star - \beta_j + \beta_j - \hat{\beta}_j)| \leq K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} |x_j^{'}|(|\tilde{\beta}_j^\star - \beta_j| + |\beta_j - \hat{\beta}_j|)$ since by Lemma 5 in GKM (87) we have $\max_{t=1,\ldots,T} H^{-1} \sum_{j=1}^{T} b_{H,|j-t|} |x_j| = O_p(1)$ and also under the null of exogeneity we have that both $|\tilde{\beta}_j^\star - \beta_j|$ and $|\beta_j - \hat{\beta}_j|$ are $o_{p^\star}(1)$ and $o_p(1)$ respectively. The part that $|\tilde{\beta}_j^\star - \beta_j| = o_{p^\star}(1)$ is shown by noting that $\tilde{\beta}_j^\star = \hat{\beta}_j + \sum_{j=1}^{T} b_{H,|j-t|} x_j^{'} u_j^\star$ and hence adding and subtracting $\hat{\beta}_j$ in the expression $||\tilde{\beta}_j^\star - \beta_j||$ we have $|\tilde{\beta}_j^\star - \hat{\beta}_j + \hat{\beta}_j - \beta_j|$ so by triangle inequality it suffices to show that $||\tilde{\beta}_j^\star - \hat{\beta}_j|| = o_{p^\star}(1)$ which is shown as follows:

$$
\begin{aligned}
||\tilde{\beta}_j^\star - \hat{\beta}_j|| &= ||\sum_{j=1}^{T} b_{H,|j-t|} x_j^{'} u_j^\star|| \\
&\leq \underbrace{\sum_{j=1}^{T} b_{H,|j-t|}}_{O(H)} ||x_j^{'} u_j^\star|| \text{ (by triangular inequality)}
\end{aligned}
$$

Now, we can employ the Cauchy-Schwarz Inequality and hence $||x_j^{'} u_j^\star|| \leq (||x_j^{'2}||\ ||u_j^{\star\ 2}||)^{1/2} \leq (\underbrace{||x_j^{'}||^2}_{C_{21}}\ \underbrace{||u_j^\star||^2}_{C_{22}})^{1/2}$.

Next, for C_{22} we have

$$
\begin{aligned}
||u_j^\star||^2 &= ||\hat{u}_j \eta_j||^2 \\
&\leq ||\hat{u}_j||^2\ ||\eta_j||^2 \\
&\leq ||u_j + x_j^{'}(\hat{\beta}_j - \beta_j)||^2\ ||\eta_j||^2
\end{aligned}
$$

where $||\eta_j||^2$, is bounded. For $||u_j + x_j^{'}(\hat{\beta}_j - \beta_j)||^2$, we first use the triangular inequality, assumption 1 in GKM about $||u_j||, ||x_j||$ and that $||\hat{\beta}_j - \beta_j|| = o_p(1)$. □

Lemma 3. *Consider the sum*

$$S_{T,t}^\star = \sum_{j=1}^{T} w_{T,j} u_j^\star$$

where $w_{T,j}$ are random $p \times 1$ vectors and $(u_j^\star)$ are scalar zero mean random variables (conditionally) independent of $w_{T,j}$ such that

$$\max_{k=1,\ldots,T} \sum_{j=1}^{T} |E^\star[u_k^\star u_j^\star]| = 0 \tag{3.A.71}$$

and suppose $t = t_T \in [1, \ldots, T]$ may vary with T. Also, assume

$$||w_{T,j}|| \leq A_T q_{T,j},\ \ j = 1, \ldots, T \tag{3.A.72}$$

where A_T does not depend on j and $Eq_{T,j}^2 \leq \infty$. Then, as $T \to \infty$,

$$||S_{T,t}^\star|| = O_{p^\star}(A_T^2) o_p^\star(1) = o_p^\star(1) \tag{3.A.73}$$

Proof *of Lemma 3:* We will show that

$$E^\star ||A_T^{-1} S_{T,t}^\star||^2 = 0$$

which implies $||A_T^{-1} S_{T,t}^\star||^2 = o_p^\star(1)$ since $S_{T,t}^\star = A_T(A_T^{-1} S_{T,t}^\star)$. We have

$$E^\star ||A_T^{-1} S_{T,t}^\star||^2 \leq \sum_{k,j=1}^{T} A_T^{-2} ||w_{T,k}||\ ||w_{T,j}|| E^\star[u_j^\star u_s^\star] \text{ (by conditional independence)}$$

where by (3.A.72),

$$A_T^{-2}||w_{T,k}||\ ||w_{T,j}|| \leq q_{T,k}q_{T,j} \leq q_{T,k}^2 + q_{T,j}^2$$

and hence,

$$\begin{aligned}E^\star||A_T^{-1}S_{T,t}^\star|| &\leq 2\sum_{j=1}^{T} q_{T,j}^2|E^\star[u_j^\star u_k^\star]| \\ &\leq 2\sum_{j=1}^{T} q_{T,j}^2|(\max_{j=1,\ldots,T})\sum_{k=1}^{T}|E^\star[u_j^\star u_k^\star]| = 0\end{aligned}$$

since from Markov's inequality and noting that $Eq_{T,j}^2 \leq \infty$, we have $||q_{T,j}^2|| < \infty$. □

Chapter 4

Bandwidth Selection for Estimators of Time-Varying Stochastic Coefficient Models

1 Introduction

Structural change and parameter instability are pervasive in relationships among economic and financial variables. To account for such instability in cases where the latter is considered to be relatively smooth rather than abrupt, various models with smoothly time-varying coefficients, have been proposed, along with suitable methods for inference on the coefficient path. These include locally linear models with parameters that vary in a continuous manner according to the values of observable variables (e.g., Terasvirta (1998)), models with deterministic coefficients that are smooth functions of a rescaled time index (e.g., Robinson (1989, 1991); Cai (2007); Zhang and Wu (2012); Chen (2015b)), and models with stochastic coefficients evolving as multivariate ARIMA processes (e.g., Nicholls and Pagan (1985)).

In more recent work, Giraitis et al. (2021) (GKM hereafter) consider linear models in which little structure is imposed on time-varying coefficients – the latter may be deterministic or stochastic, subject only to certain smoothness and boundedness conditions. In addition, GKM allow the explanatory variables in the model to be potentially endogenous, in the sense of being correlated with the unobservable errors, a setting which, like that of Chen (2015b), is often relevant in econometrics. When a set of instrumental variables (IV) is available, inference on the time-varying coefficients may be based on one of the kernel IV estimators proposed by GKM. The obvious advantage of such estimators based on local smoothing is that they do not rely on parametric specifications for the time-dependence of the parameters. However, as with all kernel-based smoothing techniques, the practical use of kernel IV or least-squares (LS) estimators requires choice of a smoothing parameter, known as the bandwidth, as well as choice of a suitable kernel function – although it is generally accepted that the former choice has by far the biggest impact on the properties of kernel smoothers in terms of bias–variance trade-off.

In the context of nonparametric regression with deterministic or random (and exogenous) explanatory variables, several automated, data-driven bandwidth selection methods have been proposed for popular kernel-type estimators such as local polynomial estimators and estimators of the Nadaraya–Watson, Priestley–Chao and Gasser–Müller type. Those most commonly used are based on cross-validation (CV) methods, undersmoothing-penalized goodness-of-fit criteria such as, for example, Akaike's information criterion (AIC) and Rice's T-criterion, bootstrap resampling methods, and so-called plug-in rules – a useful overview can be found in Köhler et al. (2014). However, as already indicated, the properties of these data-driven bandwidth selection methods have almost exclusively been studied in regression settings where the explanatory variables are uncorrelated with or independent of the unobservable errors (or even deterministic). It is, therefore, of interest to examine whether automated selectors which are known to provide effective bandwidth choices under exogeneity (or fixed-design) conditions remain successful in the presence of endogeneity, and whether the performance of such selectors is affected by the strength of correlation between explanatory variables and errors.

Our objective in this paper is to investigate some of these issues by considering the performance of several automated bandwidth selection methods for kernel IV (and LS) estimators in a general setting similar to that in GKM, that is, in linear models with time-varying coefficients and explanatory variables which may be endogenous for the parameters of interest. More specifically, we consider automated bandwidth selection by means of four different methods, namely, ordinary (leave-one-out) CV, a nonparametric variant of a bias-corrected version of AIC, and wild bootstrap (WB) and dependent wild bootstrap (DWB) procedures. The models considered are quite general, having stochastically varying coefficients, explanatory variables that may endogenous, and errors which may be heteroskedastic and serially correlated. We find that DWB and, rather remarkably, ordinary CV provide effective choices of the bandwidth under a variety of conditions that are relevant in econometrics. These data-driven selectors provide a useful and easy to implement way to overcome the hurdle of choosing bandwidths in the practical application of kernel IV estimators of time-varying coefficients like those proposed by Chen (2015b) and GKM.

The remainder of the paper is organized as follows. Section 2 introduces the model and related nonparametric kernel estimators of interest. Section 3 provides a detailed description of our data-driven procedures for the selection of the bandwidth parameter for IV and LS estimators. Section 4 provides a simulation study of the small-sample performance of automated bandwidth selectors under a variety of data-generating mechanisms. Section 5 illustrates the practical use of the automated selection procedures in the context of an empirical application. Finally, Section 6 summarizes and concludes.

2 Model and Estimation

Consider the varying-coefficient linear model given by

$$y_t = \beta_t' x_t + u_t, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

$$x_t = \Psi_t' z_t + v_t, \tag{2.2}$$

where y_t is a scalar variable, x_t is a $p \times 1$ vector of (potentially endogenous) variables, β_t is a $p \times 1$ vector of coefficients, z_t is an $n \times 1$ vector of exogenous variables ($n \geq p$), Ψ_t is an $n \times p$ matrix of coefficients, and u_t and v_t are zero-mean random errors (that may exhibit serial correlation and heteroskedasticity). As in GKM, x_t is considered to be endogenous for β_t when $E(v_t u_t) \neq 0$ for some t, and exogeneity of z_t is taken to mean that $E(z_t u_t) = 0$ and $E(z_t v_t') = 0$ for all t. The parameters β_t and Ψ_t may be deterministic or stochastic, satisfying suitable boundedness and smoothness conditions (see Giraitis et al. (2014) and GKM for details and examples).

For the model (2.1)–(2.2), the kernel IV estimator of β_t introduced by GKM is[1]

$$\tilde{\beta}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j' z_j x_j' \right)^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j' z_j y_j, \tag{2.3}$$

where $b_{H,|j-t|}$ are kernel weights, H is a bandwidth parameter, and $\hat{\Psi}_j$ is a consistent estimator of Ψ_j. A natural choice for the latter is the kernel LS estimator

$$\hat{\Psi}_t = \left(\sum_{j=1}^{T} b_{L,|j-t|} z_j z_j' \right)^{-1} \sum_{j=1}^{T} b_{L,|j-t|} z_j x_j', \tag{2.4}$$

with bandwidth parameter L (satisfying $L \geq H$). The kernel weights in (2.3) and (2.4) are obtained from a kernel function $K : [0, \infty) \to [0, \infty)$ via $b_{M,l} = K(l/M)$, with bandwidth $M > 0$ such that $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. Admissible kernel functions are those satisfying $K(w) \leq C/(1 + w^a)$ and $|(d/dw)K(w)| \leq C/(1 + w^a)$ for some $C > 0$ and $a > 3$; for example, we may take $K(w) = \alpha \exp(-w^2/2)$ or $K(w) = \alpha(1 - w)\mathbb{I}(0 \leq w \leq 1)$ for some $\alpha > 0$, where $\mathbb{I}(\cdot)$ is the indicator function.

GKM give conditions on the dependence, heretogeneity and moments of z_t, u_t and v_t, and on the variation in β_t and Ψ_t, which guarantee consistency and asymptotic normality of $\tilde{\beta}_t$. In the case where x_t is exogenous, in the sense that $E(v_t u_t) = 0$ for all t, β_t can also be consistently estimated using the kernel LS estimator

$$\hat{\beta}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} x_j x_j' \right)^{-1} \sum_{j=1}^{T} b_{H,|j-t|} x_j y_j. \tag{2.5}$$

(Throughout the paper, H is used as generic notation for the bandwidth parameter associated with an estimator

[1] This is the estimator denoted $\tilde{\beta}_{1,t}$ in GKM. Under certain conditions (see Lemma 2 in GKM), it is asymptotically equivalent to the two-stage local linear estimator of Chen (2015b).

of β_t, without implying that $\tilde{\beta}_t$ and $\hat{\beta}_t$ share the same bandwidth.)

The key issue that arises in the use of the estimators (2.1), (2.2) and (2.5) in practice is the selection of reasonable values for the bandwidth parameters H and L for a given sample size T. The choice is important because the finite-sample properties of the estimators can be affected significantly by the values of the relevant bandwidth. For example, too small a value for H and/or L may yield undersmoothed estimates which have high variance, while too large a value may result in oversmoothing and large bias. The asymptotic results in GKM offer little practical guidance beyond the requirement that $C_1 T^{(4/\vartheta)+\delta} \leq H \leq L \leq C_2 T^{1-\delta}$ for some $\delta, C_1, C_2 > 0$ and $\vartheta > 4$ such that the expected values $E(\|\zeta_t\|^{4+\vartheta})$ are bounded above uniformly in t, where $\zeta_t' = (u_t, v_t', z_t')$ and $\|\cdot\|$ denotes the Euclidean norm.[2] For practical use, it is, therefore, desirable to have data-driven rules for choosing the values of the bandwidth parameters.

3 Data-Driven Bandwidth Selection

In this section, we discuss different methods for selecting the bandwidths L and H that are required for the construction of kernel IV and LS estimator of β_t. The data-driven selectors considered are based on CV, AIC, and WB and DWB methods.

Throughout the remainder of the paper, we consider bandwidths of the form $L = T^{h_1}$ and $H = T^{h_2}$, with $0 < h_2 \leq h_1 < 1$. For any $h \in (0,1)$, we use $\hat{\Psi}_{t,h}$, $\tilde{\beta}_{t,h}$ and $\hat{\beta}_{t,h}$ to denote, respectively, the LS estimator of Ψ_t defined in (2.4) with $L = T^h$, the IV estimator of β_t defined in (2.3) with $H = T^h$, and the LS estimator of β_t defined in (2.5) with $H = T^h$.

3.1 Cross-Validation

CV is a widely used method for selecting the smoothing parameter for nonparametric estimators. The basic idea is to use part of the data for fitting and the remaining part to estimate the average squared error of the fitted model under different bandwidths, and select the bandwidth which produces the best performance. Automated CV-based bandwidth selectors for inference in varying-coefficient models have been used by Chen and Hong (2012), Zhang and Wu (2012) and Chen (2015b), among others, the latter in the context of nonparametric two-stage LS estimation.

In our IV setting, letting $\hat{\Psi}_{(-t),h}$ be the leave-one-out version of the LS estimator of Ψ_t given by

$$\hat{\Psi}_{(-t),h} = \left(\sum_{1 \leq j \leq T, j \neq t} b_{T^h,|j-t|} z_j z_j' \right)^{-1} \sum_{1 \leq j \leq T, j \neq t} b_{T^h,|j-t|} z_j x_j',$$

[2] Based on results from simulation experiments, GKM recommend setting $H = L = T^{1/2}$.

the CV choice of L is $\tilde{L}_{\text{CV}} = T^{\tilde{h}_1}$, where

$$\tilde{h}_1 = \arg\min_h \left\{ \sum_{t=1}^{T} ||x_t - \hat{\Psi}'_{(-t),h} z_t||^2 \right\}.$$

In a similar manner, letting $\tilde{\beta}_{(-t),h}$ be the leave-one-out version of the IV estimator of β_t, constructed as

$$\tilde{\beta}_{(-t),h} = \left(\sum_{1 \le j \le T, j \ne t} b_{T^h,|j-t|} \hat{\Psi}'_{j,\tilde{h}_1} z_j x'_j \right)^{-1} \sum_{1 \le j \le T, j \ne t} b_{T^h,|j-t|} \hat{\Psi}'_{j,\tilde{h}_1} z_j y_j,$$

the CV choice of H is obtained as $\tilde{H}_{\text{CV}} = T^{\tilde{h}_2}$, where

$$\tilde{h}_2 = \arg\min_{h \le \tilde{h}_1} \left\{ \sum_{t=1}^{T} |y_t - \tilde{\beta}'_{(-t),h} x_t|^2 \right\}.$$

In the case of the LS estimator $\hat{\beta}_t$, the CV choice of H is obtained as $\hat{H}_{\text{CV}} = T^{\hat{h}}$, where

$$\hat{h} = \arg\min_h \left\{ \sum_{t=1}^{T} |y_t - \hat{\beta}'_{(-t),h} x_t|^2 \right\},$$

$\hat{\beta}'_{(-t),h}$ being the leave-one-out version of the LS estimator of β_t given by

$$\hat{\beta}_{(-t),h} = \left(\sum_{1 \le j \le T, j \ne t} b_{T^h,|j-t|} x_j x'_j \right)^{-1} \sum_{1 \le j \le T, j \ne t} b_{T^h,|j-t|} x_j y_j.$$

Note that the estimator $\hat{\Psi}_{j,\tilde{h}_1}$ used to construct $\tilde{\beta}_{(-t),h}$ is based on the bandwidth chosen by CV. It is also worth noting that, although we focus on the popular leave-one-out CV method, it may be advantageous to construct CV criteria by leaving out more than one observation or blocks of consecutive observations, especially when the data and/or errors are strongly correlated (see, e.g., Burman et al. (1994) and Hall et al. (1995b)).

3.2 Information Criterion

Hurvich et al. (1998) and Cai (2007), among others, suggested selecting the bandwidth for smoothing regression methods by using a nonparametric version of AIC. In our IV context, an AIC-based procedure can be used sequentially to obtain data-driven choices of first L and then H.

To give a formal description of the procedure, let X and $\hat{X}_h$ be the $p \times T$ matrices whose t-th columns are x_t and $\hat{\Psi}'_{t,h} z_t$, respectively, and Q_h be the $pT \times pT$ matrix satisfying $\text{vec}(\hat{X}_h) = Q_h \text{vec}(X)$, where $\text{vec}(\cdot)$ is the vectorization function. The AIC choice of L is obtained as $\tilde{L}_{\text{AIC}} = T^{\bar{h}_1}$, where

$$\bar{h}_1 = \arg\min_h \left\{ \log \left(\sum_{t=1}^{n} ||x_t - \hat{\Psi}'_{t,h} z_t||^2 \right) + \frac{2[\text{tr}(Q_h) + 1]}{pT - \text{tr}(Q_h) - 2} \right\},$$

with $\mathrm{tr}(\cdot)$ being the trace function. Next, let R_h be the $T \times T$ matrix satisfying $(\tilde{\beta}'_{1,h}x_1, \ldots, \tilde{\beta}'_{T,h}x_T)' = R_h(y_1, \ldots, y_T)'$, where

$$\tilde{\beta}_{t,h} = \left(\sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}'_{j,\bar{h}_1} z_j x'_j \right)^{-1} \sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}'_{j,\bar{h}_1} z_j y_j.$$

Then, the AIC choice of H is $\tilde{H}_{\mathrm{AIC}} = T^{\bar{h}_2}$, where

$$\bar{h}_2 = \arg\min_{h \leq \bar{h}_1} \left\{ \log \left(\sum_{t=1}^{n} |y_t - \tilde{\beta}'_{t,h} x_t|^2 \right) + \frac{2[\mathrm{tr}(R_h) + 1]}{T - \mathrm{tr}(R_h) - 2} \right\}.$$

Note that, as in the CV selection procedure, the estimator $\hat{\Psi}_{j,\bar{h}_1}$ used to construct $\tilde{\beta}'_{t,h}$ is based on a data-driven bandwidth ($\tilde{L}_{\mathrm{AIC}}$) obtained by the same method. The trace of the smoother matrices Q_h and R_h associated with any given bandwidth h (as well as that of the smoother matrix S_h below) is typically viewed as the effective number of parameters involved in the smoothing procedure.

For the LS estimator $\hat{\beta}_t$, the AIC choice of H is obtained in an analogous manner as $\hat{H}_{\mathrm{AIC}} = T^{\bar{h}}$, with

$$\bar{h} = \arg\min_{h} \left\{ \log \left(\sum_{t=1}^{n} |y_t - \hat{\beta}'_{t,h} x_t|^2 \right) + \frac{2[\mathrm{tr}(S_h) + 1]}{T - \mathrm{tr}(S_h) - 2} \right\},$$

where S_h is the $T \times T$ matrix satisfying $(\hat{\beta}'_{1,h}x_1, \ldots, \hat{\beta}'_{T,h}x_T)' = S_h(y_1, \ldots, y_T)'$.

3.3 Bootstrap

The bootstrap approach to bandwidth selection amounts to choosing a bandwidth which minimizes an appropriate bootstrap estimator of the average squared error of the fitted model (e.g., Faraway (1990); Hall (1990); Hall et al. (1995b); González Manteiga et al. (2004)). In our IV context, such an approach can be employed to obtain data-driven choices of first L and then H. To allow for the possibility that the errors in the model (2.1)–(2.2) may be heteroskedastic or serially correlated, we rely on the WB and DWB schemes, originally proposed by Wu (1986) and Shao (2010), respectively. The idea behind such resampling schemes is to construct bootstrap errors by perturbing residuals by auxiliary random variables that are independent of the data; these random variables may be chosen to be mutually independent (as in WB) or correlated (as in DWB).

In the case of IV estimation, the selection procedure for L involves the following steps:

(i) Using $\tilde{L}_{\mathrm{CV}} = T^{\tilde{h}_1}$ as pilot bandwidth, generate pseudo-data x_t^* according to

$$x_t^* = \hat{\Psi}'_{t,\tilde{h}_1} z_t + \hat{v}_t \eta_{1,t}, \quad t = 1, 2, \ldots, T,$$

where $\hat{v}_t = x_t - \hat{\Psi}'_{t,\tilde{h}_1} z_t$ and $\{\eta_{1,t}\}$ are random variables, independent of $\{(y_t, x'_t, z'_t)\}$, having zero mean and unit variance. For any $h \in (0, 1)$, let $\hat{\Psi}^*_{t,h}$ be the bootstrap analogue of $\hat{\Psi}_{t,h}$, defined in the same way as the latter but using $(x_t^{*\prime}, z_t')$ in place of (x_t', z_t').

(ii) Repeating the previous step B times (with B sufficiently large), generate independent copies $\hat{\Psi}^*_{t,h,1}, \ldots, \hat{\Psi}^*_{t,h,B}$ of $\hat{\Psi}^*_{t,h}$ and obtain the bootstrap choice of L as $\tilde{L}_{\mathrm{B}} = T^{h_1^*}$, where

$$h_1^* = \arg\min_h \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} ||\hat{\Psi}^{*\prime}_{t,h,b} z_t - \hat{\Psi}'_{t,\tilde{h}_1} z_t||^2 \right\}.$$

Next, given the choice $\tilde{L}_{\mathrm{B}}$, the selection procedure for H is as follows:

(i) Using $\hat{\Psi}_{t,h_1^*}$ (the LS estimator of Ψ_t with bandwidth $\tilde{L}_{\mathrm{B}}$) and the pilot bandwidth $\tilde{H}_{\mathrm{CV}} = T^{\tilde{h}_2}$ to construct the estimator $\tilde{\beta}_{t,\tilde{h}_2}$ of β_t, generate pseudo-data $(y_t^*, x_t^{*\prime})$ according to

$$\begin{aligned} y_t^* &= \tilde{\beta}'_{t,\tilde{h}_2} x_t^* + \tilde{u}_t \eta_{2,t}, \quad t = 1, 2, \ldots, T, \\ x_t^* &= \hat{\Psi}'_{t,h_1^*} z_t + \hat{v}_t \eta_{2,t}, \end{aligned}$$

where $\tilde{u}_t = y_t - \tilde{\beta}'_{t,\tilde{h}_2} x_t$, $\hat{v}_t = x_t - \hat{\Psi}'_{t,h_1^*} z_t$, and $\{\eta_{2,t}\}$ are random variables, independent of $\{(y_t, x_t', z_t', \eta_{1,t})\}$, having zero mean and unit variance. For any $h \in (0,1)$, let $\tilde{\beta}^*_{t,h}$ be the bootstrap analogue of $\tilde{\beta}_{t,h}$ given by

$$\tilde{\beta}^*_{t,h} = \left(\sum_{j=1}^{T} b_{T^h, |j-t|} \hat{\Psi}^{*\prime}_{j,h_1^*} z_j x_j^{*\prime} \right)^{-1} \sum_{j=1}^{T} b_{T^h, |j-t|} \hat{\Psi}^{*\prime}_{j,h_1^*} z_j y_j^*.$$

(ii) Repeating the previous step B times, generate independent copies $\tilde{\beta}^*_{t,h,1}, \ldots, \tilde{\beta}^*_{t,h,B}$ of $\tilde{\beta}^*_{t,h}$ and obtain the bootstrap choice of H as $\tilde{H}_{\mathrm{B}} = T^{h_2^*}$, where

$$h_2^* = \arg\min_{h \le h_1^*} \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} |\tilde{\beta}^{*\prime}_{t,h,b} x_t - \tilde{\beta}'_{t,\tilde{h}_2} x_t|^2 \right\}.$$

Notice that, following Davidson and MacKinnon (2010) and Chen (2015b), $\tilde{u}_t$ and $\hat{v}_t$ are multiplied by the same auxiliary variable $\eta_{2,t}$ in order to preserve, as much as possible, correlation between u_t and v_t when generating bootstrap data $(x_t^{*\prime}, y_t^*)$.

In the case of the LS estimator of β_t, the selection procedure for H involves the following steps:

(i) Using $\hat{H}_{\mathrm{CV}} = T^{\hat{h}}$ as pilot bandwidth, generate pseudo-data y_t^* according to

$$y_t^* = \hat{\beta}'_{t,\hat{h}} x_t + \hat{u}_t \eta_{3,t}, \quad t = 1, 2, \ldots, T,$$

where $\hat{u}_t = y_t - \hat{\beta}'_{t,\hat{h}} x_t$ and $\{\eta_{3,t}\}$ are random variables, independent of $\{(y_t, x_t')\}$, having zero mean and unit variance. For any $h \in (0,1)$, let $\hat{\beta}^*_{t,h}$ be the bootstrap version of $\hat{\beta}_{t,h}$, defined by replacing (y_t, x_t') in the definition of $\hat{\beta}_{t,h}$ with (y_t^*, x_t').

(ii) Repeating the above step B times, generate independent copies $\hat{\beta}^{*\prime}_{t,h,1}, \ldots, \hat{\beta}^{*\prime}_{t,h,B}$ of $\hat{\beta}^{*}_{t,h}$ and obtain the bootstrap choice of H as $\hat{H}_{\mathrm{B}} = T^{h^*}$, where

$$h^* = \arg\min_h \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} |\hat{\beta}^{*\prime}_{t,h,b} x_t - \hat{\beta}^{\prime}_{t,\hat{h}} x_t|^2 \right\}.$$

The bandwidth selection procedures based on WB and DWB differ only in the choice of the correlation structure of the collections of auxiliary random variables $\{\eta_{i,t}\}$ $(i = 1, 2, 3)$. In the WB case, we take $\{\eta_{i,t}\}$ to be independent $\mathcal{N}(0, 1)$ random variables. Thus, the bootstrap errors reflect possible heterogeneity in the variance of the original errors. For the DWB, we follow Shao (2010) and Djogbenou et al. (2015) in taking $\{\eta_{i,t}\}$ to be a segment of a zero-mean Gaussian process with autocovariance function $E(\eta_{i,t}\eta_{i,\tau}) = G(\{t - \tau\}/\lambda)$, where G is the triangular Bartlett kernel $G(w) = (1 - |w|)\mathbb{I}(|w| \leq 1)$ and $\lambda > 0$ is a bandwidth controlling the extent of dependence (with $\lambda \to \infty$ and $\lambda/T \to 0$ as $T \to \infty$). Hence, the bootstrap errors reflect possible serial correlation in the original errors.

It may be noted that, if heteroskedasticity and serial correlation are not a concern, then the bootstrap errors that are required to generate x_t^* and y_t^* may be obtained by resampling from the empirical distribution of the relevant residuals. For instance, when selecting the bandwidth H for $\tilde{\beta}_t$, this amounts to choosing bootstrap errors by sampling independently and uniformly, with replacement, from the residuals $\{(\tilde{u}_t, \hat{v}_t^{\prime}), t = 1, \ldots, T\}$, after centering them around their arithmetic mean. The use of such a resampling scheme is, however, inadvisable when the original errors may be serially correlated and/or heteroskedastic (and will not be considered in the sequel).

4 Monte Carlo Simulations

In this section, simulations are used to evaluate the finite-sample performance of various data-driven bandwidth selectors for the kernel-based LS and IV estimators $\hat{\beta}_t$ and $\tilde{\beta}_t$. The Monte Carlo experiments are based on data-generating processes (DGPs) that are variants of those previously used by GKM. We consider exactly identified and overidentified models, with errors that may be independent and identically distributed (i.i.d.), heterogeneously distributed, or serially correlated. As it is generally accepted that choice of kernel (K) is of secondary importance, compared to choice of smoothing parameters (L, H), we use the Gaussian kernel $K(w) = \exp(-w^2/2)$ in all subsequent computations.

4.1 Independent, Identically Distributed Errors

The first set of experiments is based on an exactly identified version of the model (2.1)–(2.2), with $p = n = 1$, that is,

$$y_t = \beta_t x_t + u_t, \qquad x_t = \psi_t z_t + v_t, \qquad t = 1, 2, \ldots, T. \tag{4.1}$$

As in GKM, $\{z_t\}$ are i.i.d. $\mathcal{N}(0,1)$ random variables, while $\{u_t\}$ and $\{v_t\}$ are such that

$$u_t = se_{1,t} + (1-s)e_{2,t}, \qquad v_t = se_{1,t} + (1-s)e_{3,t}, \tag{4.2}$$

where $\{e_{1,t}\}$,$\{e_{2,t}\}$ and $\{e_{3,t}\}$ are mutually independent collections of i.i.d. $\mathcal{N}(0,1)$ random variables independent of $\{z_t\}$. Thus, the parameter s controls the strength of endogeneity, as measured by $\mathrm{Corr}(u_t, v_t) = s^2/[s^2 + (1-s)^2]$, with $s \in \{0, 0.2, 0.5\}$. The coefficients $\{\beta_t\}$ and $\{\psi_t\}$ vary stochastically as rescaled random walks, with

$$\beta_t = (1/T)^{1/2} \sum_{j=1}^{t} \xi_{1,j}, \qquad \psi_t = (1/T)^{1/2} \sum_{j=1}^{t} \xi_{2,j}, \tag{4.3}$$

where $\{\xi_{1,t}\}$ and $\{\xi_{2,t}\}$ are collections of i.i.d. $\mathcal{N}(0,1)$ random variables, independent of each other and of $\{(e_{1,t}, e_{2,t}, e_{3,t}, z_t)\}$. We consider two sample sizes, $T = 100$ and $T = 200$.

As indicated in the description of the bandwidth selection procedures in Section 3, the same data-driven procedure is used for the selection of both L and H in the case of the IV estimator (the only exception being the use of a CV pilot bandwidth in the construction of the IV estimator of β_t required to generate bootstrap data). For bootstrap-based selection procedures, the number of bootstrap replications is $B = 399$. In the case of DWB, we consider $\lambda \in \{2, 4, 6, 8, 10\}$ when $T = 100$ and $\lambda \in \{6, 8, 12, 16, 32\}$ when $T = 200$.[3] In all cases, the relevant objective functions are minimized over an equispaced grid of 30 points corresponding to possible bandwidths ranging from $T^{0.2}$ to $T^{0.9}$.

The properties of bandwidth selectors for IV and LS estimators of β_t are evaluated using several performance indicators. Specifically, for a kernel-based estimator of β_t, say $\check{\beta}_{t,\check{h}}$ (IV or LS), with bandwidth $T^{\check{h}}$ selected by one of the methods discussed in Section 3, we consider the following performance measures (R denotes the number of Monte Carlo replications):

(i) average ratio of selected bandwidth to optimal bandwidth, computed as

$$(1/R) \sum_{r=1}^{R} T^{\check{h}_r - h_r^{\mathrm{opt}}},$$

where $\check{h}_r$ is the value of $\check{h}$ in the r-th Monte Carlo replication and h_r^{opt} is the optimal value, i.e., the minimizer of $(1/T)\sum_{t=1}^{T} |\check{\beta}_{t,h} - \beta_t|^q$ over h, with $q = 1$ and $q = 2$ for the IV and LS estimators, respectively;[4]

(ii) average median absolute estimation error of $\check{\beta}_{t,\check{h}}$, computed as

$$(1/R) \sum_{r=1}^{R} \mathrm{median}\{|\check{\beta}_{t,\check{h}_r} - \beta_t| : t = 1, \ldots, T\};$$

[3]Recall that our choice for the covariance structure of the DWB auxiliary random variables implies that the latter are $\lceil \lambda - 1 \rceil$-dependent, where $\lceil \cdot \rceil$ denotes the least-integer function.

[4]The absolute estimation error ($q = 1$) is considered in the IV case because the finite-sample distributions of IV-type estimators tend to be heavy-tailed due to lack of finite moments.

(iii) average coverage rate of 95% two-sided confidence intervals for β_t, computed as

$$[100/(TR)]\sum_{t=1}^{T}\sum_{r=1}^{R}\mathbb{I}(|\check{\beta}_{t,\check{h}_r}-\beta_t|/\mathrm{se}(\check{\beta}_{t,\check{h}_r})\leq 1.96),$$

where $\mathrm{se}(\check{\beta}_{t,\check{h}_r})$ is an estimate of the asymptotic standard deviation of $\check{\beta}_{t,\check{h}_r}$ (obtained as in GKM).

All simulation results are based on $R = 1,000$ Monte Carlo replications.

Tables 4.1 and 4.2 present results for $T = 100$ and $T = 200$, respectively. The data-driven methods are similarly behaved when selecting the bandwidth for the LS estimator of ψ_t, DWB (with large bandwidth λ) being slightly superior in terms of the ratio of the selected bandwidth to the optimal value that minimizes the mean squared estimation error of $\hat{\psi}_t$. For the IV estimator of β_t, CV outperforms all other methods, in terms of the ratio of the selected bandwidth to the optimal value that minimizes the mean absolute estimation error of $\tilde{\beta}_t$ for $T = 100$, regardless of whether x_t is exogenous ($s = 0$) or endogenous ($s \neq 0$); it is less effective than DWB (with $\lambda \geq 16$), but only by a slight margin, when $T = 200$ and $s \neq 0$. Furthermore, CV bandwidths produce pointwise confidence intervals for β_t the average coverage of which is close to the coverage associated with the optimal bandwidth, outperforming other automatically selected bandwidths in this respect for all values of s and T. It must be pointed out, however, that even the optimal bandwidth (for the given simulated data) yields confidence intervals the average coverage of which (labelled 'optimal coverage' in the tables) falls considerably short of the nominal 95% rate.[5] The AIC-based selector is the least competitive overall, yielding bandwidths that are lower than the optimal bandwidth and associated confidence intervals for β_t which undercover considerably. There is little to choose among competing methods when considering the average median absolute estimation error of $\tilde{\beta}_t$, DWB having a slight advantage and being more successful the stronger the correlation between x_t and u_t is. It is perhaps noteworthy that DWB based on relatively large values of λ performs well (and dominates WB) even though the errors (u_t, v_t) are i.i.d. in the simulations.

[5] These findings are consistent with those of GKM (for bandwidths H and L taking the values $T^{0.4}$ or $T^{0.5}$), who also report undercoverage that becomes more pronounced as the strength of endogeneity increases.

		Estimator	CV	AIC	WB	DWB($\lambda = 2$)	DWB($\lambda = 4$)	DWB($\lambda = 6$)	DWB($\lambda = 8$)	DWB($\lambda = 10$)
	Bandwidth Ratio	$\hat{\Psi}_t$	1.084	1.132	1.117	1.111	1.097	1.087	1.072	1.072
		$\tilde{\beta}_t$	1.043	0.872	1.113	1.112	1.101	1.087	1.078	1.078
		$\hat{\beta}_t$	1.076	1.138	1.113	1.101	1.084	1.073	1.062	1.055
	Estimation Error	$\tilde{\beta}_t$	0.241	1.926	0.267	0.259	0.260	0.260	0.259	0.258
s=0		$\hat{\beta}_t$	0.157	0.153	0.156	0.157	0.157	0.157	0.157	0.157
	Coverage	$\tilde{\beta}_t$	84.713	68.742	81.303	77.679	78.004	78.396	78.462	78.93
		$\hat{\beta}_t$	70.894	70.274	70.154	70.292	70.616	70.819	71.144	71.371
	Optimal Coverage	$\tilde{\beta}_t$	86.348							
		$\hat{\beta}_t$	77.738							
		$\hat{\Psi}_t$	1.095	1.148	1.131	1.121	1.104	1.093	1.080	1.080
	Bandwidth Ratio	$\tilde{\beta}_t$	1.050	0.897	1.136	1.120	1.105	1.098	1.085	1.085
		$\hat{\beta}_t$	1.083	1.137	1.117	1.106	1.091	1.075	1.066	1.060
	Estimation Error	$\tilde{\beta}_t$	0.218	1.716	0.233	0.234	0.234	0.233	0.234	0.233
s=0.2		$\hat{\beta}_t$	0.151	0.149	0.150	0.151	0.151	0.151	0.151	0.151
	Coverage	$\tilde{\beta}_t$	83.342	65.924	80.372	76.605	76.942	77.337	77.349	77.624
		$\hat{\beta}_t$	69.850	69.270	69.107	69.322	69.578	69.982	70.202	70.422
	Optimal Coverage	$\tilde{\beta}_t$	85.090							
		$\hat{\beta}_t$	76.124							
		$\hat{\Psi}_t$	1.110	1.160	1.145	1.134	1.120	1.107	1.091	1.091
	Bandwidth Ratio	$\tilde{\beta}_t$	1.064	0.880	1.122	1.113	1.103	1.090	1.078	1.078
		$\hat{\beta}_t$	1.002	1.042	1.033	1.018	1.003	0.989	0.980	0.976
	Estimation Error	$\tilde{\beta}_t$	0.206	1.994	0.219	0.215	0.218	0.215	0.217	0.216
s=0.5		$\hat{\beta}_t$	0.245	0.244	0.245	0.245	0.245	0.245	0.245	0.245
	Coverage	$\tilde{\beta}_t$	80.794	63.457	78.913	75.171	75.486	75.673	75.745	75.992
		$\hat{\beta}_t$	47.365	46.870	46.683	47.017	47.182	47.607	47.830	47.882
	Optimal Coverage	$\tilde{\beta}_t$	83.016							
		$\hat{\beta}_t$	49.735							

Table 4.1: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.2) and $T = 100$.

		Estimator	CV	AIC	WB	DWB($\lambda = 6$)	DWB($\lambda = 8$)	DWB($\lambda = 12$)	DWB($\lambda = 16$)	DWB($\lambda = 32$)
		$\hat{\Psi}_t$	1.045	1.074	1.089	1.068	1.061	1.049	1.028	1.028
	Bandwidth Ratio	$\tilde{\beta}_t$	1.017	0.806	1.067	1.045	1.039	1.032	1.019	1.019
		$\hat{\beta}_t$	1.039	1.070	1.089	1.059	1.053	1.040	1.035	1.024
	Estimation Error	$\tilde{\beta}_t$	0.230	2.263	0.260	0.237	0.234	0.237	0.238	0.238
s=0		$\hat{\beta}_t$	0.132	0.130	0.131	0.132	0.133	0.133	0.133	0.133
	Coverage	$\tilde{\beta}_t$	87.783	70.134	84.146	82.644	82.919	83.098	83.436	83.839
		$\hat{\beta}_t$	73.520	73.227	72.182	72.807	72.963	73.484	73.645	74.175
	Optimal Coverage	$\tilde{\beta}_t$	89.089							
		$\hat{\beta}_t$	79.632							
		$\hat{\Psi}_t$	1.043	1.074	1.089	1.063	1.054	1.043	1.023	1.023
	Bandwidth Ratio	$\tilde{\beta}_t$	1.015	0.827	1.061	1.044	1.037	1.027	1.011	1.011
		$\hat{\beta}_t$	1.036	1.060	1.084	1.053	1.046	1.034	1.028	1.018
	Estimation Error	$\tilde{\beta}_t$	0.205	1.967	0.226	0.211	0.212	0.212	0.213	0.213
s=0.2		$\hat{\beta}_t$	0.129	0.127	0.128	0.129	0.129	0.129	0.129	0.129
	Coverage	$\tilde{\beta}_t$	86.928	67.088	83.990	82.367	82.586	82.983	83.086	83.462
		$\hat{\beta}_t$	72.855	72.646	71.558	72.215	72.513	72.911	73.212	73.589
	Optimal Coverage	$\tilde{\beta}_t$	88.094							
		$\hat{\beta}_t$	78.300							
		$\hat{\Psi}_t$	1.040	1.080	1.089	1.062	1.051	1.041	1.020	1.020
	Bandwidth Ratio	$\tilde{\beta}_t$	1.029	0.808	1.068	1.046	1.034	1.029	1.015	1.015
		$\hat{\beta}_t$	0.921	0.942	0.962	0.934	0.923	0.913	0.911	0.901
	Estimation Error	$\tilde{\beta}_t$	0.183	2.263	0.206	0.193	0.192	0.193	0.194	0.193
s=0.5		$\hat{\beta}_t$	0.273	0.272	0.272	0.272	0.272	0.272	0.273	0.272
	Coverage	$\tilde{\beta}_t$	85.017	70.1345	82.760	80.965	81.185	81.533	81.666	82.186
		$\hat{\beta}_t$	38.543	37.882	37.136	38.003	38.470	38.814	38.862	39.227
	Optimal Coverage	$\tilde{\beta}_t$	86.820							
		$\hat{\beta}_t$	37.367							

Table 4.2: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.2) and $T = 200$.

Turning to the LS estimator of β_t, the results in Tables 4.1 and 4.2 show that, for all bandwidth selectors, the average median absolute estimation error of $\hat{\beta}_t$ is lower than that of the IV estimator when x_t is exogenous or endogeneity is weak ($s = 0.2$), while the reverse is true under moderate endogeneity ($s = 0.5$). The bandwidths selected by the various methods tend to be somewhat higher than the optimal values when $s = 0$ or $s = 0.2$, but lower than the optimal values when $s = 0.5$ and $T = 200$. As in the IV case, CV and DWB (with λ that is not too small) generally provide the most accurate choices relative to the optimal bandwidth for $\hat{\beta}_t$, the former having a slight advantage when endogeneity is moderately strong, while AIC is the least successful. Undercoverage of confidence intervals for β_t is once again a problem, regardless of which bandwidth selector is used in the construction of $\hat{\beta}_t$. Although inaccuracy of LS confidence intervals is not surprising when x_t and u_t are correlated (coverage rates are uniformly lower than 50% when $s = 0.5$), and the use of the LS estimator is clearly not recommended in these circumstances, the problem is also present when x_t is exogenous and $\hat{\beta}_t$ is consistent.

In sum, although ordinary CV is sometimes reported to perform poorly in nonparametric regression settings (e.g., Härdle et al. (1988)), it is found to provide effective choices of the bandwidth for kernel IV and LS estimators of time-varying coefficients (at least when performance measures other than coverage of confidence intervals are considered), both in the presence and absence of endogeneity. DWB is competitive with the CV selector and consistently better than WB and AIC.

4.2 Heteroskedasticity and Serial Correlation

To assess the effect of heteroskedasticity (in the form of time-dependent error variances) on the performance bandwidth selectors, we consider artificial data from a modified version of the DGP (4.1)–(4.3) in which $\{y_t\}$ are generated according to $y_t = \beta_t x_t + \sigma_t u_t$, with the following specifications for σ_t:

$$\text{HET-1}: \quad \sigma_t = \exp(\sin(2\pi t/T) + 1); \tag{4.4}$$

$$\text{HET-2}: \quad \sigma_t = \exp(\sin(2\pi t/T) + 1) + 2\mathbb{I}(t \geq T/2). \tag{4.5}$$

Thus, the sequence $\{\sigma_t\}$ exhibits sine shape under HET-1, which is subject to a level shift under HET-2. These functional forms have also been used by Chronopoulos et al. (2021).

Table 4.3 summarizes simulation results for $s \in \{0, 0.5\}$ and $T = 100$. The performance of data-driven bandwidth selection methods for the IV estimator $\tilde{\beta}_t$ is generally similar to that documented earlier under homoskedastic designs. CV and DWB provide the best choices in terms of closeness of the automatically selected bandwidths to the optimal value and magnitude of the average median absolute estimation error of $\tilde{\beta}_t$. WB (which explicitly allows for heteroskedasticity) does as well as CV in terms of the latter performance measures and only slightly worse in terms of the former. The AIC selector also performs well, but only when considering deviations of the selected bandwidth from the optimal value and only for $s = 0$. Interestingly, when x_t is exogenous, the coverage rate of IV confidence intervals for β_t improves in comparison to the case of i.i.d. errors. However, in the presence of endogeneity, all confidence intervals, including those associated with the optimal bandwidth, have average coverage rates less than 50%. This undercoverage is, as to be expected, even more substantial in the case of the LS estimator $\hat{\beta}_t$, regardless of the bandwidth selector used. Matters improve under exogeneity, with $\hat{\beta}_t$ outperforming $\tilde{\beta}_t$ in terms of average median estimation error.

		Estimator	***CV***	***AIC***	***WB***	***DWB***$(\lambda=2)$	***DWB***$(\lambda=4)$	***DWB***$(\lambda=6)$	***DWB***$(\lambda=8)$	***DWB***$(\lambda=10)$
							HET-1			
		$\hat{\Psi}_t$	1.084	1.132	1.117	1.111	1.097	1.087	1.072	1.072
	Bandwidth Ratio	$\tilde{\beta}_t$	1.056	1.022	1.071	1.072	1.053	1.045	1.035	1.035
		$\hat{\beta}_t$	0.947	1.020	0.958	0.958	0.951	0.941	0.940	0.935
s=0	***Estimation Error***	$\tilde{\beta}_t$	0.607	2.065	0.644	0.604	0.602	0.605	0.603	0.605
		$\hat{\beta}_t$	0.339	0.336	0.339	0.340	0.340	0.340	0.341	0.341
	Coverage	$\tilde{\beta}_t$	91.377	75.564	91.366	91.100	91.083	91.079	91.150	91.081
		$\hat{\beta}_t$	79.917	78.953	79.604	79.485	79.502	79.727	79.757	79.894
	Optimal Coverage	$\tilde{\beta}_t$	92.494							
		$\hat{\beta}_t$	83.591							
		$\hat{\Psi}_t$	1.110	1.160	1.145	1.134	1.120	1.107	1.091	1.091
	Bandwidth Ratio	$\tilde{\beta}_t$	1.057	0.990	1.091	1.080	1.065	1.056	1.038	1.038
		$\hat{\beta}_t$	0.787	0.845	0.806	0.800	0.789	0.785	0.776	0.773
s=0.5	***Estimation Error***	$\tilde{\beta}_t$	0.480	2.693	0.508	0.472	0.481	0.477	0.481	0.480
		$\hat{\beta}_t$	0.661	0.671	0.665	0.664	0.662	0.660	0.659	0.657
	Coverage	$\tilde{\beta}_t$	87.539	66.476	87.613	87.198	87.246	87.216	87.153	87.254
		$\hat{\beta}_t$	42.982	41.127	42.244	42.445	42.864	43.077	43.424	43.654
	Optimal Coverage	$\tilde{\beta}_t$	89.397							
		$\hat{\beta}_t$	37.162							
							HET-2			
		$\hat{\Psi}_t$	1.084	1.132	1.117	1.111	1.097	1.087	1.072	1.072
	Bandwidth Ratio	$\tilde{\beta}_t$	1.054	1.035	1.060	1.061	1.048	1.034	1.026	1.026
		$\hat{\beta}_t$	0.964	1.011	0.972	0.971	0.965	0.957	0.954	0.951
s=0	***Estimation Error***	$\tilde{\beta}_t$	0.836	2.268	0.897	0.812	0.813	0.820	0.820	0.825
		$\hat{\beta}_t$	0.428	0.414	0.420	0.423	0.425	0.428	0.430	0.431
	Coverage	$\tilde{\beta}_t$	92.866	80.366	92.889	92.722	92.803	92.736	92.729	92.713
		$\hat{\beta}_t$	81.767	81.468	81.850	81.794	81.742	81.783	81.856	81.778
	Optimal Coverage	$\tilde{\beta}_t$	94.023							
		$\hat{\beta}_t$	87.485							
		$\hat{\Psi}_t$	1.110	1.160	1.145	1.134	1.120	1.107	1.091	1.091
	Bandwidth Ratio	$\tilde{\beta}_t$	1.066	1.018	1.087	1.077	1.063	1.054	1.043	1.043
		$\hat{\beta}_t$	0.852	0.909	0.872	0.867	0.858	0.852	0.848	0.845
s=0.5	***Estimation Error***	$\tilde{\beta}_t$	0.696	3.276	0.721	0.660	0.676	0.673	0.681	0.680
		$\hat{\beta}_t$	0.950	0.952	0.952	0.951	0.952	0.952	0.951	0.951
	Coverage	$\tilde{\beta}_t$	89.919	69.340	90.077	89.855	89.827	89.749	89.693	89.722
		$\hat{\beta}_t$	37.355	35.518	36.319	36.674	36.989	37.260	37.416	37.713
	Optimal Coverage	$\tilde{\beta}_t$	91.537							
		$\hat{\beta}_t$	32.944							

Table 4.3: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.4) and (4.5) and $T = 100$.

Next, to investigate the effect on bandwidth selection of serial correlation in the errors, we consider a variant of the DGP (4.1)–(4.3) in which

$$u_t = se_{1,t} + (1-s)(1-\varphi^2)^{1/2}\omega_t, \quad \omega_t = \varphi\omega_{t-1} + e_{2,t}, \quad v_t = se_{1,t} + (1-s)e_{3,t}. \tag{4.6}$$

Thus, for any $0 < |\varphi| < 1$, the autocovariance structure of $\{u_t\}$ is that of a causal ARMA$(1,1)$ process such that $\mathrm{Corr}(u_t, v_t) = s^2/[s^2 + (1-s)^2]$. The results obtained under this DGP, with $T = 100$ and $\varphi = 0.8$, are summarized in Table 4.4.[6]

Although leave-one-out CV is often found to experience difficulties in nonparametric regression settings with serially correlated errors (e.g., Hart (1991); Opsomer et al. (2001)), deviations from the independence assumption do not appear to have an adverse affect on CV in our varying-coefficients setting. For either value of s, CV

[6] Similar results are obtained for $\varphi = -0.8$.

outperforms other methods when selecting the bandwidth for the IV or LS estimator of β_t, yielding bandwidths that are close to the optimal values. Its performance is almost identical to that of DWB (which explicitly allows for serial correlation) and of WB in terms of the average median absolute estimation error of the estimators, while AIC is the least successful selector overall. Once again, the coverage of pointwise confidence intervals leaves much to be desired, even when the optimal bandwidth is used. It should be noted, however, that coverage results should be viewed with caution in this case since confidence intervals are based on an asymptotic normal approximation to the distribution of $\tilde{\beta}_t$ that is obtained under the assumption that $\{z_t u_t\}$ is an uncorrelated process (cf. Theorem 3(ii) in GKM).

		Estimator	*CV*	*AIC*	*WB*	*DWB*($\lambda = 2$)	*DWB*($\lambda = 4$)	*DWB*($\lambda = 6$)	*DWB*($\lambda = 8$)	*DWB*($\lambda = 10$)
		$\hat{\Psi}_t$	1.084	1.132	1.117	1.111	1.097	1.087	1.072	1.072
	Bandwidth Ratio	$\tilde{\beta}_t$	1.041	0.868	1.130	1.121	1.102	1.098	1.082	1.082
		$\hat{\beta}_t$	1.039	1.070	1.089	1.059	1.053	1.040	1.035	1.024
	Estimation Error	$\tilde{\beta}_t$	0.230	1.891	0.250	0.252	0.248	0.250	0.250	0.251
s=0		$\hat{\beta}_t$	0.153	0.150	0.154	0.153	0.153	0.153	0.153	0.153
	Coverage	$\tilde{\beta}_t$	83.833	68.791	80.535	76.861	77.139	77.53	77.653	78.027
		$\hat{\beta}_t$	69.811	69.558	69.137	69.349	69.731	70.119	70.299	70.380
	IV Optimal Coverage	$\tilde{\beta}_t$	85.305							
		$\hat{\beta}_t$	76.218							
		$\hat{\Psi}_t$	1.110	1.160	1.145	1.134	1.120	1.107	1.091	1.091
	Bandwidth Ratio	$\tilde{\beta}_t$	1.069	0.877	1.131	1.123	1.114	1.102	1.087	1.087
		$\hat{\beta}_t$	0.993	1.040	1.022	1.013	0.995	0.983	0.972	0.967
	Estimation Error	$\tilde{\beta}_t$	0.200	2.003	0.213	0.212	0.212	0.211	0.213	0.212
s=0.5		$\hat{\beta}_t$	0.247	0.245	0.246	0.246	0.246	0.247	0.246	0.247
	Coverage	$\tilde{\beta}_t$	80.685	63.963	78.608	74.959	75.316	75.397	75.715	75.886
		$\hat{\beta}_t$	46.981	46.363	46.389	46.608	46.937	47.244	47.510	47.598
	IV Optimal Coverage	$\tilde{\beta}_t$	82.991							
		$\hat{\beta}_t$	49.730							

Table 4.4: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.6) and $T = 100$.

4.3 Overidentification

In the final set of experiments, we consider an overidentified version of the model (2.1)–(2.2), with $p = n-1 = 1$, that is,

$$y_t = \beta_t x_t + \sigma_t u_t, \qquad x_t = \psi_{1,t} z_{1,t} + \psi_{2,t} z_{2,t} + v_t, \tag{4.7}$$

with $\sigma_t > 0$. As before, the coefficients $\{\beta_t\}$, $\{\psi_{1,t}\}$ and $\{\psi_{2,t}\}$ are generated as independent Gaussian random walks (rescaled by $T^{-1/2}$), and $\{z_{1,t}\}$ and $\{z_{2,t}\}$ are collections of i.i.d. $\mathcal{N}(0,1)$ random variables independent of each other and of $\{(u_t, v_t, \beta_t, \psi_{1,t}, \psi_{2,t})\}$

Simulation results, when (u_t, v_t) are generated according to (4.2) and $\sigma_t = 1$ for all t, are presented in Tables 4.5 and 4.6. As in exactly identified models, CV and DWB outperform AIC and WB in the vast majority of cases in terms of the ratio of the selected bandwidth for IV and LS estimators of β_t to the optimal value. CV and WB result in estimates of β_t that generally have the lowest median absolute estimation error, for all values of s, but the former selector has a clear advantage when considering coverage of confidence intervals relatively to the coverage associated with the optimal bandwidth value.

		Estimator	CV	AIC	WB	DWB($\lambda = 2$)	DWB($\lambda = 4$)	DWB($\lambda = 6$)	DWB($\lambda = 8$)	DWB($\lambda = 10$)
s=0	Bandwidth Ratio	$\hat{\Psi}_t$	1.045	0.662	1.103	1.101	1.099	1.095	1.094	1.094
		$\tilde{\beta}_t$	1.053	0.835	1.169	1.263	1.250	1.244	1.230	1.230
		$\hat{\beta}_t$	1.084	1.125	1.117	1.108	1.084	1.071	1.065	1.055
	Estimation Error	$\tilde{\beta}_t$	0.174	0.225	0.161	0.209	0.208	0.207	0.208	0.208
		$\hat{\beta}_t$	0.137	0.134	0.137	0.137	0.137	0.137	0.137	0.137
	Coverage	$\tilde{\beta}_t$	88.434	85.768	85.18	74.73	74.908	75.004	74.816	74.853
		$\hat{\beta}_t$	70.072	70.106	69.659	69.621	70.209	70.357	70.487	70.845
	Optimal Coverage	$\tilde{\beta}_t$	90.414							
		$\hat{\beta}_t$	76.358							
s=0.2	Bandwidth Ratio	$\hat{\Psi}_t$	1.042	0.725	1.111	1.108	1.107	1.101	1.103	1.103
		$\tilde{\beta}_t$	1.062	0.919	1.148	1.272	1.262	1.254	1.247	1.247
		$\hat{\beta}_t$	1.079	1.123	1.111	1.103	1.080	1.064	1.059	1.054
	Estimation Error	$\tilde{\beta}_t$	0.147	0.202	0.156	0.195	0.196	0.195	0.195	0.195
		$\hat{\beta}_t$	0.127	0.126	0.127	0.127	0.127	0.127	0.127	0.127
	Coverage	$\tilde{\beta}_t$	87.219	83.904	84.944	72.684	72.642	72.877	72.758	72.613
		$\hat{\beta}_t$	69.658	69.457	69.279	69.340	69.703	69.907	69.981	70.151
	Optimal Coverage	$\tilde{\beta}_t$	89.251							
		$\hat{\beta}_t$	75.195							
s=0.5	Bandwidth Ratio	$\hat{\Psi}_t$	1.038	0.766	1.106	1.107	1.105	1.104	1.104	1.104
		$\tilde{\beta}_t$	1.079	0.963	1.135	1.285	1.275	1.274	1.267	1.267
		$\hat{\beta}_t$	1.038	1.076	1.070	1.059	1.035	1.022	1.017	1.009
	Estimation Error	$\tilde{\beta}_t$	0.137	0.191	0.145	0.187	0.187	0.187	0.187	0.188
		$\hat{\beta}_t$	0.164	0.163	0.163	0.164	0.164	0.164	0.164	0.164
	Coverage	$\tilde{\beta}_t$	85.886	82.506	84.214	70.819	70.835	70.887	70.881	70.794
		$\hat{\beta}_t$	57.479	57.356	56.935	57.057	57.401	57.67	57.855	57.982
	Optimal Coverage	$\tilde{\beta}_t$	88.073							
		$\hat{\beta}_t$	60.373							

Table 4.5: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.7) and $T = 100$.

		Estimator	CV	AIC	WB	DWB($\lambda = 6$)	DWB($\lambda = 8$)	DWB($\lambda = 12$)	DWB($\lambda = 16$)	DWB($\lambda = 32$)
s=0	Bandwidth Ratio	$\hat{\Psi}_t$	1.033	0.600	1.097	1.093	1.091	1.087	1.079	1.079
		$\tilde{\beta}_t$	1.035	0.701	1.163	1.172	1.173	1.175	1.171	1.171
		$\hat{\beta}_t$	1.037	1.059	1.088	1.055	1.047	1.035	1.029	1.016
	Estimation Error	$\tilde{\beta}_t$	0.145	0.209	0.156	0.173	0.173	0.173	0.173	0.173
		$\hat{\beta}_t$	0.117	0.115	0.117	0.117	0.117	0.117	0.117	0.117
	Coverage	$\tilde{\beta}_t$	90.418	89.023	85.799	78.551	78.587	78.747	78.796	78.931
		$\hat{\beta}_t$	73.619	73.709	72.256	73.016	73.289	73.691	73.914	74.505
	Optimal Coverage	$\tilde{\beta}_t$	91.534							
		$\hat{\beta}_t$	78.519							
s=0.2	Bandwidth Ratio	$\hat{\Psi}_t$	1.034	0.669	1.104	1.100	1.100	1.096	1.092	1.092
		$\tilde{\beta}_t$	1.034	0.776	1.132	1.154	1.156	1.159	1.169	1.169
		$\hat{\beta}_t$	1.035	1.057	1.086	1.052	1.043	1.029	1.024	1.015
	Estimation Error	$\tilde{\beta}_t$	0.136	0.184	0.140	0.161	0.161	0.162	0.161	0.162
		$\hat{\beta}_t$	0.112	0.111	0.111	0.112	0.112	0.112	0.112	0.112
	Coverage	$\tilde{\beta}_t$	89.499	87.661	85.942	76.46	76.375	76.435	76.556	76.489
		$\hat{\beta}_t$	73.135	73.128	71.755	72.531	72.781	73.268	73.469	73.928
	Optimal Coverage	$\tilde{\beta}_t$	90.601							
		$\hat{\beta}_t$	77.466							
s=0.5	Bandwidth Ratio	$\hat{\Psi}_t$	1.022	0.722	1.098	1.096	1.093	1.094	1.090	1.090
		$\tilde{\beta}_t$	1.045	0.825	1.108	1.139	1.147	1.157	1.170	1.170
		$\hat{\beta}_t$	0.959	0.980	1.007	0.971	0.960	0.951	0.944	0.938
	Estimation Error	$\tilde{\beta}_t$	0.123	0.170	0.129	0.154	0.153	0.153	0.153	0.154
		$\hat{\beta}_t$	0.175	0.174	0.175	0.175	0.175	0.175	0.175	0.175
	Coverage	$\tilde{\beta}_t$	88.212	85.979	85.627	74.341	74.521	74.391	74.474	74.417
		$\hat{\beta}_t$	52.795	52.380	51.396	52.400	52.817	53.145	53.390	53.537
	Optimal Coverage	$\tilde{\beta}_t$	89.879							
		$\hat{\beta}_t$	52.805							

Table 4.6: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.7) and $T = 200$.

Table 4.7 shows results for the case where (u_t, v_t) are generated as in (4.2) and σ_t varies according to the specifications (4.4)–(4.5). CV has superior performance in terms of closeness of the selected bandwidths to the optimal value and magnitude of the average median absolute estimation error of the IV and LS estimators of β_t. All data-driven methods result in similar average median absolute estimation errors, AIC being the least successful selector in this respect. Although the coverage of LS confidence intervals leaves much to be desired even when $s = 0$, IV confidence intervals associated with bandwidths selected by CV and WB are quite accurate, and more so than in the absence of heteroskedasticity. The improved coverage of confidence intervals in heteroskedastic settings may be due to the fact that the covariance estimator used in their construction explicitly allows for heterogeneity in the error variances.

		Estimator	*CV*	*AIC*	*WB*	*DWB*$(\lambda = 2)$	*DWB*$(\lambda = 4)$	(*DWB*$\lambda = 6$)	*DWB*$(\lambda = 8)$	(*DWB*$\lambda = 10$)
							HET-1			
		$\hat{\Psi}_t$	1.045	0.662	1.103	1.101	1.099	1.095	1.094	1.094
	Bandwidth Ratio	$\tilde{\beta}_t$	1.032	0.694	1.117	1.116	1.106	1.099	1.093	1.093
		$\hat{\beta}_t$	1.004	1.071	1.022	1.019	1.009	1.000	0.996	0.991
s=0	***Estimation Error***	$\tilde{\beta}_t$	0.351	0.439	0.356	0.360	0.362	0.360	0.361	0.359
		$\hat{\beta}_t$	0.288	0.283	0.288	0.289	0.289	0.289	0.289	0.288
	Coverage	$\tilde{\beta}_t$	93.243	91.692	93.171	91.879	91.814	91.776	91.778	91.853
		$\hat{\beta}_t$	76.309	75.302	76.039	76.049	76.201	76.271	76.259	76.461
	Optimal Coverage	$\tilde{\beta}_t$	86.348							
		$\hat{\beta}_t$	77.738							
		$\hat{\Psi}_t$	1.038	0.766	1.106	1.107	1.105	1.104	1.104	1.104
	Bandwidth Ratio	$\tilde{\beta}_t$	1.039	0.811	1.101	1.095	1.090	1.088	1.086	1.086
		$\hat{\beta}_t$	0.911	0.970	0.931	0.924	0.914	0.904	0.899	0.897
s=0.5	***Estimation Error***	$\tilde{\beta}_t$	0.286	0.353	0.284	0.297	0.296	0.297	0.297	0.298
		$\hat{\beta}_t$	0.423	0.427	0.426	0.425	0.425	0.423	0.423	0.422
	Coverage	$\tilde{\beta}_t$	90.714	88.498	90.762	88.031	88.162	88.061	88.035	87.960
		$\hat{\beta}_t$	53.161	51.605	52.417	52.653	52.824	53.281	53.432	53.552
	Optimal Coverage	$\tilde{\beta}_t$	92.288							
		$\hat{\beta}_t$	53.049							
							HET-2			
		$\hat{\Psi}_t$	1.045	0.662	1.103	1.101	1.099	1.095	1.094	1.094
	Bandwidth Ratio	$\tilde{\beta}_t$	1.023	0.681	1.103	1.101	1.091	1.085	1.078	1.078
		$\hat{\beta}_t$	1.003	1.064	1.019	1.016	1.009	1.000	0.997	0.993
s=0	***Estimation Error***	$\tilde{\beta}_t$	0.461	0.626	0.453	0.440	0.444	0.441	0.442	0.443
		$\hat{\beta}_t$	0.358	0.341	0.349	0.352	0.354	0.355	0.357	0.358
	Coverage	$\tilde{\beta}_t$	95.099	93.774	94.952	93.975	93.949	93.854	93.860	93.919
		$\hat{\beta}_t$	79.135	78.692	79.147	78.931	79.053	79.143	79.200	79.274
	Optimal Coverage	$\tilde{\beta}_t$	96.105							
		$\hat{\beta}_t$	86.390							
		$\hat{\Psi}_t$	1.038	0.766	1.106	1.107	1.105	1.104	1.104	1.104
	Bandwidth Ratio	$\tilde{\beta}_t$	1.019	0.783	1.082	1.080	1.070	1.068	1.062	1.062
		$\hat{\beta}_t$	0.935	0.995	0.957	0.952	0.941	0.933	0.928	0.924
s=0.5	***Estimation Error***	$\tilde{\beta}_t$	0.373	0.478	0.365	0.360	0.360	0.360	0.361	0.362
		$\hat{\beta}_t$	0.582	0.578	0.579	0.579	0.581	0.581	0.582	0.583
	Coverage	$\tilde{\beta}_t$	93.176	91.377	93.220	91.138	91.252	91.100	91.108	91.058
		$\hat{\beta}_t$	50.656	48.683	49.974	49.992	50.283	50.668	50.873	50.883
	Optimal Coverage	$\tilde{\beta}_t$	94.685							
		$\hat{\beta}_t$	51.415							

Table 4.7: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.7), (4.4), (4.5) and $T = 100$.

Finally, Table 4.8 summarizes results when (u_t, v_t) are generated according to (4.6) and $\sigma_t = 1$ for all t. In the presence of serial correlation, CV remains the most effective method for selecting bandwidths for the IV estimator that are close to the optimal values, while DWB has the edge in the case of LS estimation. Even though CV does better than other methods in terms of coverage of IV confidence intervals, the figures are still well below the target nominal value (which is also the case for the optimal bandwidth).....ESTIMATION ERROR... As in exactly identified models, LS confidence intervals associated with any of the automated bandwidths undercover substantially even when x_t is exogeous.

		Estimator	*CV*	*AIC*	*WB*	*DWB*($\lambda=2$)	*DWB*($\lambda=4$)	*DWB*($\lambda=6$)	*DWB*($\lambda=8$)	*DWB*($\lambda=10$)
		$\hat{\Psi}_t$	1.045	0.662	1.103	1.101	1.099	1.095	1.094	1.094
	Bandwidth Ratio	$\tilde{\beta}_t$	1.053	0.835	1.169	1.176	1.190	1.196	1.205	1.205
		$\hat{\beta}_t$	1.084	1.125	1.117	1.108	1.084	1.071	1.065	1.055
s=0	***Estimation Error***	$\tilde{\beta}_t$	0.161	0.225	0.174	0.208	0.208	0.207	0.208	0.208
		$\hat{\beta}_t$	0.137	0.134	0.137	0.137	0.137	0.137	0.137	0.137
	Coverage	$\tilde{\beta}_t$	88.434	85.768	85.180	74.730	74.908	75.004	74.816	74.853
		$\hat{\beta}_t$	70.072	70.106	69.659	69.621	70.209	70.357	70.487	70.845
	Optimal Coverage	$\tilde{\beta}_t$	90.414							
		$\hat{\beta}_t$	76.358							
		$\hat{\Psi}_t$	1.038	0.766	1.106	1.107	1.105	1.104	1.104	1.104
	Bandwidth Ratio	$\tilde{\beta}_t$	1.079	0.963	1.135	1.142	1.164	1.186	1.205	1.205
		$\hat{\beta}_t$	1.038	1.076	1.070	1.059	1.035	1.022	1.017	1.009
s=0.5	***Estimation Error***	$\tilde{\beta}_t$	0.137	0.191	0.145	0.187	0.187	0.187	0.187	0.188
		$\hat{\beta}_t$	0.164	0.163	0.163	0.164	0.164	0.165	0.164	0.164
	Coverage	$\tilde{\beta}_t$	85.886	82.506	84.214	70.819	70.835	70.887	70.881	70.794
		$\hat{\beta}_t$	57.479	57.356	56.935	57.057	57.401	57.670	57.855	57.982
	Optimal Coverage	$\tilde{\beta}_t$	88.073							
		$\hat{\beta}_t$	60.373							

Table 4.8: Average ratio of automatically chosen to optimal bandwidth, average median absolute estimation error, and average pointwise coverage rates under (4.7), (4.6) and $T = 100$.

5 Empirical Application

In this section, we revisit the time-varying version of the backward-looking Phillips curve analyzed by GKM. The aim is to compare estimates of the parameters of the model obtained using different data-driven bandwidth selectors.

More specifically, we consider the model

$$\Delta\pi_t = c_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta U_t + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{5.1}$$

where π_t is price inflation, U_t is the unemployment rate, c_t are constants, ε_t is a random error, and Δ is the first-difference operator. The data (obtained from the FRED database) consist of 648 monthly observations, from January 1959 to December 2013, on U.S. consumer price inflation and the unemployment rate. Following GKM, kernel estimates are obtained using $(1, \Delta\pi_{t-1}, \Delta U_{t-1}, \Delta U_{t-2}, \Delta U_{t-3}, \Delta U_{t-4})'$ as the vector of instruments and $K(w) = \exp(-w^2/2)$ as kernel function. As noted in GKM, a Lagrange multiplier test for fourth-order serial correlation reveals no significant signs of serial correlation in IV residuals of the model (GKM set $L = H = T^{0.7}$).

Table 4.9 reports the bandwidths $H = T^h$ for LS and IV estimators that are selected by means of the data-driven methods discussed in section 3. In the case of bootstrap-based selectors, results are obtained using $B = 999$ bootstrap replications, with bandwidth $\lambda \in \{42, 59, 68, 76, 85\}$ for the DWB.[7] While the differences between the bandwidth values chosen by the various procedures do not appear to be substantial, there are some noticeable differences in the resulting coefficient estimates.

[7] These choices of λ are of the order $T^{1/3}$, which is known to be optimal in certain respects (see Shao (2010)).

Method	h	
	TV-LS	TV-IV
$\boldsymbol{CV}$	0.643	0.899
$\boldsymbol{AIC}$	0.642	0.871
$\boldsymbol{WB}$	0.690	0.823
$\boldsymbol{DWB}(\lambda = 42)$	0.597	0.853
$\boldsymbol{DWB}(\lambda = 59)$	0.643	0.783
$\boldsymbol{DWB}(\lambda = 68)$	0.597	0.807
$\boldsymbol{DWB}(\lambda = 76)$	0.690	0.877
$\boldsymbol{DWB}(\lambda = 85)$	0.620	0.899

Table 4.9: This table shows h where T^h are the bandwidths selected by the data-driven methods for the backward-looking Phillips curve.

In the LS case, these differences can be seen in Figure 4.1, which shows LS estimates of γ_t and α_t, together with corresponding 95% pointwise confidence bands, based on bandwidths obtained by CV, WB and DWB (with $\lambda = 68$). Estimates of γ_t based on the three automatically selected bandwidths are quite similar for most of the sample, the only exception being a period around 2000. The same is true for estimates of α_t, with some differences among the three sets of estimates also observed during the 1980s. In fact, its only during the latter period that the coefficient on ΔU_t appears to be statistically significant (at the 5% level), regardless of the bandwidth selector used. Needless to say, these results should be viewed with caution since LS estimates are inconsistent unless ΔU_t is exogenous in (5.1). As a matter of fact, this does not appear to be the case: a time-varying Hausman test rejects exogeneity.

Turning to IV estimation of the parameters of the model (5.1), Figure 4.2 shows IV estimates of γ_t and α_t, and associated 95% pointwise confidence bands. From the mid 1970's onwards, there is little difference between estimates of either parameter obtained using the CV, WB and DWB bandwidth choices, small differences being evident only early in the sample period. Interestingly, the coefficient on unemployment is statistically significant (at the 5% level) for all points in the sample, suggesting that a traditional unemployment–inflation trade-off is supported by the data once endogeneity of unemployment is accounted for via the use of IV.

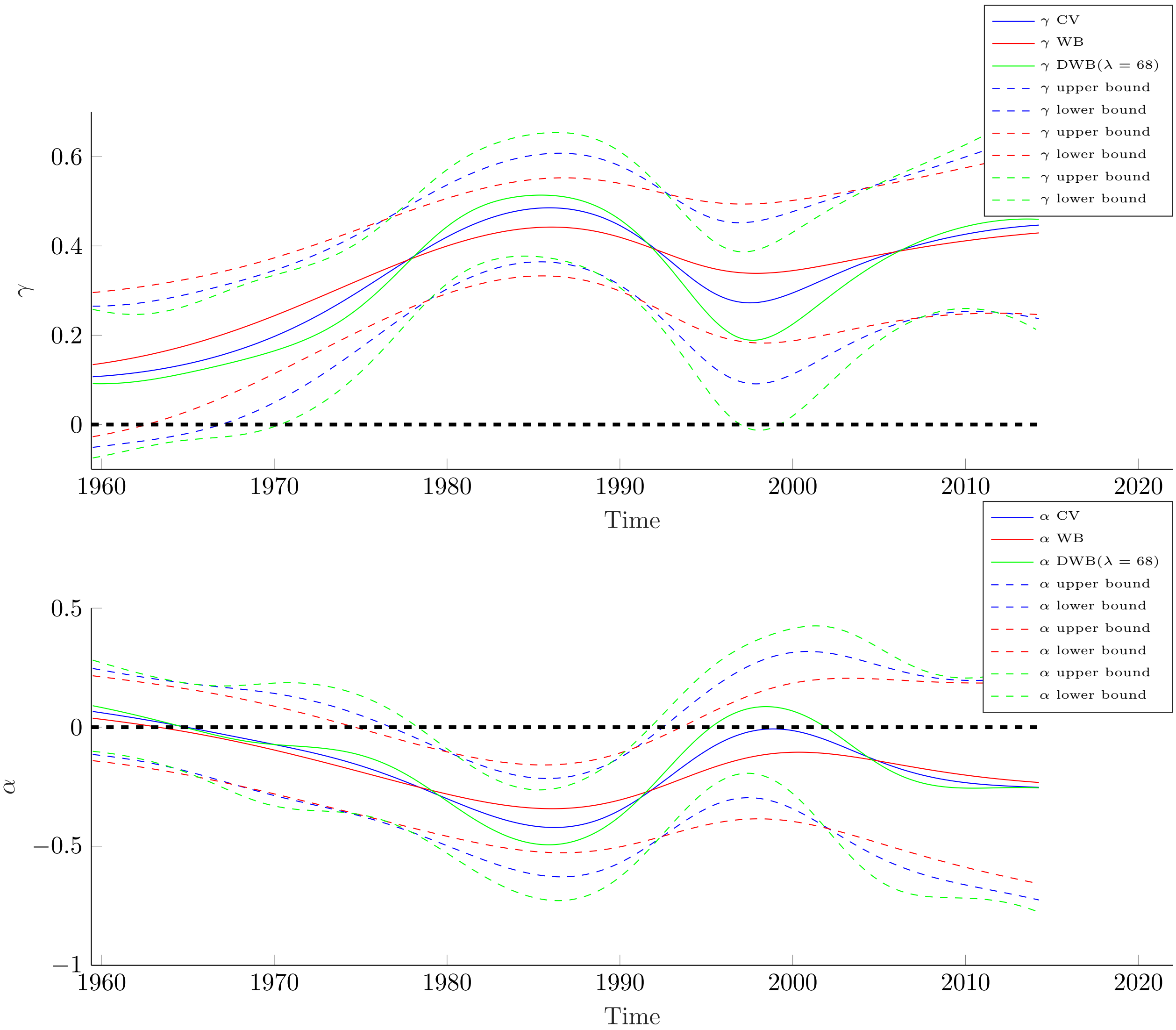


Figure 4.1: LS estimates of γ_t and α_t based on bandwidths selected by CV, WB and DWB ($\lambda = 68$).

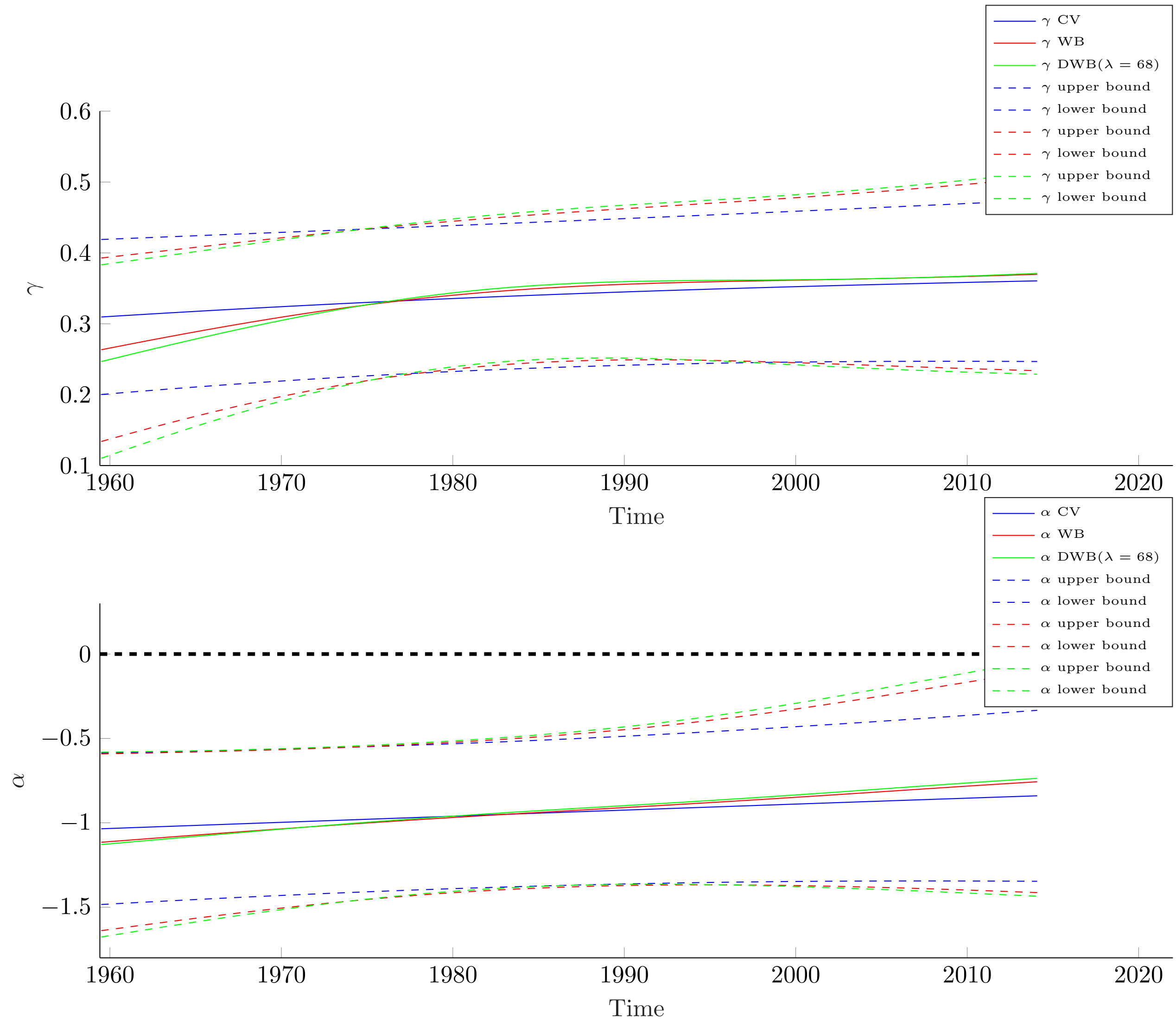


Figure 4.2: IV estimates of γ_t and α_t based on bandwidths selected by CV, WB and DWB ($\lambda = 68$).

6 Conclusion

In this paper, we have considered data-driven methods for selecting the smoothing parameter for kernel IV and LS estimators of stochastically time-varying coefficients in linear models with explanatory variables that may endogenous. Our simulation findings have revealed that CV and DWB are effective automated methods, selecting bandwidths which are close to the optimal values and yielding coefficient estimators with minimal average estimation errors. What is more, DWB and, perhaps surprisingly, ordinary CV work equally well in models with hererogeneously distributed or serially correlated errors as they do in models with identically distributed and independent errors. Our results provide valuable insights into the effectiveness of different data-driven methods for bandwidth selection and can be used to address an obvious hurdle in the practical application of kernel estimators of time-varying coefficients in a rich class of models.

A finding that should not be ignored is that, regardless of the data-driven bandwidth selector used, pointwise confidence intervals for time-varying coefficients appear to have coverage rates which are generally lower than the nominal target value, a difficulty that also arises when bandwidth values that are optimal (in the sense of minimizing the average absolute or quadratic estimation error) are used. It would be useful, therefore, to consider data-driven selectors which produce bandwidth choices that control effectively the error in coverage rates of pointwise confidence intervals, or of simultaneous confidence regions, for time-varying coefficients. The possibility of constructing such confidence intervals/regions using appropriate bootstrap approximations to the sampling distributions of kernel IV and LS estimators, instead of the asymptotic normal approximations, would also be worth exploring. These problems will be considered in detail elsewhere.

Chapter 5

Conclusions

The work in this thesis has focused on several problems relating to diagnostic tests for non-linear dependencies, to tests for endogeneity in the presence of parameter instability of general unspecified forms, and to the selection of smoothing parameters for non-parametric inference in models with stochastically or deterministically time-varying parameters. These are problems of considerable interest in econometrics given the pervasive nature of non-linearity, endogeneity, structural change and parameter instability in relationships among economic and financial variables.

In Chapter 2, we introduced a data-driven approach for a portmanteau test based on the autocorrelations of a squared time series or the autocorrelations of squared residuals from a fitted model. The selection of the number of sample autocorrelations to be used is data-dependent. The data is also allowed to select whether the choice of the number of autocorrelations is based on Akaike's information criterion or on the Bayesian information criterion. The proposed automatic test is easy to implement, has a chi-square asymptotic null distribution and, most importantly, properly controls the finite-sample probability of Type I error whilst providing higher power than the conventional portmanteau test based on a pre-specified number of autocorrelations. Moreover, if deviations from the assumption of independent and identically distributed errors in a time series model are viewed as evidence of non-linear behaviour, the automatic portmanteau test can also be used as a linearity test.

In Chapter 3, we investigated bootstrap versions of a time-varying Hausman test for exogeneity. The test compares kernel-based least squares and instrumental variables estimators of stochastic time-varying coefficients, allowing for possible changes in the endogeneity status of the regressors over time. However, when asymptotic critical values are used in the construction of the test, the latter exhibits size distortions and low power. To address these limitations, we considered using fixed-design wild bootstrap to obtain a more accurate approximation to the null sampling distribution of the test statistic. The resulting bootstrap-based test has accurate size and higher power than the asymptotic test in exactly identified models, and its size remains reasonably close to the nominal value in overidentified models. More remarkably, the size and power of the bootstrap test are insensitive with respect to the choice of the bandwidth parameters used to construct kernel estimators. This is particularly important since applied researchers usually employ a variety of ad-hoc approaches to bandwidth selection which are typically based on minimising objective functions that address estimation concerns rather than test accuracy.

In Chapter 4, we focused on the problem of selecting the bandwidth parameter for kernel least squares and instrumental variables estimators of stochastically time-varying coefficients in regression models with exogenous or endogenous regressors. We considered data-driven bandwidth selectors based on cross-validation, Akaike's information criterion, and wild bootstrap and dependent wild bootstrap procedures. Monte Carlo simulations showed that cross-validation and wild dependent bootstrap techniques perform well, yielding estimated bandwidths that are close to the optimal values. Data-driven bandwidth methods are also useful in the construction of pointwise confidence intervals for time-varying coefficients, although the coverage rates of confidence intervals based on an asymptotic normal approximation appear to be universally low.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Anderson, O. D. (1979). An Introduction to Bilinear Time Series Models. *Journal of the Royal Statistical Society Series C*, 28(3):305–306.

Anderson, T. W. (1991). *The asymptotic distributions of autoregressive coefficients.* Stanford University. Department of Statistics.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Beran, R. and Ducharme, G. (1991). Asymptotic theory for bootstrap methods in statistics. *Les Publications CRM, Université de Montréal.*

Berg, A., Paparoditis, E., and Politis, D. N. (2010). A bootstrap test for time series linearity. *Journal of Statistical Planning and Inference*, 140(12):3841–3857.

Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons.

Box, G. E. P. and Jenkins, G. (1990). *Time Series Analysis, Forecasting and Control.* Holden-Day, Inc., USA.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65.

Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods.* Springer science & business media.

Broersen, P. M. T. and de Waele, S. (2004). Finite sample properties of arma order selection. *IEEE Transactions on Instrumentation and Measurement*, 53(3):645–651.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, pages 603–618.

Bühlmann, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310.

Burman, P., Chow, E., and Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, 81(2):351–358.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Chen, B. (2015a). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.

Chen, B. (2015b). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.

Chronopoulos, I., Kapetanios, G., and Petrova, K. (2021). Kernel-based volatility generalised least squares. *Econometrics and Statistics*, 20:2–11.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169.

Davidson, R. and MacKinnon, J. G. (2000). Bootstrap tests: how many bootstraps? *Econometric Reviews*, 19(1):55–68.

Davidson, R. and MacKinnon, J. G. (2010). Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics*, 28(1):128–144.

Dendramis, Y., Giraitis, L., and Kapetanios, G. (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory*, 37(6):1100–1134.

Djogbenou, A., Gonçalves, S., and Perron, B. (2015). Bootstrap inference in regressions with estimated factors and serial correlation. *Journal of Time Series Analysis*, 36(3):481–502.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Escanciano, J. C. and Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, 151(2):140–149.

Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013). Automatic specification testing for vector autoregressions and multivariate nonlinear time series models. *Journal of Business & Economic Statistics*.

Faraway, J. J. (1990). Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *Journal of Statistical Computation and Simulation*, 37(1-2):37–44.

Fenga, L. and Politis, D. N. (2011). Bootstrap-based arma order selection. *Journal of Statistical Computation and Simulation*, 81(7):799–814.

fu Wong, K. (1996). Bootstrapping hausman's exogeneity test. *Economics Letters*, 53(2):139–143.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222.

Giannerini, S., Maasoumi, E., and Dagum, E. B. (2015). Entropy testing for nonlinear serial dependence in time series. *Biometrika*, 102(3):661–675.

Giraitis, L., Kapetanios, G., and Marcellino, M. (2021). Time-varying instrumental variable estimation. *Journal of Econometrics*, 224(2):394–415.

Giraitis, L., Kapetanios, G., and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1):46–65.

Goncalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120.

González Manteiga, W., Martínez Miranda, M. D., and Pérez González, A. (2004). The choice of smoothing parameter in nonparametric regression through wild bootstrap. *Computational Statistics & Data Analysis*, 47(3):487–515.

Granger, C. W. J. and Ap, A. (1978). *An introduction to bilinear time series models.* Vandenhoeck and Ruprecht.

Hall, A. R., Han, S., and Boldea, O. (2012). Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics*, 170(2):281–302. Thirtieth Anniversary of Generalized Method of Moments.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32(2):177–203.

Hall, P. (1992). The bootstrap and edgeworth expansion. *Springer, Berlin.*

Hall, P., HOROWITZ, J. L., and JING, B.-Y. (1995a). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Hall, P., Lahiri, S. N., and Polzehl, J. (1995b). On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Annals of Statistics*, 23(6):1921–1936.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10(1):130–145.

Härdle, W., Hall, P., and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, 83(401):86–95.

Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society B*, 53(1):173–187.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Horowitz, J. L. (2001). Chapter 52 - the bootstrap. volume 5 of *Handbook of Econometrics*, pages 3159–3228. Elsevier.

Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society B*, 60(2):271–293.

Inglot, T. and Ledwina, T. (2006). Towards data driven selection of a penalty function for data driven neyman tests. *Linear algebra and its applications*, 417(1):124–133.

Kapetanios, G. (2010). Testing for exogeneity in threshold models. *Econometric Theory*, 26(1):231–259.

Köhler, M., Schindler, A., and Sperlich, S. (2014). A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, 82(2):243–274.

Kreiss, J.-P. (1991). Estimation of the distribution function of noise in stationary processes. *Metrika*, 38(1):285–297.

Kreiss, J.-P. (1997). *Asymptotical properties of residual bootstrap for autoregressions.* Citeseer.

Kuersteiner, G. M. (2001). Optimal instrumental variables estimation for arma models. *Journal of Econometrics*, 104(2):359–405.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404.

Lawrance, A. and Lewis, P. A. (1987). Higher-order residual analysis for nonlinear time series with autoregressive correlation structures. *International Statistical Review/Revue Internationale de Statistique*, pages 21–35.

Lee, T.-H., White, H., and Granger, C. W. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of econometrics*, 56(3):269–290.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59.

Lehmann, E. L., Romano, J. P., and Casella, G. (1986). *Testing statistical hypotheses*, volume 3. Springer.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lucchetti, R. and Valentini, F. (2021). Kernel-based time-varying iv estimation: handle with care. Technical report, University Library of Munich, Germany.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*, 15(3):161–175.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255 – 285.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4):269–273.

Nicholls, D. F. and Pagan, A. (1985). 16 varying coefficient regression. *Handbook of statistics*, 5:413–449.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16(2):134–153.

Pérez, A. and Ruiz, E. (2003). Properties of the sample autocorrelations of nonlinear transformations in long-memory stochastic volatility models. *Journal of Financial Econometrics*, 1(3):420–444.

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219 – 230.

Politis, D. N. and Romano, J. P. (1991). *A circular block-resampling procedure for stationary data.* Purdue University. Department of Statistics.

Psaradakis, Z. and Vávra, M. (2019). Portmanteau tests for linearity of stationary time series. *Econometric Reviews*, 38(2):248–262.

Robinson, P. M. (1989). *Nonparametric estimation of time-varying parameters.* Springer.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change: Analysis and Forecasting*, pages 179–190. Springer.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association.*

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235.

Shao, X. and Politis, D. N. (2013). Fixed b subsampling and the block bootstrap: improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(1):161–184.

Terasvirta, T. (1998). Modelling economic relationships with smooth transition regressions. *Handbook of Applied Economic Statistics*, pages 507–552.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295.

Zhang, C., Osborn, D. R., and Kim, D. H. (2008). The new keynesian phillips curve: From sticky inflation to sticky prices. *Journal of Money, Credit and Banking*, 40(4):667–699.

Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Annals of Statistics*, 40(3):1376–1402.