CATÓLICA
UNIVERSIDADE CATÓLICA PORTUGUESA | PORTO
Faculdade de Economia e Gestão

DOCUMENTOS DE TRABALHO — WORKING PAPERS

ECONOMIA — ECONOMICS

Nº 11/2007

AUTOMATIC TESTS FOR SUPER EXOGENEITY

David Hendry
University of Oxford

Carlos Santos
Universidade Católica Portuguesa (Porto)

Automatic Tests for Super Exogeneity

David Hendry (University of Oxford, Department of Economics) and Carlos Santos (Portuguese Catholic University, FEG)*

December 3, 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 in 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 for a failure of weak exogeneity when there is a shift in the marginal 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; indicators; co-breaking

JEL classifications: C51; C22

1 Introduction

In all areas of policy which involve regime shifts or structural breaks in conditioning variables, superexogeneity 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, invariance under such shifts is equally relevant. Tests for superexogeneity 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. Stanley (2000) provides a more recent survey. Favero

*Corresponding author: csantos@porto.ucp.pt. Comments on earlier drafts by Bent Nielsen, Jurgen Doornik, Gunnar Bardesen, J.P. Urbain and Franz Palm, as well as from Conference participants at the European Meeting of the Econometric Society in Vienna, 2006, proved to be most useful. The usual disclaimer applies. 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.

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 superexogeneity 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 (1976) critique. Jansen and Teräsvirta (1996) propose self-exciting threshold models for testing constancy in the conditional model as well as superexogeneity. Krolzig and Toro (2002) developed superexogeneity 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 Massman (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 weak exogeneity (WE) and invariance.

The ability to detect outliers and shifts in a model using the dummy saturation technique (see Hendry, Johansen and Santos, 2005) opens the door to this new class of automatically computable superexogeneity and invariance tests. The marginal model (or system) is saturated with impulse indicators (entering an indicator for every observation in feasible subsets: e.g. dummy saturating the first half of the sample and storing the significant indicators; then checking the other half) and all significant outcomes are retained. The authors derive the probability of falsely retaining impulses for a location-scale IID process, and obtain the distribution of the estimated mean and variance under 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 superexogeneity of the conditioning variables for the parameters of the conditional model over a range of significance levels of the marginal model saturation tests. Moreover, it has good power to detect failures of superexogeneity. Finally, it can be computed automatically – that is without explicit user intervention, as occurs with (say) residual autocorrelation tests – once the desired nominal significance of the marginal saturation and conditional superexogeneity tests have been specified.

Five conditions should be satisfied for an automatic superexogeneity test. First the test should not require ex-ante knowledge by the investigator of the timing, signs and magnitudes of any breaks in the marginal processes for the conditioning variables. The test proposed here uses impulse saturation of 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 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 superexogeneity, as required for tests in the class proposed (say) by Revankar and Hartley (1973). Fourthly, the test must have power against any form of failure of superexogeneity or invariance in the conditional model when there are location shifts in some of the marginal processes. Below we establish the general form of the non-centrality parameters of the proposed tests in the two main cases. Finally, the test should be computed without additional user intervention. That is true for the impulse saturation test based on PcGets, although as yet the precise form of the test procedure is not yet implemented in any released version.[1]

The structure of the paper is as follows. Section 2 considers superexogeneity in a regression context to elucidate the testable hypotheses which it entails. Next, section 3 discusses three different ways in which superexogeneity can fail. Section 4 describes the impulse saturation tests (see Hendry et. al. 2005) and how these can be extended to test superexogeneity. Section 5 provides Monte Carlo evidence on the null rejection frequency (NRF) of the proposed procedure. Then, section 6 provides detailed analytic derivations for three multivariate examples of superexogeneity 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 WE with constant marginal processes, which is a case where the proposed tests may have little power. Section 7 investigates the powers of the proposed tests in an extensive set of Monte Carlo experiments related to the analysis in section 6 for a bivariate DGP. Section 8 concludes.

2 Superexogeneity in a regression context

Consider the sequentially factorized joint 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 $\boldsymbol{\phi} = \left(\boldsymbol{\phi}_1' : \boldsymbol{\phi}_2'\right)' = \mathbf{f}\left(\boldsymbol{\theta}\right) \in R^k$. 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 $\boldsymbol{\psi} = \mathbf{h}\left(\boldsymbol{\phi}_1\right)$, but that does not rule out the possibility that $\boldsymbol{\phi}_1$ may change if $\boldsymbol{\phi}_2$ is changed. Superexogeneity augments WE 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:

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

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

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):

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

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

$$\boldsymbol{\phi}_{1,t} = \left(\mu_0 : \gamma_{1,t} : \boldsymbol{\gamma}_{2,t} : \omega_t^2\right) \quad \text{and} \quad \boldsymbol{\phi}_{2,t} = \left(\boldsymbol{\mu}_{2,t} : \boldsymbol{\Sigma}_{22,t}\right)$$

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

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

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) $\omega_t^2 = \omega^2$ is constant $\forall t$;

(d) $\boldsymbol{\phi}_{1,t}$ is invariant to $\mathcal{C}^{\phi_2}$.

Condition (a) requires that $\boldsymbol{\sigma}'_{12,t}\boldsymbol{\Sigma}^{-1}_{22,t}$ is constant over time, which could occur because the two components move in tandem through being connected by $\boldsymbol{\sigma}'_{12,t} = \boldsymbol{\gamma}'_2\boldsymbol{\Sigma}_{22,t}$, 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 $\sigma_{11,t} - \boldsymbol{\sigma}'_{12,t}\boldsymbol{\Sigma}^{-1}_{22,t}\boldsymbol{\sigma}_{21,t} = \sigma_{11,t} - \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}_{22,t}\boldsymbol{\beta} = \omega^2$ 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:

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

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 $\boldsymbol{\Sigma}_{22,t}$. Under superexogeneity, the joint density is

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

so the conditional-marginal factorization is

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

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

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

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 superexogeneity

Superexogeneity 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:

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

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

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

If $\boldsymbol{\mu}_{2,t}$ were to be modelled by lagged values of $\mathbf{x}_t$, an approach that won't be pursued here, the sequential factorization would yield the augmented VAR:

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

The introduction alluded to the currently available tests for superexogeneity. The next section proposes new tests for superexogeneity based on impulse saturation, after briefly reviewing this 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. 2005). 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.

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

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

and conduct an F-test for the significance of $\left(\tau_{1,(\alpha_2)}, .., \tau_{m,(\alpha_2)}\right)$ at level α_2. Under the null of superexogeneity, 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 presents Monte Carlo evidence confirming this. Under the alternative, the test will have power in a variety of situations discussed in section 6. Crucially, such a test can be completely automated, bringing superexogeneity into the purview of hypotheses about a model that can be as easily tested as (say) residual autocorrelation.

A key feature of such a test is that the null rejection frequency of superexogeneity 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 superexogeneity. 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.

From 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:

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

so,

$$\iota_{1,t} = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \widehat{\rho}_{j,i,(\alpha_1)} 1_{\{t=t_i\}} \tag{16}$$

From (16), $\iota_{1,t}$ is the linear combination, at time t, of the estimated coefficients from the retained indicators in each of the $n-1$ marginal models.

Hence, the index test of superexogeneity is the test of the null $\varphi_1 = 0$ in:

$$y_t = \mu_0 + \boldsymbol{\beta}' \mathbf{z}_t + \varphi_{1,(\alpha_2)} \iota_{1,t} + \epsilon_t \tag{17}$$

This provides an alternative scalar test, which should be approximately t-distributed under the null of superexogeneity. Also, for testing a failure of invariance, the indices must be interacted with $\mathbf{z}_t$ as in:

$$\iota_{2,t} = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \widehat{\rho}_{j,i,(\alpha_1)} z_{j,t} 1_{\{t=t_i\}} \tag{18}$$

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

$$y_t = \mu_0 + \boldsymbol{\beta}' \mathbf{z}_t + \varphi_{1,(\alpha_2)} \iota_{1,t} + \varphi_{2,(\alpha_2)} \iota_{2,t} + \epsilon_t \tag{19}$$

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 cases.

Alternatively, if some interest resides in which of the $z_{j,t}$ is responsible for any failure of superexogeneity, then a vector test of the form in (20) could be used, which might have more or fewer degrees of freedom than the corresponding F-test in (14):

$$\boldsymbol{\iota}_{2,t} = \begin{pmatrix} \iota_{2,1,t} \\ \iota_{2,2,t} \\ \vdots \\ \iota_{2,n-1,t} \end{pmatrix} \quad \text{where} \quad \iota_{2,j,t} = \sum_{i=1}^{m_j} \widehat{\rho}_{j,i,(\alpha_1)} z_{j,t} 1_{\{t=t_i\}} \tag{20}$$

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

5 The null rejection frequency of the superexogeneity test

In these Monte Carlo experiments superexogeneity holds as the null and we consider three settings for the marginal process: where there are no breaks in 5.1; a variance change in 5.2; and a mean shift in 5.3. In each case, the baseline DGP is a bivariate system which can be expressed as (see, e.g., Hendry, (1995)):

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

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 superexogeneity 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. From bivariate normal theory we expect this to be the outcome, since the errors in the conditional are independent from those in the marginal.

We consider a constant DGP and two DGPs with changes in the z_t process, all under the null of superexogeneity, where invariance and WE hold before and after the change in the marginal process. 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 Constant marginal under the null of superexogeneity

We use the simple marginal model, defined by:

$$z_t = 1 + v_t \tag{22}$$

where $v_t \sim \mathsf{IN}[0, 5]$. This econometric model mimics the location-scale model analysis in Hendry et al. (2005). As a sample split of $T/2$ is used, the econometric models for the marginal are

$$z_t = \mu_2 + \sum_{t=1}^{T/2} \psi_t 1_t + \varsigma_t \tag{23}$$

and

$$z_t = \mu_2 + \sum_{t=T/2+1}^{T} \psi_t 1_t + \xi_t \tag{24}$$

Let S_{α_1} denote the set of significant dummies in the econometric models (23) and (24). Hence, the second stage of the extended test is to estimate

$$y_t = \beta z_t + \sum_{i \in S_{\alpha_1}} \phi_i 1_{t_i} + \nu_t \tag{25}$$

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 (25), due to belonging to

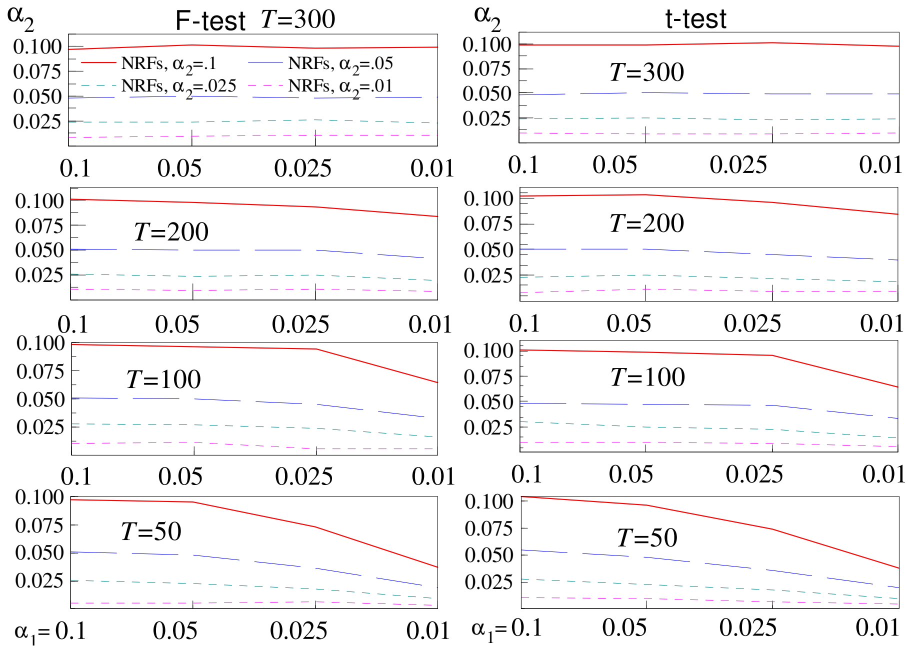

Figure 1: Null rejection frequencies in conditional when α_1 varies for a constant marginal

S_{α_1}, is retained in the conditional. Given that we have imposed superexogeneity 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 1 reports the empirical rejection frequencies of the null in the conditional model when the significant dummies from the marginal are added as in (25). 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).

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 $T \times \alpha_1 > 3$ (see Hendry and Krolzig, 2001). 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 a more constant real α_2.

5.2 Changes in the variance of z_t under the null of superexogeneity

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

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1+20\theta & 10\theta \\ 10\theta & 5\theta \end{pmatrix} \right] \tag{26}$$

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:

$$\gamma^*_{2,t} = \frac{\sigma^*_{12,t}}{\sigma^*_{22,t}} = \frac{10\theta}{5\theta} = \gamma_{2,t} = 2 = \beta \tag{27}$$

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, $\gamma_{1,t} = (\beta - \gamma_{2,t})\,\mu_{2,t} = 0$. 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, superexogeneity holds, so the null distributions of the tests should remain as in subsection 5.1.

Figure 2 reports the empirical rejection frequencies of the null in the conditional model when testing the significant dummies 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 θ, for $\alpha_1 = 0.025$, throughout. In particular, $\theta \in \{2; 5; 10\}$.

Both the F and t-tests have appropriate null rejection frequencies, even when the variance of the marginal process changes markedly. Neither test is confused between variance changes in the marginal and failure of superexogeneity, when the null holds.

5.3 Changes in the mean of z_t under the null of superexogeneity

We modify the baseline DGP (21) to:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} \beta\delta\mu_z \\ \delta\mu_z \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{28}$$

where $\delta = 1$ until $t > T_1 = 0.8T$, in both cases with $\beta = 2$. Superexogeneity holds before and after the level shift. We assume that the variance-covariance

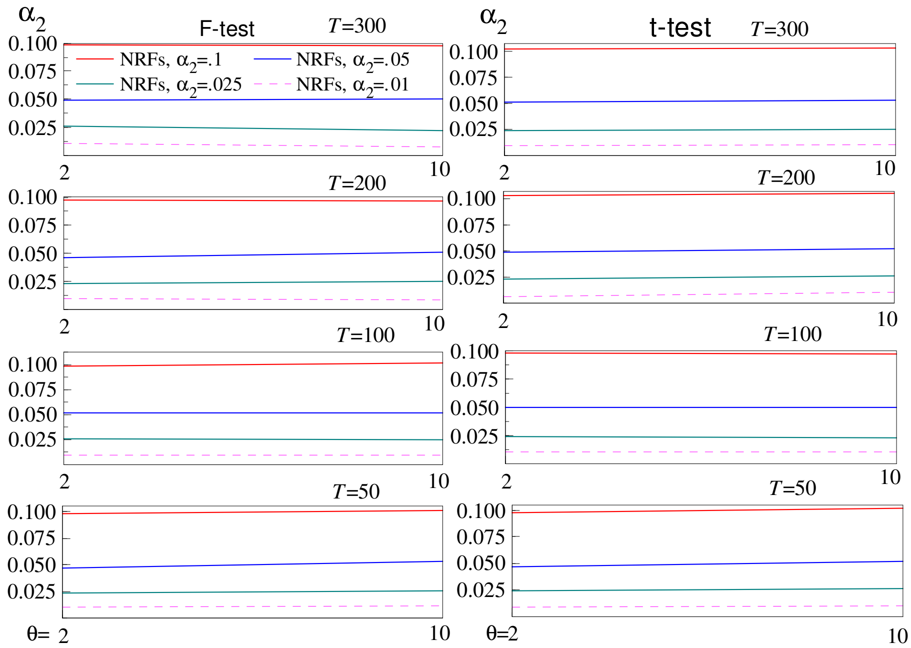

Figure 2: Null rejection frequencies of F and t-tests in conditional for a variance shift in the marginal

matrix remains the same before and after the shift, but it could be allowed to change as well, provided the values matched conditions for superexogeneity.

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 3 reports the empirical rejection frequencies where the horizontal axis corresponds to values of δ, again for $\alpha_1 = 0.025$ throughout. For large shifts, 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 superexogeneity when the null is true, but are slightly undersized at $T = 50$ for small shifts, when sometimes no impulses may be retained.

Overall, we conclude that the new tests have appropriate null rejection frequencies for both constant and changing marginal processes, so we turn to their ability to detect failures of superexogeneity. This is a two-stage process: first detect shifts in the marginal, then use those to detect shifts in the conditional. The power properties of impulse saturation in the first stage will matter but should be referred to Hendry and Santos (2006).

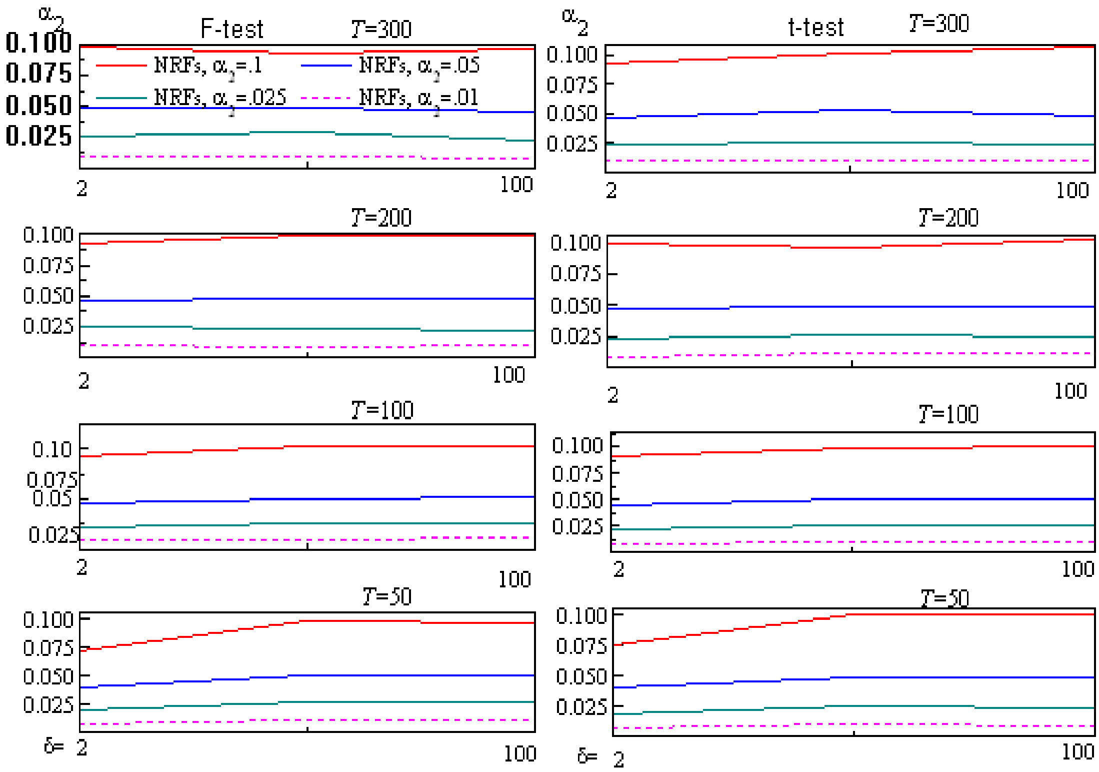

Figure 3: Null rejection frequencies of F and t-tests in the coditional when there is a mean shift in the marginal

6 Three superexogeneity failures

In this section,[2] we derive explicit outcomes for three forms of superexogeneity failure, namely weak exogeneity failure when the marginal process is non-constant in section 6.1; invariance failure in section 6.2; and weak exogeneity failure when the marginal process is constant in section 6.3. 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 (see Hendry and Santos, 2006, for this analysis).

[2]Throughout this section we shall often use the approximation $\mathsf{E}\left[\hat{\beta}\right] \simeq \left[\mathsf{E}\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1}\mathsf{E}\left(\mathbf{X}'\mathbf{Y}\right)$. The appendix to this chapter derives the power series expansion that allows for this approximation. This appendix is based on control variate theory (see, *inter alia*, Hendry (1973, 1984) and Hendry and Harrison (1974)).

6.1 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

$$\mathsf{E}[y_t \mid \mathbf{z}_t] = \mu_{1,t} + \boldsymbol{\sigma}_{12}' \boldsymbol{\Sigma}_{22}^{-1} \left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) = \mu_{1,t} + \boldsymbol{\gamma}' \left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \tag{29}$$

with conditional variance

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

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

$$\mu_{1,t} = \boldsymbol{\beta}' \boldsymbol{\mu}_{2,t} \tag{30}$$

Then,

$$y_t = \boldsymbol{\beta}' \mathbf{z}_t + (\boldsymbol{\gamma} - \boldsymbol{\beta})' \left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) + \epsilon_t \tag{31}$$

where $\epsilon_t = y_t - \mathsf{E}[y_t|\mathbf{z}_t]$ given (30), so $\mathsf{E}[\epsilon_t|\mathbf{z}_t] = 0$. Such a model is a possible example of the Lucas (1976) critique where the agents' behavioural rule depends on $\mathsf{E}[\mathbf{z}_t]$ as in (30), whereas the econometric equation uses $\mathbf{z}_t$, leading to (31).

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

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_n \left[\begin{pmatrix} \boldsymbol{\beta}' \boldsymbol{\mu}_{2,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \tag{32}$$

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

$$\mathbf{z}_t = \boldsymbol{\lambda} 1_{\{t>T_1\}} + \mathbf{v}_{2,t} \tag{33}$$

so $\mathsf{E}[\mathbf{z}_t] = \boldsymbol{\lambda} 1_{\{t>T_1\}} = \boldsymbol{\mu}_{2,t}$. 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 (33) which may affect one or more $\mathbf{z}_t$s. The relevant moments of the joint process are

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

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

$$y_t = \kappa_0 + \boldsymbol{\kappa}_1' \mathbf{z}_t + u_t \tag{34}$$

where $\mathsf{E}[\mathbf{z}_t u_t] = \mathbf{0}$. Then, in (34), letting $(T - T_1)/T = r$:

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

where

$$\mathbf{d}_r = \mathbf{H}_r^{-1}\boldsymbol{\Sigma}_{22}(\boldsymbol{\beta} - \boldsymbol{\gamma}) \tag{35}$$

and $\mathbf{H}_r^{-1}$ is the inverse of the determinant. Consequently,

$$y_t = \kappa_0 + \boldsymbol{\kappa}_1'\mathbf{z}_t + u_t = r\boldsymbol{\lambda}'\mathbf{d}_r + (\boldsymbol{\beta} - \mathbf{d}_r)'\mathbf{z}_t + u_t = \boldsymbol{\beta}'\mathbf{z}_t - \mathbf{d}_r'(\mathbf{z}_t - r\boldsymbol{\lambda}) + u_t \tag{36}$$

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

$$\mathbf{z}_t = \sum_{i=T_1+1}^{T}\widehat{\boldsymbol{\rho}}_{i,(\alpha_1)}1_{\{t=t_i\}} + \mathbf{v}_{2,t}^*$$

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

$$\widehat{\boldsymbol{\rho}}_{i,(\alpha_1)} = \boldsymbol{\lambda} + \mathbf{v}_{2,t_i} \tag{37}$$

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 $\widehat{\boldsymbol{\rho}}_{i,(\alpha_1)}$ may be omitted.

Recording which impulses matter, and adding these to (34) given (39), yields the full-sample regression (considering first the case where all relevant impulses were detected in the marginal model):

$$y_t = \tau_0 + \boldsymbol{\tau}_1'\mathbf{z}_t + \sum_{i=T_1+1}^{T}\delta_{i,(\alpha_2)}1_{\{t=t_i\}} + e_t \tag{38}$$

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

$$y_t = \varsigma_{0,t} + \boldsymbol{\varsigma}_{1,t}'\mathbf{z}_t + e_t \tag{39}$$

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

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

$$= \begin{pmatrix} \boldsymbol{\lambda}'\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) 1_{\{t>T_1\}} \\ \boldsymbol{\gamma} \end{pmatrix}$$

This suggests the model:

$$y_t = \boldsymbol{\gamma}'\mathbf{z}_t + \boldsymbol{\lambda}'\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) 1_{\{t>T_1\}} + e_t \tag{40}$$

matching (36), so that adding all the indicators selected from the marginal model should substantively improve the fit when $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$. Indeed, (40) coincides with the DGP here, so $\{e_t\}$ is an innovation process.

The power of the F-test of

$$\mathsf{H}_0\colon \delta_{i,\alpha_2} = 0 \; \forall i,$$

in (38) by an $\mathsf{F}^{T-T_1}_{T-T_1-2}$ depends on the strength of the superexogeneity violation, $|\boldsymbol{\beta} - \boldsymbol{\gamma}|$, the magnitudes of the breaks, $\boldsymbol{\lambda}$, the sample size T, the relative number of periods affected by the break, and on α_2.

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 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 an automatically computed test as

$$y_t = \tau_0 + \boldsymbol{\tau}'\mathbf{z}_t + \tau_{2,(\alpha_2)}\iota_{1,t} + \tau_{3,(\alpha_2)}\iota_{2,t} + e_t \tag{41}$$

6.1.1 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, $\mathsf{H}_0\colon \tau_2 = 0$ in

$$y_t = \boldsymbol{\tau}_1'\mathbf{z}_t + \tau_{2,(\alpha_2)} 1_{\{t>T_1\}} + u_t \tag{42}$$

where the DGP is (40) written as

$$y_t = \boldsymbol{\gamma}'\mathbf{z}_t + \left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right)' \boldsymbol{\lambda} 1_{\{t>T_1\}} + e_t \tag{43}$$

Since (43) is correctly specified, $\boldsymbol{\gamma}$ and $(\boldsymbol{\beta}-\boldsymbol{\gamma})'\boldsymbol{\lambda}$ are consistently estimated with

$$\mathsf{V}\left[\begin{pmatrix} \widehat{\boldsymbol{\lambda}'(\boldsymbol{\beta}-\boldsymbol{\gamma})} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix}\right] \simeq \sigma_e^2 \begin{pmatrix} \sum_{t=1}^{T} & \iota_{1,t}^2 & \mathsf{E}\left[\mathbf{z}_t'\iota_{1,t}\right] \\ & \mathsf{E}\left[\mathbf{z}_t\iota_{1,t}\right] & \mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] \end{pmatrix}^{-1}$$

$$= \frac{\sigma_e^2}{T} \begin{pmatrix} T & r^{-1} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda} & -\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1} \\ & -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix} \tag{44}$$

The power depends on $\boldsymbol{\lambda}$, r, T, σ_e, α_2, as well as on the departure between $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ induced by the failure of superexogeneity. 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

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

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 (42) is

$$\mathsf{E}\left[\mathsf{t}_{\tau_2,(\alpha_2)}(\varphi_{r,\alpha_2})\right] = \frac{(\boldsymbol{\beta}-\boldsymbol{\gamma})'\boldsymbol{\lambda}\sqrt{Tr}}{\sigma_e\sqrt{1+r\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda}}} = \frac{\sqrt{T}\left(\boldsymbol{\lambda}^*\right)'\mathbf{K}^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\gamma}\right)}{\sqrt{(\sigma_{11}-\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma})}\sqrt{1+\boldsymbol{\lambda}^*\left(\boldsymbol{\lambda}^*\right)'}} = \varphi_{r,\alpha_2} \tag{45}$$

The non-centrality φ_{r,α_2} in (45) 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, φ_{r,α_2} 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 $\mathsf{t}^2_{\tau_2,(\alpha_2)}(\varphi^2_{r,\alpha_2})$ by a chi-squared with 1 degree of freedom discussed in Hendry (1995), with $\mathsf{t}^2_{\tau_2,(\alpha_2)}(\varphi^2_{r,\alpha_2}) \sim \chi_1^2\left(\varphi^2_{r,\alpha_2}\right)$, and $\chi_1^2\left(\varphi^2_{r,\alpha_2}\right) \backsimeq h\chi_m^2(0)$. Then, $\mathsf{P}\left[\chi_1^2\left(\varphi^2_{r,\alpha_2}\right) > c_{\alpha_2}|\mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi_m^2(0) > h^{-1}c_{\alpha_2}\right]$. For example, when $\varphi^2_{r,\alpha_2} = 5$ for $c_{\alpha_2} = 4$, then $h = 51/26 \simeq 2$ and $m = 13$ with $\mathsf{P}\left[\chi_{13}^2(0) > 2\right] \simeq 0.9998$.

Finally, φ^2_{r,α_2} should also be the non-centrality of the corresponding F-test. However, the power may not be monotonic in the arguments of φ^2_{r,α_2}, 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 $\mathsf{F}^{T-T_1}_{T-T_1-2}\left(\varphi_{r,\alpha_2}\right)$ by its numerator $\chi^2_{k^*}\left(\varphi_{r,\alpha_2}\right)$ and that in turn using the more general formulae for $k^* = T - T_1 = Tr$. Then,

$$\mathsf{P}\left[\chi^2_{Tr}\left(\varphi_{r,\alpha_2}\right) > c_{\alpha_2} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m(0) > h^{-1}c_{\alpha_2}\right] \tag{46}$$

where

$$h = \frac{Tr + 2\varphi^2_{r,\alpha_2}}{Tr + \varphi^2_{r,\alpha_2}} \quad \text{and} \quad m = \frac{Tr + \varphi^2_{r,\alpha_2}}{h} \tag{47}$$

6.1.2 Allowing for stage 1

The above results are conditional on keeping all and only the relevant impulses from the marginal, but the analysis in Hendry and Santos (2006) revealed that was itself dependent on the parameters of the marginal DGPs. Nevertheless, we can extend the analysis 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 (40) when the DGP is (43), we have

$$\begin{aligned}
\mathsf{E}\left[\begin{array}{c} \widehat{\lambda(\beta-\gamma)} \\ \widehat{\gamma} \end{array}\right] &\simeq \left(\sum_{t=1}^{T} \begin{array}{cc} \mathsf{E}\left[\iota^2_{1,t}\right] & \mathsf{E}\left[z_t\iota_{1,t}\right] \\ \mathsf{E}\left[z_t\iota_{1,t}\right] & \mathsf{E}\left[z^2_t\right] \end{array}\right)^{-1} \left(\sum_{t=1}^{T} \begin{array}{c} \mathsf{E}\left[y_t\iota_{1,t}\right] \\ \mathsf{E}\left[y_t z_t\right] \end{array}\right) \\
&= \left(\begin{array}{cc} \mathsf{p}_d r & \lambda\mathsf{p}_d r \\ \lambda\mathsf{p}_d r & \lambda^2 r + \sigma_{22} \end{array}\right)^{-1} \left(\begin{array}{c} \beta\lambda\mathsf{p}_d r \\ \beta\lambda^2 r + \gamma\sigma_{22} \end{array}\right) \\
&= \left(\begin{array}{c} \lambda(\beta-\gamma) \\ \gamma \end{array}\right) + \frac{(\beta-\gamma)\lambda^2 r(1-\mathsf{p}_d)}{\lambda^2 r(1-\mathsf{p}_d)+\sigma_{22}} \left(\begin{array}{c} -\lambda \\ 1 \end{array}\right)
\end{aligned} \tag{48}$$

where p_d is the probability of retaining an impulse in the impulse saturated marginal (see Hendry and Santos, 2006). Comparing (48) with the consistent estimates and their variances in (44), 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

$$y_t = \kappa^*_0 z_t + \kappa^*_{1,(\alpha_2)}\iota_{1,t} + u_t \tag{49}$$

leads to a modified non-centrality corresponding to (45) when $n = 2$ but for (49), namely

$$\mathsf{E}\left[\mathsf{t}_{\kappa^*_1,(\alpha_2)}\left(\varphi^*_{r,\alpha_2}\right)\right] = \frac{\sqrt{Tr\mathsf{p}_d}\,(\beta-\gamma)\lambda}{\sigma_u\sqrt{\left(1+\lambda^2 r\sigma^{-1}_{22}\right)}} = \varphi^*_{r,\alpha_2} \tag{50}$$

so

$$\mathsf{E}\left[\mathsf{t}^2_{\kappa^*_1,(\alpha_2)}\left(\varphi^{*2}_{r,\alpha_2}\right)\right] = \frac{Tr\mathsf{p}_d(\beta-\gamma)^2\lambda^2}{\sigma^2_u\left(1+\lambda^2 r\sigma^{-1}_{22}\right)} = \frac{\mathsf{p}_d\sigma^2_e\varphi^2_{r,\alpha}}{\sigma^2_u} = \varphi^{*2}_{r,\alpha_2}$$

where

$$\sigma^2_u = \sigma^2_e + \sigma_{22}\frac{(1-\mathsf{p}_d)\lambda^2(\beta-\gamma)^2}{\lambda^2 r(1-\mathsf{p}_d)+\sigma_{22}} \tag{51}$$

Thus, the power falls directly because $\mathsf{p}_d < 1$ and indirectly because $\sigma^2_u > \sigma^2_e$. For example, combining the parameter values for the tests just above with a location shift that delivers[3] $\mathsf{p}_d = 0.16$ (where $\lambda^2 = \sigma_{22} = 5$, $\beta - \gamma = 0.25$,

[3] See the analysis in Hendry and Santos (2006) and Santos (2006).

$Tr = 20$, $\sigma_e^2 = 1$ so $\sigma_u^2 = 1.23$), $\mathsf{E}\left[\mathsf{t}^2_{\kappa_1^*,(\alpha_2)}\left(\varphi^{*2}_{r,\alpha_2}\right)\right] = 0.65$, 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 $\mathsf{E}[\mathsf{t}^2_{\kappa_1^*,(\alpha_2)}\left(\varphi^{*2}_{r,\alpha_2}\right)]$ to 11.4, so the power rises quickly towards the maximum, essentially reaching that bound by $\lambda = 4\sqrt{\sigma_{22}}$. Notice from (51) 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.

6.2 Invariance failure

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 WE, in that $\boldsymbol{\gamma} = \boldsymbol{\beta}$, but the cross-link between the means violates superexogeneity, namely $\boldsymbol{\mu}_{1,t} = \boldsymbol{\beta}_t'\boldsymbol{\mu}_{2,t}$ when

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 1_{\{t>T_1\}} \tag{52}$$

where

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{N}_n \left[\begin{pmatrix} \boldsymbol{\mu}_{1,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega^2 + \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_t & \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_t & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \tag{53}$$

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

$$\mathsf{E}[y_t \mid \mathbf{z}_t] = \boldsymbol{\beta}_t'\boldsymbol{\mu}_{2,t} + \boldsymbol{\beta}_t'\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) = \boldsymbol{\beta}_0'\mathbf{z}_t + \boldsymbol{\beta}_1'\mathbf{z}_t 1_{\{t>T_1\}} \tag{54}$$

The marginal model is the same as in the previous section, so $\widehat{\boldsymbol{\rho}}_{i,(\alpha_1)} = \boldsymbol{\lambda} + \mathbf{v}_{2,t_i}$ from (37), 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 (34), but the data moments differ for the changed DGP:

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

Hence, the implicit full-sample parameters of (34) become

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

which is similar in form to (35), and simplifies to the vector $(0:\boldsymbol{\beta}_0+\boldsymbol{\beta}_1)'$ when $\boldsymbol{\lambda}=\mathbf{0}$.

The 'instantaneous' relation is again given by

$$y_t = \varsigma_{0,t} + \boldsymbol{\varsigma}_{1,t}'\mathbf{z}_t + e_t \tag{55}$$

where

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

and

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

so as expected

$$y_t = \boldsymbol{\beta}_0'\mathbf{z}_t + \boldsymbol{\beta}_1'\mathbf{z}_t 1_{\{t>T_1\}} + e_t = \boldsymbol{\tau}_1'\mathbf{z}_t + \boldsymbol{\tau}_2'\mathbf{z}_t 1_{\{t>T_1\}} + e_t \tag{56}$$

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 superexogeneity. 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 indices, $\iota_{1,t} = 1_{\{t>T_1\}}$ and $\boldsymbol{\iota}_{2,t} = 1_{\{t>T_1\}}\mathbf{z}_t$, could be calculated and added.

6.2.1 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 (56), which is the model analogue of (54), 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 $\boldsymbol{\iota}_{2,t}$:

$$y_t = \boldsymbol{\tau}_1'\mathbf{z}_t + \boldsymbol{\tau}_{2,(\alpha_2)}'\boldsymbol{\iota}_{2,t} + u_t \tag{57}$$

The variances of the parameter estimates from (57) are approximately

$$\mathsf{V}\left[\begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1 \\ \widetilde{\boldsymbol{\tau}}_2 \end{pmatrix}\right] \simeq \sigma_e^2 \left[\sum_{t=1}^{T} \begin{pmatrix} \mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] & \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}'\right] \\ \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}\right] & \mathsf{E}\left[\boldsymbol{\iota}_{2,t}\boldsymbol{\iota}_{2,t}'\right] \end{pmatrix}\right]^{-1}$$

$$= \frac{\sigma_e^2}{T}\left[\begin{pmatrix} \left(\boldsymbol{\lambda\lambda}'r + \boldsymbol{\Sigma}_{22}\right) & \left(\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)r \\ \left(\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)r & \left[\left(\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)r\right] \end{pmatrix}\right]^{-1}$$

$$= \frac{\sigma_e^2}{Tr\left(1-r\right)}\begin{pmatrix} r\boldsymbol{\Sigma}_{22}^{-1} & -r\boldsymbol{\Sigma}_{22}^{-1} \\ -r\boldsymbol{\Sigma}_{22}^{-1} & \left((1-r)\left(\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1} + r\boldsymbol{\Sigma}_{22}^{-1}\right) \end{pmatrix}$$

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

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

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

$$\mathsf{E}\left[\mathsf{F}_{\tau_2,(\alpha_2)}\left(\phi_{r,\alpha_2}^2\right)\right] = (T-2n)\,r\,(1-r)\,\frac{\boldsymbol{\beta}_1'\left[\left(\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)\left(r\boldsymbol{\lambda\lambda}' + \boldsymbol{\Sigma}_{22}\right)^{-1}\boldsymbol{\Sigma}_{22}\right]\boldsymbol{\beta}_1}{\sigma_e^2\,(n-1)}$$

$$= (T-2n)\,(1-r)\,\frac{\boldsymbol{\beta}_1'\left(\mathbf{K}'\right)^{-1}\left(\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\prime} + r\mathbf{I}\right)\left(\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\prime} + \mathbf{I}\right)^{-1}\mathbf{K}^{-1}\boldsymbol{\beta}_1}{\omega^2\,(n-1)} = \phi_{r,\alpha_2}^2$$

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

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

$$\mathsf{E}\left[\mathsf{t}_{\tau_2,(\alpha_2)}\left(\phi_{r,\alpha_2}\right)\right] = \frac{\beta_1\sqrt{T\left(\lambda^2 + \sigma_{22}\right)r\,(1-r)\,\sigma_{22}}}{\sigma_e\sqrt{\lambda^2 r + \sigma_{22}}} = \frac{\sqrt{T\,(1-r)}\beta_1\sqrt{r + (\lambda^*)^2}}{\omega^*\sqrt{1 + (\lambda^*)^2}} = \phi_{r,\alpha_2}$$

with $\omega^* = \omega/\sqrt{\sigma_{22}}$. Again, ϕ_{r,α_2} is monotonically increasing in $\boldsymbol{\lambda}^*$, in β_1 and in r for fixed λ; and because of the form of (52), $\phi_{r,\alpha_2} \neq 0$ even if $\lambda^* = 0$.

Then, ϕ_{r,α_2}^2 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, pp. 190) that:

$$\mathsf{E}\left[\mathsf{F}_{k_2}^{k_1}\left(\phi_{r,\alpha_2}^2\right)\right] = \frac{k_2\left(k_1 + \phi_{r,\alpha_2}^2\right)}{k_1\left(k_2 - 2\right)} \tag{58}$$

6.2.2 Allowing for stage 1 effects

Returning to (57), where $\boldsymbol{\iota}_{2,t}$ reflects the power of the stage 1 selection of impulses, the estimators become

$$\mathsf{E}\left[\begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1 \\ \widetilde{\boldsymbol{\tau}}_2 \end{pmatrix}\right] \simeq \left[\sum_{t=1}^{T} \begin{pmatrix} \mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] & \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}'\right] \\ \mathsf{E}\left[\mathbf{z}_t\boldsymbol{\iota}_{2,t}\right] & \mathsf{E}\left[\boldsymbol{\iota}_{2,t}\boldsymbol{\iota}_{2,t}'\right] \end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix} \mathsf{E}\left[\mathbf{z}_t y_t\right] \\ \mathsf{E}\left[\boldsymbol{\iota}_{2,t} y_t\right] \end{pmatrix}\right]$$

$$= \begin{pmatrix} \boldsymbol{\beta}_0 - \left(\mathbf{I} - \mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\,(1-r)\right)\boldsymbol{\beta}_1 \\ \mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\,(1-r)\,\boldsymbol{\beta}_1 \end{pmatrix} \tag{59}$$

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 $\mathbf{R} = \boldsymbol{\Sigma}_{22}(1-r)$. From (59):

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

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

$$\mathsf{E}\left[\mathsf{F}_{\tau_2,(\alpha_2)}\left(\phi^{*2}_{r,\alpha_2}\right)\right] = (T-2n)(1-r)^2 \frac{\boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{22}\mathbf{R}^{-1}\left(\left[\left(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Sigma}_{22}\right)\mathsf{p}_d r\right]^{-1} + \mathbf{R}^{-1}\right)^{-1}\mathbf{R}^{-1}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_1}{\sigma_u^2(n-1)}$$

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

$$\begin{aligned}
\mathsf{E}\left[\mathsf{t}^2_{\tau_2,(\alpha_2)}\left(\phi^{*2}_{r,\alpha_2}\right)\right] &\simeq \frac{T\mathsf{p}_d\beta_1^2\left(\lambda^2+\sigma_{22}\right)r(1-r)^2\sigma_{22}^2}{\sigma_u^2\left(\lambda^2 r+\sigma_{22}\right)\left[\left(\lambda^2+\sigma_{22}\right)r\left(1-\mathsf{p}_d\right)+\sigma_{22}(1-r)\right]} \\
&= \frac{\mathsf{p}_d\sigma_{22}(1-r)\sigma_e^2}{\sigma_u^2\left[\left(\lambda^2+\sigma_{22}\right)r\left(1-\mathsf{p}_d\right)+\sigma_{22}(1-r)\right]}\phi^2_{r,\alpha_2}
\end{aligned}$$

6.2.3 Asymptotic power of the incorrect index invariance test

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

$$y_t = \left(\boldsymbol{\tau}_1^*\right)'\mathbf{z}_t + \tau^*_{2,(\alpha_2)}\iota_{1,t} + e_t^* \tag{60}$$

with average estimated parameters:

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

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 $r \neq 0$:

$$\mathsf{E}\left[\mathsf{t}^2_{\tau_2^*,(\alpha_2)}\left(\psi^2_{r,\alpha_2}\right)\right] = \frac{\sqrt{T}r\boldsymbol{\lambda}'\boldsymbol{\beta}_1(1-r)}{\sigma_{e^*}\sqrt{\left(1+r\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\lambda}\right)}} = \frac{\sqrt{T}\boldsymbol{\lambda}^{*\prime}\mathbf{K}^{-1}\boldsymbol{\beta}_1(1-r)}{\sqrt{\omega^2+\boldsymbol{\beta}_1'\left(\mathbf{K}'\right)^{-1}\mathbf{K}^{-1}\boldsymbol{\beta}_1 r(1-r)}\sqrt{\left(1+\boldsymbol{\lambda}^{*\prime}\boldsymbol{\lambda}^*\right)}} = \psi^2_{r,\alpha_2}$$

noting that

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

so

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

Thus, $\psi_{r,\alpha}$ is again monotonic in $\boldsymbol{\lambda}^*$, but need not be monotonic in r for fixed $\boldsymbol{\lambda}$. Also, $\mathsf{t}^2_{\tau_2^*,(\alpha_2)}\left(\psi^2_{r,\alpha_2}\right)$ is less powerful than $\mathsf{t}^2_{\tau_2,(\alpha_2)}\left(\phi^2_{r,\alpha_2}\right)$, as $\phi^2_{r,\alpha_2} > \psi^2_{r,\alpha_2}$. Thus, ϕ^2_{r,α_2}, the non-centrality of the F-test, which is applicable in the present setting, has an important invariance to the source of the superexogeneity failure, and should exceed ψ^2_{r,α_2}.

6.3 Weak exogeneity failure under constancy

Reconsider the bivariate example in (31) above, but where all parameters are constant, so

$$y_t = \boldsymbol{\beta}'\mathbf{z}_t + e_t = \boldsymbol{\beta}'\mathbf{z}_t - \left(\boldsymbol{\beta}-\boldsymbol{\gamma}_2\right)'\left(\mathbf{z}_t-\boldsymbol{\mu}_2\right)+\eta_t \qquad (61)$$

with

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

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

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

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

$$y_t = \tau_0 + \boldsymbol{\tau}_1'\mathbf{z}_t + u_t \qquad (62)$$

then,

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

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 (62).

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

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

Thus,

$$\begin{aligned} e_t &= y_t - \tau_0 - \boldsymbol{\tau}_1' \mathbf{z}_t - \tau_2 w_t \\ &= \boldsymbol{\beta}' \boldsymbol{\mu}_2 + \boldsymbol{\gamma}_2' (\mathbf{z}_t - \boldsymbol{\mu}_2) + \eta_t - (\boldsymbol{\beta} - \boldsymbol{\gamma}_2)' \boldsymbol{\mu}_2 - \boldsymbol{\gamma}_2' \mathbf{z}_t - \tau_2 \sum_{i=1}^{\alpha T} \widehat{\varphi}_{i,(\alpha)} 1_{\{t=t_i\}} \\ &= -\tau_2 \sum_{i=1}^{\alpha T} v_{2,t_i} 1_{\{t=t_i\}} + \eta_t \end{aligned} \tag{63}$$

Since the largest of the v_{2,t_i} in (63) 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.

7 Simulating the powers of automatic superexogeneity tests

We have undertaken simulation analyses for all three scenarios for a bivariate relation.[4] The null DGP is given by (21), implying $\gamma = 2$.

7.1 Failure of weak exogeneity under non-constancy

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

$$\mu_{2,t} = \lambda 1_{\{t \geq T_1\}} + \mu_{2,0} \tag{64}$$

and so

$$\mu_{1,t} = \beta\lambda 1_{\{t \geq T_1\}} + \beta\mu_{2,0} = \beta\lambda 1_{\{t \geq T_1\}} + \mu_{1,0} \tag{65}$$

Hence $\beta\lambda$ is the level shift of y_t at T_1. We allow $\beta\lambda$ to vary across Monte Carlo experiments, obtaining 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 WE: in particular, β takes the values 0.75, 1, 1.5 and 1.75, where the first represents the strongest departure from weak exogeneity, and the last represents the weakest. Finally, we also consider different sample sizes ($T = 100$ and $T = 300$) and different break points T_1. 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 begin by investigating the empirical power of the F-test. Table (1) reports the empirical mean rejection frequencies of the null in the joint F-test when a

[4] Monte Carlo results for a trivariate relation are available upon request. They lead to the same conclusions as the ones below.

sample size of $T = 300$ is used and 0.05 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, the power is also 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 φ^2_{r,α_2} in this bivariate case, from (50), is

$$\varphi^2_{r,\alpha_2} = \frac{k^* \mathsf{p}_d (\beta - \gamma)^2 d^2 \sigma_{22}}{\sigma^2_u (1 + d^2 r)} \tag{66}$$

with power $\mathsf{p}_\alpha = \mathsf{P}\left[\chi^2_m (0) > h^{-1} c_{\alpha_2}\right]$ where

$$h = \frac{k^* + 2\varphi^2_{r,\alpha_2}}{k^* + \varphi^2_{r,\alpha_2}} \quad \text{and} \quad m = \frac{k^* + \varphi^2_{r,\alpha_2}}{h} \tag{67}$$

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: WE failure with level shift in the marginal: $T_1 = 251$, $T = 300$, $\alpha_1 = \alpha_2 = 0.05$, F-test

In table (2), we investigate the impact of reducing the length of the second regime to $k^* = 25$. All the other defaults from the previous experiments apply.

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: WE failure with level shift in the marginal: $T_1 = 276$, $T = 300$, $\alpha_1 = \alpha_2 = 0.05$, F-test

Empirical power is never smaller when the break length diminishes: the degrees of freedom of the F-test must be playing a fundamental role here, as suggested in section 6.1.1. This is partly the motivation to look into the second class of superexogeneity tests: those based on an index replacing the indicators.

We now turn to investigate the effect on power of the sample size. Table (3) reports Monte Carlo results obtained using $T = 100$. 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 3: WE failure with level shift in the marginal: $T_1 = 81$, $T = 100$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, F-test

It is worth noticing that, 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 (empirical power is acceptable even for $\beta = 1.5$ for a break of at least 2.5σ).

The trade-off between length of the break and power is also a feature of smaller sample sizes, as illustrated in table (5), where a break of length $k^* = 30$ is assumed to begin at observation 71. The results for very small level shifts are negligible. Comparing tables (3) and (5), the increase in the length of the break reduces its empirical power.

Table (4) investigates the use of a 0.05 significance level in the marginal whilst a 0.1 significance level is used in the conditional. Clearly, the empirical rejection frequencies are never smaller that those in table (1).

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: WE failure with level shift in the marginal: $T_1 = 251$, $T = 300$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, F-test

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 5: WE failure with level shift in the marginal: $T_1 = 71$, $T = 100$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, F-test

With respect to the single-index test, table (6) reports results for a sample size of $T = 100$ ($\alpha_1 = 0.05$ and $\alpha_2 = 0.1$). The shift occurs at observation 81, implying $k^* = 20$.

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 6: WE failure with level shift in the marginal: $T_1 = 81$, $T = 100$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, index test

Comparing tables (6) and (3), although the index-based test has higher empirical power for small shifts (magnitudes σ and 2σ), the joint F-test does generally better for shifts of higher magnitudes. No test dominates the other, for the defaults used in this experiment.

Values for the empirical power of the index-based test, in table (6), are reasonable. The empirical power is decreasing as β gets closer to $\gamma = 2$, as expected. Furthermore, 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 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 (7) extends the analysis by considering $T = 300, \alpha_1 = 0.05$, $\alpha_2 = 0.1$, 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 7: WE failure with level shift in the marginal, $T_1 = 251$, $T = 300$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, index test

For this larger sample size, and these significance levels, the joint F-test dominates the index based test in terms of power. The only exceptions occur for some intermediate magnitudes, when $\beta = 1.75$.

Tables (8) and (9) refer to the double-index test. Confronting with (6) and (7), respectively, we conclude that for this type of failure of superexogeneity the single-index test dominates the double-index.

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 8: WE failure with level shift in the marginal: $T_1 = 81$, $T = 100$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, double-index test

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 9: WE failure with level shift in the marginal: $T_1 = 251$, $T = 300$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, double-index test

7.2 Failure of invariance when weak exogeneity holds

We now consider a DGP where the null hypothesis of superexogeneity is false, but weak exogeneity holds (that is: $\beta_t = \gamma_t, \forall t$). Let T_1 be such that $1 < T_1 < T$ and, for $t < T_1$, let the DGP be given by:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{68}$$

whilst for $t \geq T_1$,

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} 3\mu_2^* \\ \mu_2^* \end{pmatrix}, \begin{pmatrix} 30 & 9 \\ 9 & 3 \end{pmatrix} \right] \tag{69}$$

$\beta_t = \gamma_t$ 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 vector of parameters $\phi_{1,t}$, which contains γ_t, is not invariant to changes in the parameters' vector of the marginal model $\phi_{2,t}$, which contains $\sigma_{22,t}$.

For the Monte Carlo experiments, we work with the same settings as in previous subsection. 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.

Tables (10) and (11) refer to the joint F-test. Table (10) reports Monte Carlo results for a sample size of $T = 100$, with $\alpha_1 = 0.05$ and $\alpha_2 = 0.1$. The break dates are: $T_1 = 81$, $T_1 = 71$ and $T_1 = 61$ for $k^* = 20,\ 30,\ 40$, respectively.

Table (10) reveals that the test has good power even for a small sample. An increase in the length of the break period from $k^* = 20$ to $k^* = 30$ augments the

$T = 100$	$k^* = 20$	$k^* = 30$	$k^* = 40$
$\mu_2^* = 2.0$	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 10: Invariance failure: $T = 100$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, F-test

rejection frequency of the null, nearly always. Nonetheless, a further increase of equal absolute magnitude in the length of the break can reduce power (for greater level shifts). There is no monotonicity property.

In table (11), we consider a sample of size $T = 300$. We consider breaks at observations $T_1 = 261, T_1 = 251, T_1 = 201$ and $T_1 = 161$, matching respectively, $k^* = 40,\ 50,\ 100$ and 140.

$T = 300$	$k^* = 40$	$k^* = 50$	$k^* = 100$	$k^* = 140$
$\mu_2^* = 2.0$	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 11: Invariance failure: $T = 300$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, F-test

There is good power against failure of superexogeneity, for all break lengths, even for the smallest mean shifts considered. Notwithstanding, table (11) also highlights the problem we had just discussed for smaller sample sizes: the length of the break may adversely affect the power to detect departures from superexogeneity. In the case discussed in table (11), this is clear for $k^* = 140$.

We now address the index-based test. We consider only the case where $T = 300$. The same defaults as in the previous experiment are used. Table (12) reports the results.

$T = 300$	$k^* = 40$	$k^* = 50$	$k^* = 100$	$k^* = 140$
$\mu_2^* = 2.0$	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 12: Invariance failure: $T = 300$, $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$, index test

Comparison of tables (12) and (11) shows that the joint F-test dominates the index-based test (with the exception of the two largest unconditional mean shifts for $k^* = 140$).

However, in spite of the power dominance of the joint F-test over the index test, it is also true that power increases monotonically both with the mean shift and with the break length, for the index test.

It remains to investigate the effects on power of using a model with two indices. Adding the two indices to the conditional, with parameters φ_1 and φ_2, the null hypothesis is

$$\mathsf{H}_0 : \varphi_1 = \varphi_2 = 0 \tag{70}$$

We consider the same departures from invariance as for the previous cases in this section, and the same break lengths. Results are reported in tables (13) and (14) 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 13: Invariance failure: $T = 300$, $\alpha_1 = \alpha_2 = 0.025$, double-index test

$T = 100$	$k^* = 20$	$k^* = 30$	$k^* = 40$
$\mu_2^* = 2.0$	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 14: Invariance failure: $T = 100$, $\alpha_1 = \alpha_2 = 0.025$, double-index test

Tables (13) and (14) should be compared with tables (15) and (16), 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 0.025 are used in the marginal and 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 15: Invariance failure: $T = 300$, $\alpha_1 = \alpha_2 = 0.025$, index test

The double-index test has greater power to detect departures of superexogeneity for invariance failure, than the corresponding single-index tests. This claim is valid, irrespective of the choice of significance levels[5]. This dominance of the double-index test was to be expected on the basis of theory results from section 6.2.

[5] Complete Monte Carlo results for this comparison are not reported here but are available on request.

$T = 100$	$k^* = 20$	$k^* = 30$	$k^* = 40$
$\mu_2^* = 2.0$	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 16: Invariance failure: $T = 100$, $\alpha_1 = \alpha_2 = 0.025$, index test

7.3 Failure of weak exogeneity under constancy

Finally, we consider a departure from superexogeneity 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:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{N}_2 \left[\begin{pmatrix} \beta^* \\ 1 \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{71}$$

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 (17) reports the results for sample sizes of $T = 100$, 200 and 300.

In table (17), 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 $\alpha_2 = 0.1$.

	$T = 100$	$T = 200$	$T = 300$
$\beta^* = 0.50$	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 17: Failure of weak exogeneity under constancy: $\alpha_1 = 0.05 \wedge \alpha_2 = 0.1$

As expected, the test has virtually no power against this form of failure of the superexogeneity 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.

8 Conclusion

The concept of automatically computable tests for superexogeneity based on selecting from impulse saturation of the marginal process to test the conditional is realizable. The tests proposed here have the correct null rejection frequency in constant conditional models when the nominal NRF is not too small in the marginal at small sample sizes (e.g. 0.05), for a variety of marginal processes, both constant and with breaks. The tests also have power against failures of superexogeneity when either invariance or weak exogeneity fails and the marginal process changes. Neither class of tests uniformly dominates the other. Their theoretical powers were derived analytically for regression models and explain the simulation outcomes well.

Results of some pilot experiments not reported here suggest very good power against failures of weak exogeneity with variance breaks in the marginal. However, this set of results is still rather incomplete and lacks proper theory analysis, so we have chosen not to include it in this paper.

References

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

Chow, G. C. (1960) Tests for Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica*, 28, 591-605.

Doornik, J. A. (2001) *OX, An Object Oriented Matrix Programming Language*, 4th edition, 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.

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

Ericsson, N. R. and Irons, J. S. (1994) The Lucas Critique in Practice: Theory without Measurement, in Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, Chapter 8, 263-312 (with discussion), Boston, Massachussets: Kluwer Academic Publishers.

Favero, C. A. (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. A. and Hendry, D. F. (1992) Testing the Lucas Critique: a Review, *Econometric Reviews*, 11 (3), 265-306.

Hannan, E. J. (1970) *Multiple Time Series,* New York: John Wiley.

Hendry, D. F. (1973) On Asymptotic Theory and Finite Sample Experiments, *Economica*, 40, 210-217.

Hendry, D. F. (1984) Monte Carlo Experimemtation in Econometrics, in Griliches, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Volume 2, 652-698, Amsterdam: North-Holland.

Hendry, D. F. (1988) The Encompassing Implications of Feedback versus Feedforward Mechanisms in Econometrics, *Oxford Economic Papers*, 40, 132-149.

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

Hendry, D. F. and Harrison, R. W. (1974) Monte Carlo Methodology and the small-sample Behaviour of Ordinary and Two-Stage Least-Squares, *Journal* *of Econometrics*, 2, 151-74.

Hendry, D. F. and Krolzig, H.-M. (1999) Improving on "Data-Mining Reconsidered" by K. D. Hoover and S. J. Perez, *Econometrics Journal*, 2, 202-219.

Hendry, D. F. and Krolzig, H.-M. (2001) *Automatic Econometric Model Selection using PcGets*, London: Timberlake Consultants Press.

Hendry, D. F. and Massman, M. (2005) Co-breaking: Recent Advances and a Synopsis of the Literature, *Journal of Business and Economic Statistics*, Special Issue (forthcoming).

Hendry, D. F. and Santos, C. (2005) Regression Models with Data-Based Indicator Variables, *Oxford Bulletin of Economics and Statistics*, 67, 571-595.

Hendry, D. F. and Santos, C. (2006) On the Power of Impulse Saturation Break Tests, Unpublished paper, Department of Economics, University of Oxford.

Hendry, D. F., Johansen, S. and Santos, C. (2005) Selecting a Regression Saturated with Indicators, Unpublished Paper, Department of Economics, University of Oxford.

Jansen, E. S. and Teräsvirta, T. (1996) Testing Parameter Constancy and Super Exogeneity in Econometric Equations, *Oxford Bulletin of Economics and Statistics*, 58, 571-595.

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

Krolzig, H.-M. and Toro, J. (2002) Testing for Super Exogeneity in the Presence of Common Deterministic Shifts, *Annales d'Économie et Statistique*, 67/48, 41-71.

Lucas, R. E. (1976) Econometric Policy Evaluation: a Critique, in Brunner, K. and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conferences on Public Policy, Volume 1, 19-46, Amsterdam: North-Holland.

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. An Independence Test and Conditional Unbiased Predictions in the Context of Simultaneous Equations 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, 393-97.

Santos, C. (2006) Structural Breaks and Outliers in Economic Time Series: Modelling and Inference, D.Phil. Thesis, Department of Economics, University of Oxford.

Stanley, T. D. (2000) An Empirical Critique of the Lucas Critique, *Journal of Socio-Economics*, 29, 91-107.

Appendix: Control Variate Theory and approximating the expectation of the OLS estimator

The approximations are used in control variate theory: see Hendry (1973), Hendry and Harrison (1974) and Hendry (1984).

Let

$$\mathbf{M} = \mathsf{E}\left[T^{-1}\left(\mathbf{X}'\mathbf{X}\right)\right]$$

$$\begin{aligned}
\mathsf{E}\left[\widehat{\boldsymbol{\beta}}\right] &= \mathsf{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] = \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{M} + \left[T^{-1}\mathbf{X}'\mathbf{X} - \mathbf{M}\right]\right)^{-1} T^{-1}\mathbf{X}'\boldsymbol{\epsilon}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{M}\left\{\mathbf{I}_k + \left[\mathbf{M}^{-1}\frac{\mathbf{X}'\mathbf{X}}{T} - \mathbf{I}_k\right]\right\}\right)^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{I}_k + \left[\mathbf{M}^{-1}\frac{\mathbf{X}'\mathbf{X}}{T} - \mathbf{I}_k\right]\right)^{-1}\mathbf{M}^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{I}_k + \boldsymbol{\Delta}\right)^{-1}\mathbf{M}^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right]
\end{aligned}$$

where:

$$\mathbf{M}^{-1}\frac{\mathbf{X}'\mathbf{X}}{T} - \mathbf{I}_k = \boldsymbol{\Delta}$$

with:

$$\mathsf{E}\left[\boldsymbol{\Delta}\right] = \mathbf{0} \text{ and } \operatorname*{plim}_{T\to\infty} \boldsymbol{\Delta} = 0$$

so from Hannan (1970):

$$\boldsymbol{\Delta} = \mathsf{O}_p\left(T^{-1/2}\right),$$

and:

$$\mathsf{E}\left[\mathbf{M}^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] = \mathbf{0} \text{ where } \operatorname*{plim}_{T\to\infty}\mathbf{M}^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T} = \mathbf{0},$$

so:

$$\mathbf{M}^{-1}\frac{\mathbf{X}'\boldsymbol{\epsilon}}{T} = \mathsf{O}_p\left(T^{-1/2}\right)$$

Now expand as a power series:

$$\left(\mathbf{I}_k + \boldsymbol{\Delta}\right)^{-1} = \mathbf{I}_k - \boldsymbol{\Delta} + \mathsf{O}_p\left(T^{-1}\right) = \mathbf{I}_k - \mathsf{O}_p\left(T^{-1/2}\right)$$

so that:

$$
\begin{aligned}
\mathsf{E}\left[\widehat{\boldsymbol{\beta}}\right] &= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{I}_k + \boldsymbol{\Delta}\right)^{-1} \mathbf{M}^{-1} \frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\left(\mathbf{I}_k - \mathsf{O}_p\left(T^{-1/2}\right)\right) \mathbf{M}^{-1} \frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] \\
&= \boldsymbol{\beta} + \mathsf{E}\left[\mathbf{M}^{-1} \frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] + \mathsf{O}_p\left(T^{-1}\right) \\
&\backsimeq \boldsymbol{\beta} + \mathsf{E}\left[\mathbf{M}^{-1} \frac{\mathbf{X}'\boldsymbol{\epsilon}}{T}\right] \\
&= \boldsymbol{\beta} + \left[\mathsf{E}\left[T^{-1}\left(\mathbf{X}'\mathbf{X}\right)\right]\right]^{-1} \mathsf{E}\left[\mathbf{X}'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) T^{-1}\right] \\
&= \left[\mathsf{E}\left[T^{-1}\left(\mathbf{X}'\mathbf{X}\right)\right]\right]^{-1} \mathsf{E}\left[T^{-1}\left(\mathbf{X}'\mathbf{y}\right)\right]
\end{aligned}
$$