Serial-correlation testing in error component models with moderately small T*

Sebastian Kripfganz† Matei Demetrescu‡ Mehdi Hosseinkouchack§

June 16, 2025

Abstract

When testing for unrestricted serial correlation in linear panel data models, the number of moment restrictions under the null hypothesis of no such correlation increases quadratically in the number of time periods T. Portmanteau tests designed for fixed T can quickly lose power even for time horizons which are typically still considered as small. To circumvent this problem, we propose refinements motivated by strategies to reduce the number of instruments in the estimation of dynamic panel data models. Furthermore, we propose a new test based on covariances between first differences and encompassing longer differences. Our test yields substantial power improvements against moving-average and autoregressive alternatives. It retains high power under random-walk alternatives and high variances of the group-specific error component. Moreover, we demonstrate that serial-correlation tests based on regression residuals can suffer from severe power losses when the initial estimator is inconsistent under the alternative. Finally, we re-analyze a widely used data set for the estimation of dynamic employment equations. Contrary to previous evidence, but in line with our power comparisons, our proposed test uncovers statistical evidence for the presence of serial correlation. Taken at face value, this in turn implies that the original regression results suffer from estimator inconsistency.

Keywords: Serial correlation; Specification testing; Panel data; Dimensionality reduction; First differences; Long differences

JEL Classification: C12; C23; C52

*The authors acknowledge the use of the University of Exeter High-Performance Computing facility in carrying out this work. The paper benefited from discussions with Samuel Engle and helpful comments by seminar participants at the universities of Bristol, Exeter, and Leipzig, as well as the participants of the Workshop on Recent Advances in Panel and Network Data at the University of Oxford.

†University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, UK. Tel.: +44-1392-722110; E-mail: S.Kripfganz@exeter.ac.uk

‡TU Dortmund University, Department of Statistics, Chair of Econometrics and Statistics, CDI Building, 44221 Dortmund, Germany. Tel.: +49-231-7553125; E-mail: mdeme@statistik.tu-dortmund.de

§EBS University, EBS Business School, Gustav-Stresemann-Ring 3, 65189 Wiesbaden, Germany. Tel.: +49-611-71021221; E-mail: mehdi.hosseinkouchack@ebs.edu

1 Introduction

Serial correlation in the idiosyncratic error component is usually undesirable when estimating panel data models with an error components structure. The adverse consequences reach from inefficient to inconsistent estimation of regression coefficients, as well as inconsistent standard error estimates. When the regression model is dynamically incomplete as a consequence of omitting relevant lagged predictors, correcting standard errors for serial correlation is generally insufficient for conducting valid statistical inference. Especially in dynamic panel models, identification strategies based on instrumental variables are designed for dynamically complete models, in which the within-group correlation of the error term is entirely due to the group-specific component. Similarly, likelihood-based estimators rely on the correct specification of the error covariance structure. Tests for serial correlation are therefore an essential part of the standard specification testing toolkit. In this paper, we discuss recent advances for serial-correlation testing and propose refined tests with better power properties under a wide range of alternatives.

For models with strictly exogenous regressors – ruling out dynamic models with a lagged dependent variable – a variety of serial correlation tests has been proposed, starting with the generalized Durbin-Watson statistic of Bhargava et al. (1982) and the Lagrange multiplier statistic of Baltagi and Li (1995). A shortcoming of both tests is that they rely on the assumption of normally distributed disturbances. Wooldridge (2002) suggested a computationally simple test – operationalized by Drukker (2003) – based on first-differenced residuals. Born and Breitung (2016) summarize the properties of these tests and suggest some modifications.

The previous tests are designed for specific data-generating processes and can have poor power against other alternatives. Inoue and Solon (2006) proposed a portmanteau test against arbitrary higher-order correlation for a small number of time periods T, which was refined by Jochmans (2020a) to overcome the reliance on a regularization parameter. Since the number of moment restrictions under the null hypothesis grows quadratically in T, the cross-sectional sample size N must grow even faster for a reliable asymptotic approximation of the test statistic's distribution. To avoid a loss of power with relatively large T, the test statistic can be limited to lower-order autocovariances. Alternatively, tests can be constructed against autocorrelation of a specific order, as suggested by Born and Breitung (2016).

It is a major drawback of these approaches that they require the explanatory variables to be strictly exogenous.[1] Early tests for specific covariance restrictions in models with a lagged dependent variable

[1] Jochmans (2020a) considers an adjustment of his modified Inoue and Solon (2006) test for models with predetermined regressors.

include the Bhargava and Sargan (1983) likelihood ratio test and the Wald or minimum chi-square tests proposed by Bhargava (1987) and Arellano (1990), which treat the elements of the error covariance matrix as structural parameters. For more flexible models with any number of predetermined or endogenous regressors, the Arellano and Bond (1991) test was the empirical researcher's first choice for a long time. This is a test against second-order serial correlation of the first-differenced disturbances. It is straightforward to generalize the test for higher orders (Arellano, 2003, p. 121–123). With the aim to increase power, Yamagata (2008) proposed to test jointly against correlation of second and higher order. However, because both tests are entirely based on first differences, they lack power against random-walk alternatives. More recently, Jochmans (2020b) developed a portmanteau test against arbitrary serial correlation, which remains valid under heteroskedasticity of any form, retains power under strong autoregressive alternatives, and is applicable with as few as 3 time periods. The Arellano and Bond (1991) and Yamagata (2008) tests, which can be constructed from a subset of the portmanteau moment restrictions, require at least 4 consecutive observations.

However, if N is not sufficiently large relative to T, this portmanteau test can suffer from a similar loss of power as the Inoue and Solon (2006) test due to the large number of moment restrictions under the null hypothesis. Importantly, this can bite already at moderately small T, which empirical researchers may not consider problematic. Even when T is sufficiently small, the test's power is sensitive to a large variance of the group-specific error component.

To address the problem of moment proliferation with increasing T, we can borrow ideas from the literature on the estimation of dynamic panel models with too many instruments, especially curtailing and collapsing (Roodman, 2009; Kiviet, 2020). In many applications, the marginal signal provided by moment restrictions that are sensitive to higher-order serial correlation is of diminishing value. Limiting the focus on the most informative moment restrictions – i.e., curtailing the order of correlation that can be detected under the alternative hypothesis – can thus improve the finite-sample performance of the test. Similarly, imposing some homogeneity of the data-generating process over time – i.e., collapsing/averaging the moment restrictions over time periods – hardly limits the spectrum of alternatives that can be detected by the test.

Beyond those refinements, we propose a particular linear combination of the moment restrictions to test for significant covariances between first differences and encompassing longer differences. Our test addresses two main shortcomings of the existing tests: Unlike tests based entirely on first differencing, our moment restrictions remain informative under a random-walk alternative; in contrast to the portmanteau test, our test is invariant to high variances of the group-specific error component. We use asymptotic power calculations and Monte Carlo simulations to showcase the power improvements of our test against various alternatives.

Since serial correlation tests are typically performed on regression residuals, the estimation uncertainty affects the distribution of the test statistic – unless restrictive exogeneity assumptions are imposed on the regressors. We demonstrate analytically and with simulations that the considered tests can suffer from substantial power losses when the initial estimator is inconsistent under the alternative hypothesis – even if the estimation uncertainty is correctly accounted for. This is generally the case in dynamic panel models. In extreme cases, the signal from serially correlated errors can be entirely offset by the estimation error. For testing purposes, where possible, we therefore recommend to choose an estimator that remains consistent under the alternative, disregarding a potential efficiency loss under the null hypothesis.

As an empirical illustration, we apply the existing and new serial-correlation tests to the data set used by Arellano and Bond (1991) in their seminal paper. They did not find evidence of serial correlation after estimating dynamic employment equations in a sample of U.K. companies, affirming the correct specification of their regression model. We replicate and re-assess their results with several of the existing and new tests. While the existing tests do not reject the null hypothesis, our new test based on long and first differences strongly rejects it in favor of serially correlated errors. These findings can be explained with the power comparisons in this paper, and they imply that the original estimations are inconsistent.

2 Testing for serial correlation

2.1 Moment restrictions

We consider the familiar error components model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it}, \tag{1}$$

$i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. $\mathbf{x}_{it}$ is a $K \times 1$ vector of regressors. The error term u_{it} consists of the group-specific component α_i and the idiosyncratic component ε_{it}. It is assumed throughout that both error components are independently distributed with mean zero. Following Jochmans (2020b), suppose that an asymptotically linear estimator $\hat{\boldsymbol{\beta}}$ is available for the coefficient vector $\boldsymbol{\beta}$, such that

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\omega}_i + o_p(1), \tag{2}$$

where $\boldsymbol{\omega}_i$ is a mean zero random variable with finite variance under the null hypothesis of no serial correlation in ε_{it}. The moment restrictions jointly considered under the null hypothesis are

$$E[\varepsilon_{is}\varepsilon_{it}] = 0 \tag{3}$$

for all $s \neq t$. In total, there are $T(T-1)/2$ distinct covariances. However, since we cannot consistently estimate the group-specific component α_i under fixed-T asymptotics, we cannot back out consistent estimates for the idiosyncratic error component ε_{it} either. The test thus needs to be based on estimates of the combined errors u_{it}.

As pointed out by Jochmans (2020b), instead of the moment restrictions (3), we can work with differences of covariances:

$$E[\varepsilon_{is}\varepsilon_{it}] - E[\varepsilon_{is}\varepsilon_{ir}] = E[u_{is}(u_{it} - u_{ir})] = 0. \tag{4}$$

As long as we assume that there exists no $c \neq 0$ such that $E[\varepsilon_{is}\varepsilon_{it}] = c$ for all $s \neq t$, testing restrictions (3) is equivalent to testing restrictions (4). Notice that $T(T-1)/2 - 1$ of the latter restrictions are linearly independent. A convenient way of writing the nonredundant moment restrictions is

$$E[u_{i,t-s}\Delta u_{it}] = 0, \quad 3 \leq t \leq T, \quad 2 \leq s \leq t-1, \tag{5}$$

$$E[u_{i,t+1}\Delta u_{it}] = 0, \quad 2 \leq t \leq T-1, \tag{6}$$

where $\Delta u_{it} = u_{it} - u_{i,t-1}$. In the following, we label the moments (5) backward looking and the moments (6) forward looking.

The composite null hypothesis can then be written as

$$E[\boldsymbol{\zeta}_i] = \mathbf{0}, \tag{7}$$

where $\boldsymbol{\zeta}_i = \mathbf{H}_i'\Delta\mathbf{u}_i$, $\Delta\mathbf{u}_i = (\Delta u_{i2}, \Delta u_{i3}, \ldots, \Delta u_{iT})'$, and $\mathbf{H}_i = (\mathbf{H}_{i-}, \mathbf{H}_{i+})$, with

$$\mathbf{H}_{i-} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ u_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & u_{i1} & u_{i2} & & 0 & \cdots & 0 \\ \vdots & & & \ddots & & & \\ 0 & 0 & 0 & & u_{i1} & \cdots & u_{i,T-2} \end{pmatrix} \quad \text{and} \quad \mathbf{H}_{i+} = \begin{pmatrix} u_{i3} & 0 & \cdots & 0 \\ 0 & u_{i4} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & u_{iT} \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

2.2 Transformations

Any linear transformation of the moment conditions (7) yields a valid subset of restrictions under the null hypothesis of no serial correlation in the idiosyncratic error component:

$$E[\mathbf{R}\boldsymbol{\zeta}_i] = \mathbf{0}, \tag{8}$$

where $\mathbf{R}$ is a deterministic transformation matrix of full row rank r. If $\mathbf{R}$ is nonsingular, tests of the moment restrictions (7) and (8) are equivalent. Some relevant transformations take the block-triangular form

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_- & \mathbf{0} \\ \mathbf{R}_\pm & \mathbf{R}_+ \end{pmatrix},$$

conformable with the following partitioning of the moment functions:

$$\boldsymbol{\zeta}_i = \begin{pmatrix} \dot{\mathbf{H}}_{i-} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{H}}_{i+} \end{pmatrix}' \begin{pmatrix} \Delta\dot{\mathbf{u}}_i \\ \Delta\ddot{\mathbf{u}}_i \end{pmatrix},$$

where $\dot{\mathbf{H}}_{i-}$ equals $\mathbf{H}_{i-}$ without the first row, and $\ddot{\mathbf{H}}_{i+}$ is the diagonal matrix obtained by removing the last row from $\mathbf{H}_{i+}$. Correspondingly, $\Delta\dot{\mathbf{u}}_i$ has the initial observation Δu_{i2} excluded, and $\Delta\ddot{\mathbf{u}}_i$ is a vector without the last observation Δu_{iT}.

As an example of an equivalent set of moment restrictions, we can replace any of the moment functions $u_{i,t+1}\Delta u_{it}$ in $\ddot{\mathbf{H}}'_{i+}\Delta\ddot{\mathbf{u}}_i$ by the otherwise redundant $u_{i,t+s}\Delta u_{it}$, for some $s \geq 2$, because $(u_{i,t+s} - u_{i,t+1})\Delta u_{it}$ can be written as a linear combination of the moment functions in $\dot{\mathbf{H}}'_{i-}\Delta\dot{\mathbf{u}}_i$. More generally, we can replace $\ddot{\mathbf{H}}'_{i+}\Delta\ddot{\mathbf{u}}_i$ by any valid linear combination of the initial moment functions,

$$\check{\mathbf{H}}'_{i+}\Delta\ddot{\mathbf{u}}_i = \mathbf{R}_\pm\dot{\mathbf{H}}'_{i-}\Delta\dot{\mathbf{u}}_i + \mathbf{R}_+\ddot{\mathbf{H}}'_{i+}\Delta\ddot{\mathbf{u}}_i, \tag{9}$$

provided that $\mathbf{R}_+$ is of full rank. For example, we can let $\mathbf{R}_+ = \mathbf{I}_{T-2}$ and find a corresponding $\mathbf{R}_\pm$ such that $\check{\mathbf{H}}_{i+} = u_{iT}\mathbf{I}_{T-2}$. Another relevant case would be $\check{\mathbf{H}}_{i+} = \mathbf{K}'_+\ddot{\mathbf{H}}_{i+}$, where $\mathbf{K}_+$ is a lower-triangular transformation matrix of full rank. In particular,

$$\mathbf{K}_+ = \begin{pmatrix} \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{3} & \frac{2}{3} & & 0 \\ \vdots & \vdots & \ddots & \\ \frac{1}{T-1} & \frac{2}{T-1} & \cdots & \frac{T-2}{T-1} \end{pmatrix},$$

such that $\mathbf{K}_+\Delta\ddot{\mathbf{u}}_i$ yields a vector of "backward-orthogonal deviations" $u_{it} - \frac{1}{2}\sum_{s=1}^{t} u_{is}$, $2 \leq t \leq T-1$. It turns out that this can be achieved by setting $\mathbf{R}_+ = \mathbf{K}_+$ with an appropriate choice of $\mathbf{R}_\pm$.[2]

For the backward-looking moment functions, it follows immediately from results in Arellano and Bover (1995) that there exists a nonsingular transformation matrix $\mathbf{R}_-$, such that $\mathbf{R}_-\dot{\mathbf{H}}_{i-}\Delta\dot{\mathbf{u}}_i = \dot{\mathbf{H}}_{i-}\mathbf{K}_-\Delta\dot{\mathbf{u}}_i$, where $\mathbf{K}_-$ is an upper-triangular transformation matrix of full rank. To complement the above example, we can choose

$$\mathbf{K}_- = \begin{pmatrix} -\frac{T-2}{T-1} & \cdots & -\frac{2}{T-1} & -\frac{1}{T-1} \\ & \ddots & \vdots & \vdots \\ 0 & & -\frac{2}{3} & -\frac{1}{3} \\ 0 & \cdots & 0 & -\frac{1}{2} \end{pmatrix},$$

such that $\mathbf{K}_-\Delta\dot{\mathbf{u}}_i$ is a vector of "forward-orthogonal deviations" $u_{it} - \frac{1}{T-1}\sum_{s=t}^{T} u_{is}$, $2 \leq t \leq T-1$. Put together, we can recast the whole set of moment functions equivalently in terms of deviations from forward and backward means:

$$\mathbf{R}\boldsymbol{\zeta}_i = \begin{pmatrix} \dot{\mathbf{H}}_{i-} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{H}}_{i+} \end{pmatrix} \begin{pmatrix} \mathbf{K}_- & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_+ \end{pmatrix} \begin{pmatrix} \Delta\dot{\mathbf{u}}_i \\ \Delta\ddot{\mathbf{u}}_i \end{pmatrix}. \tag{10}$$

This result applies to all valid transformations $\mathbf{K}_-$ and $\mathbf{K}_+$ that satisfy the above requirements.

While, in general, the test statistic is invariant to such a transformation, using deviations from forward and backward means can be beneficial if there are gaps in the panel data set. Such gaps would only lead to the loss of a single observation per gap because the missing observation can simply be ignored in the calculation of the forward and backward means. In contrast, with first differences each gap leads to the loss of two consecutive observations.

In the following, we restrict our attention to balanced panel data sets without gaps. Our main focus will be on transformation matrices $\mathbf{R}$ that reduce the dimension of the moment vector by selecting or linearly combining specific moments. Such dimensionality reduction approaches are no longer invariant to the initial choice of nonredundant moments.

2.3 Test statistic

Since the regression errors u_{it} are unobserved, a feasible test needs to be based on residuals $\hat{u}_{it} = y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}$. We can expand the estimated moment functions $\hat{\boldsymbol{\zeta}}_i = \hat{\mathbf{H}}_i'\Delta\hat{\mathbf{u}}_i$ to express them as a function

[2] For details, see Appendix A.

of $\boldsymbol{\zeta}_i = (\zeta_{1i}, \zeta_{2i}, \dots, \zeta_{ri})'$ and the estimation error $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$:

$$\hat{\boldsymbol{\zeta}}_i = \boldsymbol{\zeta}_i + \mathbf{D}_{1i}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \frac{1}{2}\left(\mathbf{I}_r \otimes \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right)\mathbf{D}_{2i}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right),$$

where

$$\mathbf{D}_{1i} = \frac{\partial \boldsymbol{\zeta}_i}{\partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial \zeta_{1i}}{\partial \boldsymbol{\beta}'} \\ \frac{\partial \zeta_{2i}}{\partial \boldsymbol{\beta}'} \\ \vdots \\ \frac{\partial \zeta_{ri}}{\partial \boldsymbol{\beta}'} \end{pmatrix} \quad \text{and} \quad \mathbf{D}_{2i} = \begin{pmatrix} \frac{\partial^2 \zeta_{1i}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \\ \frac{\partial^2 \zeta_{2i}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \\ \vdots \\ \frac{\partial^2 \zeta_{ri}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \end{pmatrix}.$$

It then follows from equation (2) and Jochmans (2020b, Appendix A) that

$$\sum_{i=1}^N \hat{\boldsymbol{\zeta}}_i = \sum_{i=1}^N \boldsymbol{\zeta}_i + \left(\boldsymbol{\Gamma}_1 + \frac{1}{2}\left(\mathbf{I}_r \otimes \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right)\boldsymbol{\Gamma}_2\right)\sum_{i=1}^N \boldsymbol{\omega}_i + o_p(\sqrt{N}), \tag{11}$$

where $\boldsymbol{\Gamma}_j = E[\mathbf{D}_{ji}]$. Since under the null hypothesis $(\mathbf{I}_r \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})')\boldsymbol{\Gamma}_2 \sum_{i=1}^N \boldsymbol{\omega}_i = o_p(\sqrt{N})$ itself, we can obtain the asymptotic distribution from

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{\boldsymbol{\zeta}}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^N (\boldsymbol{\zeta}_i + \boldsymbol{\Gamma}_1 \boldsymbol{\omega}_i) + o_p(1).$$

However, the extra term involving $\boldsymbol{\Gamma}_2$ will be useful for the (finite-sample) power analysis if the estimator $\hat{\boldsymbol{\beta}}$ becomes inconsistent under the alternative hypothesis, which we explore in more detail in Section 4. As pointed out by Jochmans (2020b), $\boldsymbol{\Gamma}_1$ generally differs from the null matrix unless all regressors $\mathbf{x}_{it}$ are strictly exogenous with respect to ε_{it} and their first differences $\Delta\mathbf{x}_{it}$ are uncorrelated with α_i. The latter is satisfied in a static random-effects model but does not automatically hold under a fixed-effects assumption. In dynamic models, the lagged dependent variable violates the strict-exogeneity assumption by construction.

With estimates of the Jacobian $\hat{\boldsymbol{\Gamma}}_1$ – see Jochmans (2020b) – and the influence function $\hat{\boldsymbol{\omega}}_i$, the test statistic is computed as[3]

$$\hat{s}_{R,N} = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}\hat{\boldsymbol{\zeta}}_i\right)'\left(\frac{1}{N}\sum_{i=1}^N \mathbf{R}(\hat{\boldsymbol{\zeta}}_i + \hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\omega}}_i)(\hat{\boldsymbol{\zeta}}_i + \hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\omega}}_i)'\mathbf{R}'\right)^-\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}\hat{\boldsymbol{\zeta}}_i\right), \tag{12}$$

with a suitable choice of the transformation matrix $\mathbf{R}$; for the portmanteau test, $\mathbf{R} = \mathbf{I}_r$. Denote the noncentral χ^2-distribution with r degrees of freedom and noncentrality parameter τ by $\chi^2(r, \tau)$. The following corollary immediately follows from Theorem 2 of Jochmans (2020b):

[3] $(\cdot)^-$ denotes a generalized inverse in case of rank deficiency.

Corollary 1: Let $E[\alpha_i^4] < \infty$, $E[\varepsilon_{it}^4] < \infty$, and $E[||\mathbf{x}_{it}||^4] < \infty$, equation (2) hold, $N^{-1}\sum_{i=1}^N ||\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}_i||^2 = o_p(1)$, and $\tilde{\mathbf{V}} = E[(\boldsymbol{\zeta}_i + \boldsymbol{\Gamma}_1\boldsymbol{\omega}_i)(\boldsymbol{\zeta}_i + \boldsymbol{\Gamma}_1\boldsymbol{\omega}_i)']$ be of maximal rank r. Then, as $N \to \infty$ and with T fixed,

(i) $\hat{s}_{R,N} \overset{d}{\to} \chi^2(r, 0)$ under the null hypothesis $E[\boldsymbol{\zeta}_i] = \mathbf{0}$;

(ii) $\hat{s}_{R,N} \overset{d}{\to} \chi^2(r, \tilde{\boldsymbol{\delta}}'\mathbf{R}'(\mathbf{R}\tilde{\mathbf{V}}\mathbf{R}')^{-1}\mathbf{R}\tilde{\boldsymbol{\delta}})$ with $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta} + \boldsymbol{\Gamma}_1 w(\boldsymbol{\delta})$ under the sequence of local alternative hypotheses $E[\boldsymbol{\zeta}_i] = \boldsymbol{\delta}/\sqrt{N}$, such that $E[\boldsymbol{\omega}_i] = w(\boldsymbol{\delta})/\sqrt{N}$, where $w(\mathbf{0}) = \mathbf{0}$.[4]

To demonstrate the problem of moment proliferation, let us consider the case where $\boldsymbol{\beta}$ is known.[5] Then, the test statistic simplifies to

$$s_{R,N} = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}\boldsymbol{\zeta}_i\right)' \left(\frac{1}{N}\sum_{i=1}^N \mathbf{R}\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i'\mathbf{R}'\right)^- \left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}\boldsymbol{\zeta}_i\right), \tag{13}$$

and we can establish the following finite-sample result:[6]

Proposition 1: Let $\text{rk}\left(\sum_{i=1}^N \mathbf{R}\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i'\mathbf{R}'\right) = \min(N, r)$ with probability 1. Then, $s_{R,N} = N$ for $r \geq N$.

Thus, when the number of moment restrictions r reaches or exceeds the cross-sectional sample size N, the test statistic becomes degenerate and the null hypothesis is never rejected. In practice, this curse of dimensionality can bite quickly. For example, empirical researchers might still consider T to be small in data sets with $T = 15$ and $N = 100$, but Proposition 1 already applies for the portmanteau test because $r = 104$. Moreover, power losses occur well before reaching the degenerate case.

2.4 Dimensionality reduction

The $(T-1) \times (T-1)(T-2)/2$ matrix $\mathbf{H}_{i-}$ has the same general structure as the matrix of instruments in the Arellano and Bond (1991) GMM estimator. Because a large number of instruments can lead to overfitting problems, instrument reduction strategies are commonly applied. The leading approaches include "curtailing" and "collapsing" (Roodman, 2009; Kiviet, 2020). Corresponding strategies can be applied to reduce the dimensionality of the moment restrictions to be tested. They can be implemented as linear combinations of the full set of moment restrictions.[7]

Curtailing refers to a selection of columns of matrix $\mathbf{H}_{i-}$ corresponding to the backward-looking moment restrictions

$$E[u_{i,t-s}\Delta u_{it}] = 0, \quad 3 \leq t \leq T, \quad 2 \leq s \leq q+1, \tag{14}$$

[4] $w(\boldsymbol{\delta}) = \mathbf{0}$ for all $\boldsymbol{\delta}$ if the estimator $\hat{\boldsymbol{\beta}}$ remains consistent under the alternative.

[5] For example, with $\boldsymbol{\beta} = \mathbf{0}$, this amounts to testing for serial correlation in the idiosyncratic component of an observed variable y_{it} without the need of estimating residuals.

[6] The proof of Proposition 1 is relegated to Appendix B.

[7] Appendix C provides a characterization of matrix $\mathbf{R}$ for the cases discussed below.

for some $q \geq 1$. Curtailing thus reduces the number of these moments from $(T-1)(T-2)/2$ to $q(2T-q-3)/2$. The forward-looking moments (6) remain unchanged.

Collapsing instead leads to linear combinations of the moment restrictions of the form

$$E\left[\sum_{t=1+s}^{T} u_{i,t-s}\Delta u_{it}\right] = 0, \quad 2 \leq s \leq T-1, \tag{15}$$

$$E\left[\sum_{t=2}^{T-1} u_{i,t+1}\Delta u_{it}\right] = 0, \tag{16}$$

which reduces their number to $T-2$ backward-looking moments and a single forward-looking moment. Both curtailing and collapsing limit the growth of the moment count to a linear instead of a quadratic rate in T. If the two approaches are combined, the total number of moment restrictions becomes $q+1$, independent of T.

Further below, we will also consider a variant of the test which combines all columns of $\mathbf{H}_{i+}$ and $\mathbf{H}_{i-}$ into a single moment restriction, which we shall call "full collapsing":

$$E\left[\sum_{t=2}^{T-1} u_{i,t+1}\Delta u_{it} - \sum_{s=2}^{T-1}\sum_{t=1+s}^{T} u_{i,t-s}\Delta u_{it}\right] = 0. \tag{17}$$

2.5 First differencing

Because the levels of u_{it} are a function of the group-specific error component α_i, a high variance of α_i can adversely affect the performance of the test by overshadowing the signal coming from ε_{it}. To circumvent this issue, a test can be constructed based entirely on first-differenced errors:

$$E[\Delta u_{i,t-s}\Delta u_{it}] = 0, \quad 4 \leq t \leq T, \quad 2 \leq s \leq t-2. \tag{18}$$

By combining this idea with the collapsing approach, we obtain the test proposed by Yamagata (2008), which is a joint test against serial correlation at order 2 or higher in the first-differenced errors:

$$E\left[\sum_{t=2+s}^{T} \Delta u_{i,t-s}\Delta u_{it}\right] = 0, \quad 2 \leq s \leq T-2. \tag{19}$$

By selecting the single moment restriction for $s = q+1$, we can test against serial correlation in the first-differenced errors of order $q+1$, which would be indicative for q-th order serial correlation in the level errors. This resembles the test proposed by Arellano and Bond (1991).[8]

While the general portmanteau test is applicable with as few as $T=3$ observations per group, a

[8] Arellano and Bond (1991) compute an asymptotically standard-normally distributed version of this test.

test based on first differences requires $T \geq 4$. To be different from the Arellano-Bond test, Yamagata's test requires $T \geq 5$.

2.6 S-differencing

The tests based on first differences entirely ignore the forward-looking moment restrictions, which is inefficient and can lead to substantial power losses. In particular, when the idiosyncratic error component follows a random walk – i.e., $\varepsilon_{it} = \varepsilon_{i,t-1} + \nu_{it}$ with independently distributed innovations ν_{it} – the test has no power because Δu_{it} remains serially uncorrelated. To overcome this shortcoming while retaining the benefit of invariance to α_i, we propose an alternative differencing approach by linearly combining columns from matrix $\mathbf{H}_{i-}$ with those from $\mathbf{H}_{i+}$. This yields covariance restrictions between longer differences $\Delta_{s^*,s} u_{it} = u_{i,t+s^*} - u_{i,t-s}$, $s^* \geq 1$ and $s \geq 2$, and first differences $\Delta_{0,1} u_{it} = \Delta u_{it} = u_{it} - u_{i,t-1}$:

$$E[\Delta_{1,s} u_{it} \Delta u_{it}] = 0, \quad 4 \leq t \leq T-1, \quad 2 \leq s \leq t-2. \tag{20}$$

It suffices to focus on $s^* = 1$ because additional moment restrictions for $s^* > 1$ become redundant. Importantly, the long differences are encompassing the first differences. We refer to this transformation as S-differencing, which may be derived from "seasonal differencing"[9] or "sandwich differencing" (because of the encompassing structure). These moments remain informative under the random-walk alternative because $Cov(\varepsilon_{i,t+1}, \Delta\varepsilon_{it}) = Var(\nu_{it})$. As before, to address the issue of too many moments, we can combine the S-differencing strategy with curtailing and collapsing.

3 Power comparisons

To simplify the power calculations, let $\boldsymbol{\beta} = \mathbf{0}$ to avoid the complication of estimating $\boldsymbol{\beta}$ in a first stage. The test is then directly computed with the observed $y_{it} = u_{it}$. Analogously to Jochmans (2020b), we consider either a first-order moving-average (MA(1)) alternative,

$$\varepsilon_{it} = \nu_{it} + \theta \nu_{i,t-1}, \tag{21}$$

or a first-order autoregressive (AR(1)) alternative,

$$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + \nu_{it}. \tag{22}$$

ν_{it} is independently distributed with mean zero and variance $\sigma^2_{\nu,t}$, and the variance of α_i is σ^2_α. Initially, we focus on the stationary case, where $\sigma^2_{\nu,t} = \sigma^2_\nu$ and the initial observations ε_{i0} under the AR(1)

[9] In *Stata*, long differences are generated with the seasonal-differencing operator `S`.

Table 1: Noncentrality parameters against stationary alternatives

	dimensionality reduction	noncentrality parameter	degrees of freedom
$T=3$	–	$\tau = \frac{\kappa^2}{3}\left(\frac{2}{\eta+1}\right)$	$r=2$
	full collapsing	$\tau_{fc} = \frac{\kappa^2}{3}\left(\frac{2}{\eta+1}\right)$	$r_{fc}=1$
$T=4$	–	$\tau = \frac{\kappa^2}{6}\left(\frac{4(3+\psi^2)\eta+9+6\psi+5\psi^2}{2\eta+1}\right)$	$r=5$
	curtailing ($q=1$)	$\tau_1 = \kappa^2\left(1+\frac{1}{4\eta+5}\right)$	$r_1=4$
	collapsing	$\tau_c = \frac{\kappa^2}{2}\left(\frac{2(8-4\psi+\psi^2)\eta+16+8\psi+7\psi^2}{\eta(\eta+11)+6}\right)$	$r_c=3$
	full collapsing	$\tau_{fc} = \frac{\kappa^2}{2}\left(\frac{(4+\psi)^2}{5\eta+6}\right)$	$r_{fc}=1$
	first differences	$\tau_\Delta = \frac{\kappa^2}{4}(1-\psi)^2$	$r_\Delta=1$
	S-differences	$\tau_{\Delta_{1,2}} = \kappa^2$	$r_{\Delta_{1,2}}=1$

Note: $\eta=\frac{\sigma_\alpha^2}{\sigma_\nu^2}$; $\kappa=\theta$ and $\psi=0$ against an MA(1) alternative, $\kappa=\frac{\rho}{1+\rho}$ and $\psi=\rho$ against an AR(1) alternative.

alternative are independently distributed with mean zero and variance $\sigma_\nu^2/(1-\rho^2)$. Under the MA(1) alternative, the expected values of the moment restrictions are

$$E[u_{i,t-s}\Delta u_{it}] = \begin{cases} -\theta\sigma_\nu^2, & 3\le t\le T, \quad s=2, \\ 0, & 3\le t\le T, \quad 2<s\le t-1, \end{cases}$$
$$E[u_{i,t+1}\Delta u_{it}] = \theta\sigma_\nu^2, \quad 2\le t\le T-1.$$

Under the AR(1) alternative, they are

$$E[u_{i,t-s}\Delta u_{it}] = -\frac{\rho^{s-1}}{1+\rho}\sigma_\nu^2, \quad 3\le t\le T, \quad 2\le s\le t-1,$$
$$E[u_{i,t+1}\Delta u_{it}] = \frac{\rho}{1+\rho}\sigma_\nu^2, \quad 2\le t\le T-1.$$

The test statistic is given in equation (13). As a special case of Corollary 1, $s_{R,N} \xrightarrow{d} \chi^2(r,0)$ under the null hypothesis, and $s_{R,N} \xrightarrow{d} \chi^2(r, \boldsymbol{\delta}'\mathbf{R}'(\mathbf{RVR}')^{-1}\mathbf{R}\boldsymbol{\delta})$ under the sequence of local alternatives $E[\boldsymbol{\zeta}_i] = \boldsymbol{\delta}/\sqrt{N}$, where $\mathbf{V} = E[\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i']$ is the variance-covariance matrix of the full set of moment restrictions under the null hypothesis; see also Theorem 1 in Jochmans (2020b). For known variance-covariance matrix $\mathbf{V}$, we can compute the theoretical power as the tail probability from the noncentral χ^2-distribution.[10]

Table 1 lists the resulting noncentrality parameters for three- and four-wave panels against stationary MA(1) or AR(1) alternatives. In the following, we will use these analytical expressions to add some insights to the power comparisons carried out by Jochmans (2020b).[11] Besides considering the

[10]See Appendix D for a general characterization of matrix $\mathbf{V}$.

[11]Notice that for $T=4$ the noncentrality parameter τ for the portmanteau test without moment reduction is different from the one presented by Jochmans (2020b). Notwithstanding this discrepancy, for the special case $\sigma_\alpha^2=0$ primarily considered in the power comparison exercise by Jochmans (2020b), his expression yields the correct value. Hence, his

dimensionality reduction techniques, we also have a closer look at the performance under different variance ratios $\eta = \sigma_\alpha^2/\sigma_\nu^2$.

3.1 Three-wave panel

While our main focus is on moderately small T, it is instructive to consider the case of very small T first. When $T = 3$, there are two moments with the following expectation under the stationary alternatives:

$$E[\boldsymbol{\zeta}_i] = E\left[\begin{pmatrix} u_{i1}\Delta u_{i3} \\ u_{i3}\Delta u_{i2} \end{pmatrix}\right] = \kappa\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

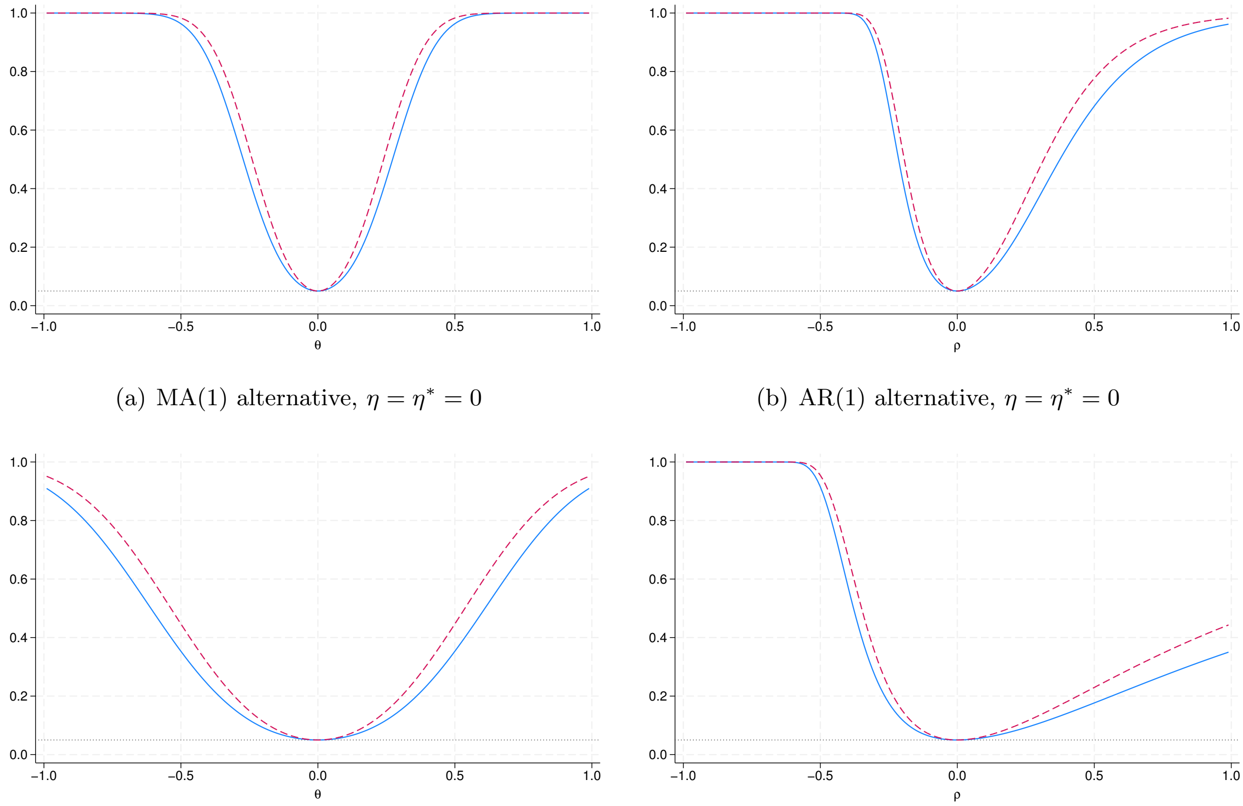
with $\kappa = \theta$ for the moving-average case, and $\kappa = \rho/(1+\rho)$ for the autoregressive case. Other alternatives, such as second-order moving-average or autoregressive processes, lead to the same expression by amending κ accordingly.

Clearly, not every linear combination yields a useful test statistic. Simply adding up both restrictions with $\mathbf{R} = (1, 1)$ yields a test with no power against either alternative, because $E[\mathbf{R}\boldsymbol{\zeta}_i] = 0$ for all values of κ. In contrast, the noncentrality parameter is maximized by $\mathbf{R} = (-1, 1)$ (up to sign and scale), which corresponds to our fully collapsed test. It turns out that the resulting noncentrality parameter is identical to the one without dimensionality reduction. However, for the same noncentrality parameter, a test with 1 degree of freedom is uniformly more powerful than one with 2 degrees of freedom.

The theoretical power improvements from linearly combining the two moment restrictions are demonstrated in Figure 1 for $\eta = 0$ (no group-specific effects) and $\eta = 4$. Local power is computed for a cross-sectional sample size of $N = 100$. The solid line depicts the test with 2 degrees of freedom, while the dashed line shows the power of the fully collapsed test. The other discussed options for dimensionality reduction are not applicable for $T = 3$. The power profiles against a moving-average alternative are symmetric, while they are asymmetric against the autoregressive alternative. In the latter case, the power improvements from full collapsing are more pronounced for positive values of ρ. A larger variance ratio η is always power reducing, which is obvious from the effect of η on the noncentrality parameter.

It needs to be noted that the fully collapsed test is generally no longer uniformly more powerful under deviations from stationarity. Exemplarily, we illustrate this for scenarios with increasing variance over time, $\sigma_{\nu,t}^2 = t/2$, or decreasing variance over time, $\sigma_{\nu,t}^2 = 2 - t/2$. We restrict our attention to $\sigma_\alpha^2 = 4$, but note that a smaller variance of the group-specific effects tends to benefit the fully

qualitative conclusions for this case continue to hold. We have numerically verified that our analytical expressions for the variance-covariance matrix and the noncentrality parameters are correct for any value of σ_α^2.



(a) MA(1) alternative, $\eta = \eta^* = 0$

(b) AR(1) alternative, $\eta = \eta^* = 0$

(c) MA(1) alternative, $\eta = 4$

(d) AR(1) alternative, $\eta = 4$

Figure 1: Theoretical power calculations for $T = 3$ against stationary alternatives

collapsed test more. The initial observations under the AR(1) alternative are set to $\varepsilon_{i0} = 0$. The implications are shown in Figure 2. The test with two degrees of freedom has a particular advantage against highly persistent autoregressive alternatives. However, it is difficult to generalize these results because they depend on the specific assumptions about the error variances.

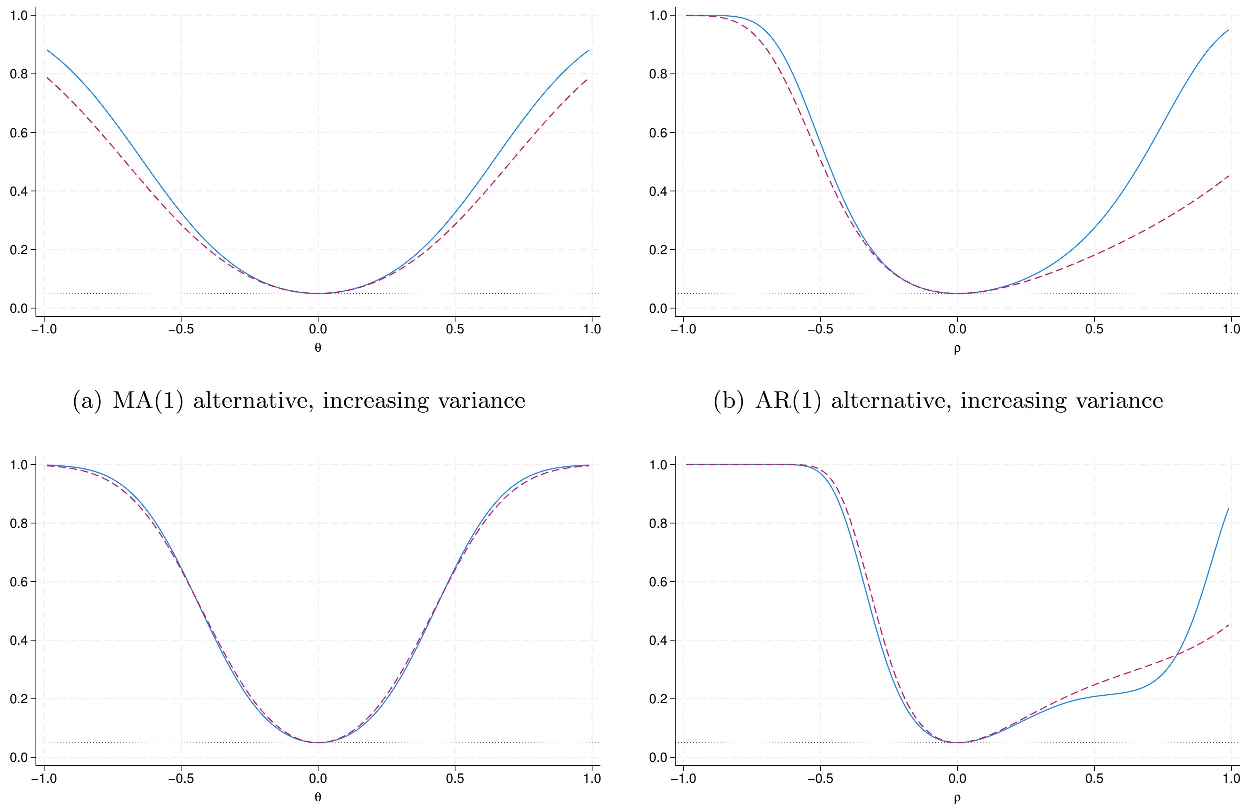
3.2 Four-wave panel

When $T = 4$, there are five moments with the following expectation under the stationary MA(1) or AR(1) alternatives:

$$E[\boldsymbol{\zeta}_i] = E\left[\begin{pmatrix} u_{i1}\Delta u_{i3} \\ u_{i1}\Delta u_{i4} \\ u_{i2}\Delta u_{i4} \\ u_{i3}\Delta u_{i2} \\ u_{i4}\Delta u_{i3} \end{pmatrix}\right] = \kappa\sigma_\nu^2 \begin{pmatrix} -1 \\ -\psi \\ -1 \\ 1 \\ 1 \end{pmatrix},$$

with $\kappa = \theta$ and $\psi = 0$ for the moving-average case, and $\kappa = \rho/(1+\rho)$ and $\psi = \rho$ for the autoregressive case.

It is interesting to note that unlike the $T = 3$ case, the Jochmans (2020b) test with all moment



(a) MA(1) alternative, increasing variance

(b) AR(1) alternative, increasing variance

(c) MA(1) alternative, decreasing variance

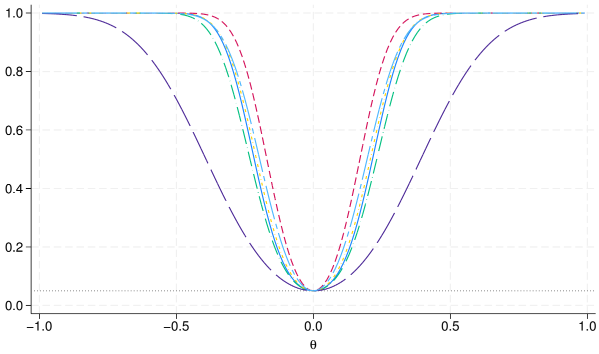
(d) AR(1) alternative, decreasing variance

Figure 2: Theoretical power calculations for $T = 3$ under heteroskedasticity

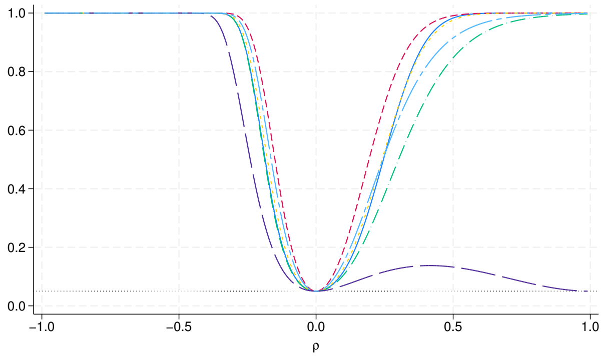
restrictions now retains power even as $\eta \to \infty$, since $\lim_{\eta\to\infty} \tau = \kappa^2(3+\psi^2)/3 > 0$.[12] Intuitively, while some power is lost, it does not vanish completely because moment restrictions in first differences or S-differences, free of α_i, can be formed from suitable linear combinations of the moments. For example, combining the second and third moment results in the Arellano and Bond (1991) test. The latter is nevertheless dominated by the Jochmans test even after taking the degrees-of-freedom reduction into account.

For the collapsed test, $\lim_{\eta\to\infty} \tau_c = 0$. The reason for this is that the collapsed moment restrictions can no longer be combined into differenced moment restrictions. The latter is still possible for the curtailed test, which just discards the second moment restriction. A linear combination of the first and last moment yields $E[(u_{i4} - u_{i1})\Delta u_{i3}]$, which again is free of the group-specific effects. While this S-differencing test is not the most powerful one when $\eta = 0$, the balance already shifts in its favor for relatively small variance ratios. Importantly, this test does not lose power against autoregressive alternatives when ρ approaches 1, unlike tests based on first differences. The fully collapsed test, which was strictly preferred for $T = 3$ against stationary alternatives, still dominates the other variants for

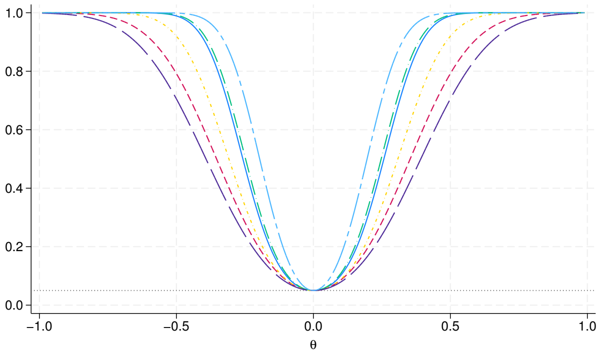
[12] This result stands in contrast to the noncentrality parameter reported by Jochmans (2020b), which incorrectly suggests that power goes to zero with increasing variance of the group-specific effects.



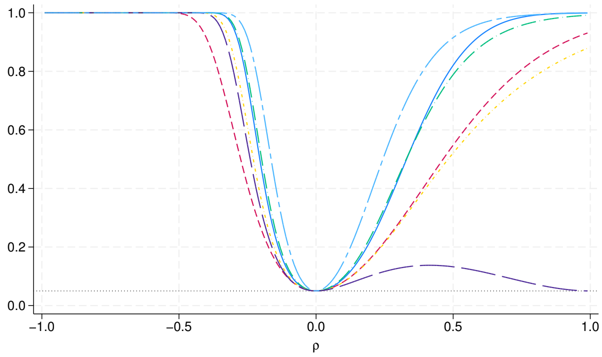
(a) MA(1) alternative, $\eta = 0$



(b) AR(1) alternative, $\eta = 0$



(c) MA(1) alternative, $\eta = 4$



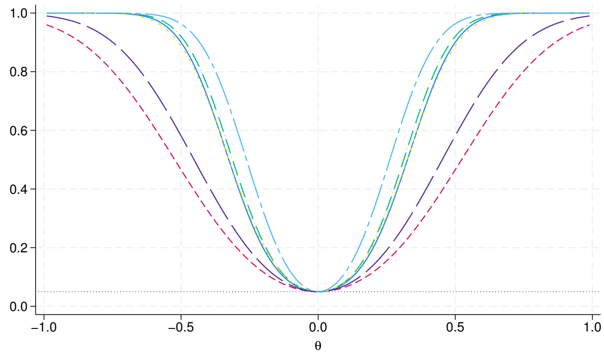
(d) AR(1) alternative, $\eta = 4$

Figure 3: Theoretical power calculations for $T = 4$ against stationary alternatives

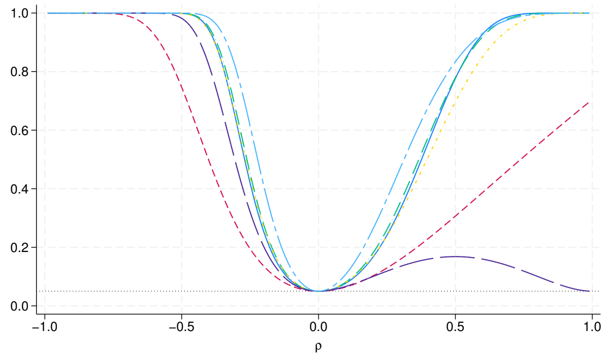
$\eta = 0$ but quickly loses power with increasing variance ratio.

Figure 3 summarizes these observations graphically. The solid and dashed lines correspond again to the Jochmans test and the fully collapsed test, respectively. The longdashed-dotted line represents the curtailed test, the shortdashed-dotted line the collapsed test, and the very-long-dashed line the Arellano-Bond test. Finally, the line alternating between very long dashes and short short dashes refers to the sandwich-differenced test.

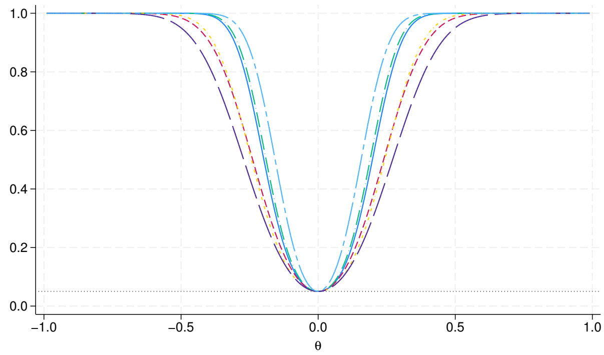
The effects of heteroskedasticity are not as pronounced anymore as in the three-wave case. We consider again scenarios with increasing variance over time, $\sigma^2_{\nu,t} = t/2 - 1/4$, or decreasing variance over time, $\sigma^2_{\nu,t} = 9/4 - t/2$. As with $T = 3$, $\sigma^2_\alpha = 4$ and $\varepsilon_{i0} = 0$. The power profiles are shown in Figure 4. Under an increasing variance, it is particularly the fully collapsed test which suffers the most. The other tests are only mildly affected. Under a decreasing variance, some of the tests no longer have monotonically increasing power functions for positive serial correlation under the AR(1) alternative. Again, it is difficult to draw general conclusions.



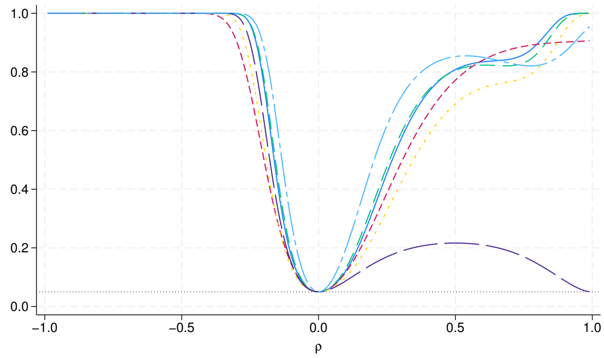
(a) MA(1) alternative, increasing variance



(b) AR(1) alternative, increasing variance



(c) MA(1) alternative, decreasing variance

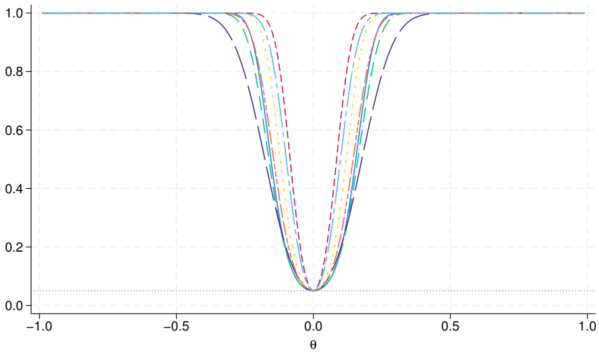


(d) AR(1) alternative, decreasing variance

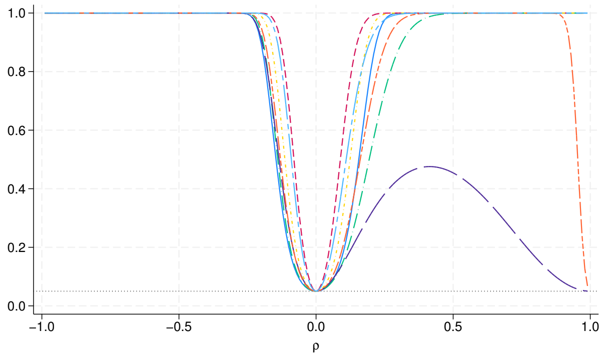
Figure 4: Theoretical power calculations for $T = 4$ under heteroskedasticity

3.3 Moderately small T

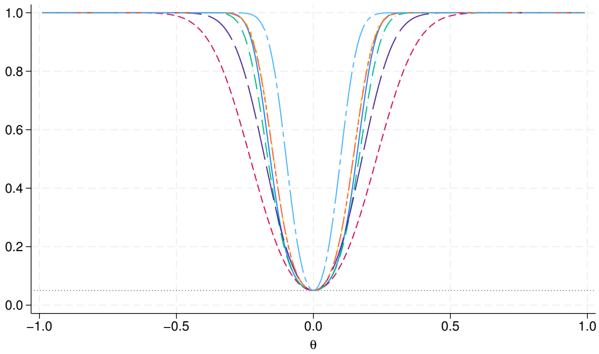
For larger T, analytical expressions of the noncentrality parameters become increasingly complex without additional insights. We thus focus on numerically obtained power profiles. Figure 5 shows the theoretical power for the same moment reduction techniques as before for $T = 10$. Additionally, the Yamagata (2008) collapsed first-difference test is included here, presented with alternating triple long and triple short dashes. The graphs generally tell a similar story as for $T = 4$. The differences between the various test versions are now less pronounced because more time periods generally improve the detectability of serial correlation. Like the Arellano and Bond (1991) test, the power of the Yamagata (2008) test eventually drops towards zero for alternatives close to a random walk, but it shows satisfactory power for most other parameter values. An important observation can be made about the Jochmans (2020b) test without moment reduction (solid line). With increasing T, it loses power relative to most of the tests that are more conservative in their use of the degrees of freedom. The results are again supportive of our new S-differencing approach, especially for higher variance ratios. Note that we have combined S-differencing with collapsing and curtailing ($q = 1$) in this analysis, thus yielding a test with 1 degree of freedom. S-differencing on its own, without further



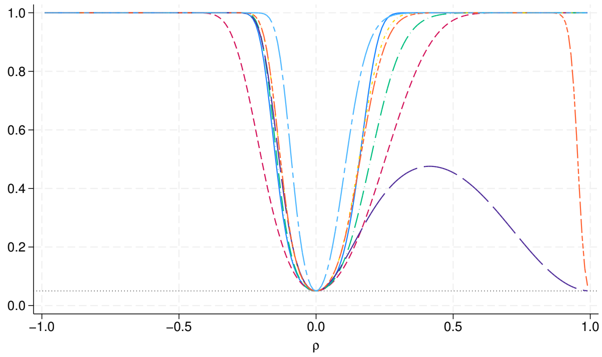
(a) MA(1) alternative, $\eta = 0$



(b) AR(1) alternative, $\eta = 0$



(c) MA(1) alternative, $\eta = 4$



(d) AR(1) alternative, $\eta = 4$

Figure 5: Theoretical power calculations for $T = 10$ against stationary alternatives

moment reductions, results in slightly lower power (not shown in Figure 5), although still favorable compared to most competitors.

4 Estimator inconsistency under the alternative hypothesis

We now turn our attention to a setting where the coefficients $\boldsymbol{\beta}$ are estimated in a first step. While the estimator $\hat{\boldsymbol{\beta}}$ is assumed to satisfy equation (2), it might be asymptotically biased under the alternative hypothesis. This is typically the case in dynamic panel models. For instance, GMM estimators (Arellano and Bond, 1991; Ahn and Schmidt, 1995; Arellano and Bover, 1995; Blundell and Bond, 1998) utilize internal instruments formed from lags of endogenous and predetermined regressors. These instruments become invalid under serially correlated errors. As a consequence, serial-correlation tests based on inconsistently estimated residuals can become inapplicable (Jung, 2005).

Recall that the noncentrality parameter is $\tilde{\boldsymbol{\delta}}'\mathbf{R}'(\mathbf{R}\tilde{\mathbf{V}}\mathbf{R}')^{-1}\mathbf{R}\tilde{\boldsymbol{\delta}}$ under local alternatives $E[\boldsymbol{\zeta}_i] = \boldsymbol{\delta}/\sqrt{N}$. Asymptotic bias of $\hat{\boldsymbol{\beta}}$ is reflected in $E[\boldsymbol{\omega}_i] = w(\boldsymbol{\delta})/\sqrt{N}$, which can have nontrivial power implications. Equation (11) implies

$$\sqrt{N}E[\hat{\boldsymbol{\zeta}}_i] = \boldsymbol{\delta} + \boldsymbol{\Gamma}_1 w(\boldsymbol{\delta}) + \frac{1}{2\sqrt{N}}\left(\mathbf{I}_r \otimes w(\boldsymbol{\delta})'\right)\boldsymbol{\Gamma}_2 w(\boldsymbol{\delta}) + o\left(\frac{1}{\sqrt{N}}\right),$$

such that $\lim_{N\to\infty}\sqrt{N}E[\hat{\boldsymbol{\zeta}}_i] = \boldsymbol{\delta}+\boldsymbol{\Gamma}_1 w(\boldsymbol{\delta}) = \tilde{\boldsymbol{\delta}}$. While the third term on the right side is asymptotically negligible, it helps to improve the approximation of the finite-sample power profiles. Moreover, this second-order effect on the noncentrality parameter can play a substantial role under fixed alternatives rather than local alternatives, as in the Monte Carlo simulations of the next section.

Exemplarily, let us consider the simple stationary AR(1) panel data model, $\mathbf{x}_{it} = y_{i,t-1}$, and the Anderson and Hsiao (1981) just-identified IV estimator

$$\hat{\beta} = \frac{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-2}\Delta y_{it}}{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-2}\Delta y_{i,t-1}}, \tag{23}$$

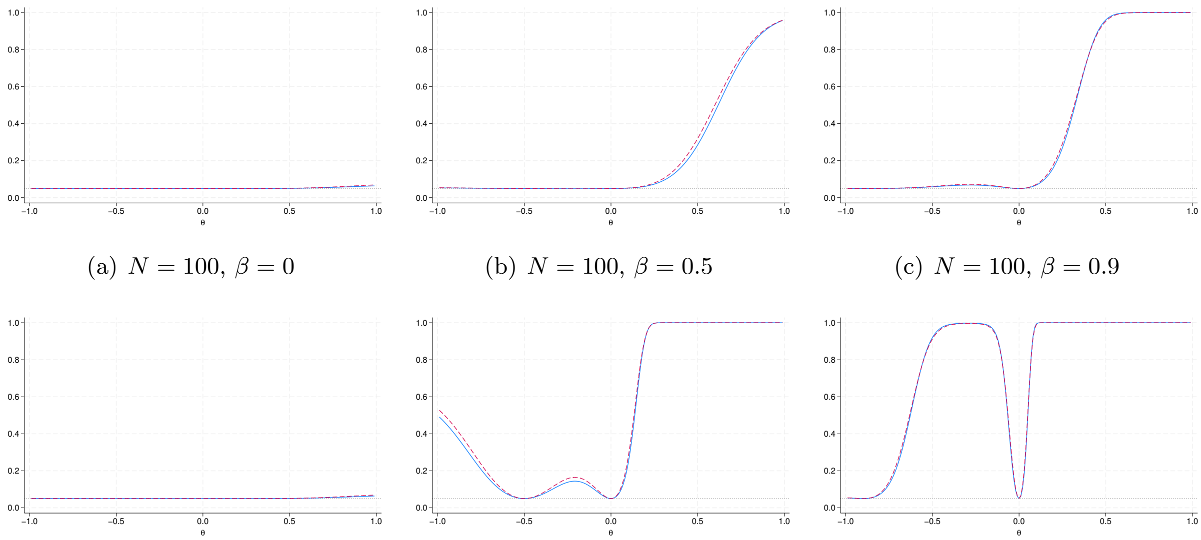
with $y_{i,t-2}$ as an instrument for the lagged dependent variable in the first-differenced regression model. In Appendix E, we obtain analytical expressions for $E[\boldsymbol{\zeta}_i]$, $E[\boldsymbol{\omega}_i]$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Gamma}_2$.

When the errors ε_{it} follow an MA(1) process, we find that $\boldsymbol{\Gamma}_1 E[\omega_i] = -E[\boldsymbol{\zeta}_i]$ and $\boldsymbol{\Gamma}_2 = \mathbf{0}$ if $\theta = -\beta$. The effect of the bias in the estimation of β on the noncentrality parameter exactly offsets the signal from the violation of the moment restrictions. This is intuitive because the data-generating process (DGP) in this case is observationally equivalent to

$$y_{it} = \frac{\alpha_i}{1-\beta} + \nu_{it},$$

where ν_{it} is serially uncorrelated. The estimator consistently estimates the pseudo-true value of the lagged dependent variable's coefficient $\beta^* = 0$ under this static representation of the DGP. This result holds for any T.

When $T = 3$, it turns out that $\boldsymbol{\Gamma}_1 E[\omega_i] = -E[\boldsymbol{\zeta}_i]$ also if $\beta = 0$. That is, a first-order expansion of $\hat{\boldsymbol{\zeta}}_i$ would suggest that the test cannot detect the serial correlation when y_{it} follows a simple MA(1) process. However, the asymptotically negligible second-order term indicates that the test remains informative in finite samples or against fixed alternatives, because $\boldsymbol{\Gamma}_2$ is now nonzero. This is highlighted in Figure 6, where we illustrate the power under a variance ratio $\eta = 1$ for two different sample sizes, $N \in \{100, 10000\}$. Panels (a) and (d) demonstrate the almost flat power curves when the true DGP is static instead of dynamic ($\beta = 0$). The minimal increase in power as $\theta \to 1$ is a result of the second-order term in the expansion of $E[\hat{\boldsymbol{\zeta}}_i]$, which does not improve with increasing sample size. Importantly, under fixed alternatives – when the second-order term is not scaled down by $N^{-1/2}$ – these tests can show substantial power in this region. For $\beta \neq 0$, we can clearly see the power loss when $\theta \to -\beta$. This is much more pronounced for large N. for small sample sizes, the power can still be virtually flat at the nominal size for a wide range of alternatives as a consequence of the estimation error in $\hat{\beta}$. It is also worth noting that there is hardly any gain from linearly combining the two



(a) $N = 100$, $\beta = 0$ (b) $N = 100$, $\beta = 0.5$ (c) $N = 100$, $\beta = 0.9$

(d) $N = 10000$, $\beta = 0$ (e) $N = 10000$, $\beta = 0.5$ (f) $N = 10000$, $\beta = 0.9$

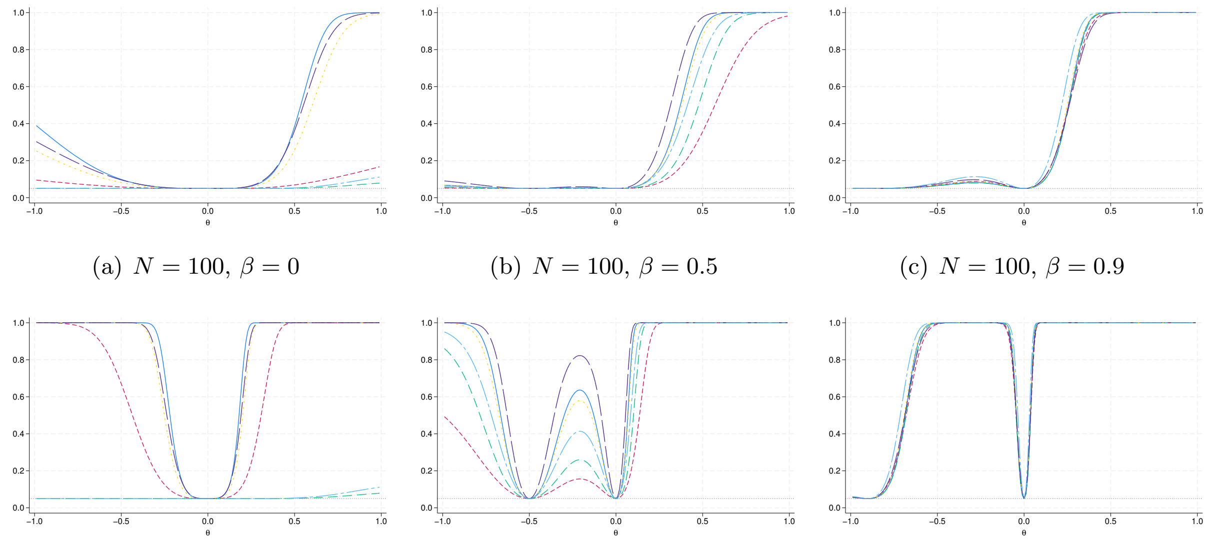
Figure 6: Theoretical power calculations for an MA(1) alternative with $T = 3$ under estimator inconsistency

moments in this scenario.

When $T \geq 4$, the "uninformative" elements of $E[\boldsymbol{\zeta}_i]$ – such as $E[u_{i1}\Delta u_{i4}]$ – lend power to the test, even under first-order asymptotics. While they are zero both under the null and the alternative hypothesis, the corresponding element of $\mathbf{\Gamma}_1 E[\omega_i]$ is nonzero due to the asymptotic bias of the estimator $\hat{\beta}$. This is particularly relevant when $\beta = 0$, in which case the "informative" moments provide virtually no power. As a logical consequence, the better performing variants of the test are those which put a large weight on the "uninformative" moments, while the picture is unchanged compared to $T = 3$ for those which do not make use of them. This is demonstrated in Figure 7.

For $\beta \neq 0$, the insights are similar to $T = 3$, although there is now more variation between the different versions of the test. When $\beta = 0.5$, it is remarkable that the highest power is achieved by the Arellano and Bond (1991) test, which had a tendency to perform worst when not accounting for the estimator inconsistency. There is still no power against $\theta = -\beta$ due to the earlier equivalence result.

The results in this section are specific to the chosen DGP and estimator $\hat{\beta}$. Including exogenous regressors in the model can already break the above implications. We analyze such models by means of Monte Carlo simulations in the next section. Nevertheless, the above results show that estimator inconsistency under the alternative hypothesis can have serious adverse effects on the power of serial-correlation tests.



(a) $N = 100,\ \beta = 0$ (b) $N = 100,\ \beta = 0.5$ (c) $N = 100,\ \beta = 0.9$

(d) $N = 10000,\ \beta = 0$ (e) $N = 10000,\ \beta = 0.5$ (f) $N = 10000,\ \beta = 0.9$

Figure 7: Theoretical power calculations for an MA(1) alternative with $T = 4$ under estimator inconsistency

5 Monte Carlo simulations

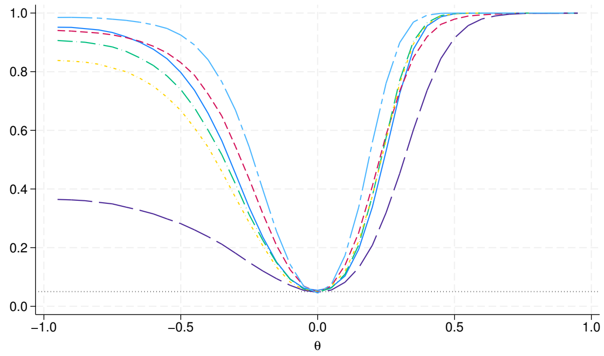
The asymptotic approximation of the test statistic's distribution used for the theoretical power comparison can be quite inaccurate when N is relatively small. We therefore resort to Monte Carlo simulations for a better understanding of the finite-sample properties of the various tests.

5.1 Static model

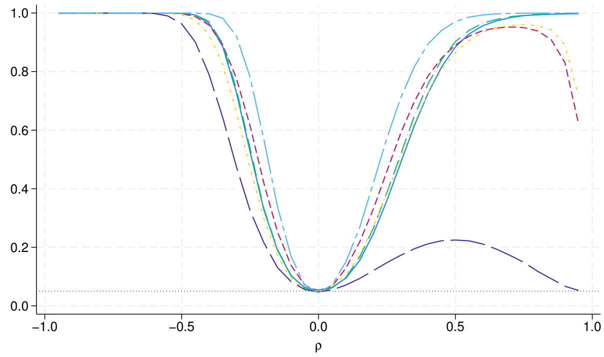
We first consider the following static DGP:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}, \tag{24}$$

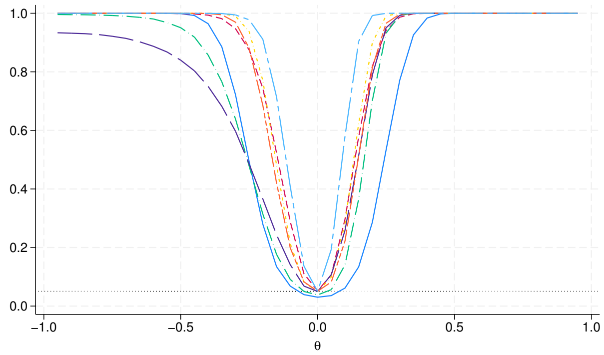
where x_{it} and α_i are independently standard-normally distributed, and ε_{it} either follows an MA(1) or AR(1) process as in equations (21) and (22), respectively. For both alternatives, we consider two scenarios. In the baseline case, the innovations ν_{it} are independently normally distributed with mean zero and constant variance σ_ν^2, and the initial observations are drawn from the stationary distribution. In the second case, we introduce time series heteroskedasticity of the form $\sigma_{\nu,t}^2 = [\tanh(2t/T)/\tanh(2)]\sigma_\nu^2$, which implies nonstationary initial observations $\varepsilon_{i0} = \nu_{i0} = 0$ and an increasing variance over time. To keep the variance ratio $Var(\alpha_i)/Var(\varepsilon_{it})$ constant at 1 in the homoskedastic case – and as $t \to T$ in the heteroskedastic case – while varying the degree of serial correlation, we set $\sigma_\nu^2 = 1/(1+\theta^2)$ under the MA(1) alternative and $\sigma_\nu^2 = (1-\rho^2)$ under the AR(1) alternative. This also ensures a constant signal-to-noise ratio for the estimation of β. Across simulations, we hold $N = 100$ fixed while varying



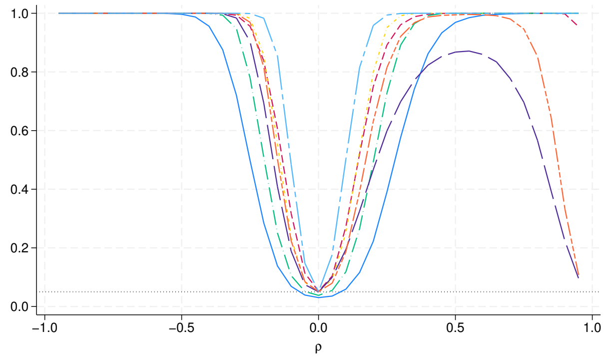
(a) MA(1) alternative, $T = 4$



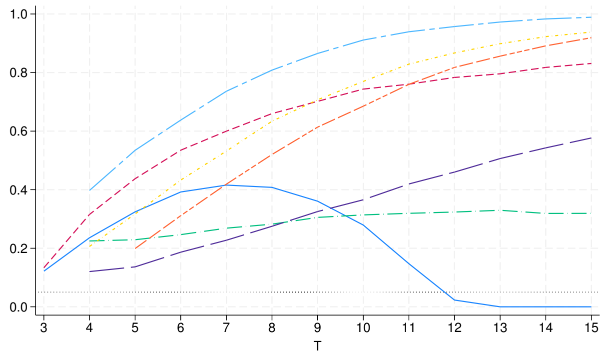
(b) AR(1) alternative, $T = 4$



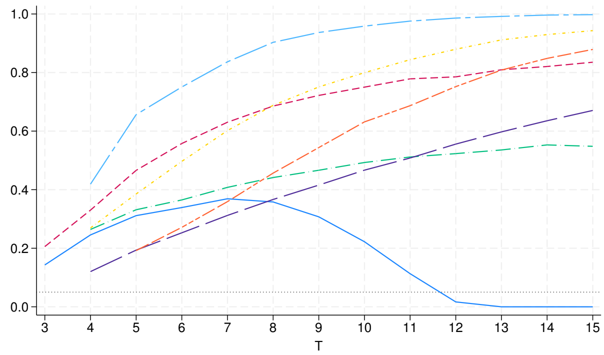
(c) MA(1) alternative, $T = 10$



(d) AR(1) alternative, $T = 10$



(e) MA(1) alternative, $\theta = -0.2$

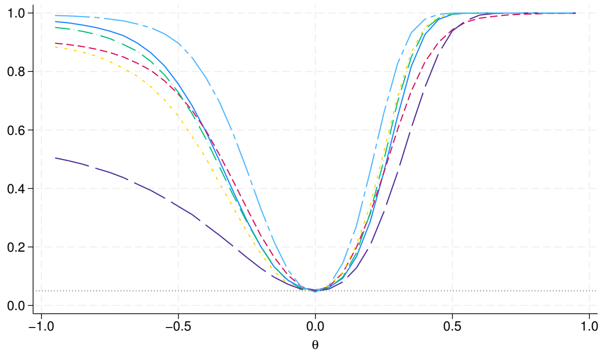


(f) AR(1) alternative, $\rho = 0.2$

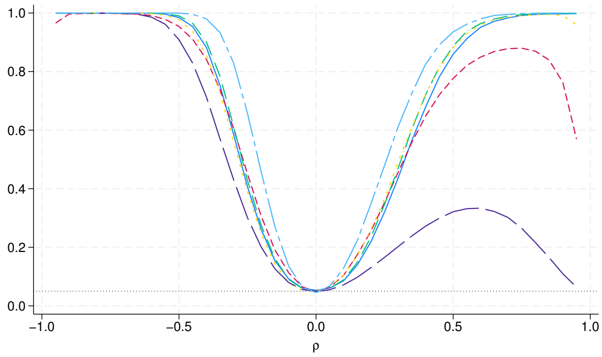
Figure 8: Simulated power against stationary alternatives

$T \in [3, 15]$. For each simulation, we perform 10,000 replications.

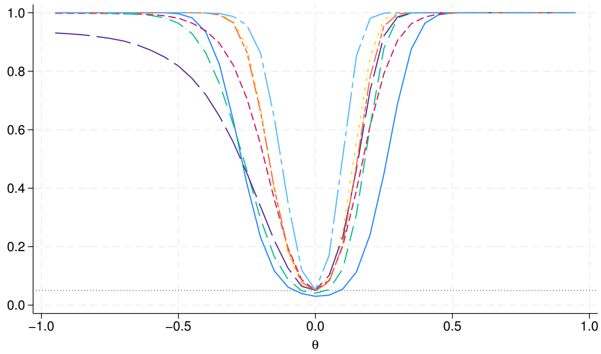
We set $\beta = 1$ and compute the same tests as in the previous sections, but this time using the residuals $\hat{u}_{it} = y_{it} - \hat{\beta} x_{it}$, where $\hat{\beta}$ is the conventional within-groups ("fixed-effects") estimator. The results are shown in Figures 8 and 9. It is apparent that the portmanteau test with all moment restrictions reaches maximum power around $T = 7$ and then quickly loses power for higher T. It also becomes undersized. Among the newly proposed tests, the S-differencing test tends to outperform all other tests both in the stationary and the heteroskedastic case. Most of the tests are largely unaffected by the heteroskedasticity. The only exception is the fully-collapsed test, which noticeably loses power.



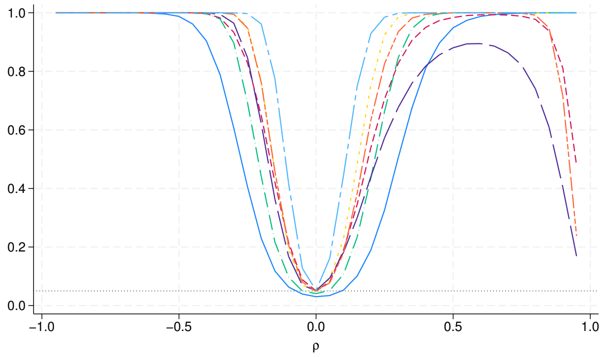
(a) MA(1) alternative, $T = 4$



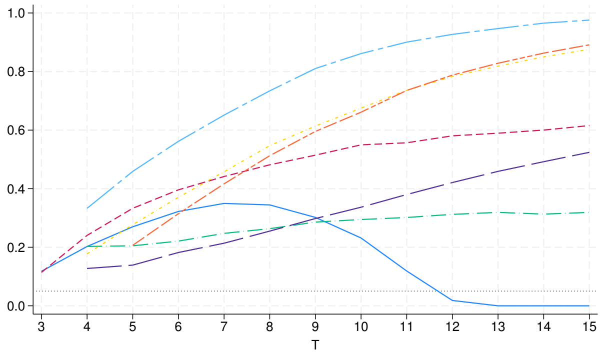
(b) AR(1) alternative, $T = 4$



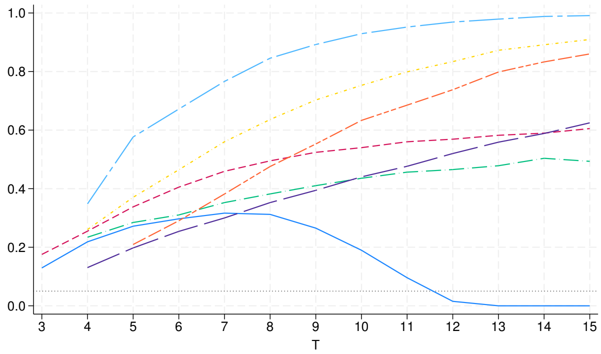
(c) MA(1) alternative, $T = 10$



(d) AR(1) alternative, $T = 10$



(e) MA(1) alternative, $\theta = -0.2$



(f) AR(1) alternative, $\rho = 0.2$

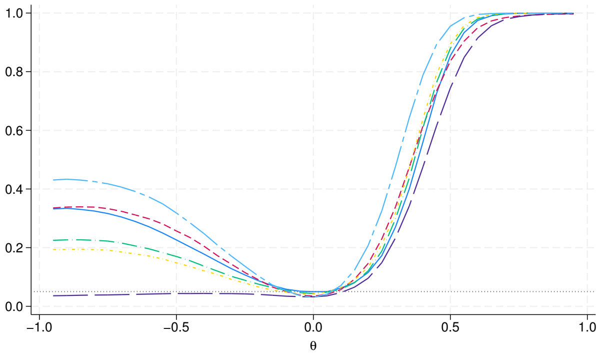
Figure 9: Simulated power against nonstationary alternatives

The Arellano-Bond test, on the other side, performs slightly better under heteroskedasticity when T is very small.

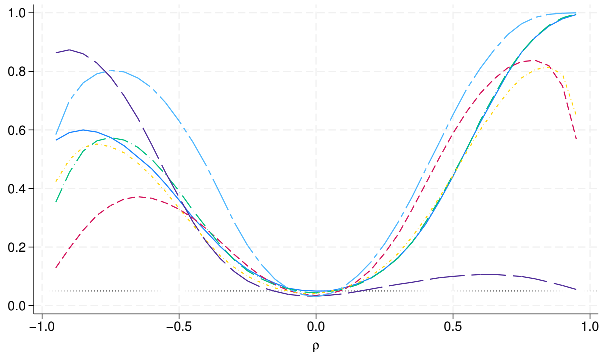
5.2 Dynamic model

To investigate the power implications of using an initial estimator that becomes inconsistent under the alternative hypothesis, we consider a dynamic DGP:

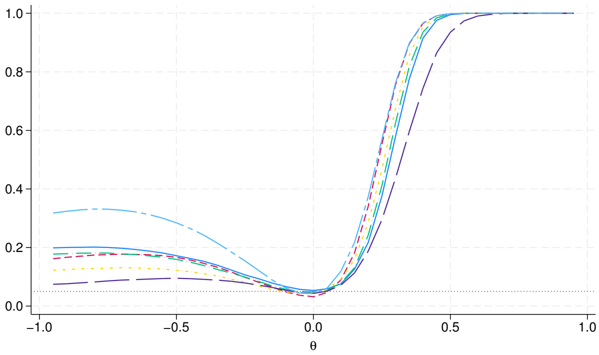
$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \omega(\alpha_i + \varepsilon_{it}). \tag{25}$$



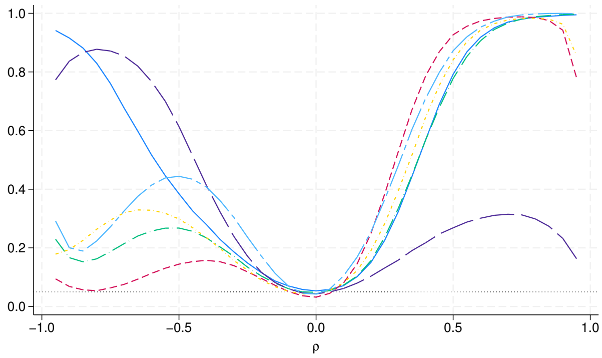
(a) MA(1) alternative, $T = 4$, $\lambda = 0$



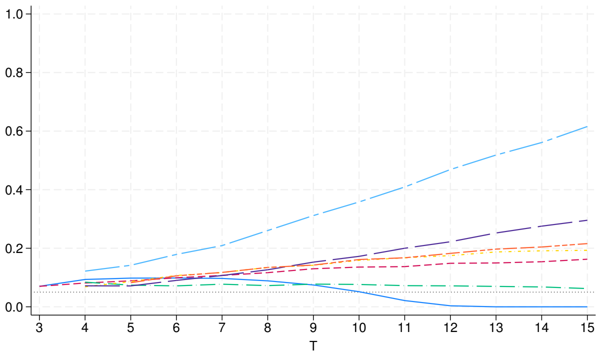
(b) AR(1) alternative, $T = 4$, $\lambda = 0$



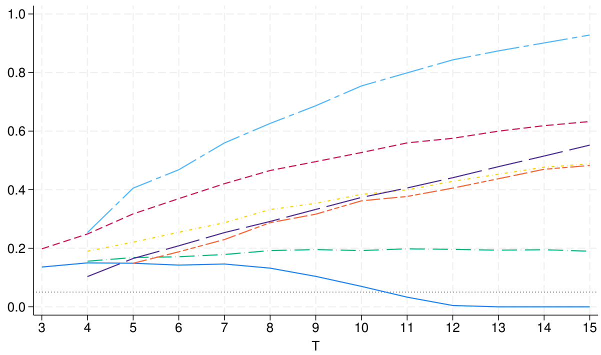
(c) MA(1) alternative, $T = 4$, $\lambda = 0.5$



(d) AR(1) alternative, $T = 4$, $\lambda = 0.5$



(e) MA(1) alternative, $\lambda = 0.5$, $\theta = -0.2$



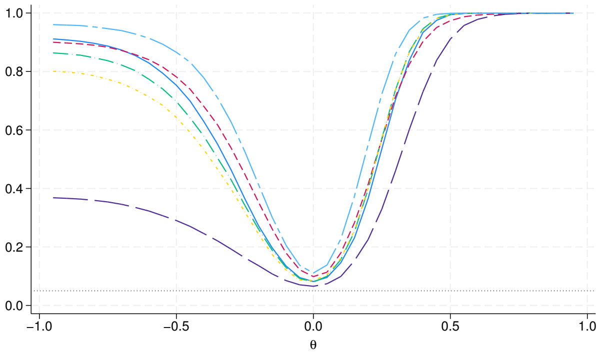
(f) AR(1) alternative, $\lambda = 0.5$, $\rho = 0.2$

Figure 10: Simulated power with inconsistent initial estimator

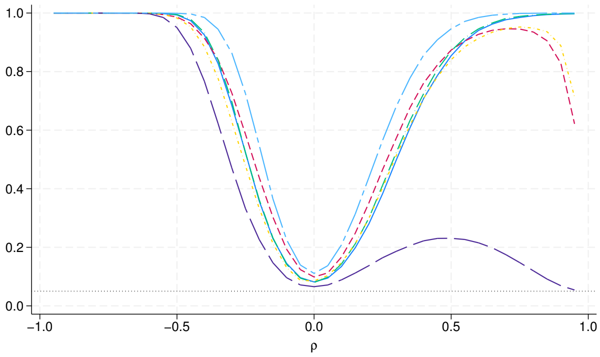
x_{it}, α_i, and ν_{it} are distributed as in the static case. We set $\beta = \omega = \sqrt{1 - \lambda^2}$ and draw the initial observations y_{i0} from the stationary distribution, conditional on the realizations of α_i:

$$y_{i0} = x_{i0} + \sqrt{\frac{1+\lambda}{1-\lambda}}\alpha_i + \varepsilon_{i0}.$$

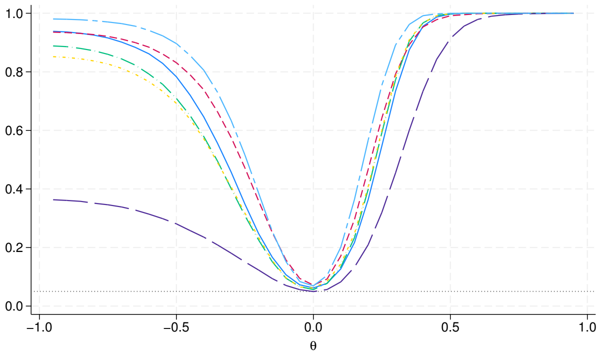
We consider $\lambda \in \{0, 0.5\}$ and restrict ourselves to the homoskedastic case. The residuals are computed from three different method-of-moments estimators, exploiting some or all of the following



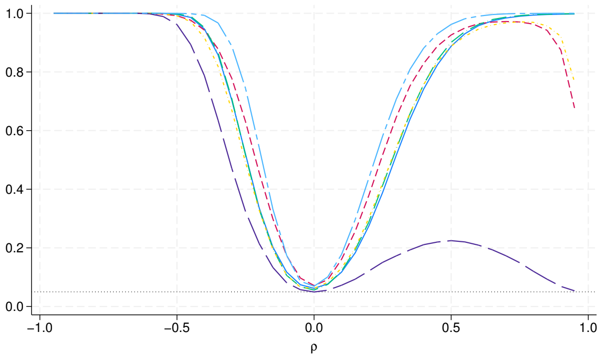
(a) MA(1) alternative, $T = 4$, $\lambda = 0$



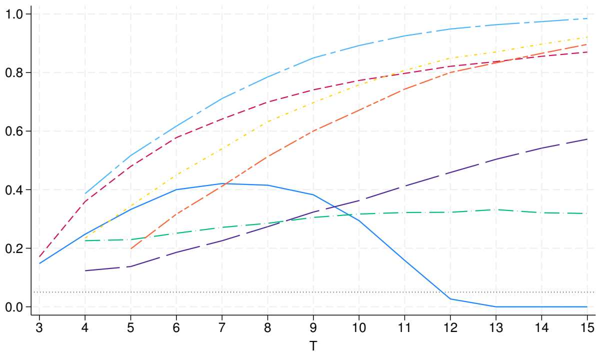
(b) AR(1) alternative, $T = 4$, $\lambda = 0$



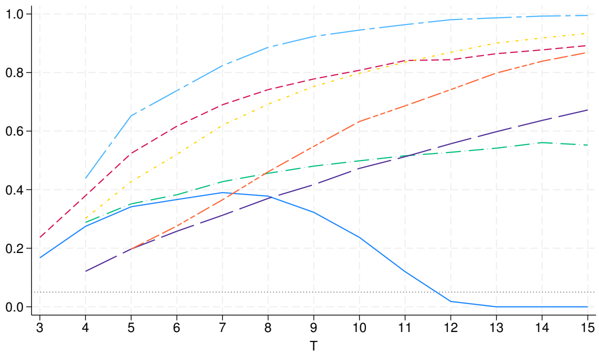
(c) MA(1) alternative, $T = 4$, $\lambda = 0.5$



(d) AR(1) alternative, $T = 4$, $\lambda = 0.5$



(e) MA(1) alternative, $\lambda = 0.5$, $\theta = -0.2$



(f) AR(1) alternative, $\lambda = 0.5$, $\rho = 0.2$

Figure 11: Simulated power with consistent initial estimator

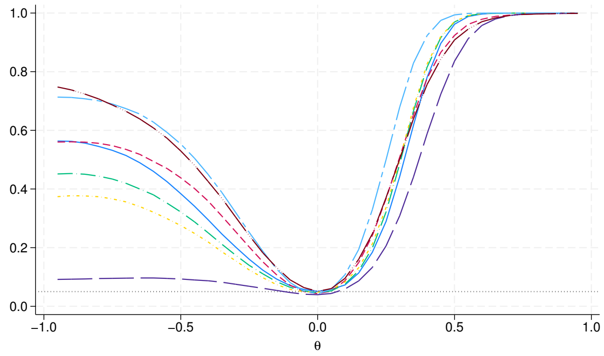
moment conditions:

$$E\left[\sum_{t=3}^{T} y_{i,t-2}\Delta u_{it}\right] = 0, \tag{26}$$

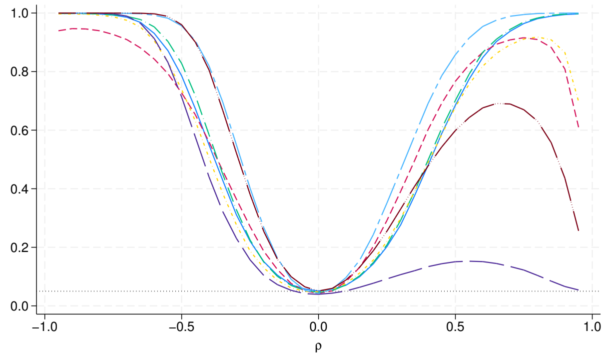
$$E\left[\sum_{t=3}^{T} x_{i,t-1}\Delta u_{it}\right] = 0, \tag{27}$$

$$E\left[\sum_{t=2}^{T} x_{it}\bar{\Delta} u_{it}\right] = 0, \tag{28}$$

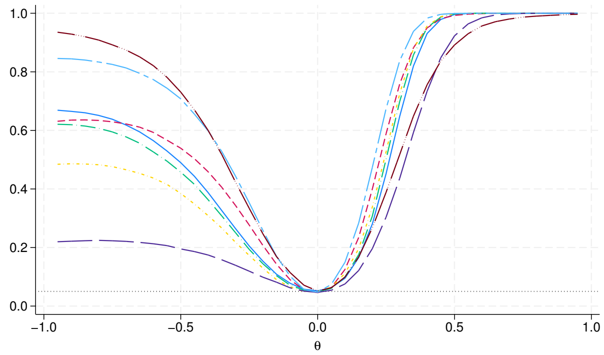
where $\bar{\Delta} u_{it} = u_{it} - (T-1)^{-1}\sum_{s=2}^{T} u_{is}$. The first estimator combines the Anderson and Hsiao (1981)



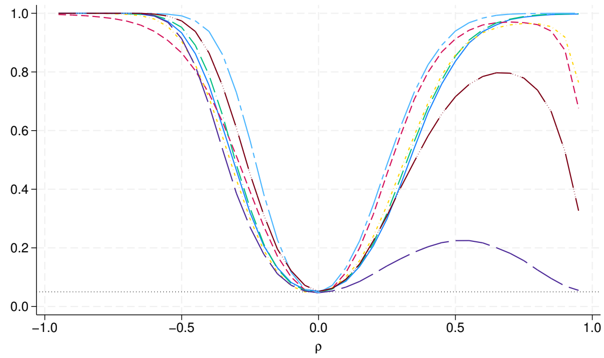
(a) MA(1) alternative, $T = 4$, $\lambda = 0$



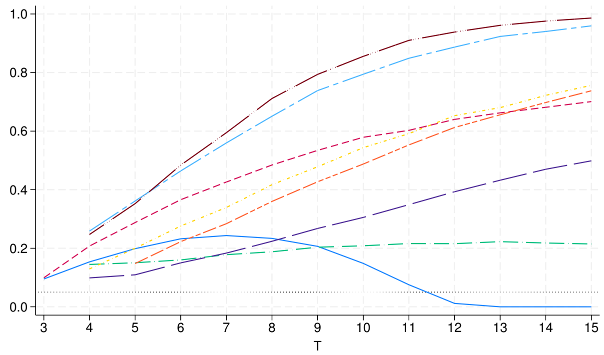
(b) AR(1) alternative, $T = 4$, $\lambda = 0$



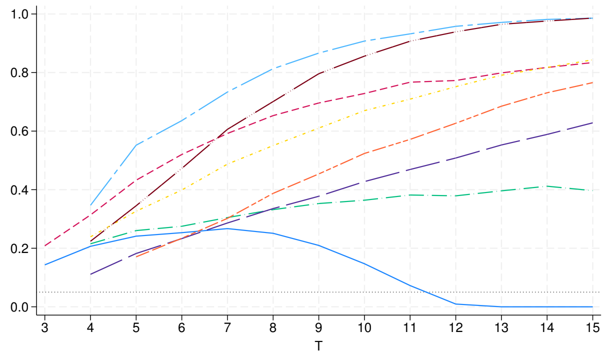
(c) MA(1) alternative, $T = 4$, $\lambda = 0.5$



(d) AR(1) alternative, $T = 4$, $\lambda = 0.5$



(e) MA(1) alternative, $\lambda = 0.5$, $\theta = -0.2$



(f) AR(1) alternative, $\lambda = 0.5$, $\rho = 0.2$

Figure 12: Simulated power with overidentified initial estimator

instrument for the first-differenced lagged dependent variable, implied by equation (26), with the fixed-effects moment condition for the exogenous regressor, equation (28). Under the alternative hypothesis, the first moment condition is violated and the estimator becomes inconsistent. Jung (2005) proposed to solve this problem by replacing invalid instruments with lags of the exogenous variables. We thus consider a second estimator that replaces moment condition (26) by (27), which remains consistent due to the strict exogeneity of x_{it}. As a third option, we look at a two-step GMM estimator utilizing all three moments together. As there are 3 instruments – two of them always valid – for 2 coefficients, we can then include the Hansen (1982) overidentification test in our comparison.

An informative selection of the results is shown in Figures 10 and 11. Compared to the static model, the power of the tests is notably lower. The widely used Arellano-Bond test suffers particularly when the initial estimator is inconsistent under the alternative. The portmanteau test and the S-differencing test tend to retain more power than the Arellano-Bond test, although they suffer from severe size distortions when $\lambda = 0.5$. This is a consequence of the IV estimator being biased in finite samples. When T becomes larger, the power of all tests vanishes under the MA(1) alternative with λ close to $-\theta$. When λ and θ have the same sign, we obtain the familiar picture with power loss only for the portmanteau test.

Using a consistent – albeit inefficient – estimator vastly improves the picture. The tests regain most of their power, although some size distortions remain. Importantly, when T becomes larger, the power of all tests but the portmanteau test approaches 1. Those results clearly indicate that it is beneficial for testing purposes to use residuals from an estimator that remains consistent under the alternative of serially correlated errors.

In the overidentified case, when using all 3 instruments, the GMM estimator is still inconsistent under the alternative, but the additional valid instrument mitigates the consequences. Expectedly, the power of the tests lies in between the two just-identified cases, as demonstrated in Figure 12. The Hansen (1982) overidentification test – depicted by long dashes and small dots – performs remarkably well. Against large negative values of θ, it is even the most powerful of the considered tests, narrowly beating the S-differencing test. However, it suffers under large positive values of ρ.

6 Empirical application

For an empirical illustration, we revisit the seminal paper by Arellano and Bond (1991) who estimate dynamic employment equations with a panel of 140 U.K. companies over the time period 1976–1984. The data set is unbalanced, although without interior gaps.[13] The dependent variable n_{it} is the natural logarithm of employment. Predictors are the logarithms of the real product wage, w_{it}, and gross capital, k_{it}. Additionally, the log of industry output, ys_{it}, and a set of time dummies are included to capture industry-level and aggregate shocks. We re-estimate the same dynamic specifications as in columns (a1) and (a2) of their Table 4:[14]

$$n_{it} = \lambda_1 n_{i,t-1} + \lambda_2 n_{i,t-2} + \mathbf{x}_{it}'\boldsymbol{\beta} + \gamma_t + \alpha_i + \varepsilon_{it}, \tag{29}$$

[13] Without gaps, there is no obvious advantage of transforming the moments into deviations from forward and backward means, as in equation (10).

[14] For the specification in column (b) of Table 4 in Arellano and Bond (1991), we can obtain similar qualitative conclusions. For the sake of brevity, we do not report those results here.

where $\mathbf{x}_{it} = (w_{it}, w_{i,t-1}, k_{it}, k_{i,t-1}, k_{i,t-2}, ys_{it}, ys_{i,t-1}, ys_{i,t-2})'$ are assumed to be strictly exogenous with respect to ε_{it}.

The coefficients are estimated with a GMM estimator for the model in first differences, which removes the incidental parameters α_i, utilizing the following moment conditions:

$$E[n_{i,t-s}\Delta\varepsilon_{it}] = 0, \quad 4 \leq t \leq T, \quad 2 \leq s \leq t-1, \tag{30}$$

$$E\left[\sum_{t=4}^{T} \Delta\mathbf{x}_{it}\Delta\varepsilon_{it}\right] = \mathbf{0}. \tag{31}$$

The estimates are reported for a one-step estimator that would be efficient under homoskedastic and serially uncorrelated idiosyncratic errors, column (a1), and a two-step estimator with a robust weighting matrix clustered at the company level, column (a2). Standard errors are always made cluster-robust. For the two-step estimator, we compute misspecification-robust doubly-corrected standard errors (Hwang et al., 2022).

Our main focus is on testing the null hypothesis that the idiosyncratic error component ε_{it} is serially uncorrelated. Importantly, if this assumption is violated, the GMM estimator becomes inconsistent because $n_{i,t-2}$ would be correlated with the first-differenced error term. Higher-order lags of the dependent variable might remain valid instruments if the serial correlation in ε_{it} is confined to first order.[15] Besides their own serial-correlation test, Arellano and Bond (1991) consider the Sargan (1958) or Hansen (1982) overidentification tests, which are expected to pick up a violation of the moment conditions (30) as long as at least 2 of those moments remain valid under the alternative hypothesis. This rules out arbitrary higher-order serial correlation, in which case all of the moment conditions (30) would be violated;[16] the remaining valid conditions (31) on their own are insufficient to identify all coefficients.

Overidentification tests can indicate model misspecification in a variety of directions, not limited to serial correlation. Moreover, given the relatively large number of overidentifying restrictions, if only a few of them are violated, this might not create a strong enough signal to push the test statistic beyond the threshold. For these reasons, Arellano and Bond (1991) also compute an incremental overidentification test in the spirit of Eichenbaum et al. (1988), which can be obtained as the difference of the overidentification test statistics for the unrestricted model and a restricted model, in which the suspicious instruments $n_{i,t-2}$ (for all t) are left out. A rejection of this test then provides specific statistical evidence against the validity of this particular group of instruments. Here, we re-

[15]Blundell and Bond (2000) present a framework in which autoregressive productivity shocks and measurement error in a Cobb-Douglas production function imply a dynamic model representation with MA(1) idiosyncratic errors.

[16]Note that autoregressive alternatives imply higher-order serial correlation by construction.

port a modified version of this test, in which the respective partition of the weighting matrix from the unrestricted model – leaving out the columns and rows associated with the instruments under investigation – is used to compute the test statistic for the restricted model. This has the benefit of guaranteeing that the difference of the test statistics is nonnegative in finite samples (Newey, 1985).

Furthermore, Arellano and Bond (1991) compute a Hausman (1978) test contrasting the estimates of λ_1 obtained from the unrestricted and the restricted estimator. Under first-order (but not higher-order) serial error correlation, the restricted estimator remains consistent, while the unrestricted estimator is more efficient under the null hypothesis. Here, we compute a generalized version of this test with a robust estimator of the covariance matrix that does not require one of the estimators to be fully efficient (White, 1982). We also add a Hausman-type test contrasting jointly the estimates of λ_1 and λ_2. One could also include the coefficients $\boldsymbol{\beta}$ in this contrast, but we might expect them to be less affected and thus to provide a less powerful signal for the test.

From the serial-correlation tests discussed in this paper, we include the Jochmans (2020b) portmanteau test, the test entirely in first differences, and our S-differencing test. For each of them, we consider a collapsed version and a version combining collapsing with curtailing ($q = 1$). For the first-differenced test, these versions resemble the Yamagata (2008) and Arellano and Bond (1991) tests, respectively.

We present the results in Table 2. In addition to the "efficient" estimates using all instruments, we present the "consistent" estimates (under first-order serial correlation) obtained by leaving out the instruments $n_{i,t-2}$. The latter form the basis for the comparison with the incremental overidentification test and the Hausman tests. Moreover, as argued earlier, serial-correlation tests might have low power when based on the efficient but possibly inconsistent estimator.

First of all, we are able to replicate the coefficient estimates reported by Arellano and Bond (1991) – with few negligible deviations at the third decimal place that can be easily explained by different levels of computational precision. The standard errors for the one-step estimator coincide as well. Notably, the doubly-corrected standard errors for the two-step estimator are substantially larger than the uncorrected ones reported by Arellano and Bond (1991). Most of this discrepancy can be explained by the finite-sample bias in the computation of the variance due to neglecting the additional variation from the initial estimates (Windmeijer, 2005).

The Sargan (1958) overidentification test strongly rejects the null hypothesis of correct model specification for the efficient one-step estimator. When applied to the consistent estimator under the alternative hypothesis of at most first-order serial correlation, the test no longer rejects. As a consequence of the significant difference in these two tests, the incremental overidentification test also

Table 2: Estimation of employment equations

n_{it}	(a1) efficient	(a1) consistent	(a2) efficient	(a2) consistent
$n_{i,t-1}$	0.686 (0.145)	0.986 (0.191)	0.629 (0.331)	0.878 (0.265)
$n_{i,t-2}$	-0.085 (0.056)	0.238 (0.181)	-0.065 (0.048)	0.381 (0.167)
w_{it}	-0.608 (0.178)	-0.683 (0.220)	-0.526 (0.166)	-0.639 (0.224)
$w_{i,t-1}$	0.393 (0.168)	0.524 (0.258)	0.311 (0.257)	0.389 (0.240)
k_{it}	0.357 (0.059)	0.317 (0.066)	0.278 (0.064)	0.254 (0.057)
$k_{i,t-1}$	-0.058 (0.073)	-0.174 (0.096)	0.014 (0.109)	-0.093 (0.102)
$k_{i,t-2}$	-0.020 (0.033)	-0.181 (0.065)	-0.040 (0.066)	-0.217 (0.065)
ys_{it}	0.609 (0.173)	0.658 (0.202)	0.592 (0.160)	0.605 (0.200)
$ys_{i,t-1}$	-0.711 (0.232)	-0.878 (0.354)	-0.566 (0.313)	-0.713 (0.342)
$ys_{i,t-2}$	0.106 (0.141)	0.060 (0.205)	0.101 (0.158)	0.026 (0.216)
Sargan/Hansen	$\chi^2_{25} = 67.6$ [0.000]	$\chi^2_{19} = 24.6$ [0.175]	$\chi^2_{25} = 31.4$ [0.177]	$\chi^2_{19} = 16.0$ [0.655]
incremental	$\chi^2_{6} = 33.5$ [0.000]		$\chi^2_{6} = 13.9$ [0.031]	
Hausman $n_{i,t-1}$	$\chi^2_{1} = 2.83$ [0.093]		$\chi^2_{1} = 1.56$ [0.209]	
+ $n_{i,t-2}$	$\chi^2_{2} = 8.36$ [0.015]		$\chi^2_{2} = 9.88$ [0.007]	
portmanteau	$\chi^2_{20} = 16.3$ [0.701]	$\chi^2_{20} = 21.6$ [0.362]	$\chi^2_{20} = 21.3$ [0.380]	$\chi^2_{20} = 27.0$ [0.135]
+ collapsing	$\chi^2_{6} = 1.98$ [0.921]	$\chi^2_{6} = 4.29$ [0.637]	$\chi^2_{6} = 3.38$ [0.760]	$\chi^2_{6} = 5.21$ [0.518]
+ curtailing	$\chi^2_{2} = 1.61$ [0.447]	$\chi^2_{2} = 2.52$ [0.284]	$\chi^2_{2} = 2.71$ [0.259]	$\chi^2_{2} = 3.92$ [0.141]
first differencing	$\chi^2_{10} = 5.21$ [0.877]	$\chi^2_{10} = 18.5$ [0.047]	$\chi^2_{10} = 12.8$ [0.237]	$\chi^2_{10} = 17.7$ [0.061]
+ collapsing	$\chi^2_{4} = 0.73$ [0.948]	$\chi^2_{4} = 16.4$ [0.002]	$\chi^2_{4} = 0.58$ [0.966]	$\chi^2_{4} = 16.5$ [0.002]
+ curtailing	$\chi^2_{1} = 0.20$ [0.652]	$\chi^2_{1} = 9.85$ [0.002]	$\chi^2_{1} = 0.18$ [0.669]	$\chi^2_{1} = 11.3$ [0.001]
S-differencing	$\chi^2_{10} = 24.9$ [0.006]	$\chi^2_{10} = 30.6$ [0.001]	$\chi^2_{10} = 28.6$ [0.001]	$\chi^2_{10} = 37.8$ [0.000]
+ collapsing	$\chi^2_{4} = 17.3$ [0.002]	$\chi^2_{4} = 19.9$ [0.001]	$\chi^2_{4} = 18.8$ [0.001]	$\chi^2_{4} = 26.8$ [0.000]
+ curtailing	$\chi^2_{1} = 8.16$ [0.004]	$\chi^2_{1} = 10.9$ [0.001]	$\chi^2_{1} = 10.2$ [0.001]	$\chi^2_{1} = 11.5$ [0.001]

(i) The regression specifications correspond to those with the same column titles, (a1) and (a2), in Table 4 of Arellano and Bond (1991). All regressions include time dummies.
(ii) The coefficients are estimated by a GMM estimator for the first-differenced regression model. Column (a1) are one-step estimates, while columns (a2) and (b) are two-step estimates. The one-step weighting matrix is optimal under homoskedastic and serially uncorrelated idiosyncratic errors.
(iii) The instruments for the efficient estimator under the null hypothesis are specified in note (vi) of Table 4 in Arellano and Bond (1991). For the consistent estimator under the alternative hypothesis of first-order serially correlated errors, the instruments $n_{i,t-2}$ (for all t) are left out.
(iii) Standard errors in parentheses are cluster-robust at the company level, computed with the doubly-corrected variance formula of Hwang et al. (2022) in columns (a2) and (b).
(iv) The first set of reported specification tests are the Sargan (1958) and Hansen (1982) overidentification test, an incremental overidentification test proposed by Eichenbaum et al. (1988), and two generalized Hausman (1978) tests for the coefficients of $n_{i,t-1}$ only and $(n_{i,t-1}, n_{i,t-2})$ jointly. p-values are in brackets.
(v) The second set of specification tests are the portmanteau test of Jochmans (2020b), a collapsed version of it, and a collapsed and curtailed ($q = 1$) version.
(vi) The third set of specification tests are tests based entirely on first differences. The collapsed version is the Yamagata (2008) test, and the collapsed and curtailed ($q = 1$) version is the χ^2 analogue of the Arellano and Bond (1991) test.
(vii) The final set of tests are our S-differencing test, a collapsed version of it, and a collapsed and curtailed ($q = 1$) version.

suggests rejection of the null hypothesis. In contrast, the Hausman test provides a less clear indication with a p-value of 0.09 when contrasting only the first coefficient. When both coefficients of the lagged dependent variables are included in the comparison, the signal becomes stronger.

When heteroskedasticity causes the one-step estimator to be inefficient, the respective overidentification tests become asymptotically invalid. With the two-step estimator, which regains asymptotic efficiency, the Hansen (1982) overidentification test no longer rejects the null hypothesis even with the efficient set of instruments. The same applies to the Hausman test for the first coefficient. The incremental overidentification test and the extended Hausman test still tend to reject the null, providing a mixed overall picture.

The portmanteau test, including its collapsed and curtailed versions, never rejects the null hypothesis in this example. The first-differencing tests also do not reject when they are based on the efficient estimators. However, they tend to reject once the residuals are taken from the respective estimator that remains consistent under first-order serially correlated errors. This would be in line with the earlier discussion that these tests tend to suffer from severe power losses when the estimator is inconsistent under the alternative, and it highlights the potential benefits from basing the test on an estimator that remains consistent under the alternative. For the portmanteau test, we also see a reduction in p-values under the consistent estimation, but it is not strong enough to trigger a rejection at conventional significance levels. With a maximum of 20 degrees of freedom, this is less likely a consequence of moment proliferation – although the portmanteau test is expected to already lose some power in a data set with $T = 7$ and $N = 140$ – but could be explained by the test's power deterioration under high variances of the company-specific error component, which would be reasonable in an application like this. Indeed, based on the consistent one-step or two-step residuals we can obtain estimates of the error components' variance ratio of $\hat{\eta} = 6.82$ or $\hat{\eta} = 7.88$, respectively.

Lastly, our S-differencing test strongly rejects the null hypothesis throughout all specifications, with and without collapsing or curtailing. The results across the different serial-correlation tests are consistent with our earlier power analysis, especially if the errors ε_{it} follow an autoregressive process with large positive autocorrelation, in addition to high variances of α_i. Notably, in this case the estimators labeled "consistent" in Table 2 would actually still be inconsistent.

7 Conclusions

Recently, Jochmans (2020b) proposed a portmanteau test against arbitrary serial correlation in the idiosyncratic error components of linear panel models. Compared to earlier tests, especially the popular Arellano and Bond (1991) test, it convinces with substantial power gains when T is (very) small. However, due to the rapid moment proliferation, this test can quickly lose its power advantage already for moderately small T, or even become degenerate. To restrain the number of moment restrictions, we borrow strategies that are widely used to address a similar problem of instrument proliferation in dynamic panel models: curtailing and collapsing.

Another shortcoming of the portmanteau test is its lack of invariance to high variances of the group-specific error component. While tests for serial correlation in the first-differenced errors avoid this issue, they have no power against a random-walk alternative. We propose a new test based on S-differencing – long differences encompassing first differences – that remains invariant to high variances of the group-specific error component and does not suffer from a power loss under highly autocorrelated alternatives. We demonstrate with power calculations and Monte Carlo simulations that the power gains from our new test can be substantial. In a real-world example, we reach different conclusions from applying the S-differencing test compared to existing tests.

Furthermore, we uncover a potentially serious deterioration in all of the tests' power when the residuals are computed from an estimator that is inconsistent under the alternative hypothesis. In extreme cases, the impact of the estimator's asymptotic bias on the tests' noncentrality parameter can fully offset the signal from a violation of the moment restrictions, leading to a complete loss of power. To circumvent this problem, we suggest for testing purposes to use an estimator that remains consistent – albeit possibly inefficient – under reasonable alternatives.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Arellano, M. (1990). Testing for autocorrelation in dynamic random effects models. *Review of Economic Studies 57*(1), 127–134.

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

Arellano, M. and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Baltagi, B. H. and Q. Li (1995). Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics 68*(2), 133–151.

Bhargava, A. (1987). Wald tests and systems of stochastic equations. *International Economic Review 28*(3), 789–808.

Bhargava, A., L. Franzini, and W. Narendranathan (1982). Serial correlation and the fixed effects model. *Review of Economic Studies 49*(4), 533–549.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*(6), 1635–1659.

Blundell, R. and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Blundell, R. and S. R. Bond (2000). GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews 19*(3), 321–340.

Born, B. and J. Breitung (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews 35*(7), 1290–1316.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal 3*(2), 168–177.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics 103*(1), 51–78.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

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

Hwang, J., B. Kang, and S. Lee (2022). A doubly corrected robust variance estimator for linear GMM. *Journal of Econometrics 229*(2), 276–298.

Inoue, A. and G. Solon (2006). A portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory 22*(5), 835–851.

Jochmans, K. (2020a). A portmanteau test for correlation in short panels. *Econometric Theory 36*(6), 1159–1166.

Jochmans, K. (2020b). Testing for correlation in error-component models. *Journal of Applied Econometrics 35*(7), 860–878.

Jung, H. (2005). A test for autocorrelation in dynamic panel data models. *Journal of the Korean Statistical Society 34*(4), 367–375.

Kiviet, J. F. (2020). Microeconometric dynamic panel data methods: Model specification and selection issues. *Econometrics and Statistics 13*, 16–45.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29*(3), 229–256.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*(3), 393–415.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*(1), 25–51.

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

Yamagata, T. (2008). A joint serial correlation test for linear panel data models. *Journal of Econometrics 146*(1), 135–145.

Appendix A Orthogonal deviations

In Section 2.2, we noted that we can replace $\ddot{\mathbf{H}}_{i+}'\Delta\ddot{\mathbf{u}}_i$ by alternative moment functions $\check{\mathbf{H}}_{i+}'\Delta\ddot{\mathbf{u}}_i$ that can be obtained as linear combinations of the initial moment functions. This follows from the fact that the set of initial moment functions $\boldsymbol{\zeta}_i$ is exhaustive. In particular, we can choose $\check{\mathbf{H}}_{i+}' = \ddot{\mathbf{H}}_{i+}'\mathbf{K}_+$, provided that $\mathbf{K}_+$ is lower triangular and of full rank. Then, analogously to equation (9), we can form alternative linear combinations with these transformed moment functions. In particular,

$$u_{iT}\mathbf{K}_+\Delta\ddot{\mathbf{u}}_i = \mathbf{R}_{\pm(\mathbf{K}_+)}\dot{\mathbf{H}}_{i-}'\Delta\dot{\mathbf{u}}_i + \ddot{\mathbf{H}}_{i+}'\mathbf{K}_+\Delta\ddot{\mathbf{u}}_i, \tag{32}$$

with an appropriate choice of $\mathbf{R}_{\pm(\mathbf{K}_+)}$. When $\mathbf{K}_+ = \mathbf{I}_{T-2}$, we obtain as a special case

$$u_{iT}\Delta\ddot{\mathbf{u}}_i = \mathbf{R}_{\pm(\mathbf{I})}\dot{\mathbf{H}}_{i-}'\Delta\dot{\mathbf{u}}_i + \ddot{\mathbf{H}}_{i+}'\Delta\ddot{\mathbf{u}}_i.$$

After premultiplication with $\mathbf{K}_+$,

$$u_{iT}\mathbf{K}_+\Delta\ddot{\mathbf{u}}_i = \mathbf{K}_+\mathbf{R}_{\pm(\mathbf{I})}\dot{\mathbf{H}}_{i-}'\Delta\dot{\mathbf{u}}_i + \mathbf{K}_+\ddot{\mathbf{H}}_{i+}'\Delta\ddot{\mathbf{u}}_i. \tag{33}$$

Combining equations (32) and (33) then yields

$$\left(\ddot{\mathbf{H}}_{i+}'\mathbf{K}_+ - \mathbf{K}_+\ddot{\mathbf{H}}_{i+}'\right)\Delta\ddot{\mathbf{u}}_i = \left(\mathbf{K}_+\mathbf{R}_{\pm(\mathbf{I})} - \mathbf{R}_{\pm(\mathbf{K}_+)}\right)\dot{\mathbf{H}}_{i-}'\Delta\dot{\mathbf{u}}_i,$$

and therefore applying the transformation $\mathbf{K}_+$ to $\ddot{\mathbf{H}}'_{i+}\Delta\ddot{\mathbf{u}}_i$ yields an equivalent test statistic as applying the same transformation directly to $\Delta\ddot{\mathbf{u}}_i$.

Appendix B Proof of Proposition 1

Define $\mathbf{Z} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_N)'$, $\mathbf{S} = \mathbf{ZR}'\left(\mathbf{RZ}'\mathbf{ZR}'\right)^{-}\mathbf{RZ}'$, and let $\boldsymbol{\iota}_N$ be an $N \times 1$ vector of ones, such that $s_{R,N} = \boldsymbol{\iota}_N'\mathbf{S}\boldsymbol{\iota}_N$. Consider the singular value decomposition $\mathbf{ZR}' = \mathbf{U}_1\mathbf{D}\mathbf{U}_2'$, where $\mathbf{U}_1$ is the $N \times N$ orthogonal eigenvector matrix of $\mathbf{S}$, and $\mathbf{U}_2$ is the $r \times r$ orthogonal eigenvector matrix of $\mathbf{RZ}'\mathbf{ZR}'$. Given $r \geq N$, $\mathbf{D}$ is an $N \times r$ rectangular diagonal matrix with singular values $d_1, d_2, \ldots, d_N$ of $\mathbf{ZR}'$ on the diagonal. All singular values are nonzero (with probability 1) due to the maximum-rank assumption $\mathrm{rk}(\mathbf{RZ}'\mathbf{ZR}') = \mathrm{rk}(\mathbf{ZR}') = N$.

With $\mathbf{U}_2'\mathbf{U}_2 = \mathbf{I}_r$, we can write the test statistic as

$$s_{R,N} = \boldsymbol{\iota}_N'\mathbf{U}_1\mathbf{D}\mathbf{U}_2'\left(\mathbf{U}_2\mathbf{D}'\mathbf{D}\mathbf{U}_2'\right)^{-}\mathbf{U}_2\mathbf{D}'\mathbf{U}_1'\boldsymbol{\iota}_N = \boldsymbol{\iota}_N'\mathbf{U}_1\mathbf{D}(\mathbf{D}'\mathbf{D})^{-}\mathbf{D}'\mathbf{U}_1'\boldsymbol{\iota}_N.$$

When $r \geq N$, $\mathbf{D}'\mathbf{D} = \mathrm{diag}(d_1^2, d_2^2, \ldots, d_N^2, 0, \ldots, 0)$, with $r - N$ eigenvalues of $\mathbf{RZ}'\mathbf{ZR}'$ equal to zero. Consequently, the diagonal matrix of eigenvalues of the idempotent matrix $\mathbf{S}$ is $\mathbf{D}(\mathbf{D}'\mathbf{D})^{-}\mathbf{D}' = \mathbf{I}_N$. Thus, with $\mathbf{U}_1\mathbf{U}_1' = \mathbf{I}_N$,

$$s_{R,N} = \boldsymbol{\iota}_N'\mathbf{U}_1\mathbf{U}_1'\boldsymbol{\iota}_N = \boldsymbol{\iota}_N'\boldsymbol{\iota}_N = N.$$

Appendix C Transformation matrices

Curtailing refers to a selection of the moment restrictions with transformation matrix

$$\mathbf{R}_q = \begin{pmatrix} \mathbf{R}_{q-} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T-2} \end{pmatrix}$$

such that $\mathbf{H}_{i-}\mathbf{R}_{q-}'$ only retains the columns of $\mathbf{H}_{i-}$ corresponding to the moment restrictions (5) for $2 \leq s \leq \max(q+1, t-1)$. For example, if $q = 1$ then

$$\mathbf{H}_{i-}\mathbf{R}_{1-}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & u_{i,T-2} \end{pmatrix}.$$

Collapsing instead corresponds to a transformation matrix

$$\mathbf{R}_c = \begin{pmatrix} \mathbf{R}_{c-} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota}'_{T-2} \end{pmatrix},$$

where $\boldsymbol{\iota}_{T-2}$ is a column vector with all $T-2$ elements equal to 1, such that

$$\mathbf{H}_{i-}\mathbf{R}'_{c-} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ u_{i1} & 0 & \cdots & 0 \\ u_{i2} & u_{i1} & & 0 \\ \vdots & \vdots & \ddots & \\ u_{i,T-2} & u_{i,T-3} & \cdots & u_{i1} \end{pmatrix} \quad \text{and} \quad \mathbf{H}_{i+}\boldsymbol{\iota}_{T-2} = \begin{pmatrix} u_{i3} \\ u_{i4} \\ \vdots \\ u_{iT} \\ 0 \end{pmatrix}.$$

For the full-collapsing approach, the respective transformation matrix is

$$\mathbf{R}_{fc} = \begin{pmatrix} -\boldsymbol{\iota}'_{(T-1)(T-2)/2} & \boldsymbol{\iota}'_{T-2} \end{pmatrix},$$

such that

$$\mathbf{H}_i\mathbf{R}'_{fc} = \begin{pmatrix} u_{i3} \\ u_{i4} - u_{i1} \\ u_{i5} - (u_{i1} + u_{i2}) \\ \vdots \\ u_{iT} - (u_{i1} + \ldots + u_{i,T-3}) \\ -(u_{i1} + \ldots + u_{i,T-2}) \end{pmatrix}.$$

The transformation matrix $\mathbf{R}_\Delta = (\mathbf{R}_{\Delta-}, \mathbf{0})$ linearly combines the columns of $\mathbf{H}_i$ to obtain first differences:

$$\mathbf{H}_i\mathbf{R}'_\Delta = \mathbf{H}_{i-}\mathbf{R}'_{\Delta-} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \Delta u_{i2} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta u_{i2} & \Delta u_{i3} & & 0 & \cdots & 0 \\ \vdots & & & \ddots & & & \\ 0 & 0 & 0 & & \Delta u_{i2} & \cdots & \Delta u_{i,T-2} \end{pmatrix}.$$

Combined with the collapsing approach, the transformation for the Yamagata (2008) test is

$$\mathbf{H}_i\mathbf{R}'_{\Delta c} = \mathbf{H}_{i-}\mathbf{R}'_{\Delta c-} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \Delta u_{i2} & 0 & \cdots & 0 \\ \Delta u_{i3} & \Delta u_{i2} & & 0 \\ \vdots & \vdots & \ddots & \\ \Delta u_{i,T-2} & \Delta u_{i,T-3} & \cdots & \Delta u_{i2} \end{pmatrix}.$$

By selecting one of the columns of $\mathbf{H}_{i-}\mathbf{R}'_{\Delta c-}$, we can replicate the Arellano and Bond (1991) test. For example, their test against second-order serial correlation in the first-differenced errors combines first differencing, collapsing, and curtailing ($q = 1$) of the original moment restrictions:

$$\mathbf{H}_i\mathbf{R}'_{\Delta c1} = \mathbf{H}_{i-}\mathbf{R}'_{\Delta c1-} = \begin{pmatrix} 0 \\ 0 \\ \Delta u_{i2} \\ \vdots \\ \Delta u_{i,T-2} \end{pmatrix}.$$

To obtain the S-differencing moment restriction, combined with curtailing ($q = 1$) and collapsing, the transformation matrix becomes $\mathbf{R}_{\Delta_{1,2}c1} = \big(-(\boldsymbol{\iota}'_{T-3}, 0)\mathbf{R}_{1-}, (0, \boldsymbol{\iota}'_{T-3})\big)$, such that

$$\mathbf{H}_i\mathbf{R}'_{\Delta_{1,2}c1} = \begin{pmatrix} 0 \\ \Delta_{1,2}u_{i3} \\ \Delta_{1,2}u_{i4} \\ \vdots \\ \Delta_{1,2}u_{i,T-1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ u_{i4} - u_{i1} \\ u_{i5} - u_{i2} \\ \vdots \\ u_{iT} - u_{i,T-3} \\ 0 \end{pmatrix}.$$

Appendix D Variance-covariance matrix under stationarity

For general T and a stationary data-generating process, under the null hypothesis we can obtain

$$\mathbf{V} = \sigma_\nu^2 \begin{pmatrix} \mathbf{A}_1 & \mathbf{B}_2' & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{C}_1' \\ \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{B}_3' & \ddots & & \mathbf{0} & \mathbf{C}_2' \\ \mathbf{0} & \mathbf{B}_3 & \mathbf{A}_3 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{B}_{T-3}' & \mathbf{0} & \mathbf{C}_{T-3}' \\ \vdots & & \ddots & \mathbf{B}_{T-3} & \mathbf{A}_{T-3} & \mathbf{B}_{T-2}' & \mathbf{C}_{T-3}' \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}_{T-2} & \mathbf{A}_{T-2} & \mathbf{C}_{T-2}' \\ \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_{T-4} & \mathbf{C}_{T-3} & \mathbf{C}_{T-2} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{A}_j$, $j = 1, 2, \ldots, T-2$, is a sequence of $j \times j$ matrices with main-diagonal elements $2(\sigma_\alpha^2 + \sigma_\nu^2)$ and off-diagonal elements $2\sigma_\alpha^2$; $\mathbf{B}_j$, $j = 2, 3, \ldots, T-2$, is a sequence of $j \times (j-1)$ matrices with main-diagonal elements $-(\sigma_\alpha^2 + \sigma_\nu^2)$ and off-diagonal elements $-\sigma_\alpha^2$; $\mathbf{C}_j$, $j = 1, 2, \ldots, T-2$, is a sequence of $(T-2) \times j$ matrices with all elements in the first $j-2$ rows and the $j-2$ leading elements in the $(j-1)$-st row (for $j > 2$) equal to zero, the last two elements in the $(j-1)$-st row (for $j > 1$) equal to σ_v^2 and $-\sigma_v^2$, all elements but the last one in the j-th row equal to $-\sigma_\alpha^2$ and the last one equal to $-(\sigma_\alpha^2 + \sigma_v^2)$, all elements in the $(j+1)$-th row (for $j < T-2$) equal to $2\sigma_\alpha^2$, all elements in the $(j+2)$-nd row (for $j < T-3$) equal to $-\sigma_\alpha^2$, and all remaining rows (if any) full of zeros; and $\mathbf{D}$ is a $(T-2) \times (T-2)$ matrix with main-diagonal elements $2(\sigma_\alpha^2 + \sigma_\nu^2)$, all elements on the diagonals directly above and below the main diagonal equal to $-\sigma_\alpha^2$, and all other elements equal to zero. For $T = 4$, this becomes

$$\mathbf{V} = \sigma_\nu^2 \begin{pmatrix} 2(\sigma_\alpha^2 + \sigma_\nu^2) & -(\sigma_\alpha^2 + \sigma_\nu^2) & -\sigma_\alpha^2 & -(\sigma_\alpha^2 + \sigma_\nu^2) & 2\sigma_\alpha^2 \\ -(\sigma_\alpha^2 + \sigma_\nu^2) & 2(\sigma_\alpha^2 + \sigma_\nu^2) & 2\sigma_\alpha^2 & \sigma_\nu^2 & -\sigma_\alpha^2 \\ -\sigma_\alpha^2 & 2\sigma_\alpha^2 & 2(\sigma_\alpha^2 + \sigma_\nu^2) & -\sigma_\nu^2 & -(\sigma_\alpha^2 + \sigma_\nu^2) \\ -(\sigma_\alpha^2 + \sigma_\nu^2) & \sigma_\nu^2 & -\sigma_\nu^2 & 2(\sigma_\alpha^2 + \sigma_\nu^2) & -\sigma_\alpha^2 \\ 2\sigma_\alpha^2 & -\sigma_\alpha^2 & -(\sigma_\alpha^2 + \sigma_\nu^2) & -\sigma_\alpha^2 & 2(\sigma_\alpha^2 + \sigma_\nu^2) \end{pmatrix}.$$

Appendix E Anderson-Hsiao IV estimation of dynamic model

For the simply dynamic panel data model

$$y_{it} = \beta y_{i,t-1} + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it},$$

with stationary representation $y_{it} = \sum_{j=0}^{\infty} \beta^j u_{i,t-j}$ for $|\beta| < 1$, the Anderson and Hsiao (1981) IV estimator (23) is inconsistent under the alternative hypothesis of serially correlated errors ε_{it}:

$$\operatorname*{plim}_{N\to\infty} \left(\hat{\beta} - \beta\right) = \frac{\operatorname{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-2} \Delta\varepsilon_{it}}{\operatorname{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-2} \Delta y_{i,t-1}}.$$

E.1 MA(1) alternative

Under the stationary MA(1) alternative (21), we have

$$E[\varepsilon_{it}\varepsilon_{i,t-j}] = \begin{cases} (1+\theta^2)\sigma_\nu^2, & j = 0 \\ \theta\sigma_\nu^2, & |j| = 1 \\ 0, & |j| > 1 \end{cases} .$$

We then obtain for the numerator

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-2} \Delta\varepsilon_{it} = -\sum_{t=2}^{T} E[\varepsilon_{i,t-2}\varepsilon_{i,t-1}] = -(T-1)\theta\sigma_\nu^2,$$

and for the denominator

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-2} \Delta y_{i,t-1} = \sum_{t=2}^{T} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \beta^{j+l} E[\varepsilon_{i,t-2-j} \Delta\varepsilon_{i,t-1-l}] = -(T-1) \frac{(1+\beta)\theta + (1-\theta)^2}{1+\beta} \sigma_\nu^2.$$

Thus,

$$E[\omega_i] = \operatorname*{plim}_{N\to\infty} \left(\hat{\beta} - \beta\right) = \frac{(1+\beta)\theta}{(1+\beta)\theta + (1-\theta)^2}.$$

When $T = 3$, $E[\boldsymbol{\zeta}_i] = \theta\sigma_\nu^2(-1, 1)'$, and

$$\begin{aligned} \boldsymbol{\Gamma}_1 &= -E\left[\begin{pmatrix} y_{i0}\Delta u_{i3} + u_{i1}\Delta y_{i2} \\ y_{i2}\Delta u_{i2} + u_{i3}\Delta y_{i1} \end{pmatrix}\right] = -\sum_{j=0}^{2} \beta^j E\left[\begin{pmatrix} \varepsilon_{i1}\Delta\varepsilon_{i,2-j} \\ \varepsilon_{i,2-j}\Delta\varepsilon_{i2} \end{pmatrix}\right] \\ &= \left((\beta^2 - \beta + 1)\theta - (1-\beta)(1+\theta^2)\right)\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \\ \boldsymbol{\Gamma}_2 &= 2E\left[\begin{pmatrix} y_{i0}\Delta y_{i2} \\ y_{i2}\Delta y_{i1} \end{pmatrix}\right] = 2\left(\beta \frac{(1+\beta)\theta + (1-\theta)^2}{1+\beta} + \theta\right)\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

Consequently,

$$E[\boldsymbol{\zeta}_i] + \boldsymbol{\Gamma}_1 E[\omega_i] = \frac{\beta\theta(\beta+\theta)(1+\beta\theta)}{(1+\beta)\theta+(1-\theta)^2}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

$$\frac{1}{2}\boldsymbol{\Gamma}_2(E[\omega_i])^2 = \frac{(1+\beta)\theta^2(\beta+\theta)(1+\beta\theta)}{((1+\beta)\theta+(1-\theta)^2)^2}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Under local alternatives of the form $\theta = \theta^*/\sqrt{N}$, we have $E[\boldsymbol{\zeta}_i] + \boldsymbol{\Gamma}_1 E[\omega_i] = o(1)$ and $\boldsymbol{\Gamma}_2(E[\omega_i])^2 = o(N^{-1/2})$. This reflects the asymptotic negligibility of the second-order term. However, it still lends some power to the test in finite samples or against fixed alternatives. This can be particularly relevant when $\beta = 0$, in which case the first term evaluates to zero while the second term remains informative. When $\theta = -\beta$ (or, trivially, $\theta = 0$), both terms equate to zero. This is intuitive because the DGP under $\theta = -\beta$ is observationally equivalent to a DGP with $\beta = \theta = 0$.

When $T = 4$, $E[\boldsymbol{\zeta}_i] = \theta\sigma_\nu^2(-1, 0, -1, 1, 1)'$, and

$$\boldsymbol{\Gamma}_1 = -E\left[\begin{pmatrix} y_{i0}\Delta u_{i3} + u_{i1}\Delta y_{i2} \\ y_{i0}\Delta u_{i4} + u_{i1}\Delta y_{i3} \\ y_{i1}\Delta u_{i4} + u_{i2}\Delta y_{i3} \\ y_{i2}\Delta u_{i2} + u_{i3}\Delta y_{i1} \\ y_{i3}\Delta u_{i3} + u_{i4}\Delta y_{i2} \end{pmatrix}\right] = -\sum_{j=0}^{2}\beta^j E\left[\begin{pmatrix} \varepsilon_{i1}\Delta\varepsilon_{i,2-j} \\ \beta\varepsilon_{i1}\Delta\varepsilon_{i,2-j} \\ \varepsilon_{i2}\Delta\varepsilon_{i,3-j} \\ \varepsilon_{i,2-j}\Delta\varepsilon_{i2} \\ \varepsilon_{i,3-j}\Delta\varepsilon_{i3} \end{pmatrix}\right] - E\left[\begin{pmatrix} 0 \\ \varepsilon_{i1}\Delta\varepsilon_{i3} \\ 0 \\ 0 \\ 0 \end{pmatrix}\right]$$

$$= \left((\beta^2-\beta+1)\theta - (1-\beta)(1+\theta^2)\right)\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix} + \theta\sigma_\nu^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma}_2 = 2E\left[\begin{pmatrix} y_{i0}\Delta y_{i2} \\ y_{i0}\Delta y_{i3} \\ y_{i1}\Delta y_{i3} \\ y_{i2}\Delta y_{i1} \\ y_{i3}\Delta y_{i2} \end{pmatrix}\right] = 2\left(\beta\frac{(1+\beta)\theta+(1-\theta)^2}{1+\beta} + \theta\right)\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix},$$

such that

$$E[\boldsymbol{\zeta}_i] + \boldsymbol{\Gamma}_1 E[\omega_i] = \frac{\beta\theta(\beta+\theta)(1+\beta\theta)}{(1+\beta)\theta+(1-\theta)^2}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix} + \frac{\theta(\beta+\theta)(1+\beta\theta)}{(1+\beta)\theta+(1-\theta)^2}\sigma_\nu^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\frac{1}{2}\boldsymbol{\Gamma}_2 (E[\omega_i])^2 = \frac{(1+\beta)\theta^2(\beta+\theta)(1+\beta\theta)}{((1+\beta)\theta+(1-\theta)^2)^2}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix}.$$

For all but the second moment restriction, the expressions are identical to the one for $T = 3$. Noteworthy, the additional term for the second moment restriction, which is asymptotically nonnegligible under local alternatives, remains nonzero when $\beta = 0$.

It should also be noted that $\boldsymbol{\Gamma}_1 \neq \mathbf{0}$ when $\theta = 0$. Consequently, the estimation error must be accounted for in the estimation of the variance matrix $\tilde{\mathbf{V}}$ even under the null hypothesis.

E.2 AR(1) alternative

Under the stationary AR(1) alternative (22), we have $E[\varepsilon_{it}\varepsilon_{i,t-j}] = \rho^{|j|}(1-\rho^2)^{-1}\sigma_\nu^2$. It follows for the numerator of the Anderson and Hsiao (1981) estimator that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}\Delta\varepsilon_{it} = \sum_{t=2}^{T}\sum_{j=0}^{\infty}\beta^j E[\varepsilon_{i,t-2-j}\Delta\varepsilon_{it}] = -\frac{(T-1)\rho}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2.$$

For the denominator, we obtain

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}\Delta y_{i,t-1} = \sum_{t=2}^{T}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\beta^{j+l} E[\varepsilon_{i,t-2-j}\Delta\varepsilon_{i,t-1-l}] = -\frac{(T-1)}{(1+\beta)(1-\beta\rho)(1+\rho)}\sigma_\nu^2.$$

Thus,

$$E[\omega_i] = \operatorname*{plim}_{N\to\infty}\left(\hat{\beta}-\beta\right) = (1+\beta)\rho.$$

When $T = 3$, $E[\boldsymbol{\zeta}_i] = \rho(1+\rho)^{-1}\sigma_\nu^2(-1, 1)'$, and

$$\boldsymbol{\Gamma}_1 = -E\left[\begin{pmatrix} y_{i0}\Delta u_{i3} + u_{i1}\Delta y_{i2} \\ y_{i2}\Delta u_{i2} + u_{i3}\Delta y_{i1} \end{pmatrix}\right] = -\sum_{j=0}^{\infty}\beta^j E\left[\begin{pmatrix} \varepsilon_{i,0-j}\Delta\varepsilon_{i3} + \varepsilon_{i1}\Delta\varepsilon_{i,2-j} \\ \varepsilon_{i,2-j}\Delta\varepsilon_{i2} + \varepsilon_{i3}\Delta\varepsilon_{i,1-j} \end{pmatrix}\right]$$
$$= \frac{\beta(1+\rho) - (1+\rho^2)}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$
$$\boldsymbol{\Gamma}_2 = 2E\left[\begin{pmatrix} y_{i0}\Delta y_{i2} \\ y_{i2}\Delta y_{i1} \end{pmatrix}\right] = 2\frac{\beta+\rho+\beta\rho}{(1+\beta)(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

It follows that

$$E[\boldsymbol{\zeta}_i] + \boldsymbol{\Gamma}_1 E[\omega_i] = \frac{\rho(\beta-\rho)(\beta+\rho+\beta\rho)}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$
$$\frac{1}{2}\boldsymbol{\Gamma}_2(E[\omega_i])^2 = \frac{(1+\beta)\rho^2(\beta+\rho+\beta\rho)}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Under local alternatives of the form $\rho = \rho^*/\sqrt{N}$, we have again $E[\boldsymbol{\zeta}_i] + \boldsymbol{\Gamma}_1 E[\omega_i] = o(1)$ and $\boldsymbol{\Gamma}_2(E[\omega_i])^2 = o(N^{-1/2})$, provided that $\beta \neq 0$. Notably, both terms equate to zero when $\rho = -\beta/(1+\beta)$, implying a complete loss of power against this alternative. Under first-order asymptotics, there is also no power against $\rho = \beta$. Importantly, when $\beta = 0$, both terms cancel each other out, which again is intuitive because a DGP with $\beta = 0$ is observationally equivalent to a DGP with $\rho = 0$ due to the exchangeability of β and ρ.

When $T = 4$, $E[\boldsymbol{\zeta}_i] = \rho(1+\rho)^{-1}\sigma_\nu^2(-1, -\rho, -1, 1, 1)'$, and

$$\boldsymbol{\Gamma}_1 = -E\left[\begin{pmatrix} y_{i0}\Delta u_{i3} + u_{i1}\Delta y_{i2} \\ y_{i0}\Delta u_{i4} + u_{i1}\Delta y_{i3} \\ y_{i1}\Delta u_{i4} + u_{i2}\Delta y_{i3} \\ y_{i2}\Delta u_{i2} + u_{i3}\Delta y_{i1} \\ y_{i3}\Delta u_{i3} + u_{i4}\Delta y_{i2} \end{pmatrix}\right] = -\sum_{j=0}^{\infty}\beta^j E\left[\begin{pmatrix} \varepsilon_{i,0-j}\Delta\varepsilon_{i3} + \varepsilon_{i1}\Delta\varepsilon_{i,2-j} \\ \varepsilon_{i,0-j}\Delta\varepsilon_{i4} + \varepsilon_{i1}\Delta\varepsilon_{i,3-j} \\ \varepsilon_{i,1-j}\Delta\varepsilon_{i4} + \varepsilon_{i2}\Delta\varepsilon_{i,3-j} \\ \varepsilon_{i,2-j}\Delta\varepsilon_{i2} + \varepsilon_{i3}\Delta\varepsilon_{i,1-j} \\ \varepsilon_{i,3-j}\Delta\varepsilon_{i3} + \varepsilon_{i4}\Delta\varepsilon_{i,2-j} \end{pmatrix}\right]$$
$$= \frac{\beta(1+\rho) - 1}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix} - \frac{\rho^2}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\rho \\ -1 \\ 1 \\ 1 \end{pmatrix} + \frac{\rho}{1+\rho}\sigma_\nu^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

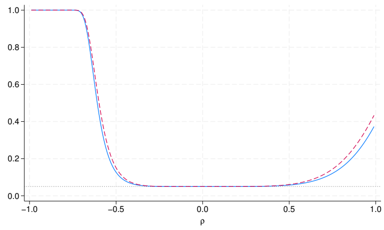
$$\mathbf{\Gamma}_2 = 2E\left[\begin{pmatrix} y_{i0}\Delta y_{i2} \\ y_{i0}\Delta y_{i3} \\ y_{i1}\Delta y_{i3} \\ y_{i2}\Delta y_{i1} \\ y_{i3}\Delta y_{i2} \end{pmatrix}\right] = 2\frac{\beta+\rho+\beta\rho}{(1+\beta)(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix} - 2\frac{\rho^2}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and thus

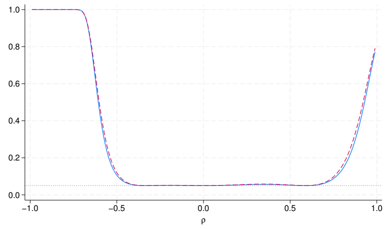
$$E[\boldsymbol{\zeta}_i] + \mathbf{\Gamma}_1 E[\omega_i] = \frac{\rho(\beta-\rho)(\beta+\rho+\beta\rho)}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} + \left(\frac{(1+\beta)\left(\rho^3-\beta(\beta+\beta\rho-1)\right)}{(1-\beta\rho)(1+\rho)} + \frac{\beta\rho}{1+\rho}\right)\sigma_\nu^2 \begin{pmatrix} 0 \\ \rho \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\frac{1}{2}\mathbf{\Gamma}_2(E[\omega_i])^2 = \frac{(1+\beta)\rho^2(\beta+\rho+\beta\rho)}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} -1 \\ -\beta \\ -1 \\ 1 \\ 1 \end{pmatrix} - \frac{(1+\beta)^2\rho^4}{(1-\beta\rho)(1+\rho)}\sigma_\nu^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

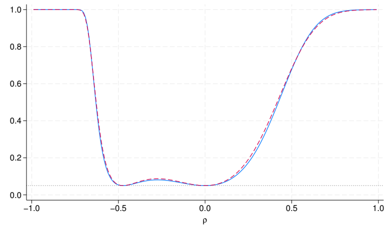
The additional terms for the second moment restriction cancel each other out under fixed alternatives when $\beta = 0$. Otherwise, they still contribute some power when the other moment restrictions evaluate to zero under the previously identified conditions. We illustrate the power profiles against AR(1) alternatives in Figures 13 and 14.



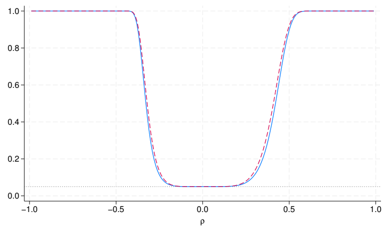
(a) $N = 100$, $\beta = 0$



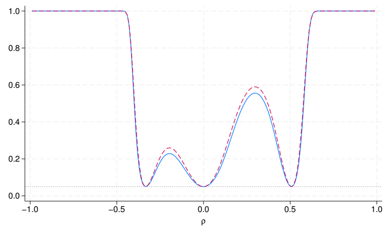
(b) $N = 100$, $\beta = 0.5$



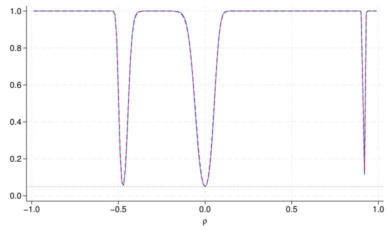
(c) $N = 100$, $\beta = 0.9$



(d) $N = 10000$, $\beta = 0$

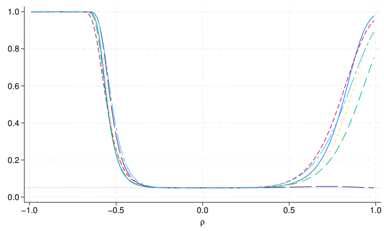


(e) $N = 10000$, $\beta = 0.5$

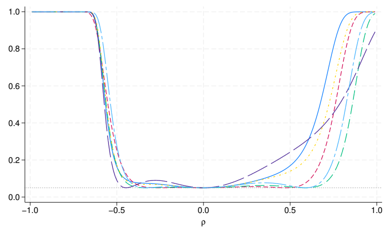


(f) $N = 10000$, $\beta = 0.9$

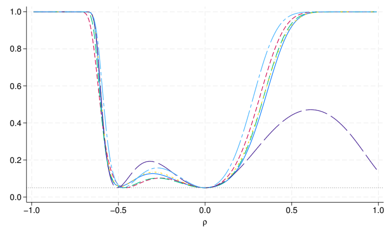
Figure 13: Theoretical power calculations for an AR(1) alternative with $T = 3$ under estimator inconsistency



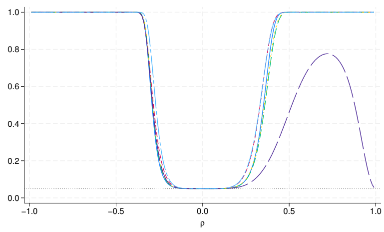
(a) $N = 100$, $\beta = 0$



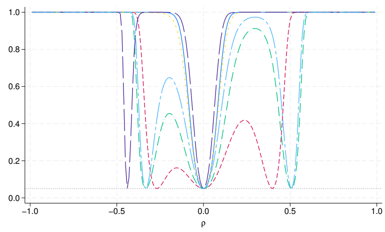
(b) $N = 100$, $\beta = 0.5$



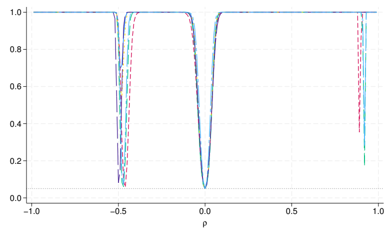
(c) $N = 100$, $\beta = 0.9$



(d) $N = 10000$, $\beta = 0$



(e) $N = 10000$, $\beta = 0.5$



(f) $N = 10000$, $\beta = 0.9$

Figure 14: Theoretical power calculations for an AR(1) alternative with $T = 4$ under estimator inconsistency