Herwartz, Helmut; Neumann, Michael H.

Working Paper

Bootstrap inference in single equation error correction models

SFB 373 Discussion Paper, No. 2000,87

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Herwartz, Helmut; Neumann, Michael H. (2000) : Bootstrap inference in single equation error correction models, SFB 373 Discussion Paper, No. 2000,87, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10048134

This Version is available at:
https://hdl.handle.net/10419/62252

Bootstrap Inference in Single Equation Error Correction Models

Helmut Herwartz and Michael H. Neumann

Institute of Statistics and Econometrics
Humboldt University Berlin
Spandauer Str. 1
D - 10178 Berlin
GERMANY
Tel.: +49 (30) 2093 5725

SFB 373
Humboldt University Berlin
Spandauer Str. 1
D - 10178 Berlin
GERMANY
Tel.: +49 (30) 2093 1458

Abstract

In the sequel of its seminal application in Davidson, Hendry, Srba and Yeo (1978) the single equation error correction model has been widely used in empirical practice. Providing a clear distinction between short- and long-run dynamics this model allows OLS-methods to be as efficient as (multivariate) full information maximum likelihood methods under a few assumptions on weak exogeneity and cointegration. We consider OLS-based tests on long-run relationships, weak exogeneity and short-run dynamics. For the latter issues it is known that common test-statistics are no longer pivotal if model errors exhibit conditional heteroskedasticity. We show that the wild bootstrap provides convenient critical values for the considered OLS-based statistics under both homoskedastic and conditionally heteroskedastic model errors. The wild bootstrap is easy to implement and turns out to improve considerably the empirical size of common test statistics compared to first order asymptotic approximations. We prove further that the wild bootstrap retains its validity for inference within a system of pooled equations exhibiting cross sectional correlation. Opposite to feasible GLS methods our approach does not require any parametric specification of cross sectional correlation, and copes with time varying patterns of contemporaneous error correlation.

Keywords: Error correction models, panel cointegration analysis, bootstrap.

JEL Classification: C12, C15, C22, C23.

The research of this paper was carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

1. Introduction

A vast body of theoretical and empirical literature on cointegration has emerged since its introduction in Granger (1981). The Engle-Granger representation theorem (Engle and Granger 1987) provides a one-to-one relationship between the framework of cointegration and the so-called error correction model (ECM) which itself has attracted an enormous interest at least following the seminal investigation in Davidson, Hendry, Srba and Yeo (1978). Within the vector ECM (VECM) transitory and long-run dynamics are separated, the latter of which are often of interest from the viewpoint of economic theory. Boswijk (1995b) advocates the conditional ECM, specifying the joint data generating process (DGP) of endogenous variables conditional on a set of variables satisfying the assumption of weak exogeneity as defined in Engle, Hendry and Richard (1983). The single equation version of this model (SECM) allows asymptotically efficient estimation and inference if a set of assumptions concerning in particular the dimension of the cointegrating space and exogeneity of the conditioning variables can be made. Compared to a VECM a particular advantage of the SECM is that efficient inference within the latter model is very close to common ordinary least squares (OLS) procedures (see e.g. Boswijk 1993, Kremers, Ericsson and Dolado 1992).

In this paper we build upon the convenience and widespread use of the SECM. Firstly we compare the performance of OLS-based inference by means of common first order asymptotic approximations on the one hand and employing a bootstrap procedure, namely the wild bootstrap, on the other hand. Providing a comprehensive toolkit we consider tests on equilibrium relationships, weak exogeneity and short-run dynamics. Throughout we allow for time series processes which are generated from conditionally heteroskedastic error terms. Secondly, we show that the recommended bootstrap scheme retains its validity if a set of SECMs exhibiting cross sectional error correlation is pooled. Therefore, apart from the literature on cointegration two further areas of research connect directly to our contribution, panel time series analysis and bootstrap methods.

Naturally, economic models become more reliable if the implied econometric hypotheses are tested across a panel of economic entities rather than relying on results of single equation analyses. Furthermore, a number of economic models directly claim identical equilibrium relations to hold for different members of a panel. Convenient examples can be found, for instance, in the exchange rate or growth and convergence literature (see e.g. Donaldson and

Mehra 1983, Edison, Gagnon and Melick 1997, Froot and Rogoff 1995, Lothian 1997, Neusser 1991). From a statistician's viewpoint an analysis of time series panel data promises a power improvement of inference procedures due to increased samples sizes which are typically implied by a consideration of pooled systems.

The growing econometric literature on the analysis of panels of nonstationary time series already addresses the issue of testing for cointegration on the pooled level. Pedroni (1998) advocates a residual based approach to test for cointegration, similar to the (two step) Engle-Granger procedure. The two step procedure, however, fails to provide information concerning the dimension of the cointegrating space, i.e. the joint cointegrating rank on the pooled level. Groen and Kleibergen (1999) generalize Johansen's (1991) maximum likelihood (ML-) procedure to apply in pooled systems assuming a fixed cross section dimension. Once a joint cointegrating rank is determined the latter approach also allows to infer against homogeneity across different members of the panel.

As a further line of research related to our paper Li and Maddala (1997) survey different methods of bootstrapping in systems of cointegrated time series variables. By means of a Monte-Carlo investigation they show for particular resampling schemes that empirical size properties of likelihood ratio (LR-) tests on long-run relationships (Johansen 1991) can be substantially improved by bootstrap schemes. Similar results are provided in Herwartz (1998) where the empirical performance of LR-type statistics in stationary time series models is analyzed. Improving the small sample properties of common test procedures is particularly relevant for the analysis of time series panels where typically a few observations are available for each member of the panel. It is known that often OLS based test statistics lose their pivotal property if the underlying model errors are heteroskedastic. In this case the wild bootstrap provides a convenient means to mimic the distribution a particular test statistic since its asymptotic distribution may be difficult or even impossible to approximate analytically.

As mentioned we address different issues of inference which are essential for (pooled) SEC modelling. Inference in SECM is assumed to be asymptotically equivalent to full information ML-methods, i.e. we regard the involved variables to be cointegrated with cointegrating rank 1 and assume further that all conditioning variables are weakly exogenous. We mainly

employ LR-type statistics to test restrictions on long-run parameters, to infer on weak exogeneity of conditioning variables, and to test on insignificance of parameters governing short-run dynamics. Note that for the latter cases OLS-based test statistics are no longer pivotal if model error terms are conditionally heteroskedastic. Compared to the WALD-statistic the LR-test has the particular advantage that it can be easily pooled across equations in the case of a finite cross section dimension. The asymptotic distribution of the pooled LR-test, however, is hardly available in presence of cross sectional error correlation. We show that the wild bootstrap is easily implemented to account for cross sectional correlation without requiring any parametric (first step) estimate of its pattern. In addition, the method copes with time dependence of contemporaneous correlation. Often used for pure regression models, a further advantage of the wild bootstrap in a (multivariate) time series framework is that this procedure does not require to specify a DGP for the involved time series variables.

To shed light on the empirical properties of competing inference procedures we perform a Monte Carlo investigation under homoskedastic and heteroskedastic error processes. Simulating a very simple bivariate process it turns out that first order asymptotic approximations might involve considerable size distortions which can be mitigated substantially or even overcome by means of the wild bootstrap. Remarkable size improvements in small samples are obtained from bootstrapping OLS-based statistics on the pooled level. If OLS-based statistics lose their pivotal property standard critical values are inappropriate involving size distortions which do not vanish asymptotically. The wild bootstrap provides valid critical values even in presence of heteroskedastic error terms.

The remainder of the paper is organized as follows: In the next Section we discuss in detail the empirical model and the investigated OLS-based test statistics. The wild bootstrap procedure is discussed in Section 4. Section 5 provides a detailed simulation study. Concluding remarks and suggestions for future work can be found in Section 6. Proofs of our results are given in Section 7.

2. Methodology

2.1. **The Model and Assumptions.** We consider the single equation conditional ECM:

$$\Delta y_t = \nu_1 + \alpha_1(y_{t-1} + \beta z_{t-1}) + \gamma_1 \Delta z_t + u_t, \quad t = 1, \dots, T, \tag{2.1}$$

where $-2 < \alpha_1 < 0$.

For the marginal process we assume the following representation:

$$\Delta z_t = \nu_2 + \alpha_2(y_{t-1} + \beta z_{t-1}) + v_t, \quad t = 1, \ldots, T. \tag{2.2}$$

The adopted single equation approach to estimate long-run equilibrium relations and adjustment dynamics is equivalent to full information ML-estimation (Johansen 1991) and thus asymptotically efficient if a set of assumptions can be made (see e.g. Johansen (1992), and Banerjee, Dolado, Galbraith and Hendry (1993)). First, the involved variables y_t and z_t are assumed to be integrated of order one. Second, there exists a linear combination of the nonstationary processes providing stationary residuals, i.e. y_t and z_t are cointegrated with cointegrating rank 1. According to the Engle-Granger representation theorem this assumption implies that at least one out of α_1 and α_2 must be nonzero. Third, z_t is weakly exogenous for the estimation of β and α_1. Note that $\alpha_2 = 0$ is a necessary and sufficient condition for weak exogeneity of z_t (see Urbain 1992, Boswijk 1995b). Fourth, following Phillips (1988) we assume that the error sequences u_t and v_t are serially uncorrelated having a positive definite covariance matrix. We consider the bivariate case, involving a scalar marginal process z_t, just for convenience of notation. The asymptotic results provided below still apply for higher dimensional marginal processes as long as the assumptions concerning the cointegrating rank and weak exogeneity are met.

In order to obtain serially uncorrelated error processes u_t and v_t equations (2.1) and (2.2) may be conveniently augmented with further (lagged) stationary explanatory variables. To derive the asymptotic distributions of a few OLS-based test statistics (e.g. testing on weak exogeneity of z_t) it is necessary to impose a stronger assumption on the model errors u_t and v_t. In this case both error processes are typically assumed to be homoskedastic. Throughout we assume that necessary presample values are available. To have single equation OLS-procedures applied to (2.1) asymptotically equivalent to an ML-analysis of the corresponding VECM one always has to test for weak exogeneity of z_t within the marginal model (2.2).

Apart from efficient estimation the assumptions provided above are also sufficient to perform asymptotically valid inference by means of common OLS-based procedures. For instance, $t-$ratios and $F-$type statistics on joint significance of selected parameters are widely used. In particular, the t-ratios ($t_{\widehat{\alpha}_i}$) of the error correction coefficients $\widehat{\alpha}_i$, $i = 1, 2$, obtained

from OLS-routines are asymptotically normally distributed if the cointegration assumption holds. Therefore the normal distribution is often used to infer on weak exogeneity of z_t which implies that the error correction parameter in the marginal process is actually zero ($\alpha_2 = 0$). If the variables fail to be cointegrated then, however, $q_t = y_t + \beta z_t$ is nonstationary. In this case the regression in (2.1) is "unbalanced" since a (lagged) nonstationary variable is employed to describe the dynamics of a stationary series. Therefore α_1 should be zero if y_t and z_t are not cointegrated. Thus, a test on significance of α_1 is implicitly also a test of the null hypothesis of cointegration. Under the alternative of no cointegration, however, $t_{\widehat{\alpha}_1}$ fails to be asymptotically normally distributed. For this case Kremers *et al.* (1992) show that the distribution of $t_{\widehat{\alpha}_1}$ is somewhere between the standard normal distribution and the distribution of the Dickey-Fuller t–statistic. Boswijk (1995a) provides a WALD-test of the null hypothesis $H_0 : \alpha_1 = 0$. Testing against the alternative of cointegration in the conditional model (2.1) such a procedure fully exploits the assumed weak exogeneity of z_t. Using Monte Carlo techniques Boswijk and Franses (1992) show that this approach outperforms competing cointegration tests in terms of size and power if the conditioning variables are actually weakly exogenous.

2.2. **The WALD-Test.** Boswijk (1993) provides the WALD-statistic to test hypotheses involving the long-run parameter β, $H_0 : \beta = \beta_0$, say. To provide a formal representation of the WALD-statistic it is convenient to reformulate the conditional equation compactly as follows:

$$\Delta y = X_{-1}\theta + W\vartheta + u \tag{2.3}$$

In (2.3), $X_{-1} = (y_{-1} \quad z_{-1})$, where $y_{-1} = (y_0, \ldots, y_{T-1})'$ and $z_{-1} = (z_0, \ldots, z_{T-1})'$, is composed of vectors of lagged nonstationary explanatory variables. Accordingly $W = (1_T \quad \Delta z)$, where $\Delta z = (\Delta z_1, \ldots, \Delta z_T)'$, contains all right-hand side variables in (2.1) which are either stationary or deterministic. The parameter vectors in (2.3) are composed as follows: $\theta = (\alpha_1, \alpha_1\beta)'$, $\vartheta = (\nu_1, \gamma_1)'$. The OLS-estimate of $\pi = (\theta', \vartheta')'$ is

$$\widehat{\pi} = \begin{pmatrix} \widehat{\theta} \\ \widehat{\vartheta} \end{pmatrix} = \begin{bmatrix} X'_{-1}X_{-1} & X'_{-1}W \\ W'X_{-1} & W'W \end{bmatrix}^{-1} \begin{bmatrix} X'_{-1}\Delta y \\ W'\Delta y \end{bmatrix}. \tag{2.4}$$

If the u_t are homoskedastic, then a natural covariance estimator is given by

$$\widehat{\Sigma}_{\widehat{\pi}} = \widehat{\sigma}_u^2 \begin{bmatrix} X'_{-1}X_{-1} & X'_{-1}W \\ W'X_{-1} & W'W \end{bmatrix}^{-1}, \qquad \widehat{\sigma}_u^2 = \frac{1}{T-K}\widehat{u}'\widehat{u}, \tag{2.5}$$

where K is the number of explanatory variables employed in (2.1). A convenient expression for the WALD-statistic is obtained after defining $R = (\beta_0 \quad -1)$ as (see Boswijk 1993):

$$\mathrm{W} = (R\widehat{\theta})'[R\widehat{\Sigma}_{\widehat{\theta}}R']^{-1}(R\widehat{\theta}) \Rightarrow \chi^2(1). \tag{2.6}$$

Instead of taking critical values for W from its asymptotic χ^2-distribution it is often preferred in investigating small samples to use the $F(1, T-K)$-distribution. In the following we refer to this strategy as the WALD-test.

2.3. **LR-type test statistics.** Since the WALD-test fails to be invariant with respect to reformulations of the null hypothesis one may regard a Gaussian LR-test as a promising alternative device. A further advantage of the LR-test is that results from different equations can often be pooled. Thus if one is interested in testing a particular hypothesis to hold in a system of univariate regression models LR-inference provides an appealing means if there is no cross sectional correlation. The purpose of the present study is at least twofold. Firstly, we provide an alternative procedure to obtain critical values for the LR-statistic, namely a bootstrap scheme. It will turn out that such an approach improves the small sample properties of the LR-test considerably. Secondly, we show that the recommended bootstrap scheme is convenient to estimate critical values for LR-type statistics even in the case of pooled equations exhibiting cross sectional error correlation. Concerning the latter issue we are interested in evaluating whether the good size properties of the bootstrap scheme in single equations carry over to the case of pooled LR-statistics.

In order to provide a comprehensive toolkit for empirical practice we consider LR-type statistics for three testing problems: First, we investigate hypothesis testing for long-run parameters providing the LR-counterpart of the WALD-statistic in (2.6). Second, we address the issue of inferring on weak exogeneity of z_t, i.e. we discuss the LR-counterpart of the standard t-ratio of α_2 in (2.2). Finally we investigate how LR-type inference may be employed to test on insignificance of further parameters governing short-run dynamics. Whereas the first two issues are of immediate interest for economists the latter two tests

support the econometrician in specifying empirical models as (2.1) or (2.2) generated by serially uncorrelated error sequences. Throughout we are interested in proving the asymptotic validity of the recommended bootstrap procedure. For this reason we provide a detailed representation of the relevant LR-statistics.

Testing restrictions on long-run parameters. The ECM in (2.1) may be given compactly as:

$$\Delta y \;=\; \mathcal{X}_1(\theta_1' \quad \vartheta')' \;+\; u, \tag{2.7}$$

where $\mathcal{X}_1 = (y_{-1} \quad z_{-1} \quad 1_T \quad \Delta z)$, $\theta_1 = (\alpha_1, \alpha_1\beta)'$ and $\vartheta = (\nu_1, \gamma_1)'$. In the context of testing the long-run parameter β we assume that $\alpha_2 = 0$, that is, (2.2) degenerates to

$$\Delta z_t \;=\; \nu_2 \;+\; v_t, \quad t = 1, \ldots, T. \tag{2.8}$$

A corresponding representation of the model is also available under particular null hypotheses. Consider first $H_0^{(1)} : \beta = \beta_0$. Defining deviations from the long-run equilibrium under this hypothesis as $q_t^0 = y_t + \beta_0 z_t$ we have

$$\Delta y \;=\; \mathcal{X}_0(\theta_0' \quad \vartheta')' + u, \tag{2.9}$$

where $\mathcal{X}_0 = (q_{-1}^0 \quad 1_T \quad \Delta z)$ and $\theta_0 = \alpha_1$.

We denote by $\Lambda_i^{(1)}$ $(i = 0, 1)$ the Gaussian likelihoods in models (2.9) and (2.7), respectively. The logarithmic likelihood ratio can be written as

$$\mathrm{LR}^{(1)} \;=\; 2T \ln\left(\frac{\Lambda_1^{(1)}}{\Lambda_0^{(1)}}\right) \;=\; T \ln\left(\frac{\mathrm{RSS}_0^{(1)}}{\mathrm{RSS}_1^{(1)}}\right), \tag{2.10}$$

where $\mathrm{RSS}_0^{(1)}$ and $\mathrm{RSS}_1^{(1)}$ are the residual sums of squares for the least squares estimates based on $\mathcal{X}_0$ and $\mathcal{X}_1$, respectively. Using $\ln(1+x) = x + O(x^2)$ for $x > 0$ we obtain that

$$\mathrm{LR}^{(1)} \;=\; T\frac{\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)}}{\mathrm{RSS}_1^{(1)}} \;+\; O\left(T\left(\frac{\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)}}{\mathrm{RSS}_1^{(1)}}\right)^2\right). \tag{2.11}$$

One usually expects that the logarithmic likelihood ratio is asymptotically $\chi^2(k)$-distributed, with k denoting the number of excess parameters under the alternative model compared to the model under the null hypothesis. In our case, the number of excess parameters is 1. We will actually show that $\mathrm{LR}^{(1)}$ is asymptotically $\chi^2(1)$-distributed.

In order to simplify the calculation of $\mathrm{RSS}_0^{(1)}$ and $\mathrm{RSS}_1^{(1)}$, we apply partial regression, that is, we project both sides of (2.7) and (2.9) first onto $(Im(W))^{\perp}$, where $W = (1_T \quad \Delta z)$, and

consider then least squares fits in these modified models. Defining the projection matrix $M = I_T - W(W'W)^{-1}W'$ we obtain from the model (2.3) that

$$M\Delta y = MX_{-1}\theta + Mu. \tag{2.12}$$

Accordingly, the residual sums of squares in (2.7) and (2.9) can be represented in terms of the variables in (2.12) as

$$\begin{aligned} \mathrm{RSS}_i^{(1)} &= (\Delta y)'(I_T - \mathcal{X}_i(\mathcal{X}_i'\mathcal{X}_i)^{-1}\mathcal{X}_i)\Delta y \\ &= (\Delta y)'M(I_T - P_i)M\Delta y, \end{aligned} \tag{2.13}$$

$i = 0, 1$, where $P_0 = \mathrm{Proj}_{\mathrm{lin}\{\mathtt{q_{-1}}\}}$ and $P_1 = \mathrm{Proj}_{\mathrm{lin}\{\mathtt{y_{-1},z_{-1}}\}}$.

In terms of the representation given in (2.12) the null hypothesis may be respecified for the parameter vector $\theta = (\alpha_1, \alpha_1\beta)'$ as follows:

$$H_0^{(1)}: \quad R\theta = 0,$$

where $R = (\beta_0 \quad -1)$ is a 1×2 matrix containing β_0, the value of the long-run parameter supposed under the null hypothesis. After standard calculations we arrive at the following representation (see e.g. Judge, Hill, Griffiths and Lütkepohl (1988)):

$$\begin{aligned} &\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)} \\ &= u'MX_{-1}(X_{-1}'MX_{-1})^{-1}R'[R(X_{-1}'MX_{-1})^{-1}R']^{-1}R(X_{-1}'MX_{-1})^{-1}X_{-1}'Mu. \end{aligned} \tag{2.14}$$

The $T \times T$ matrix given in (2.14) can be entirely specified in terms of the random variables q_{t-1}^0 and the nonstationary but weakly exogenous variables z_{t-1}. Applying some algebra we obtain the following representation:

$$\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)} = \frac{\left(z_{-1}'Mu - \frac{(q_{-1}^0)'Mz_{-1}}{(q_{-1}^0)'M(q_{-1}^0)}(q_{-1}^0)'Mu\right)^2}{z_{-1}'Mz_{-1} - ((q_{-1}^0)'Mz_{-1})^2((q_{-1}^0)'M(q_{-1}^0))^{-1}}. \tag{2.15}$$

We impose the following condition on the innovations:

(A1) $\{(u_t, v_t)', t \in \mathbb{Z}\}$ is a stationary sequence of serially uncorrelated random vectors satisfying $Eu_t = Ev_t = 0$, $0 < \sigma_u^2 = Eu_t^2 < \infty$, $0 < \sigma_v^2 = Ev_t^2 < \infty$, and $|\mathrm{Corr}(u_t, v_t)| < 1$. Moreover, we assume that

(i) $E|u_t|^{c+\varepsilon} + E|v_t|^{c+\varepsilon} < \infty$ for some $c > 2$ and $\varepsilon > 0$,

(ii) $\sum_{n=1}^{\infty} \alpha(n)^{1-2/c} < \infty$, where $\{\alpha(n)\}$ are the coefficients of strong regularity (α-mixing) defined as

$$\alpha(n) = \sup_{F,G:\, F\in\sigma(X_t,X_{t-1},\dots),\, G\in\sigma(X_{t+n},X_{t+n+1},\dots)} |P(F\cap G) - P(F)P(G)|.$$

We will see below that the limiting behavior of $\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)}$ differs in the two cases of $\nu_2 = 0$ and $\nu_2 \neq 0$. In the first case we have that $z_t = z_0 + v_1 + \cdots + v_t$, whereas we have in the second case that $z_t = \nu_2 t + O_P(T^{1/2})$. Owing to this, the limit distribution of $\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)}$ follows in the latter case from a central limit theorem, whereas we have to employ empirical process theory in the former case. In the following we sketch the steps in approximating the distribution of $\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)}$ in the case $\nu_2 = 0$. Note that u_t and v_t are not necessarily uncorrelated. Since $M\Delta z = 0$ we can replace Mu in (2.15) by Mr, where $r = (r_1, \dots, r_T)'$ and

$$r_t = u_t - \frac{Ev_tu_t}{Ev_t^2}v_t. \tag{2.16}$$

Note that r_t is chosen such that $Ev_tr_t = 0$, which will be important in the sequel. Let $\sigma_r^2 = Er_t^2$. Moreover, we have $T^{-1}\mathrm{RSS}_1^{(1)} \xrightarrow{P} \sigma_r^2$, which implies in conjunction with (2.11) and (2.15) that

$$\begin{aligned}\mathrm{LR}^{(1)} &= \left(\frac{1}{\sigma_r^2} + o_P(1)\right) \frac{\left(z_{-1}'Mr - \frac{(q_{-1}^0)'Mz_{-1}}{(q_{-1}^0)'M(q_{-1}^0)}(q_{-1}^0)'Mr\right)^2}{z_{-1}'Mz_{-1} - ((q_{-1}^0)'Mz_{-1})^2((q_{-1}^0)'M(q_{-1}^0))^{-1}} \\ &\quad + O_P\left(T^{-1}(\mathrm{RSS}_0^{(1)} - \mathrm{RSS}_1^{(1)})^2\right).\end{aligned} \tag{2.17}$$

On the basis of results of Phillips (1988) we can prove the following lemma, which will be the key to deriving the limit distribution of $\mathrm{LR}^{(1)}$.

Lemma 2.1. *Suppose that Assumption (A1) is fulfilled. Then there exist on a sufficiently rich probability space versions of the random variables $v_1, \dots, v_T$ and $r_1, \dots, r_T$, and independent Wiener processes B_v and B_r such that*

$$d\left(\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} v_t,\ s\in[0,1]\right\}, \{\sigma_v B_v(s),\ s\in[0,1]\}\right) \xrightarrow{a.s.} 0, \tag{2.18}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} r_t \xrightarrow{a.s.} \sigma_r B_r(1), \tag{2.19}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}(v_1 + \cdots + v_{t-1})r_t \xrightarrow{a.s.} \sigma_v\sigma_r \int_0^1 B_v(s)\, dB_r(s). \tag{2.20}$$

The reason for introducing a new probability space is a purely technical one. According to the construction described in the proof of Theorem IV.13 in Pollard (1984), it is enough that the new probability space embeds a uniformly distributed random variable. For simplicity of presentation, we assume here and in the following that the original probability space is already rich enough so that a transition to another space is not necessary.

In (2.18), we use the so-called Skorohod metric d. It is defined on the space $D[0,1]$ of right-continuous functions on $[0,1]$ which possess left limits as

$$\begin{aligned} & d\left(\{x(s),\ s \in [0,1]\}, \{y(s),\ s \in [0,1]\}\right) \\ &= \inf_{\varepsilon>0}\left\{\varepsilon : \quad \|\lambda\| \le \varepsilon, \quad \sup_{t\in[0,1]} |x(t) - y(\lambda(t))| \le \varepsilon\right\}, \end{aligned} \tag{2.21}$$

where λ is any continuous mapping of $[0,1]$ onto itself with $\lambda(0) = 0$ and $\lambda(1) = 1$ and

$$\|\lambda\| = \sup_{t\neq s}\left|\ln\frac{\lambda(t) - \lambda(s)}{t - s}\right|, \quad t, s \in [0,1].$$

Using these approximations we can derive the limit distribution of $\mathrm{LR}^{(1)}$ in the case $\nu_2 = 0$, whereas we get such a result for $\nu_2 \neq 0$ by the central limit theorem.

Proposition 2.1. *Suppose that Assumption (A1) is fulfilled and that $\alpha_2 = 0$. Then, under $H_0^{(1)}$,*

$$LR^{(1)} \xrightarrow{d} \chi^2(1).$$

Here the notation $X_n \xrightarrow{d} X$ means convergence in distribution of a sequence of random variables $\{X_n\}$ to X. With a slight abuse of notation we will also write $X_n \xrightarrow{d} P$ if $\{X_n\}$ converges in distribution to any random variable with distribution P. Convergence in distribution is equivalent to weak convergence of the associated probability measures which will be denoted by $\mathcal{L}(X_n) \Longrightarrow \mathcal{L}(X)$.

Testing on weak exogeneity. Now consider the problem of testing weak exogeneity of z_t for inference on β and α_1 in the conditional model (2.1). As mentioned, weak exogeneity holds if the null hypothesis $H_0^{(2)} : \alpha_2 = 0$ is true. Thus, a natural procedure (see Johansen 1992) to test on weak exogeneity is firstly to estimate equilibrium errors and secondly to infer whether this series improves a regression model explaining Δz_t. Following Boswijk (1995b) this procedure can be interpreted as a Lagrange Multiplier test.

Under $H_0^{(2)}$, equation (2.2) reduces to

$$\Delta z \;=\; \mathcal{Z}_0 \delta_0 \;+\; v, \tag{2.22}$$

$\mathcal{Z}_0 = 1_T$ and $\delta_0 = \nu_2$, while we get under $H_1^{(2)}$

$$\Delta z \;=\; \widetilde{\mathcal{Z}}_1 \widetilde{\delta}_1 \;+\; v, \tag{2.23}$$

where $\widetilde{\mathcal{Z}}_1 = (y_{-1} \quad z_{-1} \quad 1_T)$ and $\widetilde{\delta}_1 = (\alpha_2 \quad \alpha_2\beta \quad \nu_2)'$. We do not directly use the residual sum of squares obtained from the least squares estimate in (2.23) to devise a test statistic. Rather, we replace first the true long-run parameter β by the least squares estimate $\widehat{\beta}$ corresponding to the model

$$y_t \;=\; -\beta z_t \;+\; c \;+\; e_t, \quad t = 1, \ldots, T, \tag{2.24}$$

which is known to be superefficient. To be precise, we have the following result which can be found for normally distributed error terms u_t and v_t e.g. in Banerjee *et al.* (1993, Chapter 6).

Lemma 2.2. *Suppose that Assumption (A1) is fulfilled and that $\alpha_2 = 0$. Moreover, we assume that y_t and z_t are cointegrated with cointegrating rank 1.*

(i) *If $\nu_2 = 0$, then*

$$\widehat{\beta} \;-\; \beta \;=\; O_P(T^{-1}),$$

(i) *if $\nu_2 \neq 0$, then*

$$\widehat{\beta} \;-\; \beta \;=\; O_P(T^{-3/2}).$$

We consider instead of (2.23) the equation

$$\Delta z \;=\; \mathcal{Z}_1 \delta_1 \;+\; v, \tag{2.25}$$

where $\mathcal{Z}_1 = (y_{-1} + \widehat{\beta} z_{-1} \quad 1_T)$ and $\delta_1 = (\alpha_2 \quad \nu_2)'$.

Analogously to (2.10) and (2.11), the logarithmic likelihood ratio can be written as

$$\begin{aligned} \mathrm{LR}^{(2)} &= 2T \ln\left(\frac{\Lambda_1^{(2)}}{\Lambda_0^{(2)}}\right) = T \ln\left(\frac{\mathrm{RSS}_0^{(2)}}{\mathrm{RSS}_1^{(2)}}\right) \\ &= T\frac{\mathrm{RSS}_0^{(2)} - \mathrm{RSS}_1^{(2)}}{\mathrm{RSS}_1^{(2)}} + O\left(T\left(\frac{\mathrm{RSS}_0^{(2)} - \mathrm{RSS}_1^{(2)}}{\mathrm{RSS}_1^{(2)}}\right)^2\right). \end{aligned} \tag{2.26}$$

Since $H_0^{(2)}$ corresponds to the linear restriction $R\delta_1 = 0$ with $R = (1 \quad 0)$ we obtain that

$$\mathrm{RSS}_0^{(2)} - \mathrm{RSS}_1^{(2)} = v'\mathcal{Z}_1(\mathcal{Z}_1'\mathcal{Z}_1)^{-1}R'\left[R(\mathcal{Z}_1'\mathcal{Z}_1)^{-1}R'\right]^{-1}R(\mathcal{Z}_1'\mathcal{Z}_1)^{-1}\mathcal{Z}_1'v. \tag{2.27}$$

Proposition 2.2. *Suppose that Assumption (A1) is fulfilled. Moreover, we assume that y_t and z_t are cointegrated with cointegrating rank 1. Then, under $H_0^{(2)}$,*

$$LR^{(2)} \xrightarrow{d} \frac{E(q_{t-1}v_t)^2}{Eq_{t-1}^2\sigma_v^2}\chi^2(1).$$

Testing for significant short-run dynamics. Finally we consider to test on significance of additional short run dynamics entering the conditional model in (2.1). Assume that we are interested in testing whether or not lagged values of the dependent variables, Δy_{t-1} say, contribute to short-run dynamics. To be precise, we assume instead of (2.1) the following empirical model:

$$\Delta y_t = \nu_1 + \alpha_1 q_{t-1} + \gamma_1 \Delta z_t + \gamma_2 \Delta y_{t-1} + u_t, \quad t = 1, \ldots, T.$$

This can be equivalently written in matrix form as

$$\Delta y = W_1\pi_1 + W_2\pi_2 + u, \tag{2.28}$$

where $W_1 = (1 \quad q_{-1})$, $W_2 = (\Delta z \quad \Delta y_{-1})$, and the parameter vectors $\pi_1 = (\nu_1 \quad \alpha_1)'$ and $\pi_2 = (\gamma_1 \quad \gamma_2)'$ are defined accordingly. Furthermore, it is assumed that the z_t obey (2.2). The null hypothesis $H_0^{(3)} : \gamma_2 = 0$ may be specified as $R\pi_2 = 0$, where $R = (0 \quad 1)$.

Before we form the LR-test statistic, we substitute again the cointegration parameter β by the least squares estimator $\widehat{\beta}$ from the regression model (2.24).

Let $\widetilde{q}_{-1} = (\widetilde{q}_0, \ldots, \widetilde{q}_{T-1})'$, where $\widetilde{q}_t = y_t + \beta z_t + (\widehat{\beta} - \beta) z_t$. The residual sums of squares which correspond to $H_0^{(3)}$ and $H_1^{(3)}$ are derived from the model equation (2.28) where q_{t-1} is replaced by $\widetilde{q}_{t-1}$. With $\widehat{W}_1 = (1_T \quad \widetilde{q}_{-1})$, we define the projection matrix $M = I_T - \widehat{W}_1(\widehat{W}_1'\widehat{W}_1)^{-1}\widehat{W}_1'$. Applying results from partial regression we obtain under $H_0^{(3)}$ that

$$\begin{aligned} \mathrm{RSS}_0^{(3)} - \mathrm{RSS}_1^{(3)} &= u'MW_2(W_2'MW_2)^{-1}R'[R(W_2'MW_2)^{-1}R']^{-1}R(W_2'MW_2)^{-1}W_2'Mu \\ &= \frac{\left((\Delta y_{-1})'Mu - \frac{(\Delta y_{-1})'M(\Delta z)}{(\Delta z)'M(\Delta z)}(\Delta z)'Mu\right)^2}{(\Delta y_{-1})'M(\Delta y_{-1})' - ((\Delta y_{-1})'M(\Delta z))^2((\Delta z)'M(\Delta z))^{-1}}. \end{aligned} \tag{2.29}$$

As usual, we obtain that

$$\begin{aligned} \mathrm{LR}^{(3)} &= 2T\ln\left(\frac{\Lambda_1^{(3)}}{\Lambda_0^{(3)}}\right) = T\ln\left(\frac{\mathrm{RSS}_0^{(3)}}{\mathrm{RSS}_1^{(3)}}\right) \\ &= T\frac{\mathrm{RSS}_0^{(3)} - \mathrm{RSS}_1^{(3)}}{\mathrm{RSS}_1^{(3)}} + O\left(T\left(\frac{\mathrm{RSS}_0^{(3)} - \mathrm{RSS}_1^{(3)}}{\mathrm{RSS}_1^{(3)}}\right)^2\right). \end{aligned} \tag{2.30}$$

Proposition 2.3. *Suppose that Assumption (A1) is fulfilled. Moreover, we assume that y_t and z_t are cointegrated with cointegrating rank 1, and that $\alpha_2 = 0$. Let $\widetilde{\Delta y_{-1}} = M\Delta y_{-1}$. Then, under H_0,*

$$LR^{(3)} \xrightarrow{d} \frac{E\left[\left(\widetilde{\Delta y_{t-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{t-1}\right)u_t\right]^2}{E\left(\widetilde{\Delta y_{t-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{t-1}\right)^2\sigma_u^2}\chi^2(1).$$

A sufficient condition for an asymptotic χ^2-distribution of the LR-type test statistic is that the innovations $(u_t, v_t)'$ are i.i.d. If this assumption is violated, the LR-statistic loses its pivotal property. In cases with conditional heteroskedasticity we experienced considerable size distortions if critical values for the LR-statistic are taken from the $\chi^2(1)$-distribution.

3. Inference in pooled systems of equations

Now suppose a set of empirical models as in (2.1) to be under study, for example

$$\Delta y_{n,t} = \nu_{n1} + \alpha_{n1}(y_{n,t-1} + \beta_n z_{n,t-1}) + \gamma_{n1}\Delta z_{n,t} + u_{n,t}, \quad t = 1, \ldots, T, \quad n = 1, \ldots, N, \tag{3.1}$$

where N denotes the number of equations in the system. Accordingly, the marginal processes are assumed to obey the representation

$$\Delta z_{n,t} \;=\; \nu_{n2} \;+\; \alpha_{n2}(y_{n,t-1} + \beta_n z_{n,t-1}) \;+\; v_{n,t}, \quad t = 1, \ldots, T, \quad n = 1, \ldots, N. \tag{3.2}$$

We are interested in testing specific null hypotheses to hold simultaneously in the N equations. A natural generalization of the LR-tests defined in (2.11), (2.26), and (2.30) is given by

$$\begin{aligned} \mathrm{LR}_{[N]}^{(i)} &= 2T \sum_{n=1}^{N} \ln\left(\frac{\Lambda_{n1}^{(i)}}{\Lambda_{n0}^{(i)}}\right) \\ &= T \sum_{n=1}^{N} \frac{\mathrm{RSS}_{n0}^{(i)} - \mathrm{RSS}_{n1}^{(i)}}{\mathrm{RSS}_{n1}^{(i)}} + O\left(T \sum_{n=1}^{N} \left(\frac{\mathrm{RSS}_{n0}^{(i)} - \mathrm{RSS}_{n1}^{(i)}}{\mathrm{RSS}_{n1}^{(i)}}\right)^2\right), \end{aligned} \tag{3.3}$$

where $\mathrm{RSS}_{nj}^{(i)}$ and $\Lambda_{nj}^{(i)}$ are the analogs of $\mathrm{RSS}_j^{(i)}$ and $\Lambda_j^{(i)}$, respectively, in the nth model. This likelihood ratio statistic is already considered in Lütkepohl (1991) in a related context. In (3.1) and (3.2) it is implicitly assumed that T observations are available for each equation. If this is not the case, then the above statistic can be easily modified.

Before we can derive the asymptotics for the test statistics $\mathrm{LR}_{[N]}^{(i)}$, $i = 1, 2, 3$, we have to reformulate Assumption (A1) accordingly.

(A2) $\{(u_{1,t}, v_{1,t}, \ldots, u_{N,t}, v_{N,t}), t \in \mathbb{Z}\}$ is a stationary sequence of uncorrelated random vectors satisfying $Eu_{n,t} = Ev_{n,t} = 0$, $0 < \sigma_{nu}^2 := Eu_{n,t}^2 < \infty$, $0 < \sigma_{nr}^2 := Er_{n,t}^2 < \infty$, and $|\mathrm{Corr}(u_{n,t}, v_{n,t})| < 1$, for $n = 1, \ldots, N$. With $r_{n,t} = u_{n,t} - \frac{Ev_{n,t}u_{n,t}}{Ev_{n,t}^2} v_{n,t}$, we suppose that $\mathrm{Corr}(v_{n,t}, r_{m,t}) = 0$ for all n, m. Moreover, we assume that

(i) $E|u_{n,1}|^{c+\varepsilon} + E|v_{n,1}|^{c+\varepsilon} < \infty$, for $n = 1, \ldots, N$ and some $c > 2$ and $\varepsilon > 0$,

(ii) $\sum_{n=1}^{\infty} \alpha(n)^{1-2/c} < \infty$, where $\{\alpha(n)\}$ are the coefficients of strong regularity of the process $\{(u_{1,t}, v_{1,t}, \ldots, u_{N,t}, v_{N,t}), t \in \mathbb{Z}\}$.

Proposition 3.1. *Suppose that (A2) is fulfilled. Moreover we assume that $y_{n,t}$ and $z_{n,t}$ are cointegrated with cointegrating rank 1 for all $n = 1, \ldots, N$.*

(i) *Under $H_0^{(1)}$ and $\alpha_{n2} = 0$ for all $n = 1, \ldots, N$:*

(i.a) $\nu_2 = 0$

Let $(B_{1v}, \ldots, B_{Nv})'$ and $(B_{1r}, \ldots, B_{Nr})'$ be independent N-dimensional Wiener processes with covariances $Cov((v_{1,t}, \ldots, v_{N,t})')$ and $Cov((r_{1,t}, \ldots, r_{N,t})')$, respectively,

and let $\sigma_{nr}^2 = var(r_{n,t})$. *Then*

$$LR_N^{(1)} \xrightarrow{d} \sum_{n=1}^{N} \frac{1}{\sigma_{nr}^2} \frac{\left\{\int_0^1 \left[B_{nv}(t) - \int_0^1 B_{nv}(s)\,ds\right] dB_{nr}(t)\right\}^2}{\int_0^1 \left[B_{nv}(t) - \int_0^1 B_{nv}(s)\,ds\right]^2 dt}.$$

(i.b) $\nu_2 \neq 0$

$$LR_N^{(1)} \xrightarrow{d} \sum_{n=1}^{N} \frac{1}{\sigma_{nr}^2} Z_n^2,$$

where

$$(Z_1, \ldots, Z_N)' \sim \mathcal{N}(0, Cov((r_{1,t}, \ldots, r_{N,t})').$$

(ii) *Under* $H_0^{(2)}$*:*

$$LR_N^{(2)} \xrightarrow{d} \sum_{n=1}^{N} Z_n^2,$$

where

$$(Z_1, \ldots, Z_N)' \sim Diag[\frac{1}{\sqrt{Eq_{1,t-1}^2 \sigma_{1,v}^2}}, \ldots, \frac{1}{\sqrt{Eq_{N,t-1}^2 \sigma_{N,v}^2}}]\mathcal{N}(0, Cov((q_{1,t-1}v_{1,t}, \ldots, q_{N,t-1}v_{N,t})')).$$

(iii) *Under* $H_0^{(3)}$ *and* $\alpha_{n2} = 0$ *for all* $n = 1, \ldots, N$*:*

$$LR_N^{(3)} \xrightarrow{d} \sum_{n=1}^{N} Z_n^2,$$

where

$$(Z_1, \ldots, Z_N)' \sim Diag[\frac{1}{\sqrt{Ew_{1,t-1}^2 \sigma_{1,u}^2}}, \ldots, \frac{1}{\sqrt{Ew_{N,t-1}^2 \sigma_{N,u}^2}}]\mathcal{N}(0, Cov((w_{1,t-1}u_{1,t}, \ldots, w_{N,t-1}u_{N,t})'))$$

and $w_{n,t} = \widetilde{\Delta y_{n,t-1}} - \frac{Eq_{n,t-1}\widetilde{\Delta y_{n,t-1}}}{Eq_{n,t-1}^2} q_{n,t-1}$.

In all cases, the limit distributions of the test statistics depend on a number of specific properties of the time series. This clearly motivates the application of the bootstrap for determining appropriate critical values.

4. The wild bootstrap

It follows from Propositions 2.1 to 2.3 that the LR-statistics are pivotal if $\{(u_t, v_t)'\}$ forms a sequence of independent random vectors. However, Propositions 2.2 and 2.3 show that, e.g. in the presence of autoregressive conditionally heteroskedastic error sequences as introduced by Engle (1982), the LR-statistics are not pivotal in general. Note that in this case the random variables entering the factor to be multiplied with a $\chi^2(1)$ distribution fail to be independent. Hence, taking the critical value from the $\chi^2(1)$-distribution leads to a serious size distortion, which does not even vanish as $T \to \infty$.

Since the asymptotic null distributions derived in Propositions 2.2 and 2.3 depend on some nuisance parameters it is natural to apply an appropriate bootstrap method for the determination of the critical value. Moreover, although this is not visible from our first-order asymptotic results, we hope to get a better finite sample performance in comparison to the approach based on the limit distribution. In a different but related context, Herwartz (1998) obtained empirical results which showed the superiority of the bootstrap over standard asymptotics.

In the present context, one motivation for using the bootstrap is that we have to account for possible heteroskedasticity of the innovations which affects first-order asymptotic properties of two of our test statistics. One important feature of all three test statistics is that they can be approximated by squared sums of martingale differences. Therefore, the so-called wild or external bootstrap seems to be a natural choice. It goes back to a proposal of Wu (1986) and was recognized as a method fitting in the general concept of bootstrap by Beran and Efron in the discussion of Wu's paper. The validity of this particular resampling scheme has already been investigated in many papers and for statistics of different types. For example, Mammen (1993) showed its validity for F-type statistics in parametric regression models with random explanatory variables. Neumann and Kreiss (1998) showed that the validity of regression-type bootstrap procedures is often maintained for autoregressive models. The basic reason why we can neglect the dependence structure with our bootstrap method is that the random variables u_t and r_t are uncorrelated. The neglibility of the dependence is quite obvious in the case of linear statistics, and can also be shown for quadratic forms.

4.1. **Bootstrap inference in single equations.** Since the reader may not be that familiar with the wild bootstrap we describe this method in detail. We begin with that variant needed for the first test.

According to (2.17), it is enough to resample the random variables $r_1, \ldots, r_T$. Since it suffices to mimic the null distribution of $\mathrm{LR}^{(1)}$ in the case that the null hypothesis is actually true, we define approximations to the r_t as

$$(\widehat{r}_1, \ldots, \widehat{r}_T)' \;=\; (I_T \;-\; \mathcal{X}_0(\mathcal{X}_0'\mathcal{X}_0)^{-1}\mathcal{X}_0')\Delta y, \tag{4.1}$$

where, as above, $\mathcal{X}_0 = (1_T \quad q_{-1}^0 \quad \Delta z)$. Under $H_0^{(1)}$ we obtain that

$$(\widehat{r}_1, \ldots, \widehat{r}_T)' \;=\; (I_T \;-\; \mathcal{X}_0(\mathcal{X}_0'\mathcal{X}_0)^{-1}\mathcal{X}_0')u \;=\; (I_T \;-\; \mathcal{X}_0(\mathcal{X}_0'\mathcal{X}_0)^{-1}\mathcal{X}_0')r \tag{4.2}$$

Then we define bootstrap versions of the r_t with mean zero by matching their low order moments to the corresponding powers of the $\widehat{r}_t$. In our context, it suffices to mimic the first two moments of r_t in an asymptotically unbiased manner which is achieved by setting

$$r_t^* \;=\; \eta_t \widehat{r}_t, \tag{4.3}$$

where $\{\eta_t\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, also independent of the variables occurring in (2.1) and (2.2). Quite often also the third moments are matched which actually leads to better rates of convergence in the case of a studentized sample mean (see, e.g. Mammen 1992). Because of the special structure of $\mathrm{LR}^{(1)}$ it is not necessary to generate bootstrap versions Δy_t^* and Δz_t^* of Δy_t and Δz_t^*, respectively.

According to (2.10), the wild bootstrap version of $\mathrm{LR}^{(1)}$ is then

$$\mathrm{LR}^{(1),*} \;=\; T \ln\left(\frac{\mathrm{RSS}_0^{(1),*}}{\mathrm{RSS}_1^{(1),*}}\right), \tag{4.4}$$

where $\mathrm{RSS}_i^{(1),*} = (r^*)'(I_T - \mathcal{X}_i(\mathcal{X}_i'\mathcal{X}_i)^{-1}\mathcal{X}_i)r^*$ and $r^* = (r_1^*, \ldots, r_T^*)'$. The critical value for the test statistic $\mathrm{LR}^{(1)}$ is now obtained as the corresponding quantile of $\mathrm{LR}^{(1),*}$. The following lemma asserts that the bootstrap approximation is consistent.

Proposition 4.1. *Suppose that Assumption (A1) is fulfilled and that* $\alpha_2 = 0$. *Then, under* $H_0^{(1)}$,

$$\sup_{-\infty<z<\infty} \left\{ \left| P\left(LR^{(1),*} \leq z \mid \mathcal{X} \right) - P\left(LR^{(1)} \leq z \right) \right| \right\} \xrightarrow{P} 0.$$

Remark 1. At first sight, it seems to be advisable to define approximations to the innovations in the full model since they may be otherwise far away from their true values under the alternative. This can actually often lead to a bad power of a test since the critical value is then often too large under the alternative. This problem does not emerge in our case since $\mathrm{LR}^{(1)}$ is a function of a ratio of two quadratic forms in u. Hence, even a serious overestimate of the variances of the u_t does not matter much. Moreover, we experienced in our simulations a considerably worse size of the test if the critical values were taken from a bootstrap version $\mathrm{LR}^{(1),*}$ based on residuals of the full model.

In the cases of the other two tests we proceed similarly. The only difference is that for the second test we have to imitate the v_t. To this end, we approximate first the unobserved innovations v_t by least squares residuals in model (2.22), that is, we define

$$(\widehat{v}_1, \ldots, \widehat{v}_T)' = (I_T - 1_T(1_T'1_T)^{-1}1_T')\Delta z. \tag{4.5}$$

Under $H_0^{(2)}$, we have

$$(\widehat{v}_1, \ldots, \widehat{v}_T)' = (I_T - T^{-1}1_T1_T')v. \tag{4.6}$$

Then we generate bootstrap counterparts as

$$v_t^* = \eta_t \widehat{v}_t, \tag{4.7}$$

where $\{\eta_t\}$ is chosen as above. Finally, we define $\mathrm{LR}^{(2),*}$ according to (2.26), where only the v_t are replaced by their bootstrap versions v_t^*. A critical value for the test is again obtained as the corresponding quantile of $\mathrm{LR}^{(2),*}$.

Proposition 4.2. *Suppose that Assumption (A1) is fulfilled. Then, under* $H_0^{(2)}$,

$$\sup_{-\infty<z<\infty} \left\{ \left| P\left(LR^{(2),*} \leq z \mid \mathcal{X}\right) - P\left(LR^{(2)} \leq z\right) \right| \right\} \xrightarrow{P} 0.$$

A bootstrap approximation to $\mathrm{LR}^{(3)}$ is obtained by replacing the u_t in (2.30) by the r_t^* defined in (4.3).

Proposition 4.3. *Suppose that Assumption (A1) is fulfilled and that* $\alpha_2 = 0$. *Then, under* $H_0^{(3)}$,

$$\sup_{-\infty<z<\infty} \left\{ \left| P\left(LR^{(3),*} \leq z \mid \mathcal{X}\right) - P\left(LR^{(3)} \leq z\right) \right| \right\} \xrightarrow{P} 0.$$

Now it follows immediately from the Propositions 4.1 to 4.3 that the bootstrap-based tests have asymptotically the prescribed size.

4.2. **Bootstrap inference in pooled systems of equations.** We assume now that a system of time series $\{y_{n,t}\}$ and $\{z_{n,t}\}$, $n = 1, \ldots, N$, obeying (3.1) and (3.2) is observed. First we consider a bootstrap procedure for $\mathrm{LR}_{[N]}^{(1)}$. We intend to apply again a simple wild bootstrap procedure, that is, only the $r_{n,t}$ $(t = 1, \ldots, T,\ n = 1, \ldots, N)$ will be resampled. In order to mimic the joint distribution of $\mathrm{LR}_{[1]}^{(1)}, \ldots, \mathrm{LR}_{[N]}^{(1)}$, we intend to generate bootstrap vectors $(r_{1,t}^*, \ldots, r_{N,t}^*)'$ with a covariance structure similar to that of $(r_{1,t}, \ldots, r_{N,t})'$. To this end, we define in complete analogy to (4.1) first approximations to the $r_{n,t}$ as

$$(\widehat{r}_{n,1}, \ldots, \widehat{r}_{n,T})' = \left(I_T - \mathcal{X}_{n0}(\mathcal{X}_{n0}'\mathcal{X}_{n0})^{-1}\mathcal{X}_{n0}'\right) \Delta y_n, \tag{4.8}$$

where $\mathcal{X}_{n0} = (1_T \ \ q_{n,-1}^0 \ \ \Delta z_n)$. Under $H_0^{(1)}$ it holds

$$(\widehat{r}_{n,1}, \ldots, \widehat{r}_{n,T})' = \left(I_T - \mathcal{X}_{n0}(\mathcal{X}_{n0}'\mathcal{X}_{n0})^{-1}\mathcal{X}_{n0}'\right) u_n = \left(I_T - \mathcal{X}_{n0}(\mathcal{X}_{n0}'\mathcal{X}_{n0})^{-1}\mathcal{X}_{n0}'\right) r_n, \tag{4.9}$$

where $u_n = (u_{n,1}, \ldots, u_{n,T})'$ and $r_n = (r_{n,1}, \ldots, r_{n,T})'$. We take as a bootstrap version of $(r_{1,t}, \ldots, r_{N,t})'$ the quantity

$$\left(r_{1,t}^*, \ldots, r_{N,t}^*\right)' = \eta_t \left(\widehat{r}_{1,t}, \ldots, \widehat{r}_{N,t}\right)', \tag{4.10}$$

where $\{\eta_t\}$ is again a sequence of independent and identically distributed random variables with zero mean and unit variance, also independent of the random variables occurring in (3.1) and (3.2). The basic reason why this simple wild bootstrap works for vector-valued random variables can be seen from the relation

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{Cov}(r_t^*) \;=\; \frac{1}{T}\sum_{t=1}^{T}\widehat{r}_t\widehat{r}_t' \;=\; \frac{1}{T}\sum_{t=1}^{T}r_t r_t' \;+\; o_P(1) \;=\; \frac{1}{T}\sum_{t=1}^{T}\mathrm{Cov}(r_t) \;+\; o_P(1), \tag{4.11}$$

that is, the arithmetic mean of the covariance matrices is asymptotically reflected by the bootstrap. It is clear from (4.11) that this also applies in the case of time-varying covariances.

Now we obtain a bootstrap version of $\mathrm{LR}_{[N]}^{(1)}$ as

$$\mathrm{LR}_{[N]}^{(1),*} \;=\; T\sum_{n=1}^{N}\ln\left(\frac{\mathrm{RSS}_{n0}^{(1),*}}{\mathrm{RSS}_{n1}^{(1),*}}\right), \tag{4.12}$$

where $\mathrm{RSS}_{ni}^{(1),*} = (r_n^*)'(I_T - \mathcal{X}_{ni}(\mathcal{X}_{ni}'\mathcal{X}_{ni})^{-1}\mathcal{X}_{ni}')r_n^*$ and $r_n^* = (r_{n,1}^*, \ldots, r_{n,T}^*)'$. The critical value of the test statistic is given by the corresponding quantile of $\mathrm{LR}_{[N]}^{(1),*}$.

In sharp contrast to the case of a single system of equations, we will see that, in the case of $\nu_2 = 0$, the distribution of $\mathrm{LR}_{[N]}^{(1),*}$ does not necessarily approximate the unconditional distribution of $\mathrm{LR}_{[N]}^{(1)}$ consistently. Nevertheless, we can show that the bootstrap guarantees an error of the first kind tending to the prescribed value. The reason for this is that some *conditional* distribution of an asymptotic approximation to $\mathrm{LR}_{[N]}^{(1)}$ is mimicked by the bootstrap. (We cannot show without further regularity conditions that this also applies to the corresponding conditional distribution of $\mathrm{LR}_{[N]}^{(1)}$ since weak convergence to a "regular" random variable does not exclude that certain conditional distributions of $\mathrm{LR}_{[N]}^{(1)}$ behave somehow irregularly; see also the proof of Theorem 4.1.)

For the second test, we generate bootstrap versions of the vectors $(v_{1,t}, \ldots, v_{N,t})'$ as

$$\left(v_{1,t}^*, \ldots, v_{N,t}^*\right)' \;=\; \eta_t\,(\widehat{v}_{1,t}, \ldots, \widehat{v}_{N,t})', \tag{4.13}$$

where $\{\eta_t\}$ is again a sequence of i.i.d. random variables with zero mean and unit variance, also independent of the random variables occurring in (3.1) and (3.2). Approximations $\widehat{v}_{n,t}$ to the $v_{n,t}$ are obtained as

$$(\widehat{v}_{n,1}, \ldots, \widehat{v}_{n,T})' \;=\; \left(I_T \;-\; 1_T(1_T'1_T)^{-1}1_T'\right)\Delta z_n. \tag{4.14}$$

A bootstrap version of $\mathrm{LR}_{[N]}^{(2)}$ is now given as

$$\mathrm{LR}_{[N]}^{(2),*} = T \sum_{n=1}^{N} \ln \left(\frac{\mathrm{RSS}_{n0}^{(2),*}}{\mathrm{RSS}_{n1}^{(2),*}} \right), \tag{4.15}$$

where $\mathrm{RSS}_{ni}^{(2),*} = (v_n^*)'(I_T - \mathcal{Z}_{ni}(\mathcal{Z}_{ni}'\mathcal{Z}_{ni})^{-1}\mathcal{Z}_{ni}')v_n^*$, $\mathcal{Z}_{n0} = 1_T$, $\mathcal{Z}_{n1} = (y_{n,-1} + \widehat{\beta} z_{n,-1} \;\; 1_T)$ and $\widehat{\beta}$ is the least squares estimate according to (2.24).

Finally, for the third test, we can again use the $r_{n,t}^*$ defined by (4.10). This leads to a bootstrap counterpart of $\mathrm{LR}_{[N]}^{(3)}$ given by

$$\mathrm{LR}_{[N]}^{(3),*} = T \sum_{n=1}^{N} \ln \left(\frac{\mathrm{RSS}_{n0}^{(3),*}}{\mathrm{RSS}_{n1}^{(3),*}} \right), \tag{4.16}$$

where $\mathrm{RSS}_{ni}^{(3),*} = (r_n^*)'(I_T - \mathcal{V}_{ni}(\mathcal{V}_{ni}'\mathcal{V}_{ni})^{-1}\mathcal{V}_{ni}')r_n^*$ with $\mathcal{V}_{n0} = (1_T \;\; q_{n,-1} \;\; \Delta z_n)$ and $\mathcal{V}_{n1} = (1_T \;\; q_{n,-1} \;\; \Delta z_n \;\; \Delta y_{n,-1})$.

The following theorem asserts that the bootstrap-based tests have asymptotically the prescribed size.

Theorem 4.1. *Suppose that the conditions of Proposition 3.1 are fulfilled. Let $t_{N,\alpha,i}^*$ be the $(1-\alpha)$-quantile of $LR_{[N]}^{(i),*}$. Then, under $H_0^{(i)}$ $(i = 1, 2, 3)$:*

$$P\left(LR_{[N]}^{(i)} > t_{N,\alpha,i}^*\right) \longrightarrow \alpha.$$

5. Simulation Study

5.1. Monte Carlo Design.

Introductory remarks. All test procedures introduced above are justified by arguments from asymptotic theory. Thus it might be interesting to characterize their performance in finite samples and pooled systems of them. The following simulation study is mainly designed to shed some light on the performance of our tests in small samples ($T = 25, 50$). In addition, we compare two approaches to obtain critical values for a particular test statistic and try to evaluate their effects on empirical size estimates. First, we obtain critical values from a first order asymptotic approximation using the χ^2-distribution. Second, we use the wild bootstrap to mimic the distribution of the employed test statistic under the null hypothesis. Note that applying first order asymptotics may be invalid if the underlying error terms

are conditionally heteroskedastic (testing on weak exogeneity or on significance of short-run dynamics). Therefore our study also supports the analyst in evaluating the actual significance level of hypothesis tests if diagnostic checking of an empirical model yields a rejection of the homoskedasticity assumption.

Within single equation models we investigate the empirical properties of LR-type tests for restrictions on

(i) the long-run parameter in the conditional model (2.1) ($H_0^{(1)} : \beta = \beta_0$)

(ii) the error correction coefficient in the marginal model (2.2) ($H_0^{(2)} : \alpha_2 = 0$)

(iii) short-run dynamics within the conditional model (2.28) ($H_0^{(3)} : \gamma_2 = 0$)

For the first and third testing problem we also computed the WALD-statistic which is widely used in empirical practice. Note that the second issue is of importance with respect to the underlying economic model as well as the employed econometric specification. Inference on long-run parameters in single equations depends crucially on the assumption of weak exogeneity of conditioning variables or, put differently, on the acceptance of $H_0^{(2)} : \alpha_2 = 0$. The LR–statistic employed to test the hypothesis of weak exogeneity may also be used to infer against $H_0 : \alpha_1 = 0$ in the conditional model thus providing an implicit test of the cointegration hypothesis. Rejection frequencies for the null hypothesis $H_0 : \alpha_1 = 0$ and, similarly, power estimates for the third testing problem given above are not provided in the following in order to economize on space.

Apart from an investigation of the empirical properties of competing inference procedures in single equations the simulation study also addresses the issue of inference on the level of pooled equations. Inference in pooled systems is discussed considering the first and second testing problem listed above. As mentioned the LR-statistic is a convenient test statistic for inference in systems of equations since the asymptotic χ^2-distribution can be pooled easily. Note, however, that for a system of pooled equations a first order asymptotic approximation is only available in absence of cross equation error correlation.

The simulated processes. To investigate the empirical properties of inference procedures in single equations we simulated the following stochastic processes:

$$y_t = \phi_1 y_{t-1} + \phi_2 z_t + \varepsilon_{1t}, \tag{5.1}$$

$$z_t = z_{t-1} + \varepsilon_{2t}, \tag{5.2}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim N(0, I_2)$, and I_2 denotes the 2×2 identity matrix. Obviously z_t is generated as a random walk without drift. Moreover, y_t and z_t are cointegrated with cointegrating parameter $\beta = \phi_2/(\phi_1 - 1)$. The DGP in (5.1) may also be respecified in terms of an error correction equation like (2.1):

$$\Delta y_t = \nu_1 + \alpha_1(y_{t-1} + \beta z_{t-1}) + \gamma_1 \Delta z_t + u_t. \tag{5.3}$$

Comparing the specifications in (5.1) and (5.3) we arrive at the following parameter restrictions relating the empirical and the true model: $\nu_1 = 0$, $\alpha_1 = \phi_1 - 1$, $\beta = \phi_2/(\phi_1 - 1)$, $\gamma_1 = \phi_2$, and $u_t = \varepsilon_{1t}$. As given in (5.2) it is obvious that Δz_t is not affected by violations of the long-run equilibrium relation. The corresponding error correction parameter in the empirical counterpart (2.2) is actually zero ($\alpha_2 = 0$). Thus z_t is weakly exogenous for inference involving the parameters β or α_1.

In order to characterize the DGP in (5.1) and (5.2) somewhat deeper it is convenient to provide the corresponding bivariate model. We obtain the following representation which is equivalent to (5.1) and (5.2):

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} 1 & -\phi_2 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \phi_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}, \tag{5.4}$$

where $\epsilon_t \sim N(0, \Sigma_\epsilon)$, and

$$\Sigma_\epsilon = \begin{pmatrix} 1 + \phi_2^2 & \phi_2 \\ \phi_2 & 1 \end{pmatrix}.$$

The two characteristic roots of the process are found to be $\rho_1 = 1$ and $\rho_2 = 1/\phi_1$. The process always contains a unit root. The second root depends on the particular choice of ϕ_1. If $|\phi_1| > 1$ the process is explosive since then ρ_2 is located within the complex unit circle. Values of ϕ_1 smaller than but close to 1 provide a process with the second root coming close to the unit circle. Note that such a model generates only weak error correction effects and its dynamics should be similar to those of a bivariate process specified only in terms of stationary differences of y_t and z_t. Such a similarity may be even more important in the case where only small samples of the investigated variables are available.

Empirical properties of inference procedures involving the long-run coefficient β are investigated for the particular null hypothesis $H_0^{(1)} : \beta = -1$, which has become popular as

the so-called homogeneity hypothesis. To allow for different roots of the true DGP we simulated 11 alternative processes. The true error correction coefficient α_1 varied between -0.01 and -0.99, indicating borderline cases of weak and strong error correcting dynamics. The parameter $\phi_1 = \alpha_1 + 1$ in (5.1) was correspondingly chosen as follows:

$$\phi_1 = 0.99,\ 0.97,\ 0.95,\ 0.90,\ 0.70,\ 0.50,\ 0.30,\ 0.10,\ 0.05,\ 0.03,\ 0.01.$$

Since $\beta = \phi_2/(\phi_1 - 1)$ size properties for testing the homogeneity assumption can be conveniently analyzed by choosing ϕ_2 in (5.1) as $\phi_2 = (1 - \phi_1)$. To investigate empirical power properties of testing against homogeneity we employ two classes of DGPs. First, we chose $\phi_2 = 1.1(1 - \phi_1)$ obtaining an implicit long-run parameter $\beta = -1.1$. Secondly and complementary to the sequence of underlying DGPs we used DGPs assuming $\phi_1 = (1 + \alpha_1) = 0.5$ to be given and varied the long-run parameter across the interval $-0.9 \geq \beta \geq -1.1$.

The DGPs detailed above are also used to test the null hypothesis $H_0^{(2)} : \alpha_2 = 0$ in the marginal equation corresponding to the conditional model given in (5.3). Finally the processes given above are employed to infer against significance of parameters governing short-run dynamics. For this purpose we employ the augmented ECM

$$\Delta y_t = \nu_1 + \alpha_1(y_{t-1} + \beta z_{t-1}) + \gamma_1 \Delta z_t + \gamma_2 \Delta y_{t-1} + u_t \tag{5.5}$$

and consider a test of the hypothesis $H_0^{(3)} : \gamma_2 = 0$. Note that testing such a hypothesis is of interest for the specification of short-run dynamics.

As pointed out above the LR-statistics designed to test on weak exogeneity and insignificance of short-run dynamics are no longer pivotal if the underlying error terms fail to be homoskedastic. With respect to these two testing problems we are thus interested in characterizing the dependence of empirical size estimates on the error distribution of the true model. Note that in the presence of heteroskedastic error distributions only the wild bootstrap is supposed to provide suitable critical values.

Heteroskedastic error distributions. Generating the time series processes given above we distinguish two alternative error distributions. Alternatively to a Gaussian distribution with zero mean and unit covariance matrix, $\varepsilon_t \sim N(0, I_2)$, we sample the bivariate error terms from a multivariate GARCH-process,

$$\varepsilon_t = \Sigma_t^{1/2} \xi_t,\ \xi \sim N(0, I_2),$$

in order to investigate to what extent our simulation results are affected in presence of conditional heteroskedasticity. The particular multivariate GARCH-process employed for the simulations can be given in the so-called BEKK-representation, i.e.

$$\Sigma_t = C_0' C_0 + A' \varepsilon_{t-1} \varepsilon_{t-1}' A + G' \Sigma_{t-1} G. \tag{5.6}$$

In (5.6) C_0 is an upper triangular matrix determining deterministic variance components. A and G are 2×2 parameter matrices generating clusters of volatility which may be regarded as a stylized fact of time series observed e.g. on financial markets. For a discussion of the multivariate GARCH-model and its BEKK-representation see e.g. Engle and Kroner (1995).

The particular parameterization used in the simulation study has been found in Herwartz and Lütkepohl (2000) to describe volatility dynamics of a bivariate series of stock prices. The parameter matrices are chosen as follows:

$$C_0 = \begin{pmatrix} 0.0021 & 0.0015 \\ 0 & 0.0029 \end{pmatrix}, \quad A = \begin{pmatrix} 0.375 & 0.126 \\ -0.146 & 0.138 \end{pmatrix}, \quad G = \begin{pmatrix} 0.748 & -0.172 \\ 0.232 & 1.106 \end{pmatrix}.$$

After the generation step both univariate error processes ε_{1t} and ε_{2t} are standardized to have an unconditional variance of one in order to mitigate the impact of outstanding factors on the comparison of our results.

Simulating systems of pooled equations. As mentioned we also investigate the empirical performance of competing inference procedures applied to a set of pooled single equations. For simulation purposes we regard a system consisting of two single equation ECMs, i.e. $N = 2$. In the following nonstationary variables are denoted as $y_{n,t}$, $z_{n,t}, n = 1, 2$. Parameters governing dynamics relevant for particular members of the panel $\phi_{ni}, n = 1, 2$, are accordingly denoted with two indices, the first of which indicates the member of the panel. Stacking the nonstationary variables in a 4×1 column vector $\mathbf{y}_t = (y_{1t}, z_{1t}, y_{2t}, z_{2t})'$ a joint DGP for these variables, similar to (5.4), may be given as follows:

$$\mathbf{y}_t = \Gamma \mathbf{y}_{t-1} + \epsilon_t, \tag{5.7}$$

where

$$\Gamma = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \phi_{21} & \phi_{22} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \Sigma_\epsilon = \begin{pmatrix} \phi_{12}^2 + 1 & \phi_{12} & \rho & 0 \\ \phi_{12} & 1 & 0 & 0 \\ \rho & 0 & \phi_{22}^2 + 1 & \phi_{22} \\ 0 & 0 & \phi_{22} & 1 \end{pmatrix}.$$

The model in (5.7) is essentially a reduced form of the joint DGP of the involved time series variables. The covariance matrix Σ_ϵ is convenient to introduce cross sectional correlation, by selecting $\rho \neq 0$. Introducing such correlation is sensible to shed light on the dependence of LR-inference on cross equation dynamics. As outlined above first order asymptotic approximations of the distribution of the $\mathrm{LR}^{(\cdot)}_{[N]}$ statistics require innovations which are independent across equations. Applying bootstrap schemes critical values of $\mathrm{LR}^{(\cdot)}_{[N]}$ statistics can be estimated even in the case of cross sectional correlation. For the simulation of pooled systems we choose $\rho = 0.5$ throughout. The first bivariate process $(y_{1t}, z_{1t})'$ was generated using parameter values $\phi_{11} = \phi_{12} = 0.5$. The remaining parameter values ϕ_{21} and ϕ_{22} were varied similarly as ϕ_1 and ϕ_2 for the bivariate processes described before.

Final remarks. Monte Carlo experiments are in most cases performed for small sample sizes $T = 25$ and $T = 50$. We generated each time series process $Q = 2000$ times. Empirical size estimates $(\hat{\alpha})$ and the nominal size (α) of test procedures are regarded to differ significantly at the 5% level if the former fall into a critical region constructed around the latter. Formally this region is identical to the complement of the interval $(\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/2000})$. To evaluate critical values by means of the wild bootstrap we use $R = 500$ replications of each generated time series. All tests are performed at alternative nominal levels $\alpha = 0.05$ and $\alpha = 0.10$. Since the obtained results are analogous to interpret we provide only empirical results for the 5% significance level. All computations were implemented in GAUSS 3.2.7 (Unix version).

5.2. **Monte Carlo Results.**

Testing long-run relationships. Consider first the null hypothesis $H_0^{(1)} : \beta = -1$. In Figure 1 (left hand side panels) size estimates for the LR- and WALD-test (W) of this hypothesis are displayed. For these statistics critical values are obtained from the $\chi^2(1)$- and $F(1, T-K)$-distribution, respectively. In addition, rejection frequencies for the LR-statistic are given for the case where critical values are generated by means of the wild bootstrap (LR*). To facilitate the interpretation of the results all panels display an upper bound for statistical equivalence of the empirical and the assumed nominal significance level.

A few results are immediately obtained: We observe considerable size distortions for all test procedures if the underlying true error correction coefficient is small in absolute value. For $\alpha_1 = -0.1$ and $T = 50$ we obtain rejection frequencies which are up to three times larger than the nominal level of the tests. It should be noted that to obtain empirical size estimates which are close to the nominal size for these processes ($\alpha_1 \geq -0.1$) it is necessary to have samples of 500 observations and even more. Detailed results on this point are not provided here to economize on space. The largest violations of the nominal level of the test procedures are displayed for the LR-statistic with critical values taken from the $\chi^2(1)$–distribution. Computing critical values by means of the wild bootstrap improves the empirical performance of the LR-test considerably. For a few values of α_1 close to -1 bootstrap inference obtains empirical size estimates which cannot be distinguished statistically from the employed nominal level even for $T = 25$. For the smallest sample size considered bootstrap inference is preferable even in comparison to the WALD-statistic. This result underlines the convenience of the bootstrap method in small samples since the critical values used for the WALD-statistic also depend on the sample size. Turning to larger sample sizes bootstrap inference and the WALD-test perform similarly.

The right hand side panels of Figure 1 show size estimates for the homogeneity hypothesis tested in systems of $N = 2$ pooled equations. Now we concentrate on inference by means of the LR-statistic. We provide size estimates for LR and LR*. Since the variables in the first data set are generated with $\alpha_{11} = -0.5$ it is not surprising to find significant violations of the nominal level for both test procedures and all generated processes in small samples ($T = 25$). Note that we already observed considerable size distortions for the single equation methods applied to the process generated with $\alpha_1 = -0.5$. Inference by

means of the $\chi^2(2)$-distribution is clearly inferior compared to bootstrap procedures. If the underlying α_{21} parameter is sufficiently small, i.e. $\alpha_{21} \leq -0.7$, the wild bootstrap provides empirical size estimates which cannot be distinguished from their nominal counterparts. As mentioned the $\chi^2(2)$-distribution provides only valid critical values in absence of cross sectional correlation. Therefore we conjecture that our resampling scheme becomes even more fruitful in the presence of stronger contemporaneous correlation across equations. As a final advantage of cross correlation consistent resampling note that the advocated procedure does not require any first step estimation of the involved error covariance matrix. Note that such a matrix is generally of dimension $(N \times N)$. Thus in empirical panel investigations reliable covariance estimates neccessary for feasible GLS-methods may be hardly available.

Figure 2 displays single equation $(N = 1)$ power estimates obtained for samples of size $T = 50$. Comparing the power properties of the investigated testing procedures mirrors the foregoing discussion of size estimates. Rejection frequencies of the homogeneity assumption for processes with $\beta = -1.1$ are similar to the corresponding size estimates if the underlying error correction coefficient is close to zero. The power curves displayed for given $\beta = -1.1$ show that the power of the competing procedures increases with $-\alpha_1$ given that size distortions are not too severe ($\alpha_1 < -0.3$, say). For given $\alpha_1 = -0.5$ we obtain a U-shaped power curve centered at $\beta = -1$. It appears that the slope of this power curve is slightly steeper for $\beta > -1$ compared to $\beta < -1$. Throughout we observe that the LR*-test and the WALD-statistic have less power compared to the LR-test employing $\chi^2(1)$-critical values. This result can be directly related to the poor size properties of the latter procedure discussed above.

Testing on weak exogeneity. Figures 3 provides size estimates obtained from inferring on weak exogeneity by means of the LR- and LR*–statistics. The upper and middle panels are analogous to Figure 1, i.e. inference exercises in single equations and pooled systems are distinguished. Since for this testing problem the performance of standard test procedures may suffer from heteroskedastic error distributions the lower panels of Figure 3 show rejection frequencies obtained for the LR- and LR*-test applied to single equation models $(N = 1)$ in presence of conditionally heteroskedastic error distributions. To uncover the impact of heteroskedasticity we distinguish sample sizes $T = 50$ (lower left hand side panel) and $T = 200$ (lower right). Similar to the results discussed for the homogeneity hypothesis size

distortions observed for the hypothesis $H_0^{(2)} : \alpha_2 = 0$ are most severe for processes with small error correcting dynamics. Generating critical values by means of a bootstrap procedure improves the empirical size properties of the LR-test. It should be noted that the LR-statistic employed here is not the standard test statistic to infer against weak exogeneity. Regarding the standard $t-$ratio we would obtain results very close to the performance of the bootstrap procedure in single equation models. Bootstrap inference turns out to be particularly helpful if we are concerned with a pooled system of marginal processes. Considering e.g. the case of pooling 2 sets of $T = 50$ observations we obtain that the application of first order asymptotic approximations involves significant violations of the nominal size for all processes with $\alpha_{21} > -0.9$. Using the wild bootstrap to provide critical values significant violations of the nominal level are observed for a smaller set of time series processes $\alpha_{21} > -0.5$. Note that the simulated processes are free of contemporaneous correlation across marginal equations. In the presence of underlying heteroskedastic error distributions we obtain a further strong recommendation of bootstrap inference. In this case the first order asymptotic approximation fails to provide valid critical values. Compared to the nominal level the hypothesis of weak exogeneity is too often rejected and the observed size distortions do not vanish with increasing sample size T. Comparing the lower panels of Figure 3 it is seen that computing critical values by means of the wild bootstrap becomes more suitable with increasing sample sizes. Even in smaller samples $T = 50$ bootstrap inference shows empirical size estimates which are sufficiently close to the nominal values for a few DGPs ($\alpha_{21} \leq -0.7$).

Testing short-run dynamics. Finally we turn to hypothesis tests on parameters governing short-run dynamics as e.g. $H_0^{(3)} : \gamma_2 = 0$ in the augmented regression model (5.5). In the presence of homoskedastic error terms wild bootstrap inference is characterized by size estimates which are close to the properties of the WALD-statistic. Taking critical values of the LR-statistic from the $\chi^2(1)$-distribution involves considerable size distortions in small samples. To economize on space we refrain from showing detailed results on this point. For the case of heteroskedastic error distributions Figure 4 provides size estimates for the three tests which were already used to test the homogeneity hypothesis, namely LR, W, and LR*. These results may be read analogously to those reported in Herwartz (1998) for stationary time series processes. Due to heteroskedastic errors terms the WALD- and LR-statistic are no longer pivotal. Ignoring this effect and still taking critical values from

the $F(1, T-k)$- and $\chi^2(1)$-distribution, respectively, involves considerable violations of the nominal test level. Bootstrapping the LR-statistic we obtain empirical size estimates that cannot be distinguished from their nominal counterparts in the majority of considered time series processes. In addition, inference on short-run dynamics appears to be unaffected by the error correction coefficient of the true DGP.

6. Empirical scope of bootstrapping pooled ECMs - Conclusions and Outlook

The empirical relevance of bootstrap inference as outlined above reflects directly the practical scope of its key ingredients, namely EC-modelling, panel time series regression and resampling. Illuminating the importance of the ECM approach Hendry (1995) observes that most UK macro econometric models are versions of the ECM. In comparison to an asymptotically equivalent ML-analysis of a corresponding vector autoregressive model single equations are more feasible since standard OLS-procedures apply. The literature on (non-stationary) panel analysis is a further growing area of research as recently pointed out by Phillips and Moon (1999). From an economic viewpoint panel cointegration models are a natural framework to investigate issues like homogeneity or convergence within a set of economic entities, the OECD countries or members of the European (Monetary) Union, say. Since we assume N to be finite the advocated methodology will be typically interesting for empirical investigations with small or moderate cross section dimension. To provide particular applications we think of testing the purchasing power parity as, for instance, in Edison *et al.* (1997). Similarly, the implications of balanced growth theory (see e.g. Barro and Sala-I-Martin 1995, Neusser 1991) can be tested by means of the recommended procedure. Applying pooled ECMs the issue of convergence of per capita health care expenditure within the OECD is addressed in Herwartz and Theilen (2000).

Concluding this paper we recommend to use the wild bootstrap to generate critical values for OLS-based test statistics computed from single equation ECMs. Our discussion covers a wide range of possible testing problems, we consider inference on long-run equilibrium relations and weak exogeneity. Specification testing is also regarded. Bootstrap procedures show superior size properties compared to first order asymptotic approximations even in

single equations and under homoskedastic error distribution. On the pooled level wild bootstrap inference provides a means to generate valid critical values for LR-type test statistics. It can easily be modified to account for (time varying) contemporaneous correlation across equations without requiring any first step estimates as e.g. GLS-methods. If the underlying error terms are (conditionally) heteroskedastic, then first order asymptotic approximations can hardly be derived if testing on weak exogeneity and specifying short-run dynamics is considered. Using critical values from the χ^2-distribution involves size violations that do not vanish with increasing sample size. The wild bootstrap still applies for this wider class of error distributions without involving any inefficiency in the case of homoskedasticity.

7. Appendix: Proofs

Proof of Lemma 2.1. Since v_t and r_t are uncorrelated components of serially uncorrelated random vectors we obtain from equation (4) and Theorem 2.6(a) of Phillips (1988) that there exist independent Wiener processes B_v and B_r such that

$$\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \begin{pmatrix} v_t \\ r_t \end{pmatrix}, \ s \in [0,1] \right\} \xrightarrow{d} \left\{ \begin{pmatrix} \sigma_v B_v(s) \\ \sigma_r B_r(s) \end{pmatrix}, \ s \in [0,1] \right\}. \tag{7.1}$$

The symbol '$\xrightarrow{d}$' signifies convergence in distribution while '$\Longrightarrow$' denotes weak convergence of the associated probability measures. The convergence in (7.1) takes place in $D[0,1] \times D[0,1]$ endowed with the metric $d + d$. A comprehensive description of these concepts is given in Billingsley (1968) while Phillips and Durlauf (1986) provide a condensed summary.

According to Skorohod's theorem (see, for example Theorem IV.13 in Pollard (1984)) we can define on a sufficiently rich probability space versions of $v_1, \ldots, v_T$ and $r_1, \ldots, r_T$ and independent Wiener processes B_v and B_r such that

$$d\left(\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} v_t, \ s \in [0,1] \right\}, \{\sigma_v B_v(s), \ s \in [0,1]\} \right) \xrightarrow{a.s.} 0, \tag{7.2}$$

and

$$d\left(\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} r_t, \ s \in [0,1] \right\}, \{\sigma_r B_r(s), \ s \in [0,1]\} \right) \xrightarrow{a.s.} 0. \tag{7.3}$$

Assuming now (7.2) and (7.3) one can transfer successively the weak convergence results of Lemmas 2.3 and 2.5, and of Theorem 2.6(a) of Phillips (1988) into corresponding almost

sure convergence results. Hence, we obtain as a consequence of (7.2) and (7.3) that

$$\frac{1}{T}\sum_{t=1}^{T}(v_1+\cdots+v_{t-1})r_t \xrightarrow{a.s.} \sigma_v\sigma_r\int_0^1 B_v(s)\,dB_r(s). \tag{7.4}$$

Finally, since $\lambda(1)=1$ holds in (2.21) we obtain from (7.3) that (2.19) holds. □

Proof of Proposition 2.1. (i) *Explicit form of z_t and $q_t = y_t + \beta z_t$*

Before we analyze the terms occurring in (2.14), we derive explicit expressions for z_t and $q_t = y_t + \beta z_t$. It holds that

$$z_t \;=\; \nu_2 t \;+\; (v_t+\cdots+v_1) \;+\; z_0 \tag{7.5}$$

and

$$\begin{aligned}
q_t &= y_t \;+\; \beta z_t \\
&= \nu_1 \;+\; (\beta+\gamma_1)\nu_2 \\
&\quad + u_t \;+\; (\beta+\gamma_1)v_t \\
&\quad + (1+\alpha_1)q_{t-1} \\
&= \sum_{s=0}^{t-1}(1+\alpha_1)^s[\nu_1+(\beta+\gamma_1)\nu_2] \\
&\quad + \sum_{s=0}^{t-1}(1+\alpha_1)^s[u_{t-s}+(\beta+\gamma_1)v_{t-s}] \\
&\quad + (1+\alpha_1)^t q_0 \\
&= -\frac{\nu_1+(\beta+\gamma_1)\nu_2}{\alpha_1} \\
&\quad + \sum_{s=0}^{t-1}(1+\alpha_1)^s[u_{t-s}+(\beta+\gamma_1)v_{t-s}] \\
&\quad + (1+\alpha_1)^t\left\{q_0 \;+\; [\nu_1+(\beta+\gamma_1)\nu_2]/\alpha_1\right\}.
\end{aligned} \tag{7.6}$$

(ii) *Approximation of terms involving q_{-1}*

Recall that $W = (1_T \quad \Delta z)$. We obtain that

$$W'W \;=\; \begin{pmatrix} T & \sum\Delta z_t \\ \sum\Delta z_t & \sum(\Delta z_t)^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} T & \nu_2 T \\ \nu_2 T & (\nu_2^2+\sigma_v^2)T \end{pmatrix}, \tag{7.7}$$

which implies

$$(W'W)^{-1} = \frac{1}{T\sum(\Delta z_t)^2 - (\sum \Delta z_t)^2}\begin{pmatrix} \sum(\Delta z_t)^2 & -\sum \Delta z_t \\ -\sum \Delta z_t & T \end{pmatrix}$$

$$\xrightarrow{P} \frac{1}{\sigma_v^2 T}\begin{pmatrix} \nu_2^2+\sigma_v^2 & -\nu_2 \\ -\nu_2 & 1 \end{pmatrix}. \tag{7.8}$$

Let $\widetilde{q}_{-1} = (\widetilde{q}_0, \ldots, \widetilde{q}_{T-1})'$, where

$$\begin{aligned}\widetilde{q}_t &= \sum_{s=0}^{t-1}(1+\alpha_1)^s[u_{t-s} + (\beta_0+\gamma_1)v_{t-s}] \\ &\quad + (1+\alpha_1)^t\left\{q_0 + [\nu_1 + (\beta_0+\gamma_1)\nu_2]/\alpha_1\right\}.\end{aligned}$$

Then we have (recall that $M = I_T - W(W'W)^{-1}W'$)

$$Mq_{-1}^0 = M\widetilde{q}_{-1}. \tag{7.9}$$

Since

$$W'\widetilde{q}_{-1} = \begin{pmatrix} \sum \widetilde{q}_{t-1} \\ \sum(\nu_2+v_t)\widetilde{q}_{t-1} \end{pmatrix} = O_P(T^{1/2})$$

we obtain, in conjunction with (7.8), that

$$\widetilde{q}_{-1}'W(W'W)^{-1}W'\widetilde{q}_{-1} = O_P(1). \tag{7.10}$$

Therefore, we obtain

$$\begin{aligned}(q_{-1}^0)'Mq_{-1}^0 &= \widetilde{q}_{-1}'M\widetilde{q}_{-1} = E\widetilde{q}_{-1}'\widetilde{q}_{-1} + O_P(T^{1/2}) \\ &= -\frac{T}{\alpha_1}\left[\sigma_u^2 + (\beta_0+\gamma_1)^2\sigma_v^2\right] + o_P(T).\end{aligned} \tag{7.11}$$

Since $W'r = O_P(T^{1/2})$ we get

$$(q_{-1}^0)'Mu = \widetilde{q}_{-1}'Mr = \widetilde{q}_{-1}'r - \widetilde{q}_{-1}'W(W'W)^{-1}W'r = O_P(T^{1/2}). \tag{7.12}$$

(iii) *Approximation of terms involving* z_{-1}

To approximate the terms involving z_{-1}, we have to distinguish between the cases $\nu_2 = 0$ and $\nu_2 \neq 0$.

(iii.a) $\nu_2 = 0$

We obtain immediately from (2.18) that

$$z'_{-1}z_{-1} \;=\; \sigma_v^2 T^2 \int_0^1 B_v(t)^2\,dt \;+\; o_P(T^2) \tag{7.13}$$

as well as

$$\sum_{t=1}^{T} z_{t-1} \;=\; \sigma_v T^{3/2} \int_0^1 B_v(t)\,dt \;+\; o_P(T^{3/2}). \tag{7.14}$$

Using (7.8), $\sum \Delta z_t = O_P(T^{1/2})$, and $\sum \Delta z_t z_t = O_P(T)$ we obtain that

$$z'_{-1}W(W'W)^{-1}W'z_{-1} \;=\; \sigma_v^2 T^2 \left[\int_0^1 B_v(t)\,dt\right]^2 \;+\; o_P(T^2), \tag{7.15}$$

which implies, in conjunction with (7.13), that

$$z'_{-1}Mz_{-1} \;=\; \sigma_v^2 T^2 \int_0^1 \left[B_v(t) \;-\; \int_0^1 B_v(s)\,ds\right]^2 dt \;+\; o_P(T^2). \tag{7.16}$$

Using

$$z'_{-1}\tilde{q}_{-1} \;=\; O_P(T)$$

and

$$z'_{-1}W(W'W)^{-1}W'\tilde{q}_{-1} \;\leq\; \sqrt{z'_{-1}W(W'W)^{-1}W'z_{-1}}\sqrt{\tilde{q}'_{-1}W(W'W)^{-1}W'\tilde{q}_{-1}} \;=\; O_P(T)$$

we have

$$z'_{-1}Mq^0_{-1} \;=\; z'_{-1}M\tilde{q}_{-1} \;=\; O_P(T). \tag{7.17}$$

By (7.8), (2.19), (7.14), and the facts that $\sum \Delta z_t r_t = O_P(T^{1/2})$ and $\sum \Delta z_t z_{t-1} = O_P(T)$ we obtain that

$$z'_{-1}W(W'W)^{-1}W'r \;=\; \frac{1}{T}\sum_t z_{t-1}\sum_t r_t \;+\; O_P(T^{1/2}) \;=\; \sigma_v\sigma_r T \int_0^1 B_v(t)\,dt\; B_r(1) \;+\; o_P(T),$$

which implies, in conjunction with (2.20), that

$$z'_{-1}Mr \;=\; \sigma_v\sigma_r T \int_0^1 \left[B_v(t) \;-\; \int_0^1 B_v(s)\,ds\right] dB_r(t) \;+\; o_P(T). \tag{7.18}$$

(iii.b) $\nu_2 \neq 0$

(7.6) and (2.18) imply that

$$z'_{-1}z_{-1} \;=\; \nu_2^2 \sum_{t=1}^{T}(t-1)^2 \;+\; o_P(T^3) \;=\; \nu_2^2 T^3/3 \;+\; o_P(T^3). \tag{7.19}$$

For the same reason we get

$$\sum_{t=1}^{T} z_{t-1} \;=\; \nu_2 \sum_{t=1}^{T}(t-1) \;+\; o_P(T^2) \;=\; \nu_2 T^2/2 \;+\; o_P(T^2).$$

Using this, (7.8), $\sum \Delta z_t = O_P(T^{1/2})$ and

$$\sum_{t=1}^{T} z_{t-1}\Delta z_t \;=\; \nu_2 \sum z_{t-1} \;+\; \sum v_t z_{t-1} \;=\; \nu_2^2 T^2/2 \;+\; o_P(T^2)$$

we obtain that

$$\begin{aligned} z'_{-1}W(W'W)^{-1}W'z_{-1} \;&=\; \frac{1}{T\sigma_v^2}\frac{T^4}{4}\left[(\nu_2^2+\sigma_v^2)\nu_2^2 \;-\; 2\nu_2\nu_2^3 \;+\; \nu_2^2\right] \;+\; o_P(T^3) \\ &=\; \nu_2^2 T^3/4 \;+\; o_P(T^3). \end{aligned} \tag{7.20}$$

This implies

$$z'_{-1}Mz_{-1} \;=\; \nu_2^2 T^3/12 \;+\; o_P(T^3). \tag{7.21}$$

By

$$z'_{-1}\tilde{q}_{-1} \;=\; O_P(T^{3/2})$$

and

$$z'_{-1}W(W'W)^{-1}W'\tilde{q}_{-1} \;\leq\; \sqrt{z'_{-1}W(W'W)^{-1}W'z_{-1}}\sqrt{\tilde{q}'_{-1}W(W'W)^{-1}W'\tilde{q}_{-1}} \;=\; O_P(T^{3/2})$$

we obtain that

$$z'_{-1}Mq^0_{-1} \;=\; z'_{-1}M\tilde{q}_{-1} \;=\; O_P(T^{3/2}). \tag{7.22}$$

Using $z'_{-1}r = \nu_2 \sum (t-1)r_t + O_P(T)$ and

$$z'_{-1}W(W'W)^{-1}W'r \;=\; \nu_2 \frac{T}{2} \sum r_t \;+\; o_P(T^{3/2})$$

we get

$$z'_{-1}Mr \;=\; \nu_2 \sum_{t=1}^{T}(t-T/2)r_t \;+\; o_P(T^{3/2}). \tag{7.23}$$

(iv) *Limit distribution of* $LR^{(1)}$

In the case of $\nu_2 = 0$, we obtain by the approximations (2.17), (7.11), (7.12), and (7.16) to (7.18) that

$$\mathrm{LR}^{(1)} \;=\; \frac{1}{\sigma_r^2}\frac{\{\sigma_v\sigma_r T\int [B_v(t) \;-\; \int B_v(s)\,ds]\; dB_r(t)\}^2}{\sigma_v^2 T^2 \int [B_v(t) \;-\; \int B_v(s)\,ds]^2\; dt} \;+\; o_P(1). \tag{7.24}$$

According to Lemma 5.1 of Park and Phillips (1988), the right-hand side of (7.24) is asymptotically $\chi^2(1)$ distributed.

In the case of $\nu_2 \neq 0$, we obtain by the approximations (2.17), (7.11), (7.12), (7.21) to (7.23), and the fact that $\sum(t - T/2)^2 = T^3/12 + O(T^2)$ by the CLT for α-mixing heterogeneously distributed random variables (see, e.g. Theorem 5.19 in White 1984) that

$$\begin{aligned}\mathrm{LR}^{(1)} \quad &= \quad \frac{1}{\sigma_r^2}\frac{1}{T^3/12}\left[\sum_{t=1}^{T}(t - T/2)r_t\right]^2 \;+\; o_P(1) \\ &\xrightarrow{d} \quad \chi^2(1).\end{aligned}$$

□

Proof of Lemma 2.2. We have

$$\widehat{\beta} \;-\; \beta \;=\; \frac{1}{T\sum z_t^2 \;-\; (\sum z_t)^2}\left(\sum_t z_t \sum_t e_t \;-\; T\sum_t z_t e_t\right).$$

We obtain from (7.6) that

$$e_t \;=\; \sum_{s=1}^{t-1}(1+\alpha_1)^s[u_{t-s} + (\beta+\gamma_1)v_{t-s}] \;+\; (1+\alpha_1)^t q_0. \tag{7.25}$$

This implies that $\sum e_t = O_P(T^{1/2})$, regardless of whether $\nu_2 = 0$ or $\nu_2 \neq 0$.

If $\nu_2 = 0$, then $\sum z_t = O_P(T^{3/2})$, $\sum z_t e_t = O_P(T)$, and

$$T^{-2}\sum z_t^2 \;-\; T^{-3}(\sum z_t)^2 \;\xrightarrow{d}\; \sigma_v^2 \int \left[B_v(t) \;-\; \int B_v(s)\,ds\right]^2 dt,$$

which implies that $\widehat{\beta} - \beta = O_P(T^{-1})$.

If $\nu_2 \neq 0$, then $\sum z_t = O_P(T^2)$, $\sum z_t e_t = O_P(T^{3/2})$, and

$$T\sum z_t^2 \;-\; \Big(\sum z_t\Big)^2 \;=\; \nu_2^2\left(T\sum t^2 \;-\; (\sum t)^2\right) \;+\; o_P(T^4) \;=\; \nu_2^2\frac{T^4}{12} \;+\; o_P(T^4),$$

which yields that $\widehat{\beta} - \beta = O_P(T^{-3/2})$. □

Proof of Proposition 2.2. Since $\mathcal{Z}_1 = (q_{-1} + (\widehat{\beta} - \beta)z_{-1} \quad 1_T)$ we obtain

$$\mathcal{Z}_1'\mathcal{Z}_1 \;=\; \begin{pmatrix} q_{-1}'q_{-1} + 2(\widehat{\beta}-\beta)q_{-1}'z_{-1} + (\widehat{\beta}-\beta)^2 z_{-1}'z_{-1} & \sum q_{t-1} + (\widehat{\beta}-\beta)\sum z_{t-1} \\ \sum q_{t-1} + (\widehat{\beta}-\beta)\sum z_{t-1} & T \end{pmatrix}. \tag{7.26}$$

According to Lemma 2.2, we have in both of the cases $\nu_2 = 0$ and $\nu_2 \neq 0$ that $(\widehat{\beta}-\beta)\sum z_{t-1} = o_P(T^{1/2})$ and $(\widehat{\beta}-\beta)^2 z_{-1}'z_{-1} = o_P(T)$, which implies that

$$T(\mathcal{Z}_1'\mathcal{Z}_1)^{-1} \xrightarrow{P} \begin{pmatrix} 1/Eq_{t-1}^2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.27}$$

Furthermore, we have

$$\mathcal{Z}_1'v \;=\; \begin{pmatrix} q_{-1}'v + (\widehat{\beta}-\beta)z_{-1}'v \\ \sum v_t \end{pmatrix} \;=\; \begin{pmatrix} q_{-1}'v + o_P(T^{1/2}) \\ \sum v_t \end{pmatrix}. \tag{7.28}$$

¿From (7.27) and (7.28) we obtain

$$\mathrm{RSS}_1^{(2)} \;=\; v'(I_T \;-\; \mathcal{Z}_1(\mathcal{Z}_1'\mathcal{Z}_1)^{-1}\mathcal{Z}_1')v \;=\; \sigma_v^2 T \;+\; O_P(1). \tag{7.29}$$

Finally, we obtain by a CLT for α-mixing random variables that

$$\mathrm{LR}^{(2)} \;=\; \frac{1}{TEq_{t-1}^2\sigma_v^2}(q_{-1}'v)^2 \;+\; o_P(1) \;\xrightarrow{d}\; \frac{E(q_{t-1}v_t)^2}{Eq_{t-1}^2\sigma_v^2}\chi^2(1).$$

□

Proof of Proposition 2.3. First of all, we have according to (7.27) that

$$T(\widehat{W}_1'\widehat{W}_1)^{-1} \xrightarrow{P} \begin{pmatrix} 1 & 0 \\ 0 & 1/Eq_{t-1}^2 \end{pmatrix}. \tag{7.30}$$

We define

$$\widetilde{\Delta z} \;=\; M\Delta z \;=\; v \tag{7.31}$$

and

$$\widetilde{\Delta y_{-1}} \;=\; M\Delta y_{-1} \;=\; \alpha_1\widetilde{q_{-2}} \;+\; \gamma_1 v_{-1} \;+\; u_{-1}. \tag{7.32}$$

By

$$\widehat{W}_1\widetilde{\Delta z} \;=\; \begin{pmatrix} \sum v_t \\ \sum \widehat{q}_{t-1}v_t \end{pmatrix} \;=\; O_P(T^{1/2})$$

we obtain

$$\begin{aligned}(\Delta z)'M(\Delta z) &= (\widetilde{\Delta z})'M(\widetilde{\Delta z}) = (\widetilde{\Delta z})'(\widetilde{\Delta z}) + O_P(1) \\ &= \sigma_v^2 T + O_P(T^{1/2}) \end{aligned} \tag{7.33}$$

and

$$(\Delta z)'Mu = (\widetilde{\Delta z})'u + O_P(1) = O_P(T^{1/2}). \tag{7.34}$$

By

$$\widehat{W}_1(\widetilde{\Delta y_{-1}}) = \begin{pmatrix} \sum \widetilde{\Delta y}_{t-1} \\ \sum \widehat{q}_{t-1}\widetilde{\Delta y}_{t-1} \end{pmatrix} = \begin{pmatrix} O_P(T^{1/2}) \\ TEq_{t-1}\Delta y_{t-1} + o_P(T) \end{pmatrix}$$

we get

$$(\Delta y_{-1})'M(\Delta z) = (\widetilde{\Delta y_{-1}})'M(\widetilde{\Delta z}) = O_P(T^{1/2}) \tag{7.35}$$

and

$$\begin{aligned}(\Delta y_{-1})'Mu &= (\widetilde{\Delta y_{-1}})'u - (\widetilde{\Delta y_{-1}})'\widehat{W}_1(\widehat{W}_1'\widehat{W}_1)^{-1}\widehat{W}_1'u \\ &= \left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right)' u. \end{aligned} \tag{7.36}$$

Finally, we have

$$\begin{aligned}&(\Delta y_{-1})'M(\Delta y_{-1}) \\ &= (\widetilde{\Delta y_{-1}})'(\widetilde{\Delta y_{-1}}) - \frac{(q_{-1}'\widetilde{\Delta y_{-1}})^2}{q_{-1}'q_{-1}} + o_P(T) \\ &= \left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right)' \left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right) + o_P(T). \end{aligned} \tag{7.37}$$

Using the above approximations we obtain by a central limit theorem for α-mixing random variables that

$$\begin{aligned}TS_T^{(3)} &= \frac{\left[\left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right)' u\right]^2}{\left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right)' \left(\widetilde{\Delta y_{-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{-1}\right)\sigma_u^2} \\ &\xrightarrow{d} \frac{E\left[\left(\widetilde{\Delta y_{t-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{t-1}\right)u_t\right]^2}{E\left(\widetilde{\Delta y_{t-1}} - \frac{Eq_{t-1}\widetilde{\Delta y_{t-1}}}{Eq_{t-1}^2}q_{t-1}\right)^2\sigma_u^2}\chi^2(1). \end{aligned}$$

□

Proof of Proposition 4.1. We have

$$\mathcal{X}_0'\mathcal{X}_0 \;=\; \begin{pmatrix} T & 0 & 0 \\ 0 & TE(q_{t-1}^0)^2 & 0 \\ 0 & 0 & T\sigma_v^2 \end{pmatrix} + O_P(T^{1/2}),$$

which implies $\|(\mathcal{X}_0'\mathcal{X}_0)^{-1}\| = O_P(T^{-1})$. Since

$$\sum_{t=0}^{\infty} P(|q_{t-1}^0| > \sqrt{t}) \;\leq E|q_{t-1}^0|^2 \;<\; \infty$$

it follows from Borel-Cantelli that $|q_{t-1}^0|/\sqrt{t} \to 0$ almost surely, which in turn implies that $T^{-1/2}\max_{1\leq t\leq T}|q_{t-1}^0| \to 0$, again with probability 1. Hence, we obtain $\max_{1\leq t\leq T}|q_{t-1}^0| = o_P(T^{1/2})$, and analogously $\max_{1\leq t\leq T}|\Delta z_t| = o_P(T^{1/2})$. Moreover, since $\mathcal{X}_0'r = O_P(T^{1/2})$ we obtain, in conjunction with (4.2) that

$$\max_{1\leq t\leq T}|\widehat{r}_t \;-\; r_t| \;=\; o_P(1). \tag{7.38}$$

Using $T^{-1}\mathrm{RSS}_1^{(1),*} \xrightarrow{P} \sigma_r^2$ we obtain

$$\begin{aligned} \mathrm{LR}^{(1),*} \;&=\; \frac{T}{\mathrm{RSS}_1^{(1),*}}\frac{1}{z_{-1}'Mz_{-1}}\left(z_{-1}'Mr^*\right)^2 \;+\; o_P(1) \\ &=\; \left(\sum_{t=1}^{T} w_t r_t \eta_t\right)^2 \;+\; o_P(1), \end{aligned}$$

where

$$w_t \;=\; \frac{z_{t-1} \;-\; T^{-1}\sum_s z_{s-1}}{\sigma_r\sqrt{\sum_t \left(z_{t-1} \;-\; T^{-1}\sum_s z_{s-1}\right)^2}}.$$

Since $\max\{|w_t r_t|\} \xrightarrow{P} 0$ we obtain, for arbitrary $\epsilon > 0$, that

$$E\left\{I\left(|\eta_t| > \frac{\epsilon}{\max\{|w_t r_t|\}}\right)\eta_t^2\right\} \;\xrightarrow{P}\; 0.$$

Moreover, it follows from $\sum_t (w_t r_t)^2 \xrightarrow{P} 1$ that

$$\begin{aligned} &\sum_{t=1}^{T} E\left(|w_t r_t \eta_t|^2 I\left(|\eta_t| > \frac{\epsilon}{\max\{|w_t r_t|\}}\right)\middle|\, \mathcal{X}\right) \\ &\quad\leq\; \sum_{t=1}^{T}|w_t r_t|^2 E\left(I\left(|\eta_t| > \frac{\epsilon}{\max\{|w_t r_t|\}}\right)\eta_t^2\middle|\, \mathcal{X}\right) \;\xrightarrow{P}\; 0, \end{aligned}$$

that is, the classical Lindeberg condition is fulfilled. Hence, we obtain by the Lindeberg-Feller central limit theorem that

$$\mathcal{L}\left(\sum_{t=1}^{T} w_t r_t \eta_t \middle| \mathcal{X}\right) \Longrightarrow \mathcal{N}(0,1) \qquad \text{in probability.}$$

The assertion now follows from the continuous mapping theorem. □

Proof of Proposition 4.2. We have analogously to (7.38) that

$$\max_{1\le t\le T} |\widehat{v}_t - v_t| = o_P(1). \tag{7.39}$$

Hence, we obtain

$$\mathrm{LR}^{(2),*} = \left(\sum_{t=1}^{T} w_t v_t \eta_t\right)^2 + o_P(1),$$

where

$$w_t = \frac{q_{t-1}}{\sigma_v \sqrt{\sum_s q_{s-1}^2}}.$$

The assertion follows from $\sum_t (w_t v_t)^2 \xrightarrow{P} E(q_{t-1}v_t)^2/(\sigma_v^2 E q_{t-1}^2)$ and $\max\{|w_t v_t|\} \xrightarrow{P} 0$ by the CLT. □

Proof of Proposition 4.3. Using again (7.38) we obtain

$$\mathrm{LR}^{(3),*} = \left(\sum_{t=1}^{T} w_t r_t \eta_t\right)^2 + o_P(1),$$

where

$$w_t = \frac{\widetilde{\Delta y_{t-1}} - \frac{E q_{t-1}\widetilde{\Delta y_{t-1}}}{E q_{t-1}^2} q_{t-1}}{\sigma_u \sqrt{\sum_s \left(\widetilde{\Delta y_{s-1}} - \frac{E q_{s-1}\widetilde{\Delta y_{s-1}}}{E q_{s-1}^2} q_{s-1}\right)^2}}.$$

The assertion now follows from

$$\sum_{t=1}^{T} (w_t r_t)^2 \xrightarrow{P} \frac{E\left[\left(\widetilde{\Delta y_{t-1}} - \frac{E q_{t-1}\widetilde{\Delta y_{t-1}}}{E q_{t-1}^2} q_{t-1}\right) u_t\right]^2}{E\left(\widetilde{\Delta y_{t-1}} - \frac{E q_{t-1}\widetilde{\Delta y_{t-1}}}{E q_{t-1}^2} q_{t-1}\right)^2 \sigma_u^2}$$

and $\max\{|w_t r_t|\} \xrightarrow{P} 0$ by the CLT. □

Proof of Theorem 4.1. We prove the assertion only for the most difficult case, the first test with $\nu_2 = 0$.

First of all, we have a somehow untypical case in that the bootstrap distribution (i.e., the conditional distribution of $\mathrm{LR}_{[N]}^{(1),*}$ given $\mathcal{X}_T = \{\bar{v}_1, \ldots, \bar{v}_T, \bar{r}_1, \ldots, \bar{r}_T\}$, where $\bar{v}_t =$

$(v_{1,t},\ldots,v_{N,t})'$ and $\bar{r}_t = (r_{1,t},\ldots,r_{N,t})'$) does *not* consistently approximate the unconditional distribution of $\mathrm{LR}^{(1)}_{[N]}$. Rather, it seems at first sight that the bootstrap merely approximates the conditional distribution of $\mathrm{LR}^{(1)}_{[N]}$ given $V = ((z'_{n,-1}M_nM_mz_{m,-1}))_{n,m=1,\ldots,N}$. This is not exactly true; the weak convergence arguments employed below do not yield such a result for $\mathcal{L}(\mathrm{LR}^{(1)}_{[N]} \mid V)$. Rather, we actually show that $\mathrm{LR}^{(1),*}_{[N]}$ consistently approximates the conditional distribution of an asymptotic approximation to $\mathrm{LR}^{(1)}_{[N]}$.

We use the following notation: $U_n = z'_{n,-1}M_nr_n$, $V_{m,n} = z'_{n,-1}M_nM_mz_{m,-1}$, $U = (U_1,\ldots,U_N)'$, and $V = ((z'_{n,-1}M_nM_mz_{m,-1}))_{n,m=1,\ldots,N}$. According to the approximations (2.17), (7.11), (7.12), and (7.16) to (7.18), we can write the test statistic as

$$\mathrm{LR}^{(1)}_{[N]} \;=\; g(U,V) \;+\; o_P(1), \tag{7.40}$$

where $g(x,y) = \sum_{n=1}^N x_n^2/(\sigma_{nr}^2 y_{n,n})$ and $\sigma_{nr}^2 = \mathrm{var}(r_{nt})$. Let $(B_{1v},\ldots,B_{nv})'$ and $(B_{1r},\ldots,B_{nr})'$ be independent N-dimensional Wiener processes with covariances $\mathrm{Cov}(\bar{v}_t)$ and $\mathrm{Cov}(\bar{r}_t)$, respectively. We define

$$\begin{aligned} U_n^\infty &= \int_0^1 \left[B_{nv}(t) \;-\; \int_0^1 B_{nv}(s)\,ds\right] dB_{nr}(t), \\ V_{n,m}^\infty &= \int_0^1 \left[B_{nv}(t) \;-\; \int_0^1 B_{nv}(s)\,ds\right]\left[B_{mv}(t) \;-\; \int_0^1 B_{mv}(s)\,ds\right] dt, \end{aligned}$$

$U^\infty = (U_1^\infty,\ldots,U_N^\infty)'$, and $V^\infty = ((V_{n,m}^\infty))_{n,m=1,\ldots,N}$.

Since $Ev_{m,t}r_{n,t} = 0$, for $m,n = 1,\ldots,N$, we obtain by Theorem 2.6 of Phillips (1988) that

$$\begin{pmatrix} U \\ V \end{pmatrix} \xrightarrow{d} \begin{pmatrix} U^\infty \\ V^\infty \end{pmatrix}. \tag{7.41}$$

Let $t_\alpha^\infty(v)$ be the $(1-\alpha)$-quantile of the conditional distribution $\mathcal{L}(g(U^\infty,V^\infty) \mid V^\infty = v)$. Since g and $t_\alpha^\infty(\cdot)$ are continuous functions we obtain from (7.40) and (7.41) by the continuous mapping theorem (see e.g. Theorem III.6 in Pollard (1984)) that

$$\begin{pmatrix} \mathrm{LR}^{(1)}_{[N]} \\ t_\alpha^\infty(V) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} g(U^\infty,V^\infty) \\ t_\alpha^\infty(V^\infty) \end{pmatrix}. \tag{7.42}$$

Now we analyze the bootstrap statistic $\mathrm{LR}^{(1),*}_{[N]}$. With $U_n^* = z'_{n,-1}M_nr_n^*$ and $U^* = (U_1^*,\ldots,U_N^*)'$, we have that

$$\mathrm{LR}^{(1),*}_{[N]} \;=\; g(U^*,V) \;+\; o_P(1). \tag{7.43}$$

We denote by $\widetilde{t}^*_\alpha$ the $(1-\alpha)$-quantile of $\mathcal{L}(g(U^*,V) \mid \mathcal{X}_T)$. Then (7.43) implies that

$$t^*_\alpha = \widetilde{t}^*_\alpha + o_P(1). \tag{7.44}$$

We obtain by the multivariate central limit theorem for independent random variables that

$$\mathcal{L}\left(V^{-1/2}U^* \middle| \mathcal{X}_T\right) \Longrightarrow N(0, \mathrm{Cov}(\bar{r}_t)) = \mathcal{L}\left((V^\infty)^{-1/2}U^\infty\right) \qquad \text{in probability.}$$

This means that

$$\begin{aligned} g(U^*,V) &= g\left(V^{1/2}(V^{-1/2}U^*), V\right) \\ &= g\left(V^{1/2}Z, V\right) + o_P(1) \end{aligned}$$

holds with $Z \sim N(0, \mathrm{Cov}(\bar{r}_t))$ and, hence, $\mathcal{L}(g(v^{1/2}Z, v)) = \mathcal{L}(g(U^*,V^*) \mid V^* = v)$. Consequently, we obtain that

$$\widetilde{t}^*_\alpha = t^\infty_\alpha + o_P(1). \tag{7.45}$$

Now we conclude from (7.42), (7.44), and (7.45) that

$$\begin{pmatrix} \mathrm{LR}^{(1)}_{[N]} \\ t^*_\alpha \end{pmatrix} \xrightarrow{d} \begin{pmatrix} g(U^\infty, V^\infty) \\ t^\infty_\alpha(V^\infty) \end{pmatrix}. \tag{7.46}$$

Since $P(g(U^\infty, V^\infty) = t_\alpha(V^\infty)) = 0$ we obtain by the continuous mapping theorem that

$$P\left(\mathrm{LR}^{(1)}_{[N]} > t^*_\alpha\right) \longrightarrow P(g(U^\infty, V^\infty) > t_\alpha(V^\infty)) = \alpha.$$

□

References

Banerjee, A., J. Dolado, J.W. Galbraith, D.F. Hendry (1993). *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data.* Oxford University Press, Oxford.

Barro, R.J., X. Sala-I-Martin (1995). *Economic Growth.* McGraw-Hill, Boston.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Boswijk, H.P. (1993). On the formulation of Wald tests on long-run parameters. *Oxford Bulletin of Economics and Statistics* **55**, 137–144.

Boswijk, H.P. (1995a). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* **63**, 37–60.

Boswijk, H.P. (1995b). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* **69**, 133–158.

Boswijk, H.P., P.H. Franses (1992). Dynamic Specification and Cointegration. *Oxford Bulletin of Economics and Statistics* **54**, 369–381.

Davidson, J.E.H., D.F. Hendry, F. Srba, S. Yeo (1978). Econometric modelling of the aggregate time-series relationsship between consumer's expenditure and income in the United Kingdom. *Economic Journal* **88**, 661–692.

Donaldson, J.B., B. Mehra (1983). Stochastic growth with correlated productivity shocks. *Journal of Economic Theory* **29**, 282–312.

Edison, H.J., J.E. Gagnon, W.R. Melick (1997). Understanding the empirical literature on purchasing power parity: the post Bretton Woods era. *Journal of International Money and Finance* **16**, 1-17.

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. inflation. *Econometrica* **50**, 987–1008.

Engle, R.F., D.F. Hendry, J.F. Richard (1983). Exogeneity. *Econometrica* **51**, 277-304.

Engle, R.F., C.W.J. Granger (1987). Co-Integration and error correction: representation, estimation and testing. *Econometrica* **55**, 251–276

Engle, R.F., K.F. Kroner (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory* **11**, 122–150.

Froot, K.A., K. Rogoff (1995). Perspectives on PPP and long-run real exchange rates, in *Handbook of International Economics*, vol. III, G. Grossman and K. Rogoff (eds.), chapter 32, 1647-1788.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**, 121–130.

Groen, J.J.J., F. Kleibergen (1999). Likelihood based cointegration analysis in panels of vector error correction models, Erasmus University Rotterdam, University of Amsterdam, *mimeo.*

Hendry, D.F. (1995). *Dynamic Econometrics.* Oxford University Press, New York.

Herwartz, H. (1998). Testing periodicity in time series models – A recommendation of bootstrap methods. *Computational Statistics* **13**, 283–300.

Herwartz, H., H. Lütkepohl (2000). Multivariate volatility analysis of VW stock prices. *International Journal of Intelligent Systems in Accounting, Finance & Management* **9**, 35–54.

Herwartz, H., B. Theilen (2000). The determinants of health care expenditure, *Discussion Paper Humboldt University Berlin, SFB 373, 78/00* **9**, 35–54.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551-1580.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single equation analysis. *Journal of Econometrics* **52**, 389–402.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl, T.C. Lee (1988). *Introduction to the Theory and Practice of Econometrics.* Wiley, New York.

Kremers, J.J.M., N.R. Ericsson, J.J. Dolado (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* **54**, 325–348.

Li, H., G.S. Maddala (1997). Bootstrapping cointegrating regressions. *Journal of Econometrics* **80**, 297–318.

Lothian, J.R. (1997). Multi-country evidence on the behavior of purchasing power parity under the current float. *Journal of International Money and Finance* **16**, 16–35.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis.* Springer Verlag, Berlin.

Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations. Lecture Notes in Statistics 77*, Springer, New York.

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

Neumann, M.H., J.-P. Kreiss (1998). Regression-type inference in nonparametric autoregression. *Annals of Statistics* **26**, 1570–1613.

Neusser, K. (1991). Testing the long-run implications of the neoclassical growth model. *Journal of Monetary Economics* **27**, 3–27.

Park, J.O., P.C.B. Phillips (1988). Statistical inference in regressions with integrated processes: part I. *Econometric Theory* **4**, 468–497.

Pedroni, P. (1998). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, forthcoming.

Phillips, P.C.B. (1988). Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$. *Journal of Multivariate Analysis* **24**, 252–264.

Phillips, P.C.B., S.N. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies* **53**, 473–495.

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

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Urbain, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* **54**, 187–207.

White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press, San Diego, CA.

Wu, C.F.J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14**, 1261–1343.

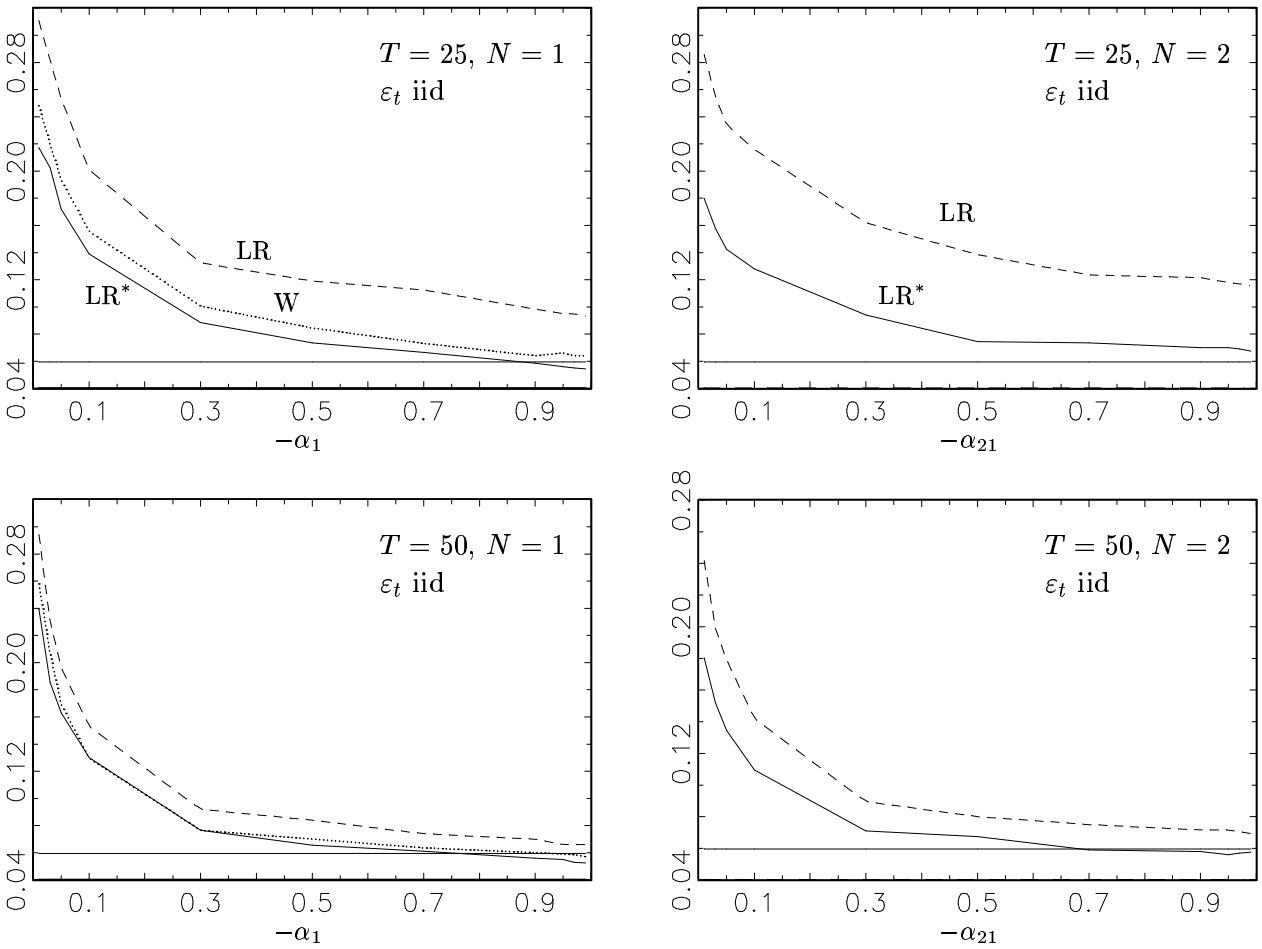

Figure 1. Size estimates for $H_0^{(1)} : \beta = -1$. Different scenarios are considered and indicated in single panels. The nominal significance level is $\alpha = 0.05$. A bound $(0.05 + 1.96\sqrt{1/2000\alpha(1-\alpha)})$ indicates the critical region where the empirical size differs from its nominal counterpart at the 5% level. LR and W are size estimates obtained from $\chi^2(1)$- and $F(1, T-K)$- critical values, respectively. LR* indicates size estimates obtained from wild bootstrap inference.

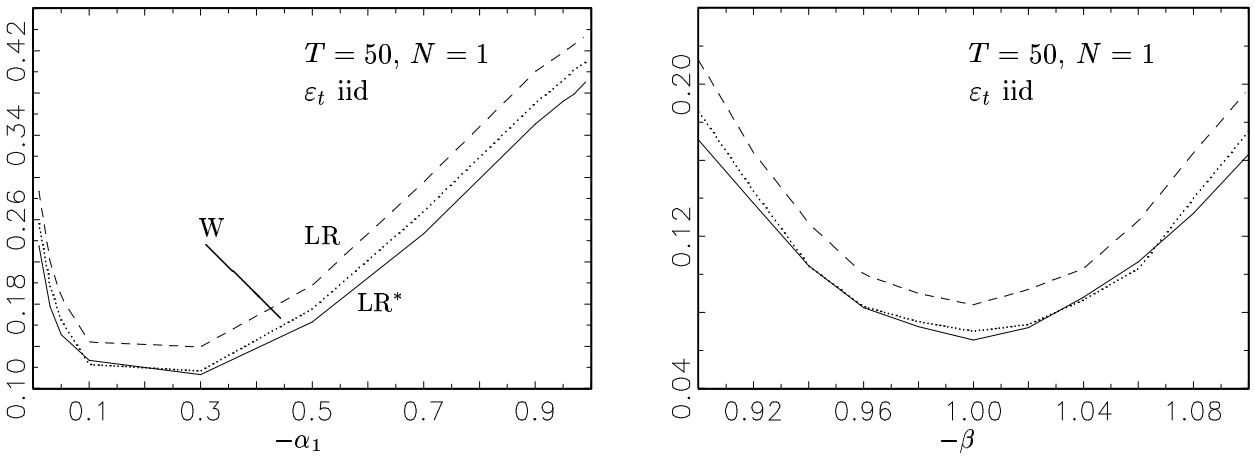

Figure 2. Power estimates of single equation inference on $H_0^{(1)} : \beta = -1$. Different scenarios are considered and indicated in single panels. The nominal significance level is $\alpha = 0.05$. LR and W are empirical rejection frequencies obtained from $\chi^2(1)$- and $F(1, T - K)$- critical values, respectively. LR* are power estimates for wild bootstrap inference.

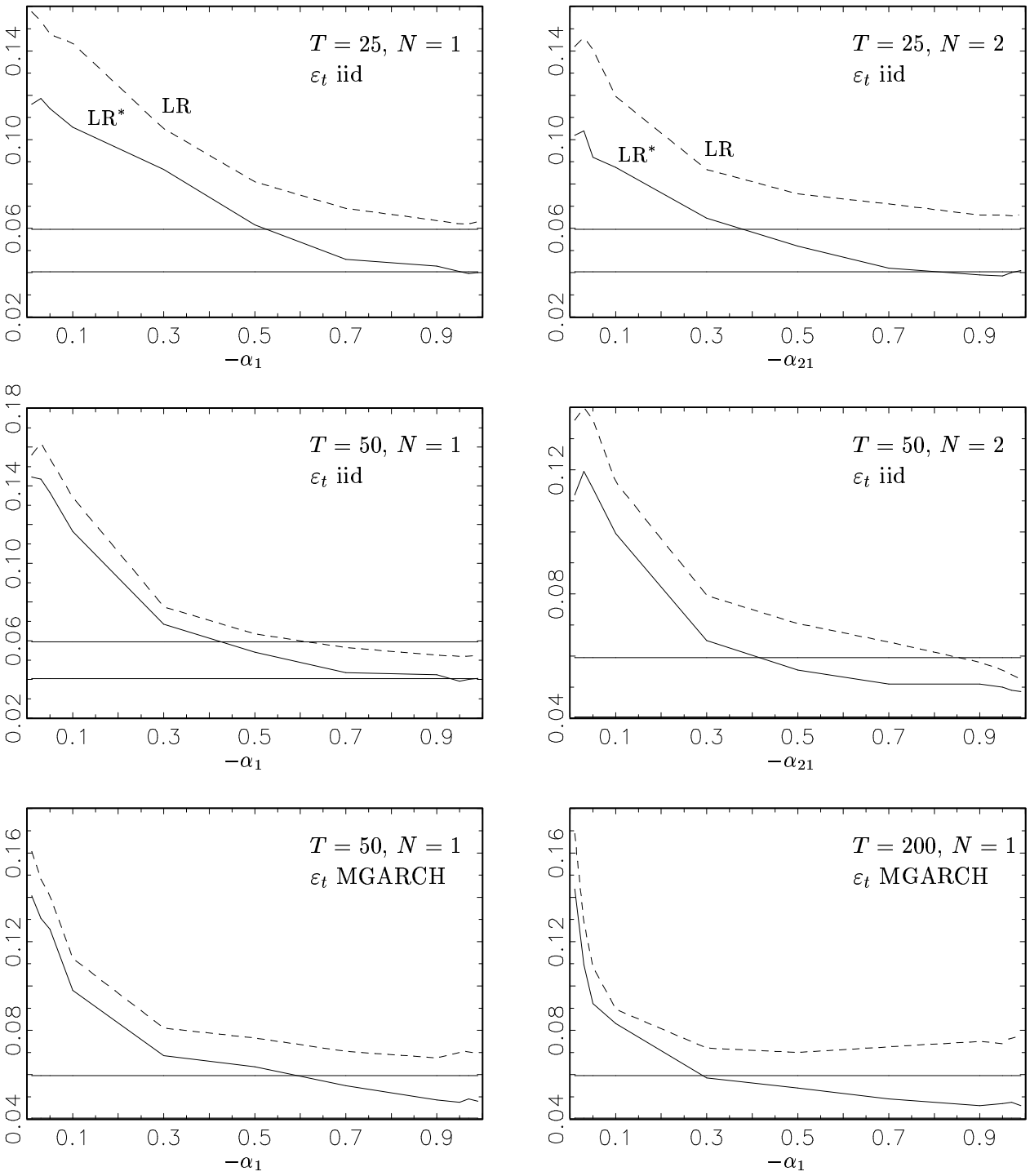

Figure 3. Size estimates for testing on weak exogeneity, i.e. $H_0^{(2)} : \alpha_2 = 0$. Different scenarios are considered and indicated in single panels. The nominal significance level is $\alpha = 0.05$. Depending on the simulation results bounds $(0.05 \pm 1.96\sqrt{1/2000\alpha(1-\alpha)})$ indicate the critical region where the empirical size differs from its nominal counterpart at the 5% level. LR indicates size estimates obtained from $\chi^2(1)$-critical values. LR* are size estimates for wild bootstrap inference.

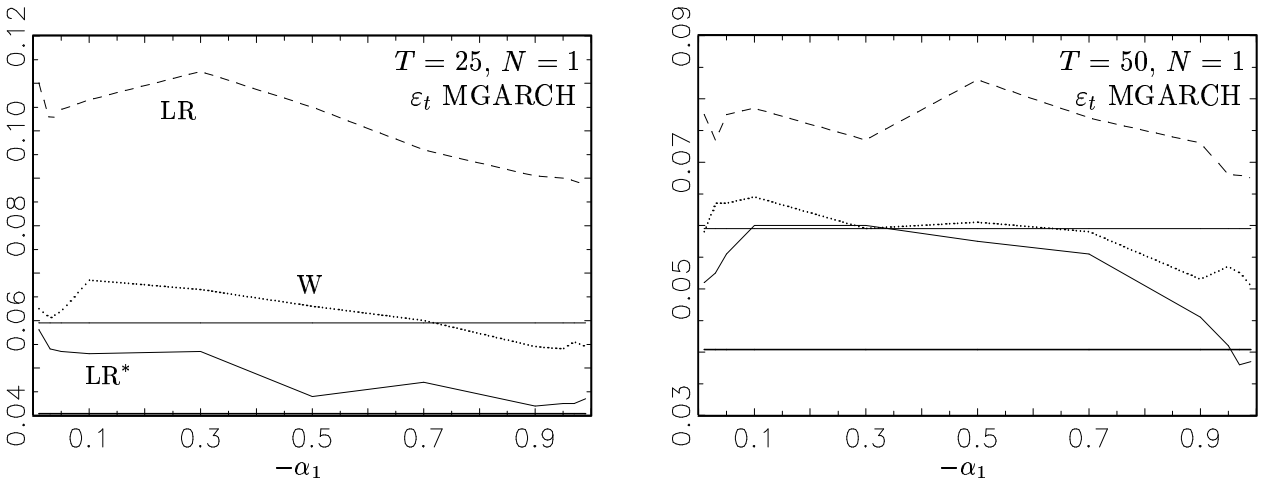

Figure 4. Size estimates for $H_0^{(3)}$: $\gamma_2 = 0$ under heteroskedastic error terms. The nominal significance level is $\alpha = 0.05$. Depending on the simulation results bounds $(0.05 \pm 1.96\sqrt{1/2000\alpha(1-\alpha)})$ indicate the critical region where the empirical size differs from its nominal counterpart at the 5% level. LR and W are size estimates obtained from $\chi^2(1)$- and $F(1, T-K)$- critical values, respectively. LR* are size estimates obtained from wild bootstrap inference.