Automatic Tests of Super Exogeneity

David F. Hendry, and Carlos Santos*
Economics Department, Oxford University.

February 9, 2006

Abstract

We develop a new automatically-computable test for super exogeneity, using a variant of general-to-specific modelling. Based on the recent developments of impulse saturation applied to marginal models under the null that no impulses matter, we select the significant impulses for testing in the conditional. The approximate analytical non-centrality of the test is derived for a failure of invariance and of weak exogeneity when there is a shift in the conditional model. Monte Carlo simulations confirm the nominal significance levels under the null, and power against the two alternatives.

Keywords: super exogeneity; general-to-specific, test power, co-breaking.
JEL classifications: C51, C22.

Contents

*Financial support from the ESRC under a Professorial Research Fellowship, RES051270035, and from the Fundação para a Ciência e a Tecnologia (Lisboa), is gratefully acknowledged by the first and second authors respectively.

1 Introduction

In all areas of policy which involve regime shifts or structural breaks in conditioning variables, super exogeneity of the parameters of conditional models under changes in the distributions of conditioning variables is of paramount importance. In models without contemporaneous conditioning variables, such as vector autoregressions (VARs), invariance under such shifts is equally relevant. Tests for super exogeneity and invariance have been proposed by Engle, Hendry and Richard (1983), Hendry (1988), Favero (1989), Favero and Hendry (1992), Engle and Hendry (1993), Psaradakis and Sola (1996), Jansen and Teräsvirta (1996) and Krolzig and Toro (2002), inter alia: Ericsson and Irons (1994) overview the literature at the time of publication. Favero and Hendry (1992), building on Hendry (1988), considered the impact of non-constant marginal processes on conditional models, and concluded that location shifts were essential for detecting violations attributable to the Lucas (1976) critique. Engle and Hendry (1993) examined the impact on a conditional model of changes in the moments of the conditioning variables, using a linear approximation: tests for super exogeneity were constructed by replacing the unobservable changing moments by proxies based on models of the process generating the conditioning variables, including models based on ARCH processes (see Engle, 1982), thereby allowing for non-constant error variances to capture changes in regimes. However, Psaradakis and Sola (1996) claim that such tests have relatively low power for rejecting the Lucas critique. Jansen and Teräsvirta (1996) propose self-exciting threshold models for testing constancy in the conditional model as well as super exogeneity. Krolzig and Toro (2002) developed super-exogeneity tests based on a reduced-rank technique for co-breaking shown by the presence of common deterministic shifts, and demonstrated that their proposal dominated existing tests (on co-breaking, see Clements and Hendry, 1999, and Hendry and Massmann, 2005). We suggest new additions to this set of possible tests, show that their rejection frequencies under the null are close to their nominal significance levels, and examine their power properties for failures of super exogeneity and invariance.

The ability to detect all outliers and shifts in a model using the dummy saturation techniques proposed by Hendry, Johansen and Santos (2004) opens the door to this new class of automatically computable super-exogeneity and invariance tests. Their approach is to saturate the marginal model (or system) with impulse indicators (namely, include an impulse for every observation, but entered in feasible subsets), and retain all significant outcomes. They derive the probability under the null of falsely retaining impulses for a location-scale IID process, and obtain the distribution of the estimated mean and variance after saturation. We extend that idea to test the relevance in the conditional model of all the retained impulses from the marginal models. As we show below, such a test has the correct size under the null of super exogeneity of the conditioning variables for the parameters of the conditional model over a range of sizes of the marginal model saturation tests. Moreover, it has power to detect failures of super exogeneity and invariance when there are location shifts in the marginal models. Finally, it can be computed automatically–that is without explicit user intervention, as occurs with (say) residual autocorrelation tests–once the desired sizes of the marginal saturation and conditional super-exogeneity tests have been specified.

Five conditions need to be satisfied for an automatic test of super exogeneity and invariance. First, the test should not require *ex ante* knowledge by the investigator of the timing, signs or magnitudes of any breaks in the marginal processes of the conditioning variables. The test proposed here uses impulse saturation techniques on the marginal equations to determine these aspects. Secondly, the correct data generation process for the marginal variables should not need to be known for the test to have the desired rejection frequency under the null. That condition is satisfied here when there are no unit roots (stochastic trends) in any of the variables: we will investigate the generalization of the approach to unit-root non-stationarity in due course. Thirdly, the conditional model should not need to be over-identified under the alternative of a failure of super exogeneity, as required for tests in the class proposed by (say) Revankar and Hartley (1973). Fourthly, the test must have power against any form of failure of super exogeneity or invariance in the conditional model when there are location shifts in some of the marginal processes. Below, we establish the general forms of the non-centrality parameters of the proposed tests in the two main cases. Finally, the test should be computable without additional user intervention. That is true of the impulse saturation test based on PcGets, although as yet the precise form of the test procedure is not implemented in any released version.[1]

The structure of the paper is as follows. Section 2 considers super exogeneity in a regression context to elucidate the testable hypotheses which it entails. Next, section 3 discusses the three different ways in which super exogeneity can fail, and how each could be tested. Section 4 describes the impulse saturation tests developed by Hendry *et al.* (2004), and how these can be extended to test super exogeneity and invariance. Section 5 provides analytic and Monte Carlo evidence on the null rejection frequency of the proposed procedure. Section 6 considers the power of the first stage to determine the location shifts in the marginal processes. Then section 7 provides detailed analytic derivations for three multivariate examples of super exogeneity failures, namely a failure of weak exogeneity under non-constant marginal processes; a failure of invariance of the conditional model parameters to shifts in those of the marginal distributions; and a failure of weak exogeneity with constant marginal processes, which is a case where the proposed tests may have little power. Section 8 investigates a co-breaking based saturation test which builds on Krolzig and Toro (2002) and Hendry and Massmann (2005). Section 9 investigates the powers of the proposed tests in an extensive set of Monte Carlo experiments related to the analysis in section 7 for a bivariate DGP. Section 10 describes similar Monte Carlo experiments with $n = 3$ variables. Section 11 concludes.

[1] PcGets is an Ox Package (see Doornik, 2001, and Hendry and Krolzig, 1999), designed for general to specific modelling.

2 Super exogeneity in a regression context

Consider the sequentially factorized joint data generation process (DGP) of an n-dimensional vector process $\{\mathbf{x}_t\}$:

$$\prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{y}|\mathsf{z}}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \mathsf{D}_{\mathsf{z}}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right) \tag{1}$$

where $\mathbf{x}'_t = (\mathbf{y}'_t : \mathbf{z}'_t)$ and $\boldsymbol{\phi} = \left(\boldsymbol{\phi}'_1 : \boldsymbol{\phi}'_2\right)' = \mathbf{f}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^k$. The parameters of the $\mathbf{y}$ and $\mathbf{z}$ processes need to be variation free for $\mathbf{z}_t$ to be weakly exogenous for the parameters of interest $\psi = \mathbf{h}\left(\boldsymbol{\phi}_1\right)$, but that does not rule out the possibility that $\boldsymbol{\phi}_1$ may change if $\boldsymbol{\phi}_2$ is changed. Super exogeneity augments weak exogeneity with such parameter invariance in the conditional model.

When $\mathsf{D}_{\mathsf{x}}\left(\cdot\right)$ is the multivariate normal, we can express (1) as the unconditional model:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n \left[\begin{pmatrix} \mu_{1,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \sigma_{11,t} & \boldsymbol{\sigma}'_{12,t} \\ \boldsymbol{\sigma}_{12,t} & \boldsymbol{\Omega}_{22,t} \end{pmatrix} \right] \tag{2}$$

where $\mu_{1,t}$ and $\boldsymbol{\mu}_{2,t}$ are possibly functions of $\mathbf{X}_{t-1}$. To define the parameters of interest, we let the economic theory formulation entail:

$$\mu_{1,t} = \mu_0 + \boldsymbol{\beta}' \boldsymbol{\mu}_{2,t} \tag{3}$$

where $\boldsymbol{\beta}$ is the primary parameter of interest. The Lucas (1976) critique explicitly considers a model where expectations (the latent decision variables given by the $\boldsymbol{\mu}_{2,t}$) are incorrectly modelled by the outcomes $\mathbf{z}_t$. From (2) and (3):

$$\begin{aligned} \mathsf{E}\left[y_t \mid \mathbf{z}_t\right] &= \mu_{1,t} + \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Omega}^{-1}_{22,t} \left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \\ &= \mu_0 + \left(\boldsymbol{\beta}' - \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Omega}^{-1}_{22,t}\right) \boldsymbol{\mu}_{2,t} + \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Omega}^{-1}_{22,t} \mathbf{z}_t \\ &= \mu_0 + \gamma_{1,t} + \boldsymbol{\gamma}'_{2,t} \mathbf{z}_t \end{aligned} \tag{4}$$

where $\boldsymbol{\gamma}'_{2,t} = \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Omega}^{-1}_{22,t}$ and $\gamma_{1,t} = \left(\boldsymbol{\beta} - \boldsymbol{\gamma}_{2,t}\right)' \boldsymbol{\mu}_{2,t}$. The conditional variance is $\omega^2_t = \sigma_{11,t} - \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Omega}^{-1}_{22,t} \boldsymbol{\sigma}_{21,t}$. Thus, the parameters of the conditional and marginal densities respectively are:

$$\boldsymbol{\phi}_{1,t} = \left(\mu_0 : \gamma_{1,t} : \boldsymbol{\gamma}_{2,t} : \omega^2_t\right) \text{ and } \boldsymbol{\phi}_{2,t} = \left(\boldsymbol{\mu}_{2,t} : \boldsymbol{\Omega}_{22,t}\right).$$

When (4) is specified as the regression model for $t = 1, \ldots, T$:

$$y_t = \mu_0 + \boldsymbol{\beta}' \mathbf{z}_t + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \omega^2\right] \tag{5}$$

four conditions must be satisfied for $\mathbf{z}_t$ to be super exogenous for $\left(\boldsymbol{\beta}, \omega^2\right)$ (see e.g., Engle and Hendry, 1993):

(a) $\boldsymbol{\gamma}_{2,t} = \boldsymbol{\gamma}_2$ is constant $\forall t$;
(b) $\boldsymbol{\beta} = \boldsymbol{\gamma}_2$;
(c) $\omega^2_t = \omega^2$ is constant $\forall t$;
(d) $\boldsymbol{\phi}_{1,t}$ is invariant to $\mathcal{C}^{\phi_2}$.

Condition (a) requires that $\boldsymbol{\sigma}'_{12,t}\boldsymbol{\Omega}^{-1}_{22,t}$ is constant over time, which could occur because the two components move in tandem through being connected by $\boldsymbol{\sigma}'_{12,t} = \boldsymbol{\gamma}'_2\boldsymbol{\Omega}_{22,t}$, as well as because the σ_{ij} happened not to change over the sample. Condition (b) then entails that $\mathbf{z}_t$ is weakly exogenous for a constant $\boldsymbol{\beta}$. Together, (a)+(b) also entail that $\gamma_{1,t} = 0$ and hence the conditional expectation in (4) is independent of $\boldsymbol{\mu}_{2,t}$. Condition (c) then entails in turn that $\sigma_{11,t} - \boldsymbol{\sigma}'_{12,t}\boldsymbol{\Omega}^{-1}_{22,t}\boldsymbol{\sigma}_{21,t} = \sigma_{11,t} - \boldsymbol{\beta}'\boldsymbol{\Omega}_{22,t}\boldsymbol{\beta} = \omega^2$ is constant. Finally, in (d), $\mathcal{C}^{\phi_2}$ is a class of interventions changing the marginal process parameters $\boldsymbol{\phi}_2$, so (d) requires no cross links between the conditional and marginal parameters. When the four conditions (a)–(d) are satisfied, then:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t, \tag{6}$$

in which case $\mathbf{z}_t$ is super exogenous for $\boldsymbol{\beta}$ in this model. That requires in turn:

$$\boldsymbol{\sigma}'_{12,t} = \boldsymbol{\beta}'\boldsymbol{\Omega}_{22,t}\ \forall t. \tag{7}$$

The necessary condition (7) requires that the means in (3) are interrelated by the same parameter $\boldsymbol{\beta}$ as the covariances $\boldsymbol{\sigma}_{12,t}$ are with the variances $\boldsymbol{\Omega}_{22,t}$. Under super exogeneity, the joint density is:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n\left[\begin{pmatrix} \mu_0 + \boldsymbol{\beta}'\boldsymbol{\mu}_{2,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega^2 + \boldsymbol{\beta}'\boldsymbol{\Omega}_{22,t}\boldsymbol{\beta} & \boldsymbol{\beta}'\boldsymbol{\Omega}_{22,t} \\ \boldsymbol{\Omega}_{22,t}\boldsymbol{\beta} & \boldsymbol{\Omega}_{22,t} \end{pmatrix}\right], \tag{8}$$

so the conditional-marginal factorization is:

$$\begin{pmatrix} y_t \mid \mathbf{z}_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n\left[\begin{pmatrix} \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega^2 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Omega}_{22,t} \end{pmatrix}\right], \tag{9}$$

Consequently, under super exogeneity, the parameters $(\boldsymbol{\mu}_{2,t}, \boldsymbol{\Omega}_{22,t})$ can change in the marginal model:

$$\mathbf{z}_t \sim \mathsf{IN}_{n-1}\left[\boldsymbol{\mu}_{2,t}, \boldsymbol{\Omega}_{22,t}\right], \tag{10}$$

without altering the parameters of (5). Deterministic-shift co-breaking will occur in (8) as $\left(1 : \boldsymbol{\beta}'\right)\mathbf{x}_t$ does not depend on $\boldsymbol{\mu}_{2,t}$. Conversely, if $\mathbf{z}_t$ is not super exogenous for $\boldsymbol{\beta}$, then changes in (10) should affect (5).

3 Failures of super exogeneity

Super exogeneity may fail for any of three reasons:

(i) $\mathbf{z}_t$ is not weakly exogenous for $\boldsymbol{\beta}$, in which case the coefficient in a regression of y_t on $\mathbf{z}_t$ will not coincide with $\boldsymbol{\beta}$;

(ii) the regression coefficient is not constant;

(iii) $\boldsymbol{\beta}$ is not invariant to changes in $\mathcal{C}^{\phi_2}$.

From (4), when $\mathbf{z}_t$ is not super exogenous for $\boldsymbol{\beta}$ but (3) holds:

$$\begin{aligned} \mathsf{E}\left[y_t \mid \mathbf{z}_t\right] &= \mu_{1,t} + \boldsymbol{\sigma}'_{12,t}\boldsymbol{\Omega}^{-1}_{22,t}\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \\ &= \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \left(\boldsymbol{\gamma}'_{2,t} - \boldsymbol{\beta}'\right)\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \\ &= \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \left(\boldsymbol{\gamma}'_{2,t} - \boldsymbol{\beta}'\right)\mathbf{v}_{2,t} \end{aligned} \tag{11}$$

where $\mathbf{v}_{2,t}$ is the error on the marginal model (10):

$$\mathbf{z}_t = \boldsymbol{\mu}_{2,t} + \mathbf{v}_{2,t} \text{ where } \mathbf{v}_{2,t} \sim \mathsf{IN}_{n-1}\left[\mathbf{0}, \boldsymbol{\Omega}_{22,t}\right].$$

Modelling $\boldsymbol{\mu}_{2,t}$ by lagged values of $\mathbf{x}_t$, to approximate the sequential factorization, yields the augmented VAR:

$$\mathbf{z}_t = \boldsymbol{\pi}_0 + \sum_{j=1}^{s} \boldsymbol{\Pi}_j \mathbf{x}_{t-j} + \mathbf{v}_{2,t} \text{ where } \mathbf{v}_{2,t} \sim \mathsf{IN}_{n-1}\left[\mathbf{0}, \boldsymbol{\Omega}_{22,t}\right]. \tag{12}$$

The introduction reviewed the currently available tests for for super exogeneity. The next section proposes new tests for super exogeneity based on impulse saturation after briefly reviewing that procedure as applied to the marginal process.

4 Impulse saturation tests

A key recent development is that of testing for non-constancy by adding a complete set of impulse indicators $\left\{1_{\{t\}}, t = 1, \ldots, T\right\}$ to a marginal model: see Hendry *et al.* (2004). Using a general-to-specific procedure, those authors analytically establish the null distribution of the estimator of the mean in a location-scale IID distribution after adding T impulse indicators when the sample size is T. A two-step process is investigated, where half the indicators are added, and all significant indicators recorded, then the other half examined, and finally the two retained sets of indicators are combined. The average retention rate of impulse indicators under the null is αT when the significance level of an individual test is set at α, so for $\alpha = 0.01$, for example, $0.01T$ indicators will be retained. Moreover, Hendry *et al.* (2004) show by simulation that other splits, such as reordering the impulses, or using three splits of size $T/3$, do not affect the retention rate under the null, or the simulation-based distribution of the estimated mean.

This procedure can be applied to the marginal models for the putative super-exogenous conditioning variables. First, the associated significant dummies in the marginal processes are recorded. Secondly, those which are retained are tested as an added variable set in the conditional model. Specifically, after the first stage when m impulse indicators are retained, a marginal model like (12) has been extended to:

$$\mathbf{z}_t = \boldsymbol{\pi}_0 + \sum_{j=1}^{s} \boldsymbol{\Pi}_j \mathbf{x}_{t-j} + \sum_{i=1}^{m} \boldsymbol{\rho}_{i,\alpha_1} 1_{\{t=t_i\}} + \mathbf{v}^*_{2,t} \tag{13}$$

where the coefficients of the significant impulses are denoted $\boldsymbol{\rho}_{i,\alpha_1}$ to emphasize their dependence on the significance level α_1 used in the marginal model. As just noted, this test has the appropriate null rejection frequency.

There is an important difference between outlier detection, which does just that, and impulse saturation which will detect outliers, but may also reveal others that are hidden by being 'picked up' incorrectly by other variables. Figure 1 illustrates for a mean shift near the mid-sample, where no outliers, as defined by $|\widehat{u}| > 2\sigma$ (say), are detected, but 40 dummies are significant in the PcGets approach (for an alternative method of tackling such problems, see Sánchez and Peña, 2003).

The second stage is to add the m retained impulses to the conditional model, yielding:

$$y_t = \mu_0 + \boldsymbol{\beta}' \mathbf{z}_t + \sum_{i=1}^{m} \tau_{i,\alpha_2} 1_{\{t=t_i\}} + \epsilon_t \tag{14}$$

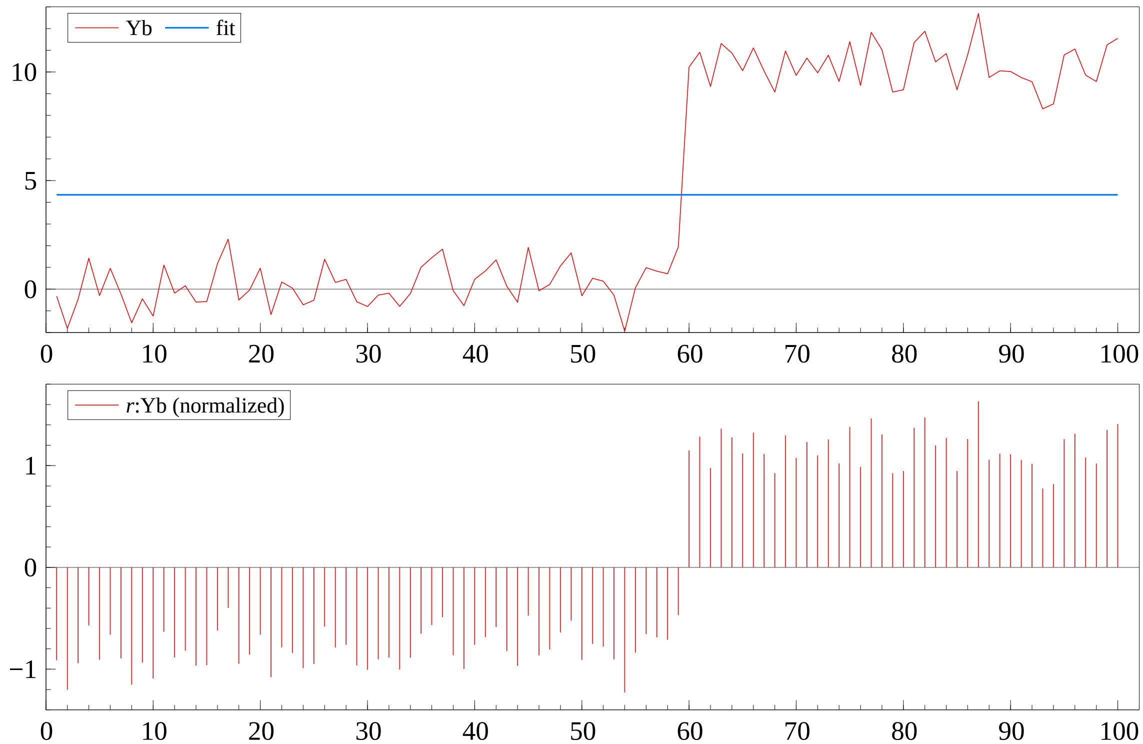

Figure 1 Absence of outliers despite a break.

and conduct an F-test for the significance of $(\tau_{1,\alpha_2} \dots \tau_{m,\alpha_2})$ at level α_2. Under the null of super exogeneity, the F-test of the joint significance of the m impulse indicators in the conditional model should have an approximate F-distribution and thereby allow an appropriately sized test: section 5 derives the null distribution and presents Monte Carlo evidence on its small-sample relevance. Under the alternative, the test will have power in a variety of situations discussed in section 7 below. Crucially, such a test can be completely automated, bringing super exogeneity into the purview of hypotheses about a model that can be as easily tested as (say) residual autocorrelation. Intuitively, if super exogeneity is invalid, so $\boldsymbol{\beta}' \neq \boldsymbol{\sigma}'_{12,t}\boldsymbol{\Omega}^{-1}_{22,t}$ in (11), then the impact of the largest values of the errors $\mathbf{v}_{2,t}$ on the conditional model should be the easiest to detect, noting that the significant impulses in (13) capture the outliers not accounted for by the regressor variables used.

A key feature of such a test is that the null rejection frequency of super exogeneity by this F-test in the conditional model should not depend on the significance level, α_1, set for each individual test in the marginal model. Monte Carlo evidence presented in section 5.1 supports that contention. Thus, the main consideration for choosing α_1 is power against reasonable alternatives to super exogeneity. Too large a value of α_1 will lead to an F-test with large degrees of freedom; too small will lead to few, or even no, impulses being retained from the marginal models. For example, with four regressors and $T = 100$ then $\alpha_1 = 0.01$ would yield four impulses in general, whereas $\alpha_1 = 0.05$ would provide 20.

Following Hendry and Santos (2005), a variant of the test in (14), discussed in more detail below, which could have different power characteristics, is to combine the m impulses detected in all the equations of (13) into an index:

$$\iota_{1,t} = \sum_{i=1}^{m} \widehat{\varrho}_{i,\alpha_1} 1_{\{t=t_i\}} \text{ where } \widehat{\varrho}_{i,\alpha_1} = \sum_{j=1}^{n-1} \widehat{\rho}_{j,i,\alpha_1} \tag{15}$$

and test the null of $\varphi_1 = 0$ in:

$$y_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \varphi_1 \iota_{1,t} + \epsilon_t. \tag{16}$$

This provides an alternative scalar test with $T-n-1$ degrees of freedom, which should be approximately distributed as t under the null of super exogeneity. In general, there should be many fewer degrees of freedom for such a test; the cost of the imposed restrictions is that the implicit null must be larger. Indeed, we show below that there are cases where its power would be low, and be dominated by the F-test. Also, for testing a failure of invariance, the indices must be interacted with $\mathbf{z}_t$ as in:

$$\iota_{2,t} = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \widehat{\rho}_{j,i,\alpha_1} z_{j,t} 1_{\{t=t_i\}} \tag{17}$$

and then test for the null of $\varphi_1 = \varphi_2 = 0$ in:

$$y_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \varphi_1 \iota_{1,t} + \varphi_2 \iota_{2,t} + \epsilon_t \tag{18}$$

using a 2-degrees-of-freedom F-test. By focusing on the empirically detected departures in the marginal process, such tests should have power under the alternative: below, we derive their large sample non-centralities in three central cases.

Alternatively, if some interest resides in which of the $z_{j,t}$ is responsible for any failure of super exogeneity, then a vector test of the form in (19) could be used, which might have more or fewer degrees of freedom than the corresponding F-test in (14):

$$\boldsymbol{\iota}_{2,t} = \begin{pmatrix} \iota_{2,1,t} \\ \iota_{2,2,t} \\ \vdots \\ \iota_{2,n-1,t} \end{pmatrix} \text{ where } \iota_{2,j,t} = \sum_{i=1}^{m_j} \widehat{\rho}_{j,i,\alpha_1} z_{j,t} 1_{\{t=t_i\}} \tag{19}$$

with m_j being the number of retailed impulses in the marginal model for $z_{j,t}$.

5 The null rejection frequency of the super-exogeneity test

Reconsider the earlier DGP:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_n \left[\begin{pmatrix} \mu_{1,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}'_{12} \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \tag{20}$$

where:

$$\mathbf{z}_t = \boldsymbol{\mu}_{2,t} + \mathbf{v}_{2,t} \tag{21}$$

and:

$$\mathbf{v}_t \sim \mathsf{N}_{n-1} \left[\mathbf{0}, \boldsymbol{\Sigma}_{22} \right]. \tag{22}$$

Then:

$$y_t = \mu_{1,t} + \varepsilon_t + \boldsymbol{\gamma}'\mathbf{v}_t \tag{23}$$

with $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$ so:

$$\varepsilon_t \mid \mathbf{v}_t \sim \mathsf{N}\left[0, \sigma_{11} - \boldsymbol{\sigma}'_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}\right]. \tag{24}$$

Under the null of super exogeneity, from (57):

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t \tag{25}$$

with:

$$y_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \varepsilon_t \tag{26}$$

where:

$$\mu_{1,t} = \mu_0 + \boldsymbol{\beta}'\boldsymbol{\mu}_{2,t} \text{ and } \boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12} = \boldsymbol{\gamma}.$$

Thus, even though the $\mathbf{z}_t$ process is non-constant, the linear relation between y_t and $\mathbf{z}_t$ in (25) is constant. Consequently, from (26), for any $k \times (n-1)$ selection matrix $\mathbf{S}$ with $k \leq (n-1)$ of rank k having elements that are zero except for unity in a different location in each column:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \mathbf{0}'\mathbf{S}\mathbf{v}_{2,t} = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \mathbf{0}'\mathbf{S}\boldsymbol{\mu}_{2,t} \tag{27}$$

so:

$$y_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \mathbf{0}'\mathbf{S}\mathbf{v}_{2,t} + \varepsilon_t. \tag{28}$$

The errors, or components of $\boldsymbol{\mu}_{2,t}$, from (21) will have zero population components when added to (26). Tests of the significance of $\mathbf{S}\mathbf{v}_{2,t}$ or $\mathbf{S}\boldsymbol{\mu}_{2,t}$ in (28) should reject at their nominal significance level. In particular, selecting which components to add by analyzing the marginal process should not alter this argument.

Consider the following baseline econometric model for $\mathbf{z}_t$:

$$\mathbf{z}_t = \boldsymbol{\Phi}\mathbf{d}_t + \mathbf{v}_{2,t} \tag{29}$$

where $\mathbf{d}_t$ is a set of impulse indicator variables with:

$$\widehat{\boldsymbol{\Phi}} = \left(\sum_{t=1}^{T}\mathbf{d}_t\mathbf{d}_t'\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{d}_t\mathbf{v}_{2,t_i}\right). \tag{30}$$

Suppose that each of the impulses is retained in the econometric model for the marginal process when:

$$\left|\mathsf{t}_{\widehat{\phi}_{i,j}}\right| > c_{\alpha_1} \tag{31}$$

where c_{α_1} is chosen according to a given significance level α_1. Now, consider the econometric model for $y_t|\mathbf{z}_t$ as in (28). Conditioning on $\mathbf{z}_t$ implies taking the $\mathbf{z}_t$s as fixed, and hence the $\mathbf{v}_{2,t}$s. Thus, the conditional econometric model remains:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\delta}'\mathbf{d}_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t. \tag{32}$$

Given a significance level α_2, indicators will be retained in the conditional econometric model, given that they are retained in the marginal if:

$$\left|\mathsf{t}_{\widehat{\boldsymbol{\delta}}_i}\right| > c_{\alpha_2}. \tag{33}$$

The probability of retaining the indicator in the conditional is:

$$\Pr\left(\left|\mathsf{t}_{\widehat{\boldsymbol{\delta}}_i}\right| > c_{\alpha_2} \mid \left|\mathsf{t}_{\widehat{\phi}_{i,j}}\right| > c_{\alpha_1}\right) = \Pr\left(\left|\mathsf{t}_{\widehat{\boldsymbol{\delta}}_i}\right| > c_{\alpha_2}\right) = \alpha_2 \tag{34}$$

since (31) holds. Moreover, this results only depends on the significance level c_{α_2} used on the conditional model and not on α_1.

5.1 Monte Carlo evidence on the null rejection frequency

In these Monte Carlo experiments, super exogeneity holds as the null, and we consider three settings for the marginal process: where there are no breaks in §5.1.1; a variance change in §5.1.2; and a mean shift in §5.1.3. In each case, the baseline DGP is a bivariate system which can be expressed as (see e.g., Hendry, 1995):

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{35}$$

which in turn implies $\beta = 2 = \gamma$ and $\omega^2 = 1$, the parameters of interest in the conditional econometric model.

The aim of the Monte Carlo experiments is to establish the null rejection frequencies of the extended super-exogeneity tests, and ascertain their dependence, if any, on the nominal significance level for impulse retention in the marginal process. Thus, impulse saturation of the marginal model and retention of the relevant indicators should not require us to change the critical values used to test such indicators in the conditional model. If so, pre-searching for the relevant dates at which shifts might have occurred in the marginal, does not affect testing for associated shifts in the conditional.

We consider a constant DGP and two DGPs with changes in the z_t process, all under the null of super exogeneity, where invariance and weak exogeneity hold before and after the change in the marginal process. For the baseline DGP in (35), the parameters of the conditional model $y_t|z_t$ are $\boldsymbol{\phi}_{1,t} = \left(\gamma_{1,t}; \gamma_{2,t}; \omega_t^2\right)$, where $\gamma_{1,t} = 0$ by virtue of weak exogeneity, and $\gamma_{2,t} = \sigma_{12,t}\sigma_{22,t}^{-1}$ with $\omega_t^2 = \sigma_{11,t} - \sigma_{12,t}^2\sigma_{22,t}^{-1} = 1$ and $\gamma_{2,t} = 2 = \beta$. The parameters of the marginal model are $\boldsymbol{\phi}_{2,t} = (\mu_{2,t}; \sigma_{22,t})$. Changes in the marginal process always occur at time $T_1 = 81$, implying $k = 20$. We examine several significance levels for testing and retaining impulses in the saturated location-scale model for the marginal, and also allow the significance levels for testing in the conditional to vary. The impulse saturation uses a partition of $T/2$ with $M = 10000$ replications conducted in the Monte Carlo experiments.

5.1.1 Constant marginal under the null of super exogeneity

We use the simplest marginal model, defined by:

$$z_t = 1 + v_t \tag{36}$$

where $v_t \sim \mathsf{IN}[0, 5]$. This econometric model mimics the location-scale model analysis in Hendry *et al.* (2004). As a sample split of $T/2$ is used, the econometric models for the marginal are:

$$z_t = \mu_2 + \sum_{t=1}^{T/2} \psi_t 1_t + \varsigma_t \tag{37}$$

and:

$$z_t = \mu_2 + \sum_{t=T/2+1}^{T} \psi_t 1_t + \xi_t. \tag{38}$$

Let S_{α_1} denote the set of significant dummies in the econometric models (37) and (38). Our test strategy entails introducing these dummies into the econometric model for the conditional. Hence, the second stage of the extended test is to estimate:

$$y_t = \beta z_t + \sum_{i \in S_{\alpha_1}} \phi_i 1_{t_i} + \nu_t \tag{39}$$

and to test the joint significance of the dummies defined by S_{α_1} in the conditional model. Averaging across the M replications, we obtain the average relative frequency with which a block of indicators included in (39), due to belonging to S_{α_1}, is retained in the conditional. Given that we have imposed super exogeneity by design, we expect such a null rejection frequency to be close to the postulated nominal significance level. This would constitute evidence that no distortion in selection of indicators was introduced by dummy saturation in the marginal model followed by testing for joint significance of the retained dummies of the marginal in the conditional.

However, the marginal tests should not use too low a probability of retaining impulses, or else the conditional must automatically have a zero null rejection frequency. At $T = 50$ and $\alpha_1 = 0.01$, about one impulse per two trials will be retained, so half the time, no impulses will be retained; on the other half of the trials, about α_2 will be retained, so roughly $0.5\alpha_2$ will be found overall, as simulation confirms (unconditional rejection frequencies were recorded throughout).

Figure 2 reports the empirical rejection frequencies of the null in the conditional model when the significant dummies from the marginal are added as in (39). As before, α_1 represents the nominal significance level used for the t-tests on each individual indicator in the marginal model (horizontal axis), and α_2 represents the significance level for the F and t tests on the retained dummies in the conditional (vertical axis).

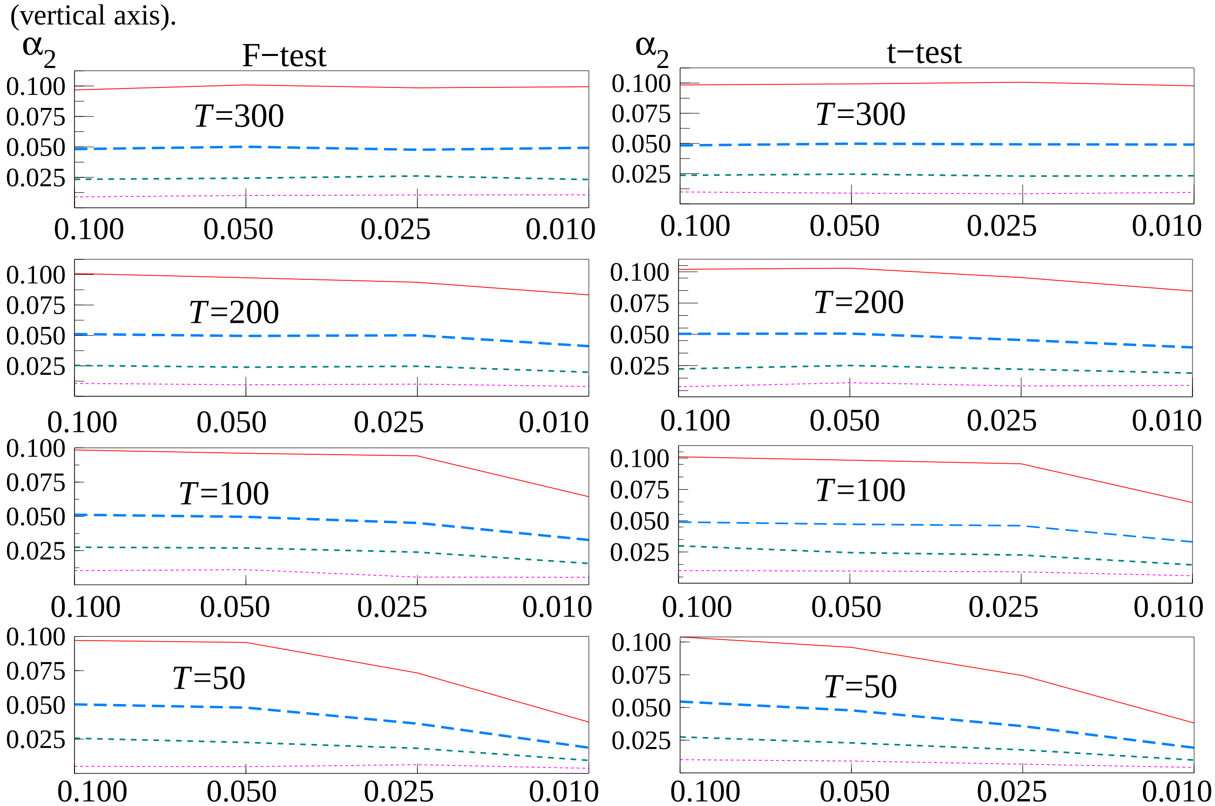

Figure 2 Null rejection frequencies of F and t tests in conditional as α_1 varies for constant marginal.

The simulated null rejection frequencies and the nominal significance levels in the conditional model are close for the F and t-tests so long as $T \times \alpha_1 > 3$. Then, there is no distortion in the number of retained dummies for either test in the conditional under the null, when t-tests are used in the marginal model. However, constant marginal processes are the 'worst-case': the next two sections consider mean and variance changes where many outliers are retained, so there are fewer cases of zero impulses to enter in the conditional leading to constant α_2 as α_1 varies.

5.1.2 Changes in the variance of z_t under the null of super exogeneity

The DGP for $T > T_1 = 0.8T$ is given by:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1+20\theta & 10\theta \\ 10\theta & 5\theta \end{pmatrix} \right] \tag{40}$$

so $\sigma_{22,t}$ is multiplied by a positive scalar θ, where $\sigma_{12,t}$ adjusts accordingly. Then, the new $\gamma^*_{2,t}$ is such that:

$$\gamma^*_{2,t} = \frac{\sigma^*_{12,t}}{\sigma^*_{22,t}} = \frac{10\theta}{5\theta} = \gamma_{2,t} = 2 = \beta. \tag{41}$$

Hence, the change in $\boldsymbol{\phi}_{2,t}$ induced by a change in $\sigma_{22,t}$ does not cause a change in $\gamma_{2,t}$. Also, $\gamma_{1,t} = (\beta - \gamma_{2,t})\,\mu_{2,t} = 0$ and, since $\sigma^*_{11,t} = 1 + 20\theta$:

$$\omega^{2*}_t = \sigma^*_{11,t} - \left(\sigma^*_{12,t}\right)^2 \left(\sigma^*_{22,t}\right)^{-1} = 1 + 20\theta - \frac{100\theta^2}{20\theta} = \omega^2_t = 1. \tag{42}$$

Thus, in this class of DGPs, $\boldsymbol{\phi}_{1,t}$ is invariant to changes in $\boldsymbol{\phi}_{2,t}$ induced by changes in $\sigma_{22,t}$. Since weak exogeneity and invariance hold, super exogeneity holds, so the null distributions of the tests should remain as in subsection 5.1.1.

However, the impulse saturation test has power to detect the variance shift in the marginal: this was presaged in Hendry and Santos (2005), who showed that impulse dummies could be used to discriminate between mixtures of distributions in marginal processes, and the variance shift here is simply a time-ordered example thereof. Thus, unlike the previous case, where only αT impulses would be retained on average, the number retained depends on the power of the impulse saturation test in the marginal. We investigate that power in subsection 6.2.

Figure 3 reports the empirical rejection frequencies of the null in the conditional model when testing the significance of the dummies selected from the marginal. Again, α_2 represents the significance level for the F and t tests on the retained dummies in the conditional (vertical axis), and the horizontal axis corresponds to the three values of $\theta = (2; 5; 10)$ for $\alpha_1 = 2.5\%$ throughout.

Both the F and t tests have appropriate null rejection frequencies for $T > 100$, even when the variance of the marginal process changes markedly, but are slightly undersized at $T = 50$ for small shifts when sometimes no impulses may be retained. Neither test is confused between variance changes in the marginal and failure of super exogeneity, when the null holds. The next sub-section assesses empirical rejection frequencies when mean shifts occur in the marginal process.

5.1.3 Changes in the mean of z_t under the null of super exogeneity

We modify the baseline DGP (35) to:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} \beta\delta\mu_z \\ \delta\mu_z \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{43}$$

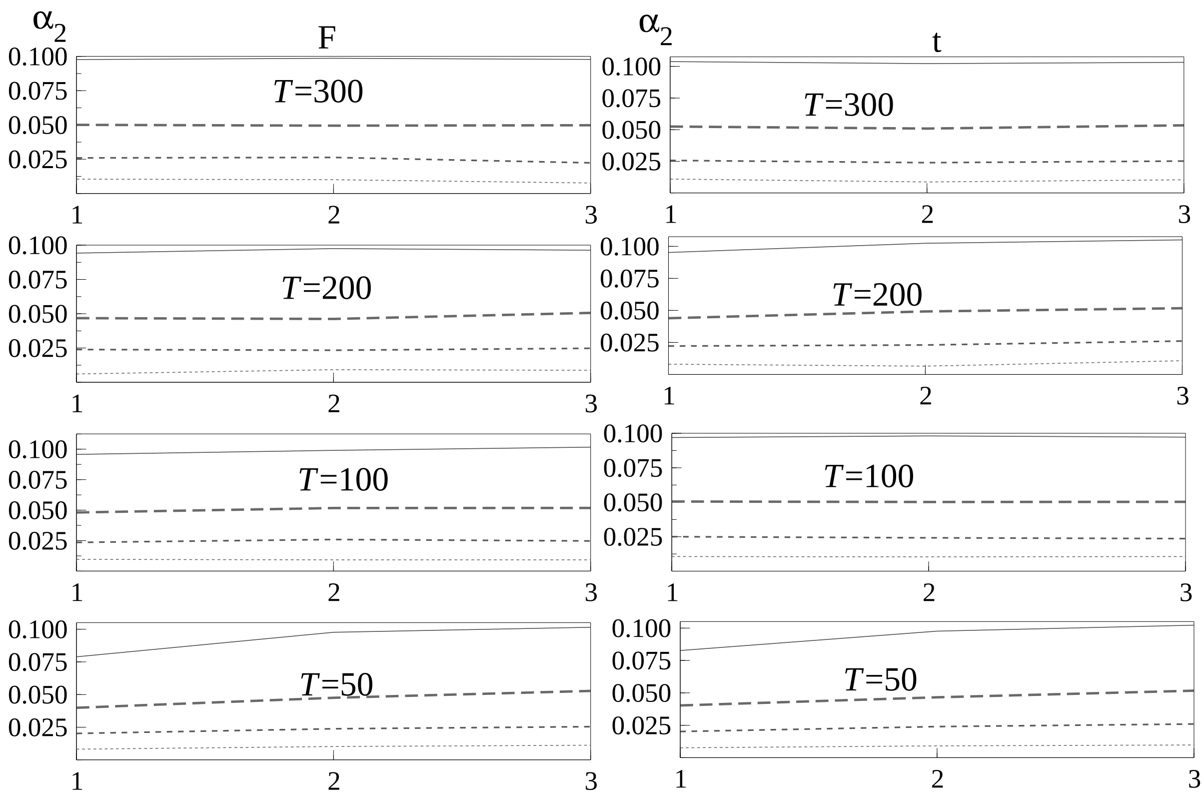

Figure 3 Null rejection frequencies of F and t tests in conditional as α_1 varies for a variance shift in the marginal.

where $\delta = 1$ until $t > T_1 = 0.8T$ when $\delta \in \mathbb{R}$, in both cases with $\beta = 2$. Super exogeneity holds before and after the level shift. We assume that the variance-covariance matrix remains the same before and after the shift, but it could be allowed to change as well, provided the values matched conditions for super exogeneity.

We consider rather extreme cases of level shifts where the current unconditional mean of z_t is multiplied by factors of $\delta = 2$, $\delta = 10$ up to $\delta = 100$. Figure 4 reports the empirical rejection frequencies where the horizontal axis corresponds to these three values of δ, again for $\alpha_1 = 2.5\%$ throughout.

In this extreme scenario, when $T > 100$ the empirical rejection frequencies are never more than two tenths of a percentage point away from the nominal significance levels postulated. Both tests do well for all larger sample sizes in failing to spuriously reject the null of super exogeneity when the null is true, but as before are slightly undersized at $T = 50$ for small shifts, when sometimes no impulses may be retained.

One might intuitively think that the length of the break matters as far as spurious rejection of super exogeneity is concerned. An experiment not reported here reveals that is not the case: even if 45% of the sample was contaminated with a level shift in the marginal of $\delta = 100$, the empirical rejection frequencies are, even at a loose nominal significance of 10% in the conditional, 9.88% when the block F- test is used, and 9.74% when the index test is used.

Overall, we conclude that the new tests have appropriate null rejection frequencies for both constant and changing marginal processes, so turn to their ability to detect failures of exogeneity. This is a two-stage process: fist detect shifts in the marginal, then use those to detect shifts in the conditional. We now consider the properties of the first stage.

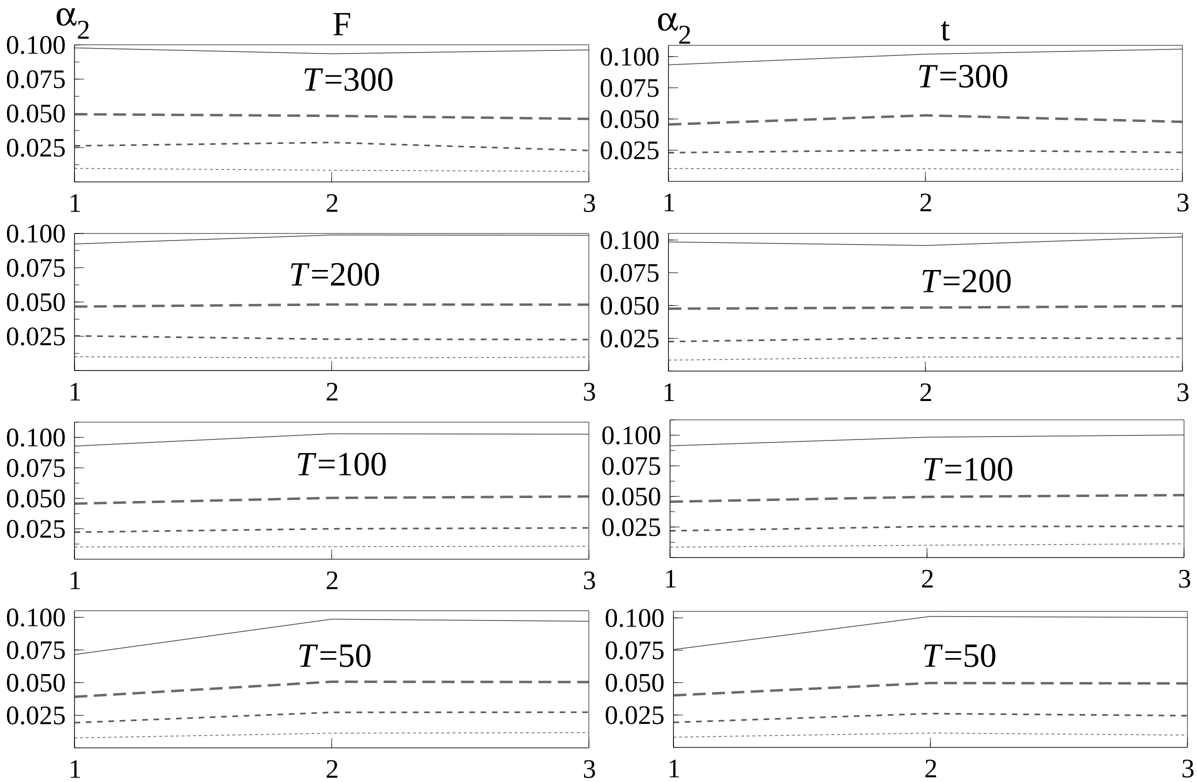

Figure 4 Null rejection frequencies of F and t tests in conditional as α_1 varies for a mean shift in the marginal.

6 Powers at stage 1

We consider the powers at stage 1 for both the mean shift and the variance change, in that order.

6.1 Detecting the mean shift in the marginal

The powers at the second stage conditional on knowing the break dates in the marginal, and hence correctly retaining every dummy, are easily calculated, but will only be accurate for large magnitude breaks, parameterized below by $\boldsymbol{\lambda}$, when the saturation approach locates all, and only, the relevant impulses. For smaller values of $\boldsymbol{\lambda}$, fewer impulses will be detected in the marginal, and indeed, although the null rejection frequency of the test does not depend on α_1, the power will, suggesting a looser α_1. Unfortunately, that in turn could lead to retaining spurious dummies (albeit fewer than $\alpha_1 T_1$ as a location shift lowers the null rejection frequency in the marginal: see Hendry *et al.*, 2004).

The power to retain each dummy in each marginal model given in its simplest form by:

$$z_{j,t} = \sum_{i \in S_{\alpha_1}} \rho_{i,j,\alpha} 1_{\{t=t_i\}} + v^*_{2,j,t}, \tag{44}$$

when the marginal process is (45), namely:

$$z_{j,t} = \lambda_j 1_{\{t>T_1\}} + v_{2,j,t} \tag{45}$$

depends on the probability of rejecting the null for the associated estimated coefficient in (44):

$$\widehat{\rho}_{i,j,\alpha} = \lambda_j + v^*_{2,j,t_i}.$$

The properties of such impulse indicators are discussed in Hendry and Santos (2005). Here, as $\mathsf{V}[\widehat{\rho}_{i,j,\alpha}] = \sigma_{22,j}$:

$$\mathsf{E}\left[\mathsf{t}_{\rho_{i,j,\alpha}=0}\left(\psi_{\lambda,\alpha}\right)\right] = \mathsf{E}\left[\frac{\widehat{\rho}_{i,j,\alpha}}{\sqrt{\widehat{\sigma}_{22,j}}}\right] \simeq \frac{\lambda_j}{\sqrt{\sigma_{22,j}}} = \psi_{\lambda,\alpha} \tag{46}$$

(say).

When $v_{2,j,t}$ is normal, the power could be computed directly from the t-distribution. However, we compute the power function here and below using an approximation to $\mathsf{t}^2_{\rho_{i,j,\alpha}=0}$ by a chi-squared with 1 degree of freedom:

$$\mathsf{t}^2_{\rho_{i,j,\alpha}=0}\left(\psi^2_{\lambda,\alpha}\right) \sim \chi^2_1\left(\psi^2_{\lambda,\alpha}\right). \tag{47}$$

Next, we relate that non-central χ^2 distribution to a central χ^2 using (see e.g., Hendry, 1995):

$$\chi^2_1\left(\psi^2_{\lambda,\alpha}\right) = h\chi^2_m(0) \tag{48}$$

such that for $k = 1$:

$$h = \frac{k + 2\psi^2_{\lambda,\alpha}}{k + \psi^2_{\lambda,\alpha}} \text{ and } m = \frac{k + \psi^2_{\lambda,\alpha}}{h}. \tag{49}$$

Finally, the power function of the $\chi^2_1\left(\psi^2_{\lambda,\alpha}\right)$ test in (47) is approximated by:

$$\mathsf{P}\left[\mathsf{t}^2_{\rho_{i,j,\alpha}=0}\left(\psi^2_{\lambda,\alpha}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_1\left(\psi^2_{\lambda,\alpha}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m(0) > h^{-1}c_{\alpha_1}\right]. \tag{50}$$

For non-integer values of m, a weighted average of the neighbouring integer values is used. As an example, when $\psi_{\lambda,\alpha} = 4$ for $c_{\alpha_1} = 3.84$, then $h = 33/17 \simeq 1.94$ and $m = 8.76$ (taking the nearest integer values as 8 and 9 with weights 0.24 and 0.76) yields $\mathsf{P}\left[\mathsf{t}^2_{\rho_{i,j,\alpha}=0}(16) > 3.84\right] \simeq 0.99$, as against the exact t-distribution outcome of 0.975. When $\lambda_j = d\sqrt{\sigma_{22,j}}$, as in the experiments reported below, $\psi^2_{\lambda,\alpha} = d^2$ and so:

$$h = \frac{1 + 2d^2}{1 + d^2} \text{ and } m = \frac{1 + d^2}{h}.$$

For $d = 1$, 2, 2.5, 3 and 4 at $c_{\alpha_1} = 3.84$ we have:

$$\mathsf{p}_d = \mathsf{P}\left[\mathsf{t}^2_{\widehat{\rho}_{i,j,\alpha}}\left(d^2\right) > c_{\alpha_1}\right] \simeq \begin{cases} 0.17 & d = 1 \\ 0.50 & d = 2 \\ 0.71 & d = 2.5 \\ 0.86 & d = 3 \\ 0.99 & d = 4 \end{cases}$$

so the power is low at $d = 1$ (the exact t-distribution outcome for $d = 1$ is 0.16), but has risen markedly by $d = 3$. Viewing these powers as the probability p_d of retaining a relevant dummy when testing the marginal model, then approximately $\mathsf{p}_d k$ relevant dummies will be retained for testing in the conditional model, attentuating the non-centrality $\varphi_{r,\alpha}$ in (70) below relative to the known break-dates' case.

Similarly, retention of irrelevant impulses, namely those corresponding to non-break related shocks in the marginal process, will also lower power relative to knowing the break dates. For the F-test, that loss will merely be an increase in its degrees-of-freedom, inducing little power reduction. However, the index will include such values at their estimated outcomes so will lose more power. These effects also differ at a given non-centrality of the test statistics by what induced the super-exogeneity failure: specifically, a larger break in the marginal with a smaller violation of the null will generally lead to a closer match of the non-null rejection frequency and the optimum for known break dates, since few irrelevant impulses will be retained when there is a large break.

6.2 Detecting the variance shift in the marginal

We consider a setting where the variance shift $\theta > 1$ occurs within one half, say at observation $T_1 > T/2$ so that:

$$z_t = 1 + \left(1_{\{t<T_1\}} + \sqrt{\theta}1_{\{t\geq T_1\}}\right) v_t. \tag{51}$$

The maximum feasible power would be from detecting and entering the set of $T - T_1 + 1$ impulses $1_{\{t\geq T_1\}}$ each of which would then equal $\sqrt{\theta}1_{\{t\geq T_1\}}v_t$ to be judged against a baseline variance of σ_v^2:

$$\mathsf{t}_{\{t\geq T_1\}} = \frac{\sqrt{\theta}1_{\{t\geq T_1\}}v_t}{\sigma_v},$$

which has a non-centrality of $\psi^2_{\theta,\alpha_1} = \theta$. Approximating by a central $\chi^2_1\left(\psi^2_{\theta,\alpha_1}\right)$ as before:

$$\mathsf{P}\left[\mathsf{t}^2_{\{t\geq T_1\}}\left(\psi^2_{\theta,\alpha_1}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_1\left(\psi^2_{\theta,\alpha_1}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m\left(0\right) > h^{-1}c_{\alpha_1}\right] \tag{52}$$

for:

$$h = \frac{1 + 2\psi^2_{\theta,\alpha_1}}{1 + \psi^2_{\theta,\alpha_1}} \text{ and } m = \frac{\left(1 + \psi^2_{\theta,\alpha_1}\right)^2}{1 + 2\psi^2_{\theta,\alpha_1}}. \tag{53}$$

Thus, for $\psi^2_{\theta,\alpha_1} = (2; 5; 10)$, power will be about (25%, 60%, 90%) respectively at $\alpha_1 = 0.05$.

7 Three super exogeneity failures

In this section, we derive explicit outcomes for three forms of super exogeneity failure, namely weak exogeneity failure when the marginal process is non-constant in section 7.1; invariance failure in section 7.2; and weak exogeneity failure when the marginal process is constant in section 7.3. In each case, we obtain the non-centralities and approximate powers of the tests for a known break, then modify these in light of the stage 1 pre-test for indicators. Section 9 reports the simulation outcomes.

7.1 Weak exogeneity failure under non-constancy

Consider the normally-distributed $n \times 1$ vector random variable $\mathbf{x}_t = (y_t : \mathbf{z}'_t)'$ where the conditional expectation of y_t is:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_{1,t} + \boldsymbol{\sigma}'_{12}\boldsymbol{\Sigma}^{-1}_{22}\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) = \mu_{1,t} + \boldsymbol{\gamma}'\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \tag{54}$$

with conditional variance:

$$\mathsf{E}\left[\left(y_t - \mathsf{E}\left[y_t|\mathbf{z}_t\right]\right)^2 \mid \mathbf{z}_t\right] = \left(\sigma_{11} - \boldsymbol{\sigma}'_{12}\boldsymbol{\Sigma}^{-1}_{22}\boldsymbol{\sigma}_{12}\right),$$

where the parameter of interest is $\boldsymbol{\beta}$ in the theoretical model (ignoring intercepts for simplicity of exposition):

$$\mu_{1,t} = \boldsymbol{\beta}'\boldsymbol{\mu}_{2,t}. \tag{55}$$

Then:

$$y_t = \boldsymbol{\beta}'\mathbf{z}_t + (\boldsymbol{\gamma} - \boldsymbol{\beta})'\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) + \epsilon_t \tag{56}$$

where $\epsilon_t = y_t - \mathsf{E}[y_t|\mathbf{z}_t]$ given (55), so $\mathsf{E}[\epsilon_t|\mathbf{z}_t] = 0$. Such a model is a possible example of the Lucas critique where the agents' behavioural rule depends on $\mathsf{E}[\mathbf{z}_t]$ as in (55), whereas the econometric equation uses $\mathbf{z}_t$, leading to (56).

The joint distribution of $\mathbf{x}_t$ is:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_n\left[\begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{\mu}_{2,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}'_{12} \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right] \tag{57}$$

To complete the model, we postulate an explicit breaking process for $\{\mathbf{z}_t\}$ which will induce a violation in super, as well as weak, exogeneity through $\boldsymbol{\gamma} \neq \boldsymbol{\beta}$, where $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$, namely:

$$\mathbf{z}_t = \boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t} \tag{58}$$

so $\mathsf{E}\left[\mathbf{z}_t\right] = \boldsymbol{\lambda}1_{\{t>T_1\}} = \boldsymbol{\mu}_{2,t}$. In general, there could be breaks in the different marginal processes at different times, but little additional insight is gleaned over the one-off break in (58) which may affect one or more $\mathbf{z}_t$s. The relevant moments of the joint process are:

$$\begin{aligned}
\mathsf{E}\left[\mathbf{z}_t\right] &= \boldsymbol{\lambda}1_{\{t>T_1\}} \\
\mathsf{E}\left[y_t\right] &= \boldsymbol{\beta}'\mathsf{E}\left[\mathbf{z}_t\right] = \boldsymbol{\beta}'\boldsymbol{\lambda}1_{\{t>T_1\}} \\
\mathsf{E}\left[\mathbf{z}_t\mathbf{z}'_t\right] &= \mathsf{E}\left[\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)'\right] = \boldsymbol{\lambda}\boldsymbol{\lambda}'1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22} \\
\mathsf{E}\left[\mathbf{z}_t y_t\right] &= \mathsf{E}\left[\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)\left(\boldsymbol{\beta}'\boldsymbol{\lambda}1_{\{t>T_1\}} + v_{1,t}\right)\right] = \boldsymbol{\lambda}\left(\boldsymbol{\beta}'\boldsymbol{\lambda}\right)1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}.
\end{aligned}$$

If the break is not handled, the fitted model is the regression:

$$y_t = \kappa_0 + \boldsymbol{\kappa}'_1\mathbf{z}_t + u_t \tag{59}$$

where $\mathsf{E}[\mathbf{z}_t u_t] = \mathbf{0}$. Then, in (59), letting $\left(T - T_1\right)/T = r$:

$$\begin{aligned}
\mathsf{E}\left[\begin{pmatrix} \widehat{\kappa}_0 \\ \widehat{\boldsymbol{\kappa}}_1 \end{pmatrix}\right] &\simeq \left[\sum_{t=1}^{T}\begin{pmatrix} 1 & \mathsf{E}\left[\mathbf{z}_t\right]' \\ \mathsf{E}\left[\mathbf{z}_t\right] & \mathsf{E}\left[\mathbf{z}_t\mathbf{z}'_t\right] \end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix} \mathsf{E}\left[y_t\right] \\ \mathsf{E}\left[\mathbf{z}_t y_t\right] \end{pmatrix}\right] \\
&= \begin{pmatrix} 1 & r\boldsymbol{\lambda}' \\ r\boldsymbol{\lambda} & r\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1}\begin{pmatrix} r\boldsymbol{\beta}'\boldsymbol{\lambda} \\ r\boldsymbol{\lambda}\left(\boldsymbol{\beta}'\boldsymbol{\lambda}\right) + \boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \boldsymbol{\beta} \end{pmatrix} - \begin{pmatrix} -r\boldsymbol{\lambda}' \\ \mathbf{I} \end{pmatrix}\mathbf{d}_r
\end{aligned}$$

where:

$$\mathbf{d}_r = \mathbf{H}_r^{-1}\boldsymbol{\Sigma}_{22}\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right). \tag{60}$$

Consequently:

$$y_t = \kappa_0 + \boldsymbol{\kappa}'_1\mathbf{z}_t + u_t = r\boldsymbol{\lambda}'\mathbf{d}_r + \left(\boldsymbol{\beta} - \mathbf{d}_r\right)'\mathbf{z}_t + u_t = \boldsymbol{\beta}'\mathbf{z}_t - \mathbf{d}'_r\left(\mathbf{z}_t - r\boldsymbol{\lambda}\right) + u_t \tag{61}$$

showing that the coefficients are a function of the proportion r of the sample affected by the shift in the marginal process. Recursive estimation and testing for constancy could reveal that problem, but here we consider the extent to which adding the impulse indicators from the marginal process will also do so.

Adding the impulse dummies to the marginal model at best would yield:

$$\mathbf{z}_t = \sum_{i=T_1+1}^{T} \widehat{\boldsymbol{\rho}}_{i,\alpha} 1_{\{t=t_i\}} + \mathbf{v}_{2,t}^{*}$$

for $t_i = T_1+1, \ldots, T$ where:

$$\widehat{\boldsymbol{\rho}}_{i,\alpha} = \boldsymbol{\lambda} + \mathbf{v}_{2,t_i} \tag{62}$$

with:

$$\mathbf{v}_{2,t}^{*} = \mathbf{0} \ \ \forall t > T_1,$$

noting that:

$$1_{\{t>T_1\}} = \sum_{i=T_1+1}^{T} 1_{\{t=t_i\}}.$$

Potentially, some irrelevant impulses may be retained and some relevant ones omitted, both of which could lower the power derived below. However, when a break occurs, few non-break impulses are retained, although for small values of $\boldsymbol{\lambda}$ some of the $\widehat{\boldsymbol{\rho}}_{i,\alpha}$ may be omitted as noted in section 6 above.

Recording which impulses matter, and adding these to (59) given (64), yields the full-sample regression (considering first the case where all relevant impulses were detected in the marginal model):

$$y_t = \tau_0 + \boldsymbol{\tau}_1' \mathbf{z}_t + \sum_{i=T_1+1}^{T} \delta_{i,\alpha} 1_{\{t=t_i\}} + e_t. \tag{63}$$

To see whether such a regression will have any power to detect failures of super exogeneity, consider the 'instantaneous' relation given by:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \varsigma_{0,t} + \boldsymbol{\varsigma}_{1,t}' \mathbf{z}_t$$

so that:

$$y_t = \varsigma_{0,t} + \boldsymbol{\varsigma}_{1,t}' \mathbf{z}_t + e_t \tag{64}$$

where $\mathsf{E}\left[e_t\right] = 0$ and $\mathsf{E}\left[\mathbf{z}_t e_t\right] = \mathbf{0}$ implying:

$$\begin{aligned} \mathsf{E}\left[\begin{pmatrix} \varsigma_{0,t} \\ \boldsymbol{\varsigma}_{1,t} \end{pmatrix}\right] &\simeq \begin{pmatrix} 1 & \boldsymbol{\lambda}' 1_{\{t>T_1\}} \\ \boldsymbol{\lambda} 1_{\{t>T_1\}} & \boldsymbol{\lambda}\boldsymbol{\lambda}' 1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{\lambda} 1_{\{t>T_1\}} \\ \boldsymbol{\lambda}\left(\boldsymbol{\beta}'\boldsymbol{\lambda}\right) 1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\lambda}'\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) 1_{\{t>T_1\}} \\ \boldsymbol{\gamma} \end{pmatrix}. \end{aligned}$$

This suggests the model:

$$y_t = \boldsymbol{\gamma}' \mathbf{z}_t + \boldsymbol{\lambda}'\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) 1_{\{t>T_1\}} + e_t \tag{65}$$

matching (61), so that adding all the indicators selected from the marginal model should substantively improve the fit when $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$. Indeed, (65) coincides with the DGP here, so $\{e_t\}$ is an innovation process.

The power of the F-test of:

$$\mathsf{H}_0\colon \delta_{i,\alpha} = 0 \ \forall i,$$

in (63) by an $\mathsf{F}_{T-T_1-2}^{T-T_1}$ depends on the strength of the super-exogeneity violation, $(\boldsymbol{\beta} - \boldsymbol{\gamma})$, the sizes of the breaks, $\boldsymbol{\lambda}$, the sample size T, the relative number of periods affected by the break, and on α_2. For example, at $T = 100$ with a mid-point break inducing 50 impulse dummies, since $\mathsf{P}\left(\mathsf{F}_{48}^{50} \geq 1.97 | \mathsf{H}_0\right) \simeq$

0.01 and $\mathsf{P}\left(\mathsf{F}^{50}_{48} \geq 1.62|\mathsf{H}_0\right) \simeq 0.05$, these relatively low values of c_{α_2} suggest that such a test is likely to have some power.

Before deriving that power, we noted above that test power could potentially be increased by forming indices of the impulses found in the marginal model (see e.g., Hendry and Santos, 2005). Thus, instead of adding the $T - T_1$ individual $1_{\{t=t_i\}}$, one could add the composite variables $\iota_{1,t}$ and $\iota_{2,t}$ as in (15). This always results in a two degree-of-freedom test, which again can be computed automatically as:

$$y_t = \tau_0 + \boldsymbol{\tau}'\mathbf{z}_t + \tau_2\iota_{1,t} + \tau_3\iota_{2,t} + e_t. \tag{66}$$

The size and power properties are checked by simulation below, and contrasted with the optimal, but generally infeasible, index $1_{\{t>T_1\}}$. Note that there should not be any selection of which dummies to retain in the conditional model, simply a one-off test of the joint null.

7.1.1 Asymptotic power of the index test

A case where theoretical analysis is feasible is when $1_{\{t>T_1\}}$ is known, and the test only depends on the index $1_{\{t>T_1\}}$. In this specific case, the index-based test is equivalent to a Chow (1960) test for a known break point (see Salkever, 1976), but that equivalence will not hold in general for (say) intermittent changes. Then, the index-based test is of the null, H_0: $\tau_2 = 0$ in:

$$y_t = \tau_0 + \boldsymbol{\tau}_1'\mathbf{z}_t + \tau_2 1_{\{t>T_1\}} + u_t \tag{67}$$

where the DGP is (65) written as:

$$y_t = \boldsymbol{\gamma}'\mathbf{z}_t + (\boldsymbol{\beta} - \boldsymbol{\gamma})'\boldsymbol{\lambda}1_{\{t>T_1\}} + e_t. \tag{68}$$

Since (68) is correctly specified, $\boldsymbol{\gamma}$ and $(\boldsymbol{\beta} - \boldsymbol{\gamma})'\boldsymbol{\lambda}$ are consistently estimated with:

$$\begin{aligned}
\mathsf{V}\left[\begin{pmatrix} \boldsymbol{\lambda}'\widehat{(\boldsymbol{\beta} - \boldsymbol{\gamma})} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix}\right] &\simeq \sigma_e^2 \left(\sum_{t=1}^{T} \begin{matrix} \iota_{1,t}^2 & \mathsf{E}\left[\mathbf{z}_t'\iota_{1,t}\right] \\ \mathsf{E}\left[\mathbf{z}_t\iota_{1,t}\right] & \mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] \end{matrix}\right)^{-1} \\
&= \frac{\sigma_e^2}{T}\left(T \begin{matrix} r^{-1} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda} & -\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda} & \boldsymbol{\Sigma}_{22}^{-1} \end{matrix}\right).
\end{aligned} \tag{69}$$

The power depends on $\boldsymbol{\lambda}$, r, T, σ_e, α, as well as the departure between $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ induced by the failure of super exogeneity. Since:

$$e_t = v_{1,t} - \boldsymbol{\gamma}'\mathbf{v}_{2,t},$$

then:

$$\sigma_e^2 = \sigma_{11} - \boldsymbol{\sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}.$$

Let:

$$\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{K}\mathbf{K}' \text{ so } \mathbf{K}'\boldsymbol{\Sigma}_{22}\mathbf{K} = \mathbf{I}_{n-1}$$

where:

$$\mathbf{K}'\mathbf{z}_t = \mathbf{K}'\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{K}'\mathbf{v}_{2,t}$$

and $\boldsymbol{\lambda}^* = \sqrt{r}\mathbf{K}'\boldsymbol{\lambda}$ is the normalized break impact. Then the non-centrality of a t-test of H_0: $\tau_2 = 0$ in (67) is:

$$\mathsf{E}\left[\mathsf{t}_{\tau_2=0}\right] = \frac{(\boldsymbol{\beta} - \boldsymbol{\gamma})'\boldsymbol{\lambda}\sqrt{Tr}}{\sigma_e\sqrt{1 + r\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda}}} = \frac{\sqrt{T}\left(\boldsymbol{\lambda}^*\right)'\mathbf{K}^{-1}(\boldsymbol{\beta} - \boldsymbol{\gamma})}{\sqrt{(\sigma_{11} - \boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma})}\sqrt{1 + \boldsymbol{\lambda}^*\left(\boldsymbol{\lambda}^*\right)'}} = \varphi_{r,\alpha}, \tag{70}$$

The non-centrality $\varphi_{r,\alpha}$ in (70) would be zero if $\boldsymbol{\beta} = \boldsymbol{\gamma}$ (no failure of weak exogeneity), or if $\boldsymbol{\lambda} = \mathbf{0}$ or $r = 0$ (no shift in the marginal process). Otherwise, $\varphi_{r,\alpha}$ is monotonically increasing in $\sqrt{T}$, $(\boldsymbol{\beta} - \boldsymbol{\gamma})$ and in $\boldsymbol{\lambda}^*$ (even though increasing λ^* also increases the denominator), and monotonically decreasing in σ_e and $\boldsymbol{\Sigma}_{22}$ *ceteris paribus*.

We compute the power function using the approximation to $\mathsf{t}^2_{\tau_2=0}$ by a chi-squared with 1 degrees of freedom discussed in section 6 above with $\mathsf{t}^2_{\tau_2=0}\left(\varphi^2_{r,\alpha}\right) \sim \chi^2_1\left(\varphi^2_{r,\alpha}\right) \simeq h\chi^2_m(0)$ from (49). Then, from (50), $\mathsf{P}\left[\chi^2_1\left(\varphi^2_{r,\alpha}\right) > c_{\alpha_2}|\mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m(0) > h^{-1}c_{\alpha_2}\right]$. For example, when $\varphi^2_{r,\alpha} = 5$ for $c_{\alpha_2} = 4$, then $h = 51/26 \simeq 2$ and $m = 13$ with $\mathsf{P}\left[\chi^2_{13}(0) > 2\right] \simeq 0.9998$.

Finally, $\varphi^2_{r,\alpha}$ should also be the non-centrality of the corresponding F-test, a conjecture that can be checked by its mean value in the Monte Carlo simulations. However, the power may not be monotonic in the arguments of $\varphi^2_{r,\alpha}$ since the degrees of freedom of the F-test alter with r: a given value of λ^*_1 achieved by a larger $\sqrt{r}$ will have lower power than that from a smaller $\sqrt{r}$. More precisely, we approximate the $\mathsf{F}^{T-T_1}_{T-T_1-2}(\varphi_{r,\alpha})$ by its numerator $\chi^2_k(\varphi_{r,\alpha})$ and that in turn by (48) using the more general formulae in (49) for $k = T - T_1 = Tr$. Then:

$$\mathsf{P}\left[\chi^2_{Tr}(\varphi_{r,\alpha}) > c_{\alpha_2} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m(0) > h^{-1}c_{\alpha_2}\right] \tag{71}$$

where:

$$h = \frac{Tr + 2\varphi^2_{r,\alpha}}{Tr + \varphi^2_{r,\alpha}} \text{ and } m = \frac{Tr + \varphi^2_{r,\alpha}}{h}. \tag{72}$$

In comparison to the numerical example following (70), when $Tr = 20$ (say) for $T = 100$, then $h = 70/45 \simeq 1.56$ and $m \simeq 29$ with $c_{\alpha_2} \simeq 31.4$ so $\mathsf{P}\left[\chi^2_{29}(0) > 20.1\right] \simeq 0.89$, delivering a somewhat lower power.

7.1.2 Allowing for stage 1

The above results are conditional on keeping all and only the relevant impulses from the marginal, but the analysis in section 6 revealed that was itself dependent on the parameters of the marginal DGPs. Nevertheless, we can extend the analysis roughly to allow for such an effect by distinguishing the number of elements in the index $\iota_{1,t}$ from the length of the break. In a bivariate setting, corresponding to (65) when the DGP is (68), we have:

$$\begin{aligned}
\mathsf{E}\begin{bmatrix} \lambda\widehat{(\beta - \gamma)} \\ \widehat{\gamma} \end{bmatrix} &\simeq \left(\sum_{t=1}^{T} \begin{matrix} \mathsf{E}\left[\iota^2_{1,t}\right] & \mathsf{E}\left[z_t\iota_{1,t}\right] \\ \mathsf{E}\left[z_t\iota_{1,t}\right] & \mathsf{E}\left[z^2_t\right] \end{matrix}\right)^{-1} \left(\sum_{t=1}^{T} \begin{matrix} \mathsf{E}\left[y_t\iota_{1,t}\right] \\ \mathsf{E}\left[y_t z_t\right] \end{matrix}\right) \\
&= \begin{pmatrix} \mathsf{p}_d r & \lambda\mathsf{p}_d r \\ \lambda\mathsf{p}_d r & \lambda^2 r + \sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} \beta\lambda\mathsf{p}_d r \\ \beta\lambda^2 r + \gamma\sigma_{22} \end{pmatrix} \\
&= \begin{pmatrix} \lambda(\beta - \gamma) \\ \gamma \end{pmatrix} + \frac{(\beta - \gamma)\lambda^2 r(1 - \mathsf{p}_d)}{\lambda^2 r(1 - \mathsf{p}_d) + \sigma_{22}} \begin{pmatrix} -\lambda \\ 1 \end{pmatrix}.
\end{aligned} \tag{73}$$

Comparing (73) with the consistent estimates and their variances in (69) which result when the break date is known, the effect of stage 1 selection is bound to be a loss of power. More precisely, letting the estimated stage 2 model be:

$$y_t = \kappa^*_0 z_t + \kappa^*_1 \iota_{1,t} + u_t \tag{74}$$

leads to a modified non-centrality corresponding to (70) when $n = 2$ but for (74), namely:

$$\mathsf{E}\left[\mathsf{t}_{\kappa_1^*=0}\right] = \frac{\sqrt{Tr\mathsf{p}_d}\,(\beta-\gamma)\,\lambda}{\sigma_u\sqrt{\left(1+\lambda^2 r\sigma_{22}^{-1}\right)}}, \tag{75}$$

so:

$$\mathsf{E}\left[\mathsf{t}^2_{\kappa_1^*=0}\right] = \frac{Tr\mathsf{p}_d\,(\beta-\gamma)^2\,\lambda^2}{\sigma_u^2\left(1+\lambda^2 r\sigma_{22}^{-1}\right)} = \frac{\mathsf{p}_d\sigma_e^2\varphi_{r,\alpha}^2}{\sigma_u^2},$$

where:

$$\sigma_u^2 = \sigma_e^2 + \sigma_{22}\frac{(1-\mathsf{p}_d)\,\lambda^2\,(\beta-\gamma)^2}{\lambda^2 r\,(1-\mathsf{p}_d)+\sigma_{22}}. \tag{76}$$

Thus, the power falls directly because $\mathsf{p}_d < 1$ and indirectly because $\sigma_u^2 > \sigma_e^2$. For example, combining the parameter values for the tests just above with the location shift that delivered $\mathsf{p}_d = 0.16$ for each impulse in section 6 yields (where $\lambda^2 = \sigma_{22} = 5$, $\beta-\gamma = 0.25$, $Tr = 20$, $\sigma_e^2 = 1$ so $\sigma_u^2 = 1.23$) $\mathsf{E}\left[\mathsf{t}^2_{\kappa_1^*=0}\right] = 0.65$, which is a notable reduction in the non-centrality. However, increasing to $\lambda = 2.5\sqrt{\sigma_{22}}$ raises p_d to 0.71 and $\mathsf{E}[\mathsf{t}^2_{\kappa_1^*=0}]$ to 11.4, so the power rises quickly towards the maximum, essentially reaching that bound by $\lambda = 4\sqrt{\sigma_{22}}$. Notice from (76) that σ_u^2 need not tend monotonically to σ_e^2 as λ increases, although it eventually converges since $\mathsf{p}_d \to 1$ as λ gets sufficiently large.

7.2 Invariance failure

The previous subsection concerned a model where the regression lacked invariance to a location shift in the marginal model because of a failure of weak exogeneity induced by $\boldsymbol{\gamma} \neq \boldsymbol{\beta}$. Nevertheless, the test had some power, since the non-centrality was non-zero under the alternative of no weak exogeneity with a shift in the marginal process. We now allow the parameters of the marginal and conditional to be directly cross-linked, where the marginal remains:

$$\mathbf{z}_t = \boldsymbol{\lambda}_{0,t} + \mathbf{v}_{2,t} = \boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t},$$

with $\mathsf{E}[\mathbf{z}_t] = \boldsymbol{\lambda}1_{\{t>T_1\}} = \boldsymbol{\mu}_{2,t}$. Moreover, there is no 'direct' violation of weak exogeneity, in that $\boldsymbol{\gamma} = \boldsymbol{\beta}$, but the cross-link between the means violates super exogeneity, namely $\boldsymbol{\mu}_{1,t} = \boldsymbol{\beta}_t'\boldsymbol{\mu}_{2,t}$ when:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 1_{\{t>T_1\}}, \tag{77}$$

where:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_n\left[\begin{pmatrix} \boldsymbol{\mu}_{1,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega^2 + \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_t & \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_t & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right]. \tag{78}$$

Thus, the parameters of the conditional distribution shift when those of the marginal process alter. Since $\boldsymbol{\gamma}_t = \boldsymbol{\beta}_t$:

$$\mathsf{E}[y_t \mid \mathbf{z}_t] = \boldsymbol{\beta}_t'\boldsymbol{\mu}_{2,t} + \boldsymbol{\beta}_t'\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) = \boldsymbol{\beta}_0'\mathbf{z}_t + \boldsymbol{\beta}_1'\mathbf{z}_t 1_{\{t>T_1\}}. \tag{79}$$

The marginal model is the same as in the previous section, so $\widehat{\boldsymbol{\rho}}_{i,\alpha} = \boldsymbol{\lambda} + \mathbf{v}_{2,t_i}$ from (62), and hence a test based on adding the associated $\left\{1_{\{t=t_i\}}\right\}$ and $\left\{1_{\{t=t_i\}}z_{j,t_i}\right\}$, or their matching summaries as in (15), should also have power against violations of invariance, as we now show.

The regression equation postulated by the econometrician is the same as (59), but the data moments differ for the changed DGP:

$$
\begin{aligned}
\mathsf{E}\left[\mathbf{z}_t\right] &= \boldsymbol{\lambda}1_{\{t>T_1\}} \\
\mathsf{E}\left[y_t\right] &= \left(\boldsymbol{\beta}_0' + \boldsymbol{\beta}_1' 1_{\{t>T_1\}}\right)\mathsf{E}\left[\mathbf{z}_t\right] = \left(\boldsymbol{\beta}_0' + \boldsymbol{\beta}_1'\right)\boldsymbol{\lambda}1_{\{t>T_1\}} \\
\mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] &= \mathsf{E}\left[\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)'\right] = \boldsymbol{\lambda}\boldsymbol{\lambda}'1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22} \\
\mathsf{E}\left[\mathbf{z}_t y_t\right] &= \mathsf{E}\left[\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right)\left(\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 1_{\{t>T_1\}}\right)'\boldsymbol{\lambda}1_{\{t>T_1\}} + v_{1,t}\right)\right] \\
&= \left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\right)1_{\{t>T_1\}} + \boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_0\left(1 - 1_{\{t>T_1\}}\right) \\
\mathsf{E}\left[\boldsymbol{\iota}_{2,t} y_t\right] &= \mathsf{E}\left[\left(1_{\{t>T_1\}}\left[\boldsymbol{\lambda} + \mathbf{v}_{2,t}\right]\right)\left(\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 1_{\{t>T_1\}}\right)'\boldsymbol{\lambda}1_{\{t>T_1\}} + v_{1,t}\right)\right] \\
&= \left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\right)1_{\{t>T_1\}}
\end{aligned}
$$

Hence, the implicit full-sample parameters of (59) become:

$$
\begin{aligned}
\mathsf{E}\left[\begin{pmatrix} \widetilde{\kappa}_0 \\ \widetilde{\boldsymbol{\kappa}}_1 \end{pmatrix}\right] &\simeq \left[\sum_{t=1}^{T}\begin{pmatrix} 1 & \mathsf{E}\left[\mathbf{z}_t\right] \\ \mathsf{E}\left[\mathbf{z}_t\right] & \mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] \end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix} \mathsf{E}\left[y_t\right] \\ \mathsf{E}\left[\mathbf{z}_t y_t\right] \end{pmatrix}\right] \\
&= \begin{pmatrix} 1 & \boldsymbol{\lambda}'r \\ \boldsymbol{\lambda}r & \boldsymbol{\lambda}\boldsymbol{\lambda}'r + \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1}\begin{pmatrix} \left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\right)'\boldsymbol{\lambda}r \\ \boldsymbol{\lambda}\boldsymbol{\lambda}'\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\right)r + \boldsymbol{\Sigma}_{22}\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 r\right) \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \end{pmatrix} - \begin{pmatrix} -\boldsymbol{\lambda}'r \\ \mathbf{I}_{n-1} \end{pmatrix}\left(\boldsymbol{\lambda}\boldsymbol{\lambda}'r\left(r-1\right) + \boldsymbol{\Sigma}_{22}\right)^{-1}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_1\left(1-r\right),
\end{aligned}
$$

which is similar in form to (60), and simplifies to the vector $(0, \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1)'$ when $\boldsymbol{\lambda} = \mathbf{0}$.

The 'instantaneous' relation is again given by:

$$
y_t = \varsigma_{0,t} + \boldsymbol{\varsigma}_{1,t}'\mathbf{z}_t + e_t \tag{80}
$$

where:

$$
\varsigma_{0,t} = \mathsf{E}\left[y_t\right] - \boldsymbol{\varsigma}_{1,t}'\mathsf{E}\left[\mathbf{z}_t\right] = \left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\right)'\boldsymbol{\lambda}1_{\{t>T_1\}} - \boldsymbol{\varsigma}_{1,t}'\boldsymbol{\lambda}1_{\{t>T_1\}} = 0,
$$

and:

$$
\begin{aligned}
\boldsymbol{\varsigma}_{1,t} &\simeq \left(\mathsf{E}\left[\left(\mathbf{z}_t - \mathsf{E}\left[\mathbf{z}_t\right]\right)\left(\mathbf{z}_t - \mathsf{E}\left[\mathbf{z}_t\right]\right)'\right]\right)^{-1}\mathsf{E}\left[\left(y_t - \mathsf{E}\left[y_t\right]\right)\left(\mathbf{z}_t - \mathsf{E}\left[\mathbf{z}_t\right]\right)\right] \\
&= \boldsymbol{\Sigma}_{22}^{-1}\mathsf{E}\left[v_{1,t}\mathbf{v}_{2,t}\right] \\
&= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 1_{\{t>T_1\}}
\end{aligned}
$$

so as expected:

$$
y_t = \boldsymbol{\beta}_0'\mathbf{z}_t + \boldsymbol{\beta}_1'\mathbf{z}_t 1_{\{t>T_1\}} + e_t = \boldsymbol{\tau}_1'\mathbf{z}_t + \boldsymbol{\tau}_2'\mathbf{z}_t 1_{\{t>T_1\}} + e_t. \tag{81}
$$

Since:

$$
1_{\{t>T_1\}}\mathbf{z}_t = \sum_{i=T_1+1}^{T} 1_{\{t=t_i\}}\mathbf{z}_t = \sum_{i=T_1+1}^{T}\widehat{\boldsymbol{\rho}}_i 1_{\{t=t_i\}}
$$

(say), then adding a complete set of impulses from the marginal model should detect departures from super exogeneity. The index equivalent here requires adding the impulses from the marginal model times $\mathbf{z}_t$, so differs from the previous case, albeit that both indexes, $\iota_{1,t} = 1_{\{t>T_1\}}$ and $\boldsymbol{\iota}_{2,t} = 1_{\{t>T_1\}}\mathbf{z}_t$ could be calculated and added.

7.2.1 Asymptotic power of the test of invariance

Two issues where theoretical analysis can shed light concern the power of the test based on $\boldsymbol{\tau}_2$ (adding $\boldsymbol{\iota}_{2,t}$ as in (81), which is the model analogue of (79), so that $\mathsf{E}[e_t] = 0 = \mathsf{E}[e_t|\mathbf{z}_t]$), and just adding the index $1_{\{t>T_1\}}$. First, for adding $\boldsymbol{\iota}_{2,t}$:

$$y_t = \boldsymbol{\tau}_1'\mathbf{z}_t + \boldsymbol{\tau}_2'\boldsymbol{\iota}_{2,t} + u_t. \tag{82}$$

The variances of the parameter estimates from (82) are approximately:

$$\begin{aligned}
\mathsf{V}\left[\begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1 \\ \widetilde{\boldsymbol{\tau}}_2 \end{pmatrix}\right] &\simeq \sigma_e^2 \left[\sum_{t=1}^{T} \begin{pmatrix} \mathsf{E}[\mathbf{z}_t\mathbf{z}_t'] & \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}'\right] \\ \mathsf{E}[\mathbf{z}_t\boldsymbol{\iota}_{2,t}] & \mathsf{E}\left[\boldsymbol{\iota}_{2,t}\boldsymbol{\iota}_{2,t}'\right] \end{pmatrix}\right]^{-1} \\
&= \frac{\sigma_e^2}{T}\left[\begin{pmatrix} (\boldsymbol{\lambda}\boldsymbol{\lambda}'r + \boldsymbol{\Sigma}_{22}) & (\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22})\,r \\ (\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22})\,r & [(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22})\,r] \end{pmatrix}\right]^{-1} \\
&= \frac{\sigma_e^2}{Tr\,(1-r)} \begin{pmatrix} r\boldsymbol{\Sigma}_{22}^{-1} & -r\boldsymbol{\Sigma}_{22}^{-1} \\ -r\boldsymbol{\Sigma}_{22}^{-1} & \left((1-r)\left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1} + r\boldsymbol{\Sigma}_{22}^{-1}\right) \end{pmatrix}.
\end{aligned}$$

Consequently, as $\boldsymbol{\lambda}^* = \sqrt{r}\mathbf{K}'\boldsymbol{\lambda}$, and noting that:

$$\left((1-r)\left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1} + r\boldsymbol{\Sigma}_{22}^{-1}\right)^{-1} = \left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)\left(r\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1}\boldsymbol{\Sigma}_{22},$$

an F-test of $\boldsymbol{\tau}_2 = 0$ is:

$$\begin{aligned}
\mathsf{E}\left[\mathsf{F}_{\tau_2=0}\right] &= (T-2n)\,r\,(1-r)\,\frac{\boldsymbol{\beta}_1'\left(\left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)\left(r\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1}\boldsymbol{\Sigma}_{22}\right)\boldsymbol{\beta}_1}{\sigma_e^2\,(n-1)} \\
&= (T-2n)\,(1-r)\,\frac{\boldsymbol{\beta}_1'\left(\mathbf{K}'\right)^{-1}\left(\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\prime} + r\mathbf{I}\right)\left(\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\prime} + \mathbf{I}\right)^{-1}\mathbf{K}^{-1}\boldsymbol{\beta}_1}{\omega^2\,(n-1)} = \phi_{r,\alpha}^2,
\end{aligned}$$

as $e_t = y_t - \mathsf{E}[y_t|\mathbf{z}_t]$ so $\sigma_e^2 = \omega^2$.

In a scalar setting, so $n = 2$:

$$\mathsf{E}\left[\mathsf{t}_{\tau_2=0}\right] = \frac{\beta_1\sqrt{T\,(\lambda^2 + \sigma_{22})\,r\,(1-r)\,\sigma_{22}}}{\sigma_e\sqrt{\lambda^2 r + \sigma_{22}}} = \frac{\sqrt{T\,(1-r)}\beta_1\sqrt{r + (\lambda^*)^2}}{\omega^*\sqrt{1 + (\lambda^*)^2}} = \phi_{r,\alpha},$$

with $\omega^* = \omega/\sqrt{\sigma_{22}}$. Again $\phi_{r,\alpha}$ is monotonically increasing in $\boldsymbol{\lambda}^*$, in β_1 and in r for fixed λ; and because of the form of (77), $\phi_{r,\alpha} \neq 0$ even if $\lambda^* = 0$.

The power can be calculated as in (47)–(50) above.

Then $\phi_{r,\alpha}^2$ probably represents the non-centrality of the F-test: this can be checked by the mean value in the Monte Carlo simulations using the formula in Johnson and Kotz (1970, p.190) that:

$$\mathsf{E}\left[\mathsf{F}_{k_2}^{k_1}\left(\phi_{r,\alpha}^2\right)\right] = \frac{k_2\left(k_1 + \phi_{r,\alpha}^2\right)}{k_1\,(k_2 - 2)}. \tag{83}$$

The power can be calculated as in (50)–(72).

7.2.2 Allowing for stage 1 effects

Returning to (82), where $\boldsymbol{\iota}_{2,t}$ reflects the power of the stage 1 selection of impulses, the estimators become:

$$\begin{aligned}\mathsf{E}\left[\begin{pmatrix}\widetilde{\boldsymbol{\tau}}_1\\ \widetilde{\boldsymbol{\tau}}_2\end{pmatrix}\right] &\simeq \left[\sum_{t=1}^{T}\begin{pmatrix}\mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] & \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}'\right]\\ \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}\right] & \mathsf{E}\left[\boldsymbol{\iota}_{2,t}\boldsymbol{\iota}_{2,t}'\right]\end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix}\mathsf{E}\left[\mathbf{z}_t y_t\right]\\ \mathsf{E}\left[\boldsymbol{\iota}_{2,t}y_t\right]\end{pmatrix}\right]\\ &= \begin{pmatrix}\boldsymbol{\beta}_0-\left(\mathbf{I}-\mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\left(1-r\right)\right)\boldsymbol{\beta}_1\\ \mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\left(1-r\right)\boldsymbol{\beta}_1\end{pmatrix}\end{aligned}\tag{84}$$

where:

$$\mathbf{R}=\left(\boldsymbol{\lambda}\boldsymbol{\lambda}'+\boldsymbol{\Sigma}_{22}\right)r\left(1-\mathsf{p}_d\right)+\boldsymbol{\Sigma}_{22}\left(1-r\right).$$

The bias effect vanishes when $\mathsf{p}_d=1$ as $\mathbf{R}=\boldsymbol{\Sigma}_{22}\left(1-r\right)$. From (84):

$$\mathsf{V}\left[\widetilde{\boldsymbol{\tau}}_2\right]=\frac{\sigma_u^2}{T}\left(\left[\left(\boldsymbol{\lambda}\boldsymbol{\lambda}'+\boldsymbol{\Sigma}_{22}\right)\mathsf{p}_d r\right]^{-1}+\mathbf{R}^{-1}\right),$$

so the F-test of $\boldsymbol{\tau}_2=\mathbf{0}$ has an expected value of:

$$\mathsf{E}\left[\mathsf{F}_{\tau_2=0}\right]=\left(T-2n\right)\left(1-r\right)^2\frac{\boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{22}\mathbf{R}^{-1}\left(\left[\left(\boldsymbol{\lambda}\boldsymbol{\lambda}'+\boldsymbol{\Sigma}_{22}\right)\mathsf{p}_d r\right]^{-1}+\mathbf{R}^{-1}\right)^{-1}\mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_1}{\sigma_u^2\left(n-1\right)}.$$

It is difficult to simply this further, but in the bivariate case, we have:

$$\begin{aligned}\mathsf{E}\left[\mathsf{t}^2_{\tau_2=0}\right] &\simeq \frac{T\mathsf{p}_d\beta_1^2\left(\lambda^2+\sigma_{22}\right)r\left(1-r\right)^2\sigma_{22}^2}{\sigma_u^2\left(\lambda^2 r+\sigma_{22}\right)\left[\left(\lambda^2+\sigma_{22}\right)r\left(1-\mathsf{p}_d\right)+\sigma_{22}\left(1-r\right)\right]}\\ &= \frac{\mathsf{p}_d\sigma_{22}\left(1-r\right)\sigma_e^2}{\sigma_u^2\left[\left(\lambda^2+\sigma_{22}\right)r\left(1-\mathsf{p}_d\right)+\sigma_{22}\left(1-r\right)\right]}\phi^2_{r,\alpha}.\end{aligned}$$

7.2.3 Asymptotic power of the incorrect index invariance test

Now the fitted conditional model is the incorrect specification, assuming a known break:

$$y_t=\left(\boldsymbol{\tau}_1^*\right)'\mathbf{z}_t+\tau_2^*\iota_{1,t}+e_t^*\tag{85}$$

with average estimated parameters:

$$\begin{aligned}\mathsf{E}\left[\begin{pmatrix}\overline{\boldsymbol{\tau}}_1^*\\ \overline{\tau}_2^*\end{pmatrix}\right] &= \left[\sum_{t=1}^{T}\begin{pmatrix}\mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] & \mathsf{E}\left[\mathbf{z}_t\iota_{1,t}\right]\\ \mathsf{E}\left[\mathbf{z}_t\iota_{1,t}\right] & \iota_{1,t}^2\end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix}\mathsf{E}\left[\mathbf{z}_t y_t\right]\\ \mathsf{E}\left[y_t\iota_{1,t}\right]\end{pmatrix}\right]\\ &= \begin{pmatrix}\left(\boldsymbol{\lambda}\boldsymbol{\lambda}'r+\boldsymbol{\Sigma}_{22}\right) & \boldsymbol{\lambda}r\\ \boldsymbol{\lambda}'r & r\end{pmatrix}^{-1}\begin{pmatrix}r\boldsymbol{\lambda}\boldsymbol{\lambda}'\left(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1\right)+\boldsymbol{\Sigma}_{22}\left(\boldsymbol{\beta}_0+r\boldsymbol{\beta}_1\right)\\ r\left(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1\right)'\boldsymbol{\lambda}\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{\beta}_0+r\boldsymbol{\beta}_1\\ \boldsymbol{\lambda}'\boldsymbol{\beta}_1\left(1-r\right)\end{pmatrix}.\end{aligned}$$

Although these estimators are inconsistent for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ respectively, the important issue is the power of the test on the relevance of $\iota_{1,t}$ which yields for $r\neq 0$:

$$\mathsf{E}\left[\mathsf{t}_{\tau_2^*=0}\right]=\frac{\sqrt{Tr}\boldsymbol{\lambda}'\boldsymbol{\beta}_1\left(1-r\right)}{\sigma_{e^*}\sqrt{\left(1+r\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda}\right)}}=\frac{\sqrt{T}\boldsymbol{\lambda}^{*\prime}\mathbf{K}^{-1}\boldsymbol{\beta}_1\left(1-r\right)}{\sqrt{\omega^2+\boldsymbol{\beta}_1'\left(\mathbf{K}'\right)^{-1}\mathbf{K}^{-1}\boldsymbol{\beta}_1 r\left(1-r\right)}\sqrt{\left(1+\boldsymbol{\lambda}^{*\prime}\boldsymbol{\lambda}^*\right)}}=\psi_{r,\alpha},$$

noting that:

$$\begin{aligned} e_t^* &= \boldsymbol{\beta}_0'\mathbf{z}_t + \boldsymbol{\beta}_1'\mathbf{z}_t 1_{\{t>T_1\}} + e_t - (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 r)'\,\mathbf{z}_t - \boldsymbol{\beta}_1'\boldsymbol{\lambda}\,(1-r)\,1_{\{t>T_1\}} \\ &= \boldsymbol{\beta}_1'\left(\mathbf{z}_t\left(1_{\{t>T_1\}} - r\right) - \boldsymbol{\lambda}\,(1-r)\,1_{\{t>T_1\}}\right) + e_t \\ &= \boldsymbol{\beta}_1'\left(\left(1_{\{t>T_1\}} - r\right)\left(\boldsymbol{\lambda}1_{\{t>T_1\}} + \mathbf{v}_{2,t}\right) - \boldsymbol{\lambda}\,(1-r)\,1_{\{t>T_1\}}\right) + e_t \\ &= \boldsymbol{\beta}_1'\mathbf{v}_{2,t}\left(1_{\{t>T_1\}} - r\right) + e_t \end{aligned}$$

so:

$$\sigma_{e^*}^2 = \mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_1'\mathbf{v}_{2,t}\left(1_{\{t>T_1\}} - r\right) + e_t\right)^2\right] = \omega^2 + \boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_1 r\,(1-r)\,.$$

Thus, $\psi_{r,\alpha}$ is again monotonic in $\boldsymbol{\lambda}^*$, but need not be monotonic in r for fixed $\boldsymbol{\lambda}$. Also, $\mathsf{t}_{\tau_2^*=0}$ is less powerful than $\mathsf{t}_{\tau_2=0}$, as $\phi_{r,\alpha}^2 > \psi_{r,\alpha}^2$. Thus, $\phi_{r,\alpha}^2$, the non-centrality of the F-test, which is applicable in the present setting, has an important invariance to the source of the super-exogeneity failure, and should exceed $\psi_{r,\alpha}^2$: again this can be checked by the mean value in the Monte Carlo simulations, and the power function calculations documented above.

7.3 Weak exogeneity failure under constancy

Reconsider the bivariate example in (56) above, but where all parameters are constant, so:

$$y_t = \boldsymbol{\beta}'\mathbf{z}_t + e_t = \boldsymbol{\beta}'\mathbf{z}_t - (\boldsymbol{\beta} - \boldsymbol{\gamma}_2)'\,(\mathbf{z}_t - \boldsymbol{\mu}_2) + \eta_t \tag{86}$$

with:

$$\mathbf{z}_t = \boldsymbol{\mu}_2 + \mathbf{v}_{2,t},$$

but $\mathsf{E}[e_t|\mathbf{z}_t] \neq 0$ as:

$$e_t = \eta_t + (\boldsymbol{\gamma}_2 - \boldsymbol{\beta})'\,\mathbf{v}_{2,t}$$

and $\mathsf{E}[\eta_t|\mathbf{z}_t] = 0$. One mode of generating such a model is when $y_t = \boldsymbol{\beta}'\mathbf{z}_t^e + \eta_t$, but the outcome $\mathbf{z}_t$ is used in place of the expectation $\mathbf{z}_t^e$. Writing the fitted model as:

$$y_t = \tau_0 + \boldsymbol{\tau}_1'\mathbf{z}_t + u_t \tag{87}$$

then:

$$\begin{aligned} \mathsf{E}\left[\begin{array}{c} \widehat{\tau}_0 \\ \widehat{\boldsymbol{\tau}}_1 \end{array}\right] &\simeq \left(\begin{array}{cc} 1 & -\boldsymbol{\mu}_2' \\ -\boldsymbol{\mu}_2 & (\boldsymbol{\mu}_2\boldsymbol{\mu}_2' + \boldsymbol{\Sigma}_{22}) \end{array}\right)^{-1}\left(\begin{array}{c} \boldsymbol{\beta}'\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2\boldsymbol{\mu}_2'\boldsymbol{\beta} + \boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}_2 \end{array}\right) \\ &= \left(\begin{array}{c} (\boldsymbol{\beta} - \boldsymbol{\gamma}_2)'\,\boldsymbol{\mu}_2 \\ \boldsymbol{\gamma}_2 \end{array}\right) \end{aligned}$$

so $\widehat{\boldsymbol{\tau}}_1$ estimates the regression coefficient $\boldsymbol{\gamma}_2$ rather than the structural parameter $\boldsymbol{\beta}$, and correspondingly, $\mathsf{E}[\mathbf{z}_t u_t] = 0$ in (87).

Now only impulses corresponding to randomly large $\mathbf{v}_{2,t}$ will be retained, of which there will be αT on average. The index of these impulses again has the form:

$$w_t = \sum_{i=1}^{\alpha T}\widehat{\varphi}_{i,\alpha}1_{\{t=t_i\}},$$

where:

$$\widehat{\varphi}_{i,\alpha} = v_{2,t_i} \text{ when } |v_{2,t_i}| > c_\alpha.$$

Thus:

$$\begin{aligned} e_t &= y_t - \tau_0 - \boldsymbol{\tau}_1' \mathbf{z}_t - \tau_2 w_t \\ &= \boldsymbol{\beta}' \boldsymbol{\mu}_2 + \boldsymbol{\gamma}_2' (\mathbf{z}_t - \boldsymbol{\mu}_2) + \eta_t - (\boldsymbol{\beta} - \boldsymbol{\gamma}_2)' \boldsymbol{\mu}_2 - \boldsymbol{\gamma}_2' \mathbf{z}_t - \tau_2 \sum_{i=1}^{\alpha T} \widehat{\varphi}_{i,\alpha} 1_{\{t=t_i\}} \\ &= -\tau_2 \sum_{i=1}^{\alpha T} v_{2,t_i} 1_{\{t=t_i\}} + \eta_t. \end{aligned} \tag{88}$$

Since the largest of the v_{2,t_i} in (88) are eliminated by setting $\tau_2 = 0$ to deliver the innovation component η_t, there will be essentially no detectability of the failure of weak exogeneity.

8 Co-breaking based tests

A key assumption underlying the above tests is that the power of impulse saturation tests to detect breaks and outliers was not applied to the conditional. In many situations, investigators will have done precisely that, vitiating the power of the direct super-exogeneity tests to detect failures. Conversely, one can utilize such results for a deterministic co-breaking based test of super exogeneity.

Again considering the simplest case for exposition, consider adding impulses to the conditional model, such that after saturation:

$$y_t = \beta z_t + \sum_{j=1}^{s} \phi_j 1_{t_j} + \nu_t. \tag{89}$$

At the same time, if S_{α_1} denotes the significant dummies in the marginal model:

$$z_t = \mu + \sum_{i \in S_{\alpha_1}} \delta_i 1_{t_i} + u_t \tag{90}$$

then the test tries to ascertain whether the timing of the impulses in (89) and (90) overlaps. For example, a perfect match would be strong evidence against super exogeneity, corresponding to the result above that the significance of the marginal-model impulses in the conditional would reject super exogeneity.

9 Simulating the powers of automatic super-exogeneity tests

We undertaken simulation analyses for all three scenarios, first for a bivariate relation, then trivariate.

9.1 Failure of weak exogeneity under non-constancy

We begin by considering violations of super exogeneity due to a failure of weak exogeneity, that is $\beta \neq \gamma$, although invariance holds. Further we consider a level shift. The relationship $\mu_{1,t} = \beta \mu_{2,t}$ holds both in the first regime and in the second regime, but:

$$\mu_{2,t} = \lambda 1_{\{t > T_1\}} + \mu_{2,0} \tag{91}$$

and so:

$$\mu_{1,t} = \beta\lambda 1_{\{t>T_1\}} + \beta\mu_{2,0} = \beta\lambda 1_{\{t>T_1\}} + \mu_{1,0} \tag{92}$$

Hence $\beta\lambda$ is the level shift in the mean of y_t at T_1. We allow $\beta\lambda$ to vary across our Monte Carlo experiments to obtain results associated with level shifts of different magnitudes. In particular, $d = \lambda/\sqrt{\sigma_{22}}$ takes the values $1,\ 2,\ \ 2.5,\ \ 3$ and 4. We also allow β to vary across experiments to obtain different degrees of departure from the weak exogeneity condition: in particular, β takes the values 0.75, 1, 1.5 and 1.75, where the first represents the strongest departure from the weak exogeneity condition relating β and γ and the last represents the weakest violation of that condition. Finally, we also consider different sample sizes ($T = 100$ and $T = 300$), break points T_1, and the impact of different choices of the significance level in the marginal and in the conditional. Throughout all Monte Carlo experiments, $M = 10000$ replications were conducted. For the impulse saturation in the marginal model, a partition of $T/2$ was always used.

We investigate the three new types of automatic super exogeneity tests:

(1) a joint F-test in the conditional model, on the set of dummies added because they were significant in the marginal (after single impulse indicator saturation of the marginal and retention of the statistically significant indicators);

(2) a t-test on the individual significance in the conditional model of an index formed using the retained single impulse indicators in the marginal after its impulse saturation (the index weights are the estimated coefficients of the respective indicators in the marginal model).

(3) an F-test on the joint significance in the conditional of two indexes: one formed as described in the previous paragraph and another one whose weights are the product of the weights in the previous index, for each observation, by the value of the regressor we are conditioning on, for that same observation.

We begin by investigating the first test. Table 1 reports the empirical mean rejection frequencies of the null in the joint F-test when a sample size of $T = 300$ is used and 5% significance levels are employed both in the marginal and in the conditional models. The level shift occurs at observation 251. Hence, the second regime has a length of $k = 50$, so $r = 1/6$. The power of the test increases with the decrease in β, as expected, since a smaller β indicates a stronger violation of the weak exogeneity condition. Furthermore, also as expected, the power of the test is increasing with the magnitude of the level shift. Even mild violations of the null are easily detected for level shifts of 2.5σ. The non-centrality $\varphi^2_{r,\alpha}$ in this bivariate case, from (75), is:

$$\varphi^2_{r,\alpha} = \frac{k\mathsf{p}_d\left(\beta-\gamma\right)^2 d^2\sigma_{22}}{\sigma_u^2\left(1+d^2r\right)} \tag{93}$$

with power $\mathsf{p}_\alpha = \mathsf{P}\left[\chi^2_m\left(0\right) > h^{-1}c_\alpha\right]$ where:

$$h = \frac{k+2\varphi^2_{r,\alpha}}{k+\varphi^2_{r,\alpha}} \text{ and } m = \frac{k+\varphi^2_{r,\alpha}}{h}. \tag{94}$$

In table 2, we investigate the impact of reducing the length of the second regime to $k = 25$. All the other defaults of the experiments leading to table 1 apply.

The previous conclusions still apply, but the empirical power is never smaller when the break length diminishes, contrary to the prior theory: the degrees of freedom of the F-test must be playing a fundamental role here. This is partly the motivation to look into the second class of super exogeneity tests: those based on an index replacing the indicators (see Hendry and Santos, 2005).

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1910	0.1527	0.0777	0.0539
2.0	0.9722	0.9362	0.5289	0.1497
2.5	0.9999	0.9930	0.9173	0.3388
3.0	1.0000	1.0000	0.9985	0.6527
4.0	1.0000	1.0000	1.0000	0.9672

Table 1 Level shift at $T_1 = 251$, $T = 300$, 5% used in marginal and conditional.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.3767	0.2742	0.0969	0.0601
2.0	0.9999	0.9968	0.8026	0.2376
2.5	1.0000	1.0000	0.9902	0.5045
3.0	1.0000	1.0000	1.0000	0.7970
4.0	1.0000	1.0000	1.0000	0.9839

Table 2 Level shift at $T_1 = 276$, $T = 300$, 5% used in marginal and conditional.

To assess the impact of the choice of the significance levels, both in the marginal and in the conditional, we consider the case where a more stringent significance is used in the marginal ($\alpha_1 = 1\%$), whilst 5% is still used in the conditional. We assume the remaining default settings, namely $k = 50$ (where these are the last 50 observations of the sample). Table 3 reports results for only two values of β as these are clear enough to highlight the conclusions.

d	$\beta = 1$	$\beta = 1.75$
1.0	0.0850	0.0520
2.0	0.2894	0.0766
2.5	0.6802	0.1192
3.0	0.9647	0.2680
4.0	1.0000	0.8627

Table 3 Level shift at $T_1 = 251$, $T = 300$, 1% used in marginal and 5% in conditional.

Results should be compared with the matching columns in table 1. The choice of a 5% significance level in the marginal, instead of a 1% significance level, leads to a more powerful test, other things being equal.

Table 4 investigates the use of a 5% significance level in the marginal whilst a 10% significance level is used in the conditional, assuming the same defaults as table 3.

This choice of significance levels yields a more powerful test for super exogeneity. Empirical rejection frequencies are never smaller than the ones in table 1.

For these significance levels, the effect observed in table 2 still occurs, namely a trade-off between power and the length of the break that leads to cases where this type of failure of super exogeneity is more difficult to detect by the F-test for longer breaks. Table 5 illustrates this for the case where $\beta = 1$. The break in the underlying DGP is assumed to occur at observation $T_1 = 201$ and hence $k = 100$. We neglect the results for small level shifts.

We now turn to investigate the effect on power of the sample size. Table 6 reports the Monte Carlo

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.3063	0.2486	0.1462	0.1099
2.0	0.9878	0.9690	0.6604	0.2512
2.5	1.0000	0.9996	0.9570	0.4743
3.0	1.0000	1.0000	0.9920	0.7700
4.0	1.0000	1.0000	1.0000	0.9847

Table 4 Level shift at $T_1 = 251$, $T = 300$, 5% used in marginal and 10% in conditional .

d	$\beta = 1$
2.5	0.0972
3.0	0.7757
4.0	1.0000

Table 5 Level shift at $T_1 = 200$, $T = 300$.

results obtained using the same settings as previously (namely with a significance level of 5% in the marginal model and a significance level of 10% in the conditional model) when $T = 100$. The level shift is assumed to have occurred at observation 81, yielding $k = 20$.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1408	0.1306	0.1057	0.0960
2.0	0.5553	0.4944	0.2805	0.1487
2.5	0.8613	0.8189	0.5484	0.2306
3.0	0.9816	0.9719	0.8447	0.3910
4.0	0.9999	0.9999	0.9972	0.7267

Table 6 Level shift at $T_1 = 81$, $T = 100$.

First, even for a sample size of $T = 100$, the test has good power against mild violations of weak exogeneity provided there is at least a level shift, even if not too steep (power is acceptable even for $\beta = 1.5$ for a break of at least 2.5σ).

Although results are not directly comparable (given that the percentage of observations in the second regime differs, even if not greatly), there is a loss of power with the reduction of sample size. Further, power continues to increase monotonically with the magnitude of the level shift and with the decrease in β. The trade off between length of the break and power is also a feature of smaller sample sizes, as illustrated in table 7, where a break of length $k = 30$ is assumed to begin at observation 71. Again, the results for very small level shifts are negligible.

From table 7, the increase in the length of the break is reducing its power. However, moderate level shifts (say 3σ) allow the detection of violations of weak exogeneity with a high relative frequency.

We now investigate the empirical power of the index-based test. Table 11 reports the results for a sample size of $T = 100$, and for the DGP same parameter values: the shift occurs at observation 81, implying $k = 20$.

The first conclusions concern comparing tables 11 and 6. First, although the index-based test has higher empirical power for small shifts (magnitudes σ and 2σ), the joint F test tends to do better for shifts of higher magnitudes. Differences are, however, often too small to be significant, and on that basis

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
2.5	0.2602	0.2447	0.1736	0.1182
3.0	0.7078	0.6805	0.4857	0.2212
4.0	0.9969	0.9955	0.9672	0.5758

Table 7 Level shift at $T_1 = 71$, $T = 100$.

$T = 300$	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
σ	0.0805	0.0654	0.0350	0.0263
2σ	0.7169	0.6119	0.2196	0.0618
2.5σ	0.9770	0.9534	0.6159	0.1431
3σ	0.9999	0.9995	0.9526	0.3724
4σ	1.0000	1.0000	1.0000	0.9084

Table 8 Level shift at $T_1 = 250$, 2.5% used in marginal and conditional, F-test.

$T = 100$	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
σ	0.0267	0.0270	0.0257	0.0225
2σ	0.1144	0.0982	0.0540	0.0345
2.5σ	0.3920	0.3493	0.1590	0.0550
3σ	0.7572	0.7153	0.4336	0.1118
4σ	0.9956	0.9939	0.9491	0.4181

Table 9 Level shift at $T_1 = 80$, 2.5% used in marginal and conditional, F-test.

$T = 300$	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
σ	0.0646	0.0530	0.0339	0.0266
2σ	0.6161	0.5565	0.2920	0.1032
2.5σ	0.9176	0.8905	0.6633	0.2578
3σ	0.9937	0.9903	0.9209	0.5307
4σ	1.0000	1.0000	0.9994	0.9085

Table 10 Level shift at $T_1 = 250$, 2.5% used in marginal and conditional, t-test.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1477	0.1390	0.1169	0.1094
2.0	0.5860	0.5462	0.3463	0.1860
2.5	0.8322	0.8024	0.5971	0.3025
3.0	0.9522	0.9444	0.8170	0.4730
4.0	0.9969	0.9951	0.9686	0.7228

Table 11 Level shift at $T_1 = 81$, $T = 100$: index test.

it would seem the index-based test does better. However, no test clearly dominates the other, for the defaults used in this experiment.

Table 11 also deserves to be analyzed by itself. First, values for the power of the index-based test (for shocks of magnitudes greater than σ) are reasonable. The empirical power is decreasing as β gets closer to $\gamma = 2$, as expected. Furthermore, for the range considered (magnitudes up to 4σ) the power is monotonically increasing with the size of the shift, for any β.

The index-based test is a t-test on a single parameter, so its degrees of freedom do not depend on the number of single-impulse indicators 'picked up' from the marginal model. Hence, it is not to be expected that the test would face similar problems to those detected with the joint F-test, where a smaller break length could be associated with higher power. Table 12 extends the analysis by considering $T = 300$ and $k = 50$. Results are to be compared with table 4.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1932	0.1711	0.1217	0.1000
2.0	0.8661	0.8370	0.6209	0.3139
2.5	0.9874	0.9817	0.9011	0.5720
3.0	0.9997	0.9997	0.9891	0.8015
4.0	1.0000	1.0000	1.0000	0.9687

Table 12 Level shift at $T_1 = 251$, $T = 300$: index test.

For this larger sample size, the joint F-test dominates the index based test in terms of power. The only exceptions occur for some intermediate magnitudes when $\beta = 1.75$. In the following subsection, we shall again come to the conclusion that the index test is generally less powerful than the joint F-test for larger samples. On the basis of this, there seems to be no clear decision as to which test is better: one will dominate in some cases, the other will dominate in other settings.

Table 14 reports empirical power for the case where two indexes are used. A sample size of $T = 100$ is considered. Results are to be compared with table 13 where a single index is used, since sample sizes are the same, and so are the significance levels used (2.5% both in the marginal and in the conditional). It is clear that the single index dominates the use of two indexes (in terms of power) for this type of failure of super exogeneity.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.0327	0.0314	0.0287	0.0244
2.0	0.2331	0.2069	0.1025	0.0475
2.5	0.5148	0.4712	0.2572	0.0930
3.0	0.7888	0.7541	0.5198	0.1918
4.0	0.9871	0.9813	0.9007	0.4924

Table 13 Level shift at $T_1 = 81$, $T = 100$: index test, 2.5% significance.

If, instead, 5% significance levels are used in the marginal and 10% in the conditional, the empirical rejection frequencies when $T = 100$ and $T = 300$ are reported in tables 15 and 16, for the two indexes test. Again, both refer to violations of super exogeneity due to a failure in weak exogeneity when invariance holds but there is a level shift.

Even for these more liberal model selection strategies, power to detect this type of departure from

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.0273	0.0272	0.0253	0.0230
2.0	0.1524	0.1345	0.0707	0.0351
2.5	0.4284	0.3862	0.1956	0.0706
3.0	0.7417	0.7027	0.4425	0.1513
4.0	0.9821	0.9750	0.8774	0.4205

Table 14 Level shift at $T_1 = 251$, $T = 300$: two-index test, 2.5% significance.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1277	0.1235	0.1087	0.1052
2.0	0.4855	0.4453	0.2742	0.1515
2.5	0.7668	0.7328	0.5089	0.2452
3.0	0.9323	0.9153	0.7572	0.3956
4.0	0.9951	0.9928	0.9546	0.6484

Table 15 Level shift at $T_1 = 81$, $T = 100$: two-index test, 5% significance in marginal and 10% in conditional.

d	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
1.0	0.1763	0.1575	0.1139	0.0980
2.0	0.8135	0.7772	0.5317	0.2494
2.5	0.9779	0.9681	0.8510	0.4820
3.0	0.9992	0.9986	0.9788	0.7200
4.0	1.0000	1.0000	0.9997	0.9469

Table 16 Level shift at $T_1 = 251$, $T = 300$: two-index test, 5% significance in marginal and 10% in conditional.

super exogeneity is higher with the single index than with the two indexes (compare with tables 11 and 12 respectively).

9.2 Failure of invariance when weak exogeneity holds

We now consider a DGP where the null hypothesis of super exogeneity is false, but weak exogeneity holds (that is: $\beta_t = \gamma_t, \forall t$). Let T^* be such that $1 < T^* < T$, and, for $t < T^*$, let the DGP be given by:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{95}$$

whilst for $t \geq T^*$,

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 3\mu_2^* \\ \mu_2^* \end{pmatrix}, \begin{pmatrix} 28 & 9 \\ 9 & 3 \end{pmatrix} \right]. \tag{96}$$

$\beta_t = \gamma_t$ even after the break, but since $\gamma_t = \sigma_{12,t}\sigma_{22,t}^{-1}$, and the change in σ_{22} is not offset by the change in σ_{12}, the parameter $\phi_{1,t}$, which contains γ_t, is not invariant to changes in the parameter space of the marginal $\phi_{2,t}$, which contains $\sigma_{22,t}$.

For the Monte Carlo experiments, we work with the same settings as in the previous subsection. We consider the same sample sizes as before: $T = 100$ and $T = 300$. We allow μ_2^* to take values from the set $\{2, 2.5, 3, 4\}$ implying a certain set of pairs of unconditional means. Finally, we also allow the break length, k, to vary.

Table 17 reports the Monte Carlo results for a sample size of $T = 100$. We always consider breaks from a certain T^* until the end of the sample. Hence the break dates are: $T = 81,\ T = 71$ and $T = 61$ for $k = 20,\ 30,\ 40$ respectively.[2]

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2.0$	0.3982	0.4939	0.5503
$\mu_2^* = 2.5$	0.5438	0.5998	0.5628
$\mu_2^* = 3.0$	0.6810	0.6926	0.5605
$\mu_2^* = 4.0$	0.9108	0.8527	0.5342

Table 17 Invariance failure, $T = 100$.

Table 17 reveals that the test has good power even for a small sample. An increase in the length of the mean shift from $k = 20$ to $k = 30$ increases the rejection frequency of the null. Nonetheless, a further increase of equal absolute magnitude in the length of the break can reduce power (for greater level shifts) and loses the monotonicity property (power does not increase with the size of the shift for $k = 40$).

In table 18, we maintain the default settings of the Monte Carlo experiment of the previous table, but we consider a sample of size $T = 300$. We consider breaks at observations $T^* = 261, T^* = 251, T^* = 201$ and $T^* = 161$, matching respectively the values $k = 40,\ 50,\ 100$ and 140 considered in table 18.

The remarkable feature is the good power against this failure of super exogeneity, for all break lengths (which are in some cases always less or equal to a third of the sample size) and even for the smallest mean shifts considered.

[2]Tables 17 and 18 use a 5% significance level in the marginal model and a 10% significance level in the conditional.

$T = 300$	$k = 40$	$k = 50$	$k = 100$	$k = 140$
$\mu_2^* = 2.0$	0.5413	0.6008	0.8161	0.8980
$\mu_2^* = 2.5$	0.7404	0.7968	0.8907	0.8157
$\mu_2^* = 3.0$	0.8963	0.9288	0.9456	0.6800
$\mu_2^* = 4.0$	0.9973	0.9991	0.9910	0.2878

Table 18 Invariance failure, $T = 300$.

Notwithstanding, table 18 also highlights the problem we had discussed for smaller sample sizes: the length of the break adversely affects the power to detect departures from super exogeneity when k becomes 'too big'. In the case discussed in table 18, this is clear for $k = 140$. A set of results not reported here indicates that for $k = 125$, the same reverse effects on power might already be present.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2$	0.1874	0.2181	0.2400
$\mu_2^* = 2.5$	0.2864	0.2891	0.2496
$\mu_2^* = 3$	0.4063	0.3712	0.2558
$\mu_2^* = 4$	0.6730	0.5352	0.2584

Table 19 Invariance failure, $T_1 = 0.8T = 80$, F-test, 2.5%.

$T = 300$	$k = 40$	$k = 50$	$k = 100$	$k = 140$
$\mu_2^* = 2$	0.2885	0.3326	0.4506	0.5746
$\mu_2^* = 2.5$	0.4643	0.5062	0.5719	0.4930
$\mu_2^* = 3$	0.6726	0.7138	0.6969	0.4152
$\mu_2^* = 4$	0.9527	0.9644	0.8795	0.2382

Table 20 Invariance failure, $T_1 = 240$ $T = 300$, 2.5 per cent, F-test .

We now look at the empirical rejection frequencies of the null for the index-based test. We consider only the case where $T = 300$. The same defaults as in the previous experiment are used in the Monte Carlo, namely significance levels of 5% and 10%. Table 21 reports the results.

Comparing tables 21 and 18 shows that the joint F-test dominates the index-based test (with the exception of the two largest unconditional mean shifts for $k = 140$). The power of the index-based test is only acceptable for level shifts of length greater than some threshold.

However, in spite of the power dominance of the joint F-test over the index test, it is nonetheless true that the power properties of the index test are worth while: power increases monotonically both with the mean shift and with the break length (this last claim could not be made for the joint F-test as referred to above).

It remains to investigate the effects on power of using a conditional model with two indexes: as explained earlier, the first index, ι_1, is the usual index where the weights are the dummy coefficients for the significant dummies in the marginal; whilst the second index, ι_2, uses as weights the products, for each observation, of the respective indicator coefficient estimate in the marginal (if significant) and its observed value for the variable we are conditioning on, z_t.

$T = 300$	$k = 40$	$k = 50$	$k = 100$	$k = 140$
$\mu_2^* = 2.0$	0.2016	0.2095	0.2591	0.3167
$\mu_2^* = 2.5$	0.2802	0.2916	0.4034	0.5691
$\mu_2^* = 3.0$	0.3602	0.3836	0.6483	0.8358
$\mu_2^* = 4.0$	0.5725	0.6608	0.9793	0.9967

Table 21 Invariance failure, index test, $T = 300$.

Adding the two indexes to the conditional model, with parameters φ_1 and φ_2, the null hypothesis is:

$$\mathsf{H}_0 : \varphi_1 = \varphi_2 = 0 \tag{97}$$

which can be tested using the usual statistic with null distribution F with 2 degrees of freedom.

We consider the same departures from invariance as for the previous cases in this section, and the same break lengths. A significance level of 2.5% is used in both the marginal model and in the conditional model. Results are reported in tables 22 and 23 with respect to sample sizes of $T = 300$ and $T = 100$, respectively.

$T = 300$	$k = 40$	$k = 50$	$k = 100$	$k = 140$
$\mu_2^* = 2.0$	0.2257	0.2604	0.4329	0.5453
$\mu_2^* = 2.5$	0.3321	0.3838	0.5285	0.6640
$\mu_2^* = 3.0$	0.4518	0.5020	0.6521	0.8363
$\mu_2^* = 4.0$	0.6300	0.6653	0.9339	0.9939

Table 22 Invariance failure, two-index test, $T = 300$.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2.0$	0.2440	0.3471	0.4874
$\mu_2^* = 2.5$	0.3526	0.4562	0.5923
$\mu_2^* = 3.0$	0.4829	0.5765	0.7087
$\mu_2^* = 4.0$	0.7630	0.8243	0.9176

Table 23 Invariance failure, two-index test, $T = 100$.

Tables 22 and 23 should be compared with tables 25 and 24, respectively, which report the empirical rejection frequency of the null for the cases where the simpler index is used in the conditional and where significance levels of 2.5% are used in the marginal and the conditional.

The two-index test has greater power to detect departures of super exogeneity for invariance failure, than the corresponding single index tests. This claim is valid, irrespective of the choice of significance levels. Tables 26 and 27 confirm this by reporting the empirical power for a sample size of $T = 100$ when a 5% significance level is used in the marginal model and a 10% significance level is used in the conditional.

$T = 300$	$k = 40$	$k = 50$	$k = 100$	$k = 140$
$\mu_2^* = 2.0$	0.0957	0.1016	0.1249	0.1601
$\mu_2^* = 2.5$	0.1490	0.1576	0.2060	0.3335
$\mu_2^* = 3.0$	0.2260	0.2472	0.3964	0.6473
$\mu_2^* = 4.0$	0.4241	0.4755	0.8746	0.9829

Table 24 Invariance failure, index test, $T = 300$, 2.5% significance.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2.0$	0.1008	0.1175	0.1270
$\mu_2^* = 2.5$	0.1478	0.1557	0.1570
$\mu_2^* = 3.0$	0.2106	0.2139	0.2480
$\mu_2^* = 4.0$	0.3642	0.4276	0.5742

Table 25 Invariance failure, index test, $T = 100$, 2.5% significance.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2.0$	0.6929	0.8346	0.9268
$\mu_2^* = 2.5$	0.7883	0.8985	0.9603
$\mu_2^* = 3.0$	0.8721	0.9435	0.9791
$\mu_2^* = 4.0$	0.9767	0.9932	0.9982

Table 26 Invariance failure, two-index test, $T = 100$, 5% significance in marginal and 10% in conditional.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2.0$	0.2074	0.2424	0.2664
$\mu_2^* = 2.5$	0.2734	0.3067	0.3414
$\mu_2^* = 3.0$	0.3429	0.4017	0.4928
$\mu_2^* = 4.0$	0.5403	0.7031	0.8344

Table 27 Invariance failure, index test, $T = 100$, 5% significance in marginal and 10% in conditional.

9.3 Failure of weak exogeneity under constancy

Finally, we consider a departure from super exogeneity due to a failure in weak exogeneity ($\beta \neq \gamma$) alone, when invariance holds and there is no level shift. We consider the following alternative DGP:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} \beta^* \\ 1 \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right]. \quad (98)$$

We allow β^* to take values from the set $\{0.5, 0.75, 1, 1.25, 1.5, 1.75\}$, $\beta^* \neq \gamma$ when $\gamma = 2$. All the default settings from previous experiments apply. Table 28 reports the results for sample sizes of $T = 100,\ 200$ and 300.

In table 28, apart from the empirical rejection frequencies, we also include the empirical significance level in the conditional (α_c) for each sample size, when the nominal significance level in the conditional is 10%.

	$T = 100$	$T = 200$	$T = 300$
$\beta^* = 0.50$	0.096	0.0974	0.1009
$\beta^* = 0.75$	0.096	0.0974	0.1009
$\beta^* = 1.00$	0.096	0.0974	0.1009
$\beta^* = 1.25$	0.096	0.0974	0.1009
$\beta^* = 1.50$	0.096	0.0974	0.1009
$\beta^* = 1.75$	0.096	0.0974	0.1009
α_c	**0.096**	**0.0974**	**0.1009**

Table 28 Failure of weak exogeneity under constancy, $T = 100$.

As expected, the test has virtually no power against this form of failure of the weak exogeneity hypothesis. Indeed, averaging across $M = 10000$ replications, we conclude that the mean rejection frequency is the same for any value of β^* considered, and virtually the same as it would be the case for $\beta^* = \beta = \gamma = 2$, the value under the null of weak exogeneity. Hence, the empirical power is equal to the empirical significance level, meaning the test has no power to detect this form of failure of super exogeneity.

9.4 Optimal infeasible-test power

The optimal infeasible test differs from those analyzed above in that the location of the breaks in the marginal proccess is known. Thus, there is no need to impulse saturate the marginal and retain the relevant impulses. Rather (for the joint F-test, say), one tests in the conditional a set of single impulse indicators, each of which corresponds to an observation within the break period in the marginal proccess. In the tables below we use a 2.5% significance level for testing in the conditional.

Hence, the empirical rejection frequencies are the empirical proxies of maximum achievable power for the relevant sample sizes here $T = 100$. The break is known to be a mean shift starting at observation $T_1 = 80$, so 20 single impulse indicators are included in the conditional model. Table (29) refers to the case of no weak exogeneity (and no constancy), but with invariance holding. Table (30) refers to the case of invariance failure but with weak exogeneity holding.

Relative to the optimal infeasible test, the automatic tests based on saturation of the marginal naturally lose a signifcant power for breaks of small magnitudes.

$T = 100$	$\beta = 0.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
σ	0.9999	0.9941	0.4040	0.0831
2σ	1.0000	1.0000	0.9301	0.2467
2.5σ	1.0000	1.0000	0.9730	0.3258
3σ	1.0000	1.0000	0.9851	0.3803
4σ	1.0000	1.0000	0.9881	0.432

Table 29 Level shift at $T_1 = 80$, 2.5% test in conditional, F-test: break location known.

$T = 100$	$k = 20$	$k = 30$	$k = 40$
$\mu_2^* = 2$	0.9977	0.9985	0.9964
$\mu_2^* = 2.5$	0.9995	0.9998	0.9983
$\mu_2^* = 3$	0.9999	0.9999	0.9990
$\mu_2^* = 4$	1.0000	1.0000	0.9992

Table 30 Invariance failure, $T_1 = 80$, 2.5% F-test: break location known.

The optimal test has power increasing with the length of the break for breaks of magnitude σ and 2σ, and a failure of weak exogeneity where $\beta = 1.75$ At $T = 100$, where k is the length of the break, we obtain table (31).

$T = 100$	$k = 45$	$k = 40$	$k = 30$	$k = 20$	$k = 15$	$k = 10$	$k = 5$
σ	0.5720	0.5633	0.5145	0.4232	0.3477	0.2590	0.0725
2σ	0.9418	0.9376	0.9199	0.8801	0.8280	0.7198	0.4837

Table 31 Weak exogeneity failure, $T = 100$, optimal test, 2.5%, break location known.

The optimal test exhibits the predicted behaviour from the theory section: power increases with the break length.

10 Monte Carlo experiments with $n = 3$

In this section, we conduct some Monte Carlo experiments to assess what happens with the tests when the DGP is the three-dimensional normal distribution:

$$\begin{pmatrix} y_t \\ z_{1,t} \\ z_{2,t} \end{pmatrix} \sim \mathsf{N}_3 \left[\begin{pmatrix} \mu_y \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 54/16 & 4 & 1 \\ 4 & 5 & 2 \\ 1 & 2 & 4 \end{pmatrix} \right] \tag{99}$$

which could also be expressed as:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_3 \left[\begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Omega}_{22} \end{pmatrix} \right] \tag{100}$$

where $\mathbf{z}_t$ and $\boldsymbol{\mu}_2$ are 2×1 vectors mn 2×1 vectors, $\boldsymbol{\sigma}_{12}$ is a 1×2 row vector and $\boldsymbol{\Omega}_{22}$ is the variance-covariance matrix of $\mathbf{z}_t$. Under the null of super exogeneity, weak exogeneity and invariance must hold.

Weak exogeneity entails that $\boldsymbol{\beta} = \boldsymbol{\gamma} = \boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\sigma}_{12}$, which in this case is:

$$\boldsymbol{\beta} = \begin{pmatrix} 14/16 \\ -3/16 \end{pmatrix}. \tag{101}$$

As $\mu_y = \boldsymbol{\beta}'\boldsymbol{\mu}_2$, the unconditional mean of y_t is $\mu_y = 19/16 = 1.1875$, whereas the conditional variance of $y_t|\mathbf{z}_t$ is given by:

$$\sigma_{11} - \boldsymbol{\sigma}'_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\sigma}_{12} = \omega^2 = 1/16. \tag{102}$$

These conditions guarantee super exogeneity of the parameters of the conditional model $y_t|\mathbf{z}_t$, with respect to changes in the parameters of the marginal model for $\mathbf{z}_t$.

In the Monte Carlo experiments, $M = 10000$ replications with sample sizes of $T = 100$ and 300 (albeit we focus on $T = 100$ to highlight the crucial features of the tests when $\mathbf{z}_t$ is not a scalar). Throughout, we use 2.5% significance levels for both in the impulse saturation stage in the marginal, and in testing in the conditional.

10.1 Weak exogeneity failure under invariance for level shifts

We consider the block F-test on the dummies retained from each of the marginal processes, and the single-index test, using two index variables: one for each of the impulses retained from the location-scale models for $z_{1,t}$ and $z_{2,t}$. For the block F-test, the relevant econometric model is:

$$y_t = \eta + \beta_1 z_{1,t} + \beta_2 z_{2,t} + \sum_{S_1} \phi_t D_t + \sum_{S_2} \zeta_t D_t + u_t \tag{103}$$

where it is assumed that $u_t \sim \mathsf{N}\left[0, \sigma_u^2\right]$, S_1 and S_2 are the sets of indicators retained in the location-scale models for the first and second elements of the $\mathbf{z}_t$ vector.

For the single-index based test, we use the econometric model:

$$y_t = \theta + \delta_1 z_{1,t} + \delta_2 z_{2,t} + \pi_1 I_{1,t} + \pi_2 I_{2,t} + v_t \tag{104}$$

The weights in $I_{1,t}$ and $I_{2,t}$ are the estimated single impulse indicators' coefficients in the impulse-saturated location-scale models of $z_{1,t}$ and $z_{2,t}$ respectively. Again, we assume that $v_t \sim \mathsf{N}\left[0, \sigma_v^2\right]$. In (104), the super-exogeneity test based on the single-index is also an F-test, with null hypothesis:

$$\mathsf{H}_0 : \pi_1 = \pi_2 = 0 \tag{105}$$

hence having two degrees of freedom. For the first super-exogeneity test considered above, the null hypothesis is that all single impulse indicators' coefficients are zero.

Here, we violate super exogeneity by losing weak exogeneity, so consider DGPs where $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$, using three alternative values for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}_1 = \begin{pmatrix} 2.5 \\ -2 \end{pmatrix}; \quad \boldsymbol{\beta}_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}; \quad \boldsymbol{\beta}_3 = \begin{pmatrix} 0.5 \\ -0.25 \end{pmatrix} \tag{106}$$

implying, respectively, $\mu_y = -1$, $\mu_y = 1$ and $\mu_y = 0.25$, so that $\mu_y = \boldsymbol{\beta}'\boldsymbol{\mu}_2$ and invariance hold. We also allow for the existence of level shifts in the marginal proccess. For simplicity, we assume the level shifts occur for both $z_{1,t}$ and $z_{2,t}$ on the same date T_1, are of the same magnitude, and only look at a

d	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$
1.0	0.0474	0.0773	0.0902
2.0	0.4776	0.6782	0.7455
2.5	0.8422	0.9428	0.9614
3.0	0.9856	0.9968	0.9987
4.0	0.9999	1.0000	1.0000

Table 32 Index test conditioning on two regressors, $T = 100$.

sample size of $T = 100$ with the break start at $T_1 = 81$, implying the break length $k = 20$. We consider break magnitudes of $\sqrt{5},\ 2\sqrt{5},\ 2.5\sqrt{5},\ 3\sqrt{5}$ and $4\sqrt{5}$ as $\sigma_{22} = 5$. Table 32 reports the rejection frequencies for these values of $\boldsymbol{\beta}$ and magnitudes of the level shift.

The results in table 32 reflect prior expectations. In particular, the empirical power of the test is increasing with the magnitude of the level shift. Let the magnitudes of the level shifts be stored in a column vector $\boldsymbol{\lambda}$ of dimensions 2×1 (in our case, where $n = 3$), then power is increasing with the absolute value of $(\boldsymbol{\beta} - \boldsymbol{\gamma})' \boldsymbol{\lambda}$. Hence, noting that:

$$\left|(\boldsymbol{\beta}_1 - \boldsymbol{\gamma})'\right| < \left|(\boldsymbol{\beta}_2 - \boldsymbol{\gamma})'\right| < \left|(\boldsymbol{\beta}_3 - \boldsymbol{\gamma})'\right| \tag{107}$$

the test behaves according to theory: for each column vector of magnitude $\boldsymbol{\lambda}$, the empirical power of the test increases from $\boldsymbol{\beta}_1$to $\boldsymbol{\beta}_2$ and then to $\boldsymbol{\beta}_3$.

Table 33 reports the rejection frequencies for the block F-test. A preliminary Monte Carlo experiment, under the null hypothesis, showed that the nominal significance of 2.5% was well approximated by an empirical rejection frequency of the null of 2.34%. In table 33, we only consider values for the break magnitude greater or equal to 2σ, as smaller magnitudes yield very small values for the empirical power.

d	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$
2.0	0.0663	0.1738	0.2292
2.5	0.2343	0.5058	0.5983
3.0	0.5512	0.8303	0.8869
4.0	0.9238	0.9933	0.9960

Table 33 Block F-test conditioning on two regressors, $T = 100$.

Results in table 33 again confirm that the properties of the block F-test do not seem to be affected by the existence of additional regressors to condition on. The empirical power of the test behaves as it should from a theory point of view: increasing with the magnitude of the level shift for each $\boldsymbol{\beta}$; and increasing from $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_2$ and from this to $\boldsymbol{\beta}_3$ for each $\boldsymbol{\lambda}$. The empirical rejection frequencies are high for shifts of at least 2.5σ. For bigger magnitudes, the test has quite good power against failures of super exogeneity due to lack of weak exogeneity.

In conclusion, both the single-index test and the block F-test behave well in terms of power when $n = 3$, that is when the single equation model is conditioning y_t on more than one variable. This holds when the alternative of no super exogeneity is due to a failure in weak exogeneity. We only look at these two tests because it was the case for $n = 2$ that the double-index test did not add power for this type of failure of super exogeneity.

10.2 Failure of invariance with weak exogeneity

Consider the following DGP for the first $T-k$ observations in the sample:

$$\begin{pmatrix} y_t \\ z_{1,t} \\ z_{2,t} \end{pmatrix} \sim \mathsf{N}_3 \left[\begin{pmatrix} 1.1875 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 54/16 & 4 & 1 \\ 4 & 5 & 2 \\ 1 & 2 & 4 \end{pmatrix} \right] \tag{108}$$

where $k = 20$, and hence $T_1 = 81$. From T_1 to T (which we assume to be $T = 100$), the DGP will instead be:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_3 \left[\begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{\mu}_2^* \\ \boldsymbol{\mu}_2^* \end{pmatrix}, \begin{pmatrix} 48/13 & 4 & 1 \\ 4 & 6 & 2 \\ 1 & 2 & 5 \end{pmatrix} \right] \tag{109}$$

Then:

$$\boldsymbol{\gamma}' = \boldsymbol{\sigma}_{12}'\boldsymbol{\Omega}_{22}^{-1} = \begin{pmatrix} 9/13 & -1/13 \end{pmatrix} \tag{110}$$

and since weak exogeneity holds:

$$\boldsymbol{\gamma} = \boldsymbol{\beta} = \begin{pmatrix} 9/13 \\ -1/13 \end{pmatrix}. \tag{111}$$

As $\boldsymbol{\Omega}_{22}$ changes, $\boldsymbol{\gamma}$ changes. Since, $\boldsymbol{\gamma} \in \boldsymbol{\phi}_{1,t}$, the parameter space of the conditional model, and $\boldsymbol{\Omega}_{22} \in \boldsymbol{\phi}_{2,t}$, the parameter spaces of the variables we are conditioning on, $\boldsymbol{\phi}_{1,t}$ is not invariant to $\mathcal{C}^{\boldsymbol{\phi}_{2,t}}$, which is the class of interventions on $\boldsymbol{\phi}_{2,t}$. Since invariance does not hold, super exogeneity fails from T_1 onwards.

In the DGP for the last 20 observations, the unconditional vector of means of $\mathbf{z}_t$ has changed to $\boldsymbol{\mu}_2^*$. We consider four possible DGPs for these last 20 observations. The variance-covariance matrix is always as in (109), but with four possible vectors for $\boldsymbol{\mu}_2^*$, and since weak exogeneity holds, four possible values for μ_y. In particular, we consider as vectors of unconditional means for the DGP in (109) those in table 34.

Shift	
$\boldsymbol{\mu}_2^* = 2.0\boldsymbol{\mu}_2$	$\begin{pmatrix} 2.307692308 & 4 & 6 \end{pmatrix}$
$\boldsymbol{\mu}_2^* = 2.5\boldsymbol{\mu}_2$	$\begin{pmatrix} 2.884615385 & 5 & 7.5 \end{pmatrix}$
$\boldsymbol{\mu}_2^* = 3.0\boldsymbol{\mu}_2$	$\begin{pmatrix} 3.461538462 & 6 & 9 \end{pmatrix}$
$\boldsymbol{\mu}_2^* = 4.0\boldsymbol{\mu}_2$	$\begin{pmatrix} 4.615384615 & 8 & 10.2 \end{pmatrix}$

Table 34 Vectors of unconditional means in four DGPs.

We now consider three tests for invariance: the block F-test; one based on the single-index (which makes use of two indexes as we are conditioning on $n-1>1$ variables); another based on the double index: along with the indexes in the previous section, another two indexes are considered (making a total of four index variables), each of these refers to one of the marginal regressors and is the product, for each observation, of the respective value of that marginal regressor and the estimate of the single impulse coefficient for that observation in tha marginal location-scale model for that regressor (if this was found to be significant when the location scale model was impulse saturated).

The econometric model on which the double index test is based is given by:

$$y_t = \kappa_0 + \kappa_1 z_1^t + \kappa_2 z_2^t + \rho_1 I_{1,t} + \rho_2 I_{2,t} + \rho_3 I_{3,t} + \rho_2 I_{4,t} + \nu_t \tag{112}$$

where $I_{3,t}$ and $I_{4,t}$ are the new indexes with respect to the single-index version of the test. The null hypothesis is

$$\mathsf{H}_0 : \rho_1 = \rho_2 = \rho_3 = \rho_4 = 0 \tag{113}$$

which entails 4 restrictions instead of 2.

Table 35 reports the empirical rejection frequencies of the null for the single-index test, the double-index test and the block F-test, for each of the four DGPs, given the sample size we are considering and the use, in the marginal and in the conditional, of a 2.5% significance level.

$T = 100$	single	double	block F
$\boldsymbol{\mu}_2^* = 2.0\boldsymbol{\mu}_2$	0.2702	0.3515	0.6067
$\boldsymbol{\mu}_2^* = 2.5\boldsymbol{\mu}_2$	0.3359	0.4448	0.8712
$\boldsymbol{\mu}_2^* = 3.0\boldsymbol{\mu}_2$	0.3639	0.5056	0.9733
$\boldsymbol{\mu}_2^* = 4.0\boldsymbol{\mu}_2$	0.384	0.5679	0.9985

Table 35 Invariance tests when weak exogeneity holds, $T = 100$.

The noticeable conclusion from table 35 is that, as occurred with $n-1 = 1$, the double index performs better in terms of empirical power than the single index, when testing the null of super exogeneity in a setting where the alternative is due to a failure of invariance rather than weak exogeneity.

The empirical powers observed for the single index are relatively low. Notwithstanding, for moderate level shifts (say, $3\boldsymbol{\mu}_2$) the empirical power is acceptable when the double-index test is used.

Table 35 also clarifies that the block F-test dominates the others in terms of empirical rejection frequencies. It exhibits very reasonable empirical power for level shifts of all magnitudes considered. This dominance did not occur for $n = 2$.[3]

11 Conclusion

The concept of automatically computable tests for super exogeneity based on selecting from impulse saturation of the marginal process to test the conditional is clearly realisable. The tests proposed here have the correct null rejection frequency in constant conditional models when the nominal size is not too small in the marginal at small sample sizes (e.g. 5%), for a variety of marginal processes, both constant and with breaks. The tests also have power against failures of super exogeneity when either of invariance or weak exogeneity fails and the marginal process changes. Neither class of tests uniformly dominates the other. Their approximate power functions were derived analytically for regression models and explain the simulation outcomes well.

While all the derivations and Monte Carlo experiments here have been for static regression equations, the principles are general, and should apply to dynamic equations (probably with more approximate null rejection frequencies) and to non-stationary settings: these are the focus of our present research.

[3] It should be stressed that with $n > 2$, neither of these tests behaves well with the break length. Results not reported here reveal lower empirical rejection frequencies for $k > 20$, for all level shifts.

References

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox* 4th edn. London: Timberlake Consultants Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994; and in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Ericsson, N. R., and Irons, J. S. (eds.)(1994). *Testing Exogeneity*. Oxford: Oxford University Press.

Favero, C. (1989). Testing for super exogeneity: The case of the term structure of interest rates. Discussion paper 67, Institute of Economics and Statistics, Oxford.

Favero, C., and Hendry, D. F. (1992). Testing the Lucas critique: A review. *Econometric Reviews*, **11**, 265–306.

Hendry, D. F. (1988). The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers*, **40**, 132–149. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994; and in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

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

Hendry, D. F., Johansen, S., and Santos, C. (2004). Selecting a regression saturated by indicators. Unpublished paper, Economics Department, University of Oxford.

Hendry, D. F., and Krolzig, H.-M. (1999). General-to-specific model selection using PcGets for Ox. Unpublished paper, Economics Department, Oxford University.

Hendry, D. F., and Massmann, M. (2005). Co-breaking: Recent advances and a synopsis of the literature. forthcoming, Journal of Business and Economic Statistics, Special Issue.

Hendry, D. F., and Santos, C. (2005). Regression models with data-based indicator variables. *Oxford Bulletin of Economics and Statistics*, **67**, 571–595.

Jansen, E. S., and Teräsvirta, T. (1996). Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, **58**, 735–763.

Johnson, N. L., and Kotz, S. (1970). *Continuous Univariate Distributions – 2*. New York: John Wiley.

Krolzig, H.-M., and Toro, J. (2002). Testing for super-exogeneity in the presence of common deterministic shifts. *Annales d'Économie et de Statistique*, **67/68**, 41–71.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Brunner, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Psaradakis, Z., and Sola, M. (1996). On the power of tests for superexogeneity and structural invariance. *Journal of Econometrics*, **72**, 151–175.

Revankar, N. S., and Hartley, M. J. (1973). An independence test and conditional unbiased predictions in the context of simultaneous equation systems. *International Economic Review*, **14**, 625–631.

Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, **4**, 393–397.

Sánchez, M. J., and Peña, D. (2003). The identification of multiple outliers in ARIMA models. *Communications in Statistics: Theory and Methods*, **32**, 1265–1287.