CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

An Automatic Test for Super Exogeneity

David F. Hendry and Carlos Santos

CEA@Cass Working Paper Series

WP–CEA–10-2008

An Automatic Test of Super Exogeneity

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

May 28, 2008

Abstract

We develop a new automatically-computable test for super exogeneity, using a variant of general-to-specific modelling. Based on the recent developments of impulse saturation applied to marginal models under the null that no impulses matter, we select the significant impulses for testing in the conditional. Since zero-mean changes are relatively undetectable in both VARs and conditional equations, we focus on location shifts although we also discuss variance changes. The approximate analytical non-centrality of the test is derived for a failure of weak exogeneity when there are shifts in the marginal process. Monte Carlo simulations confirm the empirical accuracy of the nominal significance levels under the null, and show rejections for this failure of super exogeneity. An empirical application to UK M1 delivers new results for this much-studied data set.

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

1 Introduction

It is a real pleasure to contribute to a volume in honor of Rob Engle, who has greatly advanced our understanding of exogeneity, among many other aspects of econometrics, and published with the first author on the topic of this paper. At the time of writing Engle, Hendry and Richard (1983), or even Engle and Hendry (1993), we could not have imagined that an approach based on handling more variables than observations would have been possible, let alone lead to an automatic test.

In all areas of policy which involve regime shifts or structural breaks in conditioning variables, the invariance of the parameters of conditional models under changes in the distributions of conditioning variables is of paramount importance, and was called super exogeneity by Engle *et al.* (1983). Even in models without contemporaneous conditioning variables, such as vector equilibrium systems (EqCMs), invariance under such shifts is equally relevant. Tests for super exogeneity have been proposed by Engle *et al.* (1983), Hendry (1988), Favero and Hendry (1992), Engle and Hendry (1993), Psaradakis and Sola (1996), Jansen and Teräsvirta (1996) and Krolzig and Toro (2002), inter alia: Ericsson and Irons (1994) overview the literature at the time of publication. Favero and Hendry (1992), building on Hendry (1988), considered the impact of non-constant marginal processes on conditional models, and concluded that location shifts were essential for detecting violations attributable to the Lucas (1976) critique. Engle and Hendry (1993) examined the impact on a conditional model of changes in the moments of the

*Financial support from the ESRC under Research Grant RES-062-23-0061, and from the Fundação para a Ciência e a Tecnologia (Lisboa), is gratefully acknowledged by the first and second authors respectively. We are indebted to Jennifer Castle, Jurgen Doornik and Bent Nielsen for helpful comments on an earlier draft, and to Jurgen and James Reade for providing results based on *Autometrics*.

conditioning variables, using a linear approximation: tests for super exogeneity were constructed by replacing the unobservable changing moments by proxies based on models of the processes generating the conditioning variables, including models based on ARCH processes (see Engle, 1982), thereby allowing for non-constant error variances to capture changes in regimes. However, Psaradakis and Sola (1996) claim that such tests have relatively low power for rejecting the Lucas critique. Jansen and Teräsvirta (1996) propose self-exciting threshold models for testing constancy in the conditional model as well as super exogeneity. Krolzig and Toro (2002) developed super-exogeneity tests based on a reduced-rank technique for co-breaking shown by the presence of common deterministic shifts, and demonstrated that their proposal dominated existing tests (on co-breaking, see Clements and Hendry, 1999, and Hendry and Massmann, 2007). We propose a new addition to this set of possible tests, show that its rejection frequency under the null is close to the nominal significance level in static settings, and examine its rejection frequencies when super exogeneity does not hold.

The ability to detect outliers and shifts in a model using the dummy saturation techniques proposed by Hendry, Johansen and Santos (2008) opens the door to this new class of automatically computable super-exogeneity tests. Their approach is to saturate the marginal model (or system) with impulse indicators (namely, include an impulse for every observation, but entered in feasible subsets), and retain all significant outcomes. They derive the probability under the null of falsely retaining impulses for a location-scale IID process, and obtain the distribution of the estimated mean and variance after saturation. Johansen and Nielsen (2008) extend that analysis to dynamic regression models which may have unit roots. Building on the ability to detect shifts in marginal models, we consider testing the relevance of all the retained impulses in conditional models. As we show below, such a test has the correct size under the null of super exogeneity of the conditioning variables for the parameters of the conditional model over a range of null-rejection frequencies in the marginal model saturation tests. Moreover, it can detect failures of super exogeneity when there are location shifts in the marginal models. Finally, it can be computed automatically–that is without explicit user intervention, as occurs with (say) tests for residual autocorrelation–once the desired nominal sizes of the marginal saturation and conditional super-exogeneity tests have been specified.

Five conditions need to be satisfied for an automatic test of super exogeneity. First, the test should not require *ex ante* knowledge by the investigator of the timing, signs or magnitudes of any breaks in the marginal processes of the conditioning variables. The test proposed here uses impulse saturation techniques on the marginal equations to determine these aspects. Secondly, the correct data generation process for the marginal variables should not need to be known for the test to have the desired rejection frequency under the null. That condition is satisfied here for the impulse-saturation stage when there are no explosive roots in any of the variables. Thirdly, the conditional model should not need to be over-identified under the alternative of a failure of super exogeneity, as required for tests in the class proposed by (say) Revankar and Hartley (1973). Fourthly, the test must have power against a large class of potential failures of super exogeneity in the conditional model when there are location shifts in some of the marginal processes. Below, we establish the non-centrality parameter of the proposed test in a canonical case. Finally, the test should be computable without additional user intervention, as holds for both impulse saturation and the conditional test proposed here.[1]

The structure of the paper is as follows. Section 2 considers which shifts in vector autoregressions (VARs) are detectable, and derives the implications for testing for breaks in conditional representations. Section 3 considers super exogeneity in a regression context to elucidate its testable hypotheses, and discusses how super exogeneity can fail. Section 4 describes the impulse saturation tests in Hendry *et al.* (2008) and Johansen and Nielsen (2008), and how to extend these to test super exogeneity. Section 5 provides analytic and Monte Carlo evidence on the null rejection frequencies of that procedure. Section

[1] *PcGets* and *Autometrics* are *Ox*-based programs designed for general to specific modelling (see Hendry and Krolzig, 2001, and Doornik, 2007, 1999).

6 considers the power of the first stage to determine location shifts in marginal processes. Section 7 analyses a failure of weak exogeneity under a non-constant marginal process. Section 8 notes a co-breaking saturation-based test which builds on Krolzig and Toro (2002) and Hendry and Massmann (2007). Section 9 investigates the powers of the proposed automatic test in Monte Carlo experiments for a bivariate data generation process (DGP) based on section 7. Section 10 tests super exogeneity in UK money demand; and section 11 concludes.

2 Detectable shifts

Consider the n-dimensional I(0) VAR(1) data generation process (DGP) of $\{\mathbf{x}_t\}$ over $t = 1, \ldots, T$:

$$\mathbf{x}_t = \boldsymbol{\phi} + \boldsymbol{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\nu}_t \;\text{ where }\; \boldsymbol{\nu}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\nu}}\right] \tag{1}$$

so $\boldsymbol{\Pi}$ has all its eigenvalues less than unity in absolute value, with unconditional expectation $\mathsf{E}\left[\mathbf{x}_t\right]$:

$$\mathsf{E}\left[\mathbf{x}_t\right] = \left(\mathbf{I}_n - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\phi} = \boldsymbol{\varphi} \tag{2}$$

hence:

$$\mathbf{x}_t - \boldsymbol{\varphi} = \boldsymbol{\Pi}\left(\mathbf{x}_{t-1} - \boldsymbol{\varphi}\right) + \boldsymbol{\nu}_t \tag{3}$$

At time T_1, however, $(\boldsymbol{\phi} : \boldsymbol{\Pi})$ changes to $(\boldsymbol{\phi}^* : \boldsymbol{\Pi}^*)$, so for $h \geq 1$ the data are generated by:

$$\mathbf{x}_{T_1+h} = \boldsymbol{\phi}^* + \boldsymbol{\Pi}^*\mathbf{x}_{T_1+h-1} + \boldsymbol{\nu}_{T_1+h} \tag{4}$$

where $\boldsymbol{\Pi}^*$ still has all its eigenvalues less than unity in absolute value. Letting $\boldsymbol{\varphi}^* = \left(\mathbf{I}_n - \boldsymbol{\Pi}^*\right)^{-1}\boldsymbol{\phi}^*$:

$$\mathbf{x}_{T_1+h} - \boldsymbol{\varphi}^* = \boldsymbol{\Pi}^*\left(\mathbf{x}_{T_1+h-1} - \boldsymbol{\varphi}^*\right) + \boldsymbol{\nu}_{T_1+h} \tag{5}$$

Clements and Hendry (1994), Hendry and Doornik (1997), and Hendry (2000) show that changes in $\boldsymbol{\varphi}$ are easy to detect, whereas those in $\boldsymbol{\phi}$ and $\boldsymbol{\Pi}$ are not when $\boldsymbol{\varphi}$ is unchanged. This delimits the class of structural breaks and regime changes that any test for super exogeneity can detect.

The forecast errors from $T_1 + 1$ onwards would be $\widehat{\boldsymbol{\nu}}_{T_1+h|T_1+h-1} = \mathbf{x}_{T_1+h} - \widehat{\mathbf{x}}_{T_1+h|T_1+h-1}$ where:

$$\widehat{\boldsymbol{\nu}}_{T_1+h|T_1+h-1} = \left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right) + \boldsymbol{\Pi}^*\left(\mathbf{x}_{T_1+h-1} - \boldsymbol{\varphi}^*\right) - \boldsymbol{\Pi}\left(\mathbf{x}_{T_1+h-1} - \boldsymbol{\varphi}\right) + \boldsymbol{\nu}_{T_1+h}. \tag{6}$$

Finite-sample biases in estimators and estimation uncertainty are neglected here as negligible relative to the sizes of the effects we seek to highlight. Unconditionally, using (2):

$$\mathsf{E}\left[\mathbf{x}_{T_1+h} - \widehat{\mathbf{x}}_{T_1+h|T_1+h-1}\right] = \mathsf{E}\left[\widehat{\boldsymbol{\nu}}_{T_1+h|T_1+h-1}\right] = \left(\mathbf{I}_n - \boldsymbol{\Pi}^*\right)\left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right) \tag{7}$$

Consequently, $\mathsf{E}\left[\widehat{\boldsymbol{\nu}}_{T_1+h|T_1+h-1}\right] = \mathbf{0}$ when $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}$, however large the changes in $\boldsymbol{\Pi}$ or $\boldsymbol{\phi}$. For a given change in $(\boldsymbol{\varphi}^* - \boldsymbol{\varphi})$, the effect is smaller (larger) as $(\mathbf{I}_n - \boldsymbol{\Pi}^*)$ moves closer to (further from) zero. The mean square error (MSE) matrix $\mathsf{E}\left[\widehat{\boldsymbol{\nu}}_{T_1+h|T_1+h-1}\widehat{\boldsymbol{\nu}}'_{T_1+h|T_1+h-1}\right] = \mathbf{Q}^*$ from (6) is:

$$\mathbf{Q}^* = \left(\mathbf{I}_n - \boldsymbol{\Pi}^*\right)\left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right)\left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right)'\left(\mathbf{I}_n - \boldsymbol{\Pi}^*\right)' + \left(\boldsymbol{\Pi}^* - \boldsymbol{\Pi}\right)\mathbf{M}\left(\boldsymbol{\Pi}^* - \boldsymbol{\Pi}\right)' + \boldsymbol{\Omega}_{\boldsymbol{\nu}}$$

when:

$$\mathsf{E}\left[\left(\mathbf{x}_{T_1+h-1} - \boldsymbol{\varphi}^*\right)\left(\mathbf{x}_{T_1+h-1} - \boldsymbol{\varphi}^*\right)'\right] = \mathbf{M}^* = \boldsymbol{\Pi}^*\mathbf{M}^*\left(\boldsymbol{\Pi}^*\right)' + \boldsymbol{\Omega}_{\boldsymbol{\nu}}$$

where $(\cdot)^v$ denotes column vectorizing and $\otimes$ is a Kronecker product, so:

$$\mathbf{M}^* = \left[\left(\mathbf{I}_n - \boldsymbol{\Pi}^* \otimes \boldsymbol{\Pi}^*\right)^{-1}\boldsymbol{\Omega}^v_{\boldsymbol{\nu}}\right]^m$$

when $(\cdot)^m$ is reforming as the associated square matrix. Then $\mathbf{Q}^*$ has to be judged relative to $\mathbf{Q} = \boldsymbol{\Omega}_{\boldsymbol{\nu}}$. Detectability also depends indirectly on the magnitudes of shifts relative to $\boldsymbol{\Omega}_{\boldsymbol{\nu}}$, as there are data variance shifts following unmodelled breaks, but these are hard to detect when $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}$, as the next section illustrates.

2.1 Simulation outcomes

To illustrate, let $n = 2$, and for the baseline case (a):

$$\mathbf{\Pi} = \begin{pmatrix} \boldsymbol{\pi}_1' \\ \boldsymbol{\pi}_2' \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 \\ -0.2 & 0.6 \end{pmatrix}, \quad \boldsymbol{\phi} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \tag{8}$$

where $\mathbf{\Pi}$ has eigenvalues of $0.65 \pm 0.19i$ with modulus 0.68, and for $|\rho| < 1$:

$$\mathbf{\Omega}_{\boldsymbol{\nu}} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = (0.01)^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

so the error standard deviations are 1% for $\mathbf{x}_t$ interpreted as logs, with:

$$\boldsymbol{\varphi} = \begin{pmatrix} 1 - 0.7 & -0.2 \\ 0.2 & 1 - 0.6 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3.75 \\ 0.625 \end{pmatrix}. \tag{9}$$

At time T_1, $\mathbf{\Pi}$ and $\boldsymbol{\phi}$ change to $\mathbf{\Pi}^*$ and $\boldsymbol{\phi}^*$ leading to case (b):

$$\mathbf{\Pi}^* = \begin{pmatrix} 0.5 & -0.2 \\ 0.1 & 0.5 \end{pmatrix}, \quad \boldsymbol{\phi}^* = \begin{pmatrix} 2.0 \\ -0.0625 \end{pmatrix} \tag{10}$$

where the eigenvalues of $\mathbf{\Pi}^*$ are $0.5 \pm 0.14i$ with modulus 0.27. The coefficients in $\mathbf{\Pi}$ are shifted at $T_1 = 0.75T$ by -20σ, -40σ, $+30\sigma$ and $+10\sigma$, so the standardized impulse responses are radically altered between $\mathbf{\Pi}$ and $\mathbf{\Pi}^*$. Figure 1 shows the data outcomes on a randomly-selected experiment in the first column, with the Chow test rejection frequencies on 1000 replications in the second (we will discuss the third below):
(a) for the baseline DGP in (8);
(b) for the changed DGP in (10);
(c) for the intercept-shifted DGP in (11) below;
(d) for the intercept-shifted DGP in (11) below, changed for one period.
The data over 1 to T_1 are the same in the four cases, and although the DGPs differ over $T_1 + 1$ to T in (a) and (b), it is hard to tell their data apart. The changes in $\boldsymbol{\phi}$ in (b) are approximately $\pm 100\sigma$, vastly larger than any likely shifts in real-world economies. Nevertheless, the rejection frequencies on the Chow test are under 13% at a 1% nominal significance.

However, keeping $\mathbf{\Pi}$ constant in (8), and changing only $\boldsymbol{\phi}$ by $\pm 5\sigma$ to $\boldsymbol{\phi}^{**}$ yields case (c):

$$\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.2 \\ -0.2 & 0.6 \end{pmatrix}, \ \boldsymbol{\phi}^{**} = \begin{pmatrix} 1.05 \\ 0.95 \end{pmatrix} \tag{11}$$

which leads to massive forecast failure. Indeed, changing the DGP in (11) for just one period is quite sufficient to reveal the shift almost 100% of the time. The explanation for such dramatic differences between the second and third rows–where the former had every parameter greatly changed and the latter only had a small shift in the intercept–is that $\boldsymbol{\varphi}$ is unchanged from (a) to (b) at:

$$\boldsymbol{\varphi}^* = \begin{pmatrix} 1 - 0.5 & 0.2 \\ -0.1 & 1 - 0.5 \end{pmatrix}^{-1} \begin{pmatrix} 2.0 \\ -0.0625 \end{pmatrix} = \begin{pmatrix} 3.75 \\ 0.625 \end{pmatrix} = \boldsymbol{\varphi} \tag{12}$$

whereas in (c):

$$\boldsymbol{\varphi}^{**} = \begin{pmatrix} 1 - 0.7 & -0.2 \\ 0.2 & 1 - 0.6 \end{pmatrix}^{-1} \begin{pmatrix} 1.05 \\ 0.95 \end{pmatrix} = \begin{pmatrix} 3.8125 \\ 0.46875 \end{pmatrix} \tag{13}$$

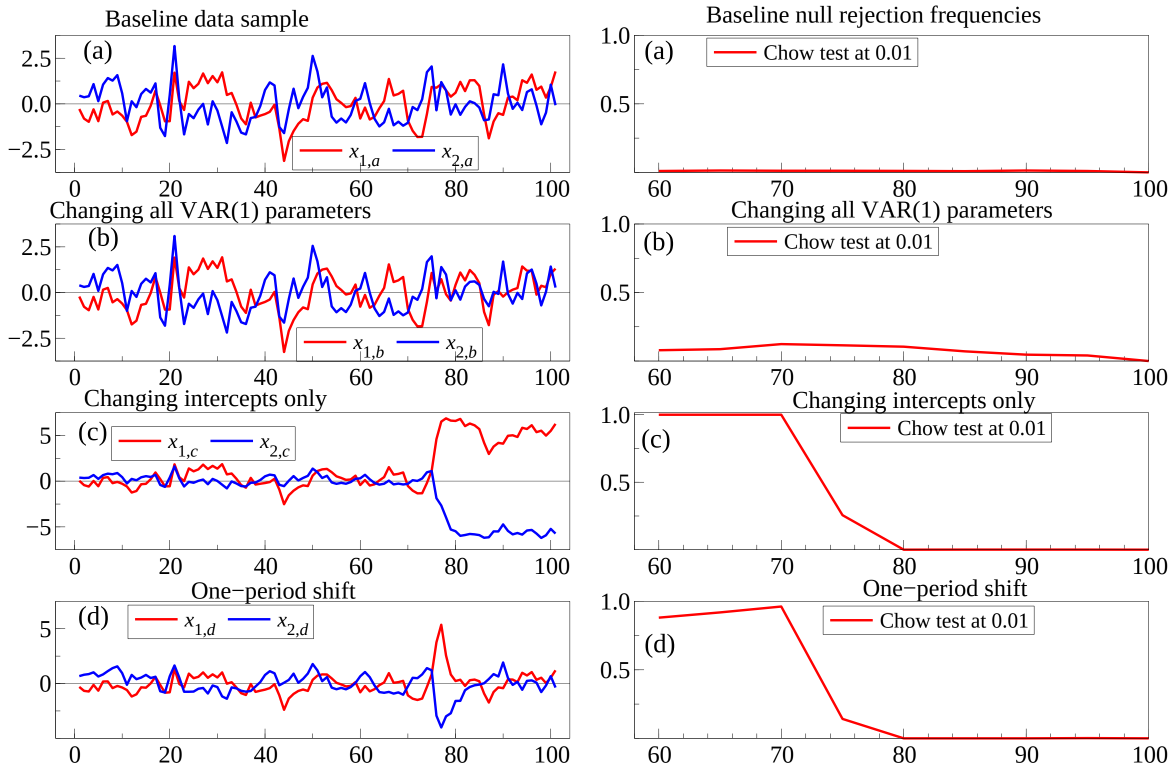

Figure 1: Data graphs and constancy test rejection frequencies

inducing shifts of a little over 6σ and 16σ in the locations of $x_{1,t}$ and $x_{2,t}$ respectively relative to the in-sample $\mathsf{E}[\mathbf{x}_t]$. Case (d) may seem the most surprising–it is far easier to detect a 1-period intercept shift of 5σ than when radically changing every parameter in the system for a quarter of the sample, but where the long-run mean is unchanged: indeed the rejection frequency is essentially 100% versus less than 15%. The entailed impacts of such shifts in the marginal distributions on conditional model are considered in the next section.

2.2 Detectability in conditional models

From the analysis in section 2, letting $\mathbf{x}'_t = (\mathbf{y}'_t : \mathbf{z}'_t)$ to match the notation below, then in case (a):

$$\begin{aligned}\mathsf{E}[y_t \mid z_t, \mathbf{x}_{t-1}] &= \phi_1 + \boldsymbol{\pi}'_1\mathbf{x}_{t-1} + \rho\left(z_t - \phi_2 - \boldsymbol{\pi}'_2\mathbf{x}_{t-1}\right) \\ &= \varphi_1 + \rho\left(z_t - \varphi_2\right) + \left(\boldsymbol{\pi}_1 - \rho\boldsymbol{\pi}_2\right)'\left(\mathbf{x}_{t-1} - \boldsymbol{\varphi}\right)\end{aligned}$$

as:

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \varphi_1 - \boldsymbol{\pi}'_1\boldsymbol{\varphi} \\ \varphi_2 - \boldsymbol{\pi}'_2\boldsymbol{\varphi} \end{pmatrix}.$$

After the shift in case (b), so $t > T_1$:

$$\begin{aligned}\mathsf{E}[y_t \mid z_t, \mathbf{x}_{t-1}] &= \left(\phi_1^* - \rho\phi_2^*\right) + \rho z_t + \left(\boldsymbol{\pi}_1^* - \rho\boldsymbol{\pi}_2^*\right)'\mathbf{x}_{t-1} \\ &= \varphi_1 + \rho\left(z_t - \varphi_2\right) + \left(\boldsymbol{\pi}_1^* - \rho\boldsymbol{\pi}_2^*\right)'\left(\mathbf{x}_{t-1} - \boldsymbol{\varphi}\right)\end{aligned} \tag{14}$$

and hence the conditional model is constant only if:

$$\boldsymbol{\pi}_1 - \boldsymbol{\pi}_1^* = \rho\left(\boldsymbol{\pi}_2 - \boldsymbol{\pi}_2^*\right) \tag{15}$$

which is strongly violated by the numerical values used here:

$$\boldsymbol{\pi}_1 - \boldsymbol{\pi}_1^* = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix} \neq \begin{pmatrix} -0.15 \\ 0.05 \end{pmatrix} = \rho \left(\boldsymbol{\pi}_2 - \boldsymbol{\pi}_2^*\right).$$

Nevertheless, as the shift in (14) anyway depends on changes in the coefficients of zero-mean variables, detectability will be low. In case (c) when $t >> T_1$:

$$\mathsf{E}\left[y_t \mid z_t, \mathbf{x}_{t-1}\right] = \varphi_1^* + \rho\left(z_t - \varphi_2^*\right) + \left(\boldsymbol{\pi}_1 - \rho\boldsymbol{\pi}_2\right)'\left(\mathbf{x}_{t-1} - \boldsymbol{\varphi}^*\right) \tag{16}$$

where $\mathsf{E}\left[z_t\right] = \varphi_2^*$ and $\mathsf{E}\left[\mathbf{x}_{t-1}\right] = \boldsymbol{\varphi}^*$ so there is a location shift of $\varphi_1^* - \varphi_1$. The third column of graphs in figure 1 confirms that the outcomes in the four cases above carry over to conditional models, irrespective of exogeneity: cases (a) and (b) are closely similar and low, yet rejection is essentially 100% in cases (c) and (d). Notice that there is no shift at all in (14) when (15) holds, however large the changes to the VAR. Consequently, we focus the super exogeneity test to have power for location shifts in the marginal distributions which 'contaminate' the conditional model.

2.2.1 Moving window estimation

One mechanism that would certainly detect that breaks of type (a) had occurred is the use of a moving window for estimation, since a purely post-break sample would deliver the second regime parameters. Thus, if impulse response analysis is to play a role in policy advice, it would seem advisable to check on a relatively small final period sample that the estimated parameters and error variances have not changed.

3 Super exogeneity in a regression context

Consider the sequentially-factorized DGP of the n-dimensional $\mathsf{I}(0)$ vector process $\{\mathbf{x}_t\}$:

$$\prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{y}|\mathsf{z}}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \mathsf{D}_{\mathsf{z}}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right) \tag{17}$$

where $\mathbf{x}_t' = (\mathbf{y}_t' : \mathbf{z}_t')$, $\mathbf{X}_{t-1} = (\mathbf{X}_0\ \mathbf{x}_1 \ldots \mathbf{x}_{t-1})$ for initial conditions $\mathbf{X}_0$, and $\boldsymbol{\phi} = \left(\boldsymbol{\phi}_1' : \boldsymbol{\phi}_2'\right)' \in \Phi$ with $\boldsymbol{\phi} = \mathbf{f}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^k$. The parameters $\boldsymbol{\phi}_1 \in \Phi_1$ and $\boldsymbol{\phi}_2 \in \Phi_2$ of the $\{\mathbf{y}_t\}$ and $\{\mathbf{z}_t\}$ processes need to be variation free, so that $\Phi = \Phi_1 \times \Phi_2$, if $\mathbf{z}_t$ is to be weakly exogenous for the parameters of interest $\boldsymbol{\psi} = \mathbf{h}\left(\boldsymbol{\phi}_1\right)$ in the conditional model. However, such a variation-free condition by itself does not rule out the possibility that $\boldsymbol{\phi}_1$ may change if $\boldsymbol{\phi}_2$ is changed. Super exogeneity augments weak exogeneity with parameter invariance in the conditional model such that:

$$\frac{\partial \boldsymbol{\phi}_1}{\partial \boldsymbol{\phi}_2'} = \mathbf{0}\ \forall \boldsymbol{\phi}_2 \in \mathcal{C}^{\phi_2} \tag{18}$$

where $\mathcal{C}^{\phi_2}$ is a class of interventions changing the marginal process parameters $\boldsymbol{\phi}_2$, so (18) requires no cross links between the parameters of the conditional and marginal processes. No DGPs can be invariant for all possible changes, hence the limitation to $\mathcal{C}^{\phi_2}$, the 'coverage' of which will vary with the problem under analysis.

When $\mathsf{D}_{\mathsf{x}}\left(\cdot\right)$ is the multivariate normal, we can express (17) as the unconditional model:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n \left[\begin{pmatrix} \mu_{1,t} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \sigma_{11,t} & \boldsymbol{\sigma}_{12,t}' \\ \boldsymbol{\sigma}_{12,t} & \boldsymbol{\Sigma}_{22,t} \end{pmatrix} \right] \tag{19}$$

where $\mathsf{E}[y_t] = \mu_{1,t}$ and $\mathsf{E}[\mathbf{z}_t] = \boldsymbol{\mu}_{2,t}$ are usually functions of $\mathbf{X}_{t-1}$. To define the parameters of interest, we let the economic theory formulation entail:

$$\mu_{1,t} = \mu + \boldsymbol{\beta}' \boldsymbol{\mu}_{2,t} + \boldsymbol{\alpha}' \mathbf{x}_{t-1} \tag{20}$$

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 (19) and (20):

$$\mathsf{E}[y_t \mid \mathbf{z}_t, \mathbf{x}_{t-1}] = \mu_{1,t} + \boldsymbol{\sigma}'_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1} \left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t} \right) = \mu + \gamma_{1,t} + \boldsymbol{\gamma}'_{2,t} \mathbf{z}_t + \boldsymbol{\alpha}' \mathbf{x}_{t-1} \tag{21}$$

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

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

When (21) is specified as a constant-parameter regression model over $t = 1, \ldots, T$:

$$y_t = \mu + \boldsymbol{\beta}' \mathbf{z}_t + \boldsymbol{\alpha}' \mathbf{x}_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \omega_t^2\right] \tag{22}$$

Three conditions are required for $\mathbf{z}_t$ to be super exogenous for $\left(\mu, \boldsymbol{\beta}, \boldsymbol{\alpha}, \omega_t^2 \right)$ (see Engle and Hendry, 1993):

(i) $\boldsymbol{\gamma}_{2,t} = \boldsymbol{\gamma}_2$ is constant $\forall t$;

(ii) $\boldsymbol{\beta} = \boldsymbol{\gamma}_2$;

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

Condition (i) requires that $\boldsymbol{\sigma}'_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1}$ is constant over time, which could occur because the σ_{ij} happened not to change over the sample, as well as because the two components move in tandem through being connected by $\boldsymbol{\sigma}'_{12,t} = \boldsymbol{\gamma}'_2 \boldsymbol{\Sigma}_{22,t}$. Condition (ii) then entails that $\mathbf{z}_t$ is weakly exogenous for a constant $\boldsymbol{\beta}$. Together, (i)+(ii) entail that $\gamma_{1,t} = 0$ and $\alpha = \mu_0$ in (21), so the conditional expectation does not depend on $\boldsymbol{\mu}_{2,t}$. Finally, (iii) requires the absence of links between the conditional and marginal parameters. Each of these conditions can be valid or invalid separately: for example, $\boldsymbol{\beta}_t = \boldsymbol{\gamma}_{2,t}$ is possible when (i) is false, and vice versa. A fully-constant regression also requires that $\omega_t^2 = \sigma_{11,t} - \boldsymbol{\beta}' \boldsymbol{\Sigma}_{22,t} \boldsymbol{\beta} = \omega^2$ is constant, so that the observed variation in $\sigma_{11,t}$ is derived from changes in $\boldsymbol{\Sigma}_{22,t}$, but the non-constancy of ω_t^2 $\forall t$ can be due to factors other than a failure of super exogeneity, so is only tested below as a requirement for congruency.

When conditions (i)–(iii) are satisfied:

$$\mathsf{E}[y_t \mid \mathbf{z}_t] = \mu + \boldsymbol{\beta}' \mathbf{z}_t + \boldsymbol{\alpha}' \mathbf{x}_{t-1} \tag{23}$$

in which case $\mathbf{z}_t$ is super exogenous for $\left(\mu, \boldsymbol{\beta}, \boldsymbol{\alpha}, \omega_t^2 \right)$ in this conditional model. Consequently:

$$\boldsymbol{\sigma}'_{12,t} = \boldsymbol{\beta}' \boldsymbol{\Sigma}_{22,t} \ \forall t \tag{24}$$

where condition (24) requires that the means in (20) are interrelated by the same parameter $\boldsymbol{\beta}$ as the covariances $\boldsymbol{\sigma}_{12,t}$ are with the variances $\boldsymbol{\Sigma}_{22,t}$. Under super exogeneity, the joint density is:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n \left[\begin{pmatrix} \mu + \boldsymbol{\beta}' \boldsymbol{\mu}_{2,t} + \boldsymbol{\alpha}' \mathbf{x}_{t-1} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega_t^2 + \boldsymbol{\beta}' \boldsymbol{\Sigma}_{22,t} \boldsymbol{\beta} & \boldsymbol{\beta}' \boldsymbol{\Sigma}_{22,t} \\ \boldsymbol{\Sigma}_{22,t} \boldsymbol{\beta} & \boldsymbol{\Sigma}_{22,t} \end{pmatrix} \right] \tag{25}$$

so the conditional-marginal factorization is:

$$\begin{pmatrix} y_t \mid \mathbf{z}_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n \left[\begin{pmatrix} \mu + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} \\ \boldsymbol{\mu}_{2,t} \end{pmatrix}, \begin{pmatrix} \omega_t^2 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22,t} \end{pmatrix} \right] \tag{26}$$

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

$$\mathbf{z}_t \sim \mathsf{IN}_{n-1}\left[\boldsymbol{\mu}_{2,t}, \boldsymbol{\Sigma}_{22,t}\right] \tag{27}$$

without altering the parameters of (22). Deterministic-shift co-breaking will then occur in (25), as $\left(1 : \boldsymbol{\beta}'\right)\mathbf{x}_t$ does not depend on $\boldsymbol{\mu}_{2,t}$: see §8. Conversely, if $\mathbf{z}_t$ is not super exogenous for $\boldsymbol{\beta}$, then changes in (27) should affect (22).

3.1 Failures of super exogeneity

Super exogeneity may fail for any of the three reasons corresponding to (i)–(iii) above:

(a) $\mathbf{z}_t$ is not weakly exogenous for $\boldsymbol{\beta}$;

(b) the regression coefficient $\boldsymbol{\gamma}_2$ is not constant when $\boldsymbol{\beta}$ is;

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

From (21), when $\mathbf{z}_t$ is not super exogenous for $\boldsymbol{\beta}$ but (20) holds, then:

$$\begin{aligned} \mathsf{E}\left[y_t \mid \mathbf{z}_t\right] &= \mu_{1,t} + \boldsymbol{\sigma}_{12,t}'\boldsymbol{\Sigma}_{22,t}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) \\ &= \mu + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} + \left(\boldsymbol{\gamma}_{2,t} - \boldsymbol{\beta}\right)'\mathbf{v}_{2,t} \end{aligned} \tag{28}$$

where, under the null, $\mathbf{v}_{2,t} \sim \mathsf{IN}_{n-1}\left[\mathbf{0}, \boldsymbol{\Sigma}_{22,t}\right]$ is the error on the marginal model (27), where we model $\boldsymbol{\mu}_{2,t}$ by lagged values of $\mathbf{x}_t$, to approximate the sequential factorization in (17):

$$\mathbf{z}_t = \boldsymbol{\mu}_{2,t} + \mathbf{v}_{2,t} = \boldsymbol{\pi}_0 + \sum_{j=1}^{s} \boldsymbol{\Gamma}_j \mathbf{x}_{t-j} + \mathbf{v}_{2,t} \tag{29}$$

Section 2 established that the detectable breaks in (29) are location shifts, so the next section considers impulse saturation applied to the marginal process, then derives the distribution under the null in §5, and its behaviour under the alternative in §6. Section 7 proposes the test for super exogeneity based on including the significant impulses from such marginal-model analyses in conditional equations.

4 Impulse saturation

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, where $1_{\{t\}} = 1$ for observation t, and zero otherwise: see Hendry *et al.* (2008) and Johansen and Nielsen (2008). Using a general-to-specific procedure, those authors analytically establish the null distribution of the estimator of regression parameters after adding T impulse indicators when the sample size is T. A two-step process is investigated, where half the indicators are added, and all significant indicators recorded, then the other half examined, and finally the two retained sets of indicators are combined. The average retention rate of impulse indicators under the null is αT when the significance level of an individual test is set at α, so for $\alpha = 0.01$, for example, $0.01T$ indicators will be retained. Moreover, Hendry *et al.* (2008) show that other splits, such as using three splits of size $T/3$, do not affect the retention rate under the null, or simulation-based distributions.

This procedure can be applied to the marginal models for the 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 (29) has been extended to:

$$\mathbf{z}_t = \boldsymbol{\pi}_0 + \sum_{j=1}^{s} \boldsymbol{\Gamma}_j \mathbf{x}_{t-j} + \sum_{i=1}^{m} \boldsymbol{\tau}_{i,\alpha_1} 1_{\{t=t_i\}} + \mathbf{v}^*_{2,t} \tag{30}$$

where the coefficients of the significant impulses are denoted $\boldsymbol{\tau}_{i,\alpha_1}$ to emphasize their dependence on the significance level α_1 used in testing the marginal model.

There is an important difference between outlier detection, which does just that, and impulse saturation which will detect outliers, but may also reveal others shifts that are hidden by being 'picked up' incorrectly by other variables. Figure 2 illustrates for a mean shift near the mid-sample, where no outliers, as defined by $|\widehat{u}_{i,t}| > 2\sigma_{ii}$ (say), are detected (for an alternative approach, see Sánchez and Peña, 2003).

By way of comparison, figure 3 shows impulse saturation for the same data, where the columns show the outcomes for the first half, second half, then combined respectively, and each row shows in order the impulses included at that stage, their plot against the data, and the impulses retained. Overall, 20 impulses are significant, spanning the break.

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

$$y_t = \mu + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} + \sum_{i=1}^{m} \delta_{i,\alpha_2} 1_{\{t=t_i\}} + \epsilon_t \tag{31}$$

and conduct an F-test for the significance of $(\delta_{1,\alpha_2} \dots \delta_{m,\alpha_2})$ at level α_2. Under the null of super exogeneity, the F-test of the joint significance of the m impulse indicators in the conditional model should have an approximate F-distribution and thereby allow an appropriately sized test: section 5 derives the null distribution and presents Monte Carlo evidence on its small-sample relevance. Under the alternative, the test will have power in a variety of situations discussed in section 7 below. Such a test can be automated, bringing super exogeneity into the purview of hypotheses about a model that can be as easily tested as (say) residual autocorrelation. Intuitively, if super exogeneity is invalid, so $\boldsymbol{\beta}' \neq \boldsymbol{\sigma}'_{12,t}\boldsymbol{\Omega}^{-1}_{22,t}$ in (28), then the impact on the conditional model of the largest values of the errors $\mathbf{v}_{2,t}$ should be the easiest to detect, noting that the significant impulses in (30) capture the outliers not accounted for by the regressor variables used.

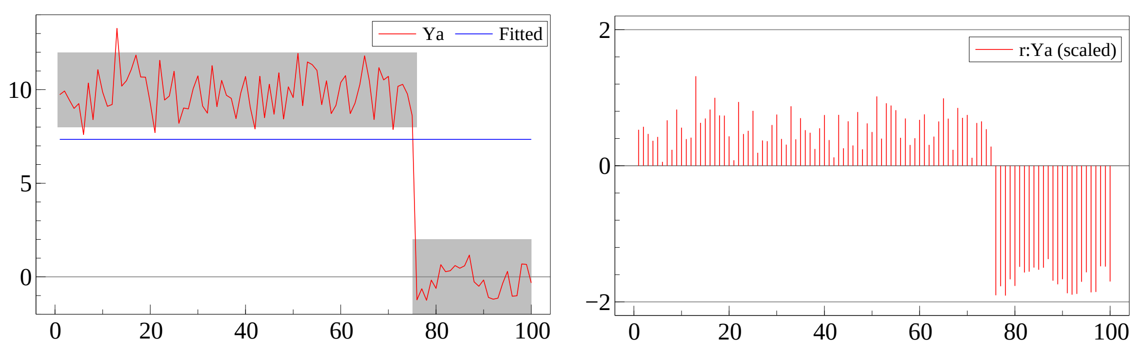

Figure 2: Absence of outliers despite a break

The null rejection frequency of this F-test of super exogeneity in the conditional model should not depend on the significance level, α_1, set for each individual test in the marginal model. However, 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

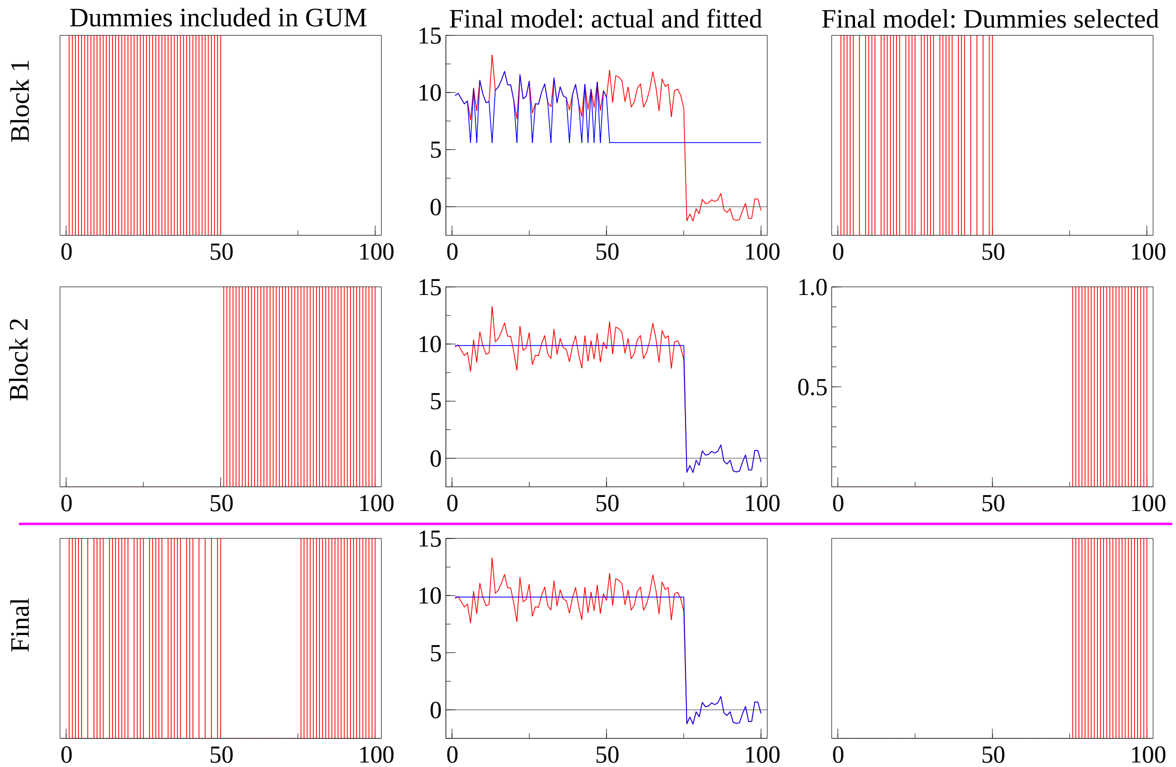

Figure 3: Impulse saturation in action

even no, impulses being retained from the marginal models. Monte Carlo evidence presented in section 5.1 supports that contention. For example, with four conditioning variables and $T = 100$, then under the null, $\alpha_1 = 0.01$ would yield four impulses in general, whereas $\alpha_1 = 0.025$ would provide 10. Otherwise, the main consideration for choosing α_1 is to allow power against reasonable alternatives to super exogeneity.

A variant of the test in (31), which builds on Hendry and Santos (2005) and has different power characteristics, is to combine the m impulses detected in (30) into an index (see Hendry and Santos, 2007).

5 Null rejection frequency of the impulse-based test

Reconsider the earlier sequentially factorized DGP in (19), where under the null of super exogeneity, from (23):

$$y_t = \mu + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} + \epsilon_t \tag{32}$$

so although the $\{\mathbf{z}_t\}$ process is non-constant, the linear relation between y_t and $\mathbf{z}_t$ in (23) is constant.

Let $\mathcal{S}_{\alpha_1}$ denote the dates of the significant impulses $\left\{1_{\{t_i\}}\right\}$ retained in the model for the marginal process (30) where:

$$\left|\mathsf{t}_{\widehat{\tau}_{i,t_i}}\right| > c_{\alpha_1} \tag{33}$$

when c_{α_1} is the critical value for significance level α_1. In the model (32) for $y_t|\mathbf{z}_t, \mathbf{x}_{t-1}$, conditioning on $\mathbf{z}_t$ implies taking the $\mathbf{v}_{2,t}$s as fixed, so stacking the impulses in $\left\{1_{\{t_i\}}\right\}$ in the vector $\mathbf{1}_t$:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} + \boldsymbol{\delta}'\mathbf{1}_t \tag{34}$$

where $\boldsymbol{\delta} = \mathbf{0}$ under the null. Given a significance level α_2, a subset of the indicators $\{\mathbf{1}_t\}$ will be retained in the conditional econometric model, given that they were retained in the marginal when:

$$\left| \mathsf{t}_{\widehat{\boldsymbol{\delta}}_j} \right| > c_{\alpha_2}. \tag{35}$$

Thus, when (33) occurs, the probability of retaining any indicator in the conditional is:

$$\Pr\left(\left| \mathsf{t}_{\widehat{\boldsymbol{\delta}}_j} \right| > c_{\alpha_2} \mid \left| \mathsf{t}_{\widehat{\tau}_{i,t_i}} \right| > c_{\alpha_1} \right) = \Pr\left(\left| \mathsf{t}_{\widehat{\boldsymbol{\delta}}_j} \right| > c_{\alpha_2} \right) = \alpha_2 \tag{36}$$

since (33) holds, which only depends on the significance level c_{α_2} used on the conditional model and not on α_1.

5.1 Monte Carlo evidence on the null rejection frequency

The Monte Carlo experiments estimate the empirical null rejection frequencies of the super-exogeneity test for a variety of settings, sample sizes, and nominal significance levels, and check if there is any dependence of these on the nominal significance levels for impulse retention in the marginal process. If there is dependence, then searching for the relevant dates at which shifts might have occurred in the marginal would affect testing for associated shifts in the conditional. In the following subsections, super exogeneity is the null, and we consider three settings for the marginal process: where there are no breaks in §5.1.1; a mean shift in §5.1.2; and a variance change in §5.1.3. Because the 'size' of a test statistic has a definition which is only precise for a similar test, and the word is anyway ambiguous in many settings (such as sample size), we use the term 'gauge' to denote the empirical null rejection frequency of the test procedure.

The general form of DGP is the bivariate system:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} \mu + \beta\xi_{(t)}\mu_{z_t} + \boldsymbol{\alpha}'\mathbf{x}_{t-1} \\ \xi_{(t)}\mu_{z_t} \end{pmatrix}, \sigma_{22} \begin{pmatrix} \sigma_{22}^{-1}\sigma_{11} + \beta^2\theta_{(t)} & \beta\theta_{(t)} \\ \beta\theta_{(t)} & \theta_{(t)} \end{pmatrix} \right] \tag{37}$$

where $\xi_{(t)} = 1 + \xi 1_{\{t>T_1\}}$ and $\theta_{(t)} = 1 + \theta 1_{\{t>T_2\}}$ so throughout:

$$\gamma_2^* = \frac{\sigma_{12,t}^*}{\sigma_{22,t}^*} = \frac{\beta\sigma_{22}\theta_{(t)}}{\sigma_{22}\theta_{(t)}} = \beta = \gamma_2 \tag{38}$$

$$(\omega^*)^2 = \sigma_{11,t}^* - \left(\sigma_{12,t}^*\right)^2 \left(\sigma_{22,t}^*\right)^{-1} = \sigma_{11} + \beta^2\sigma_{22}\theta - \frac{\beta^2\sigma_{22}^2\theta^2}{\sigma_{22}\theta} = \sigma_{11} = \omega^2 \tag{39}$$

and from (37):

$$\begin{aligned} \mathsf{E}\left[y_t \mid z_t, \mathbf{x}_{t-1}\right] &= \mu + \beta\xi_{(t)}\mu_{z_t} + \boldsymbol{\alpha}'\mathbf{x}_{t-1} + \gamma_2\left(z_t - \xi_{(t)}\mu_{z_t}\right) \\ &= \mu + \beta z_t + \boldsymbol{\alpha}'\mathbf{x}_{t-1} \end{aligned} \tag{40}$$

Three cases of interest are $\xi = \theta = 0$, $\xi = 0$, and $\theta = 0$ in all of which super exogeneity holds but for different forms of change in the marginal process. In all cases, $\beta = 2 = \gamma_2$ and $\omega^2 = 1$, which are the constant and invariant parameters of interest in the conditional model, with $\sigma_{22} = 5$. Any changes in the marginal process occur at time $T_1 = 0.8T$. The impulse saturation uses a partition of $T/2$ with $M = 10000$ replications. Sample sizes of $T = (50, 100, 200, 300)$ are investigated, and we examine four significance levels for α_1 equal to (0.1, 0.05, 0.025, 0.01) for testing impulses in the marginal, where the significance levels α_2 for testing in the conditional vary over the same range.

5.1.1 Constant marginal

The baseline DGP is (37) with $\xi = \theta = 0$, $\mu_{z_t} = 1$ and $\boldsymbol{\alpha} = \mathbf{0}$. Thus, the parameters of the conditional model $y_t|z_t$ are $\boldsymbol{\phi}_1' = (\mu; \gamma_2; \omega^2) = (0; 2; 1)$ and the parameters of the marginal are $\boldsymbol{\phi}_{2,t}' = (\mu_{2,t}; \sigma_{22,t}) = (1; 5)$. The conditional representation is:

$$y_t = \beta z_t + \sum_{i \in \mathcal{S}_{\alpha_1}} \delta_i 1_{t_i} + \epsilon_t \tag{41}$$

and testing super exogeneity is based on the F-test of the null $\boldsymbol{\delta} = \mathbf{0}$ in (41).

The first column in figure 4 reports the test's gauges where α_1 is the nominal significance level used for the t-tests on each individual indicator in the marginal model (horizontal axis), and α_2 is the significance level for the F-test on the retained dummies in the conditional (vertical axis). Unconditional rejection frequencies are recorded throughout.

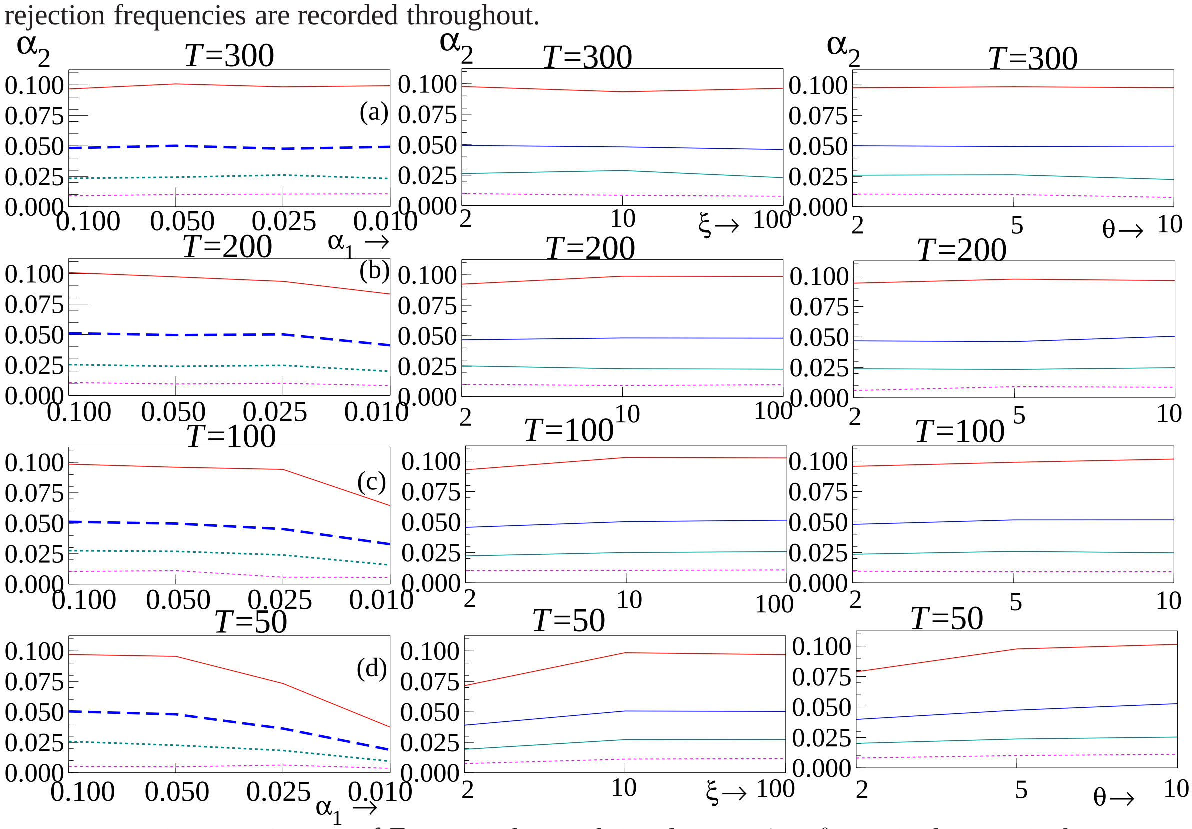

Figure 4: Gauges of F-tests in the conditional as α_1, ξ or θ vary in the marginal

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. For example, 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. The simulated gauges and nominal null rejection frequencies are close so long as $\alpha_1 T > 3$. Then, there is no distortion in the number of retained dummies in the conditional. However, constant marginal processes are the 'worst-case': the next two sections consider mean and variance changes where many more impulses are retained, so there are fewer cases of zero impulses to enter in the conditional.

5.1.2 Changes in the mean of z_t

The second DGP is given by (37) where $\xi = 2, 10, 100$ with $\theta = 0$, $\mu_{z_t} = 1$ and $\boldsymbol{\alpha} = \mathbf{0}$ as before. Super exogeneity holds irrespective of the level shift in the marginal. The variance-covariance matrix is constant, but could be allowed to change as well, provided the values matched the conditions for super exogeneity as in §5.1.3.

The second column of graphs in figure 4 reports the test's gauges where the horizontal axis now corresponds to the three values of ξ, whereas $\alpha_1 = 2.5\%$ throughout.

Despite large changes in ξ, when $T > 100$ the gauges are close to the nominal significance levels. Importantly, the test does not spuriously reject the null, but now is slightly undersized at $T = 50$ for small shifts, as again sometimes no impulses are retained.

5.1.3 Changes in the variance of z_t

The third DGP is given by (37) where $\theta = 2, 5, 10$ with $\xi = 0$, $\mu_{z_t} = 1$ and $\boldsymbol{\alpha} = \mathbf{0}$, so $\boldsymbol{\phi}_{1,t}$ is again invariant to changes in $\boldsymbol{\phi}_{2,t}$ induced by changes in $\sigma_{22,t}$. The impulse saturation test has power to detect variance shifts in the marginal, so like the previous case, more than αT impulses should be retained on average, depending on the magnitude of the marginal variance change (see §6.2).

The third column of graphs in figure 4 reports the test's gauges as before. Again, the vertical axis reports α_2, the nominal significance level for the F-test on the retained impulses in the conditional, but now the horizontal axis corresponds to the three values of θ, using $\alpha_1 = 2.5\%$ throughout.

The F-test has gauge close to the nominal for $T > 100$, even when the variance of the marginal process changes markedly, but the test is again slightly undersized at $T = 50$ for small shifts. As in §5.1.2, the test is not 'confused' by variance changes in the marginal to falsely imply a failure of super exogeneity even though the null holds.

Overall, the proposed test has appropriate empirical null rejection frequencies for both constant and changing marginal processes, so we now turn to its ability to detect failures of exogeneity. Being a selection procedure, this hardly corresponds to the conventional notion of 'power', so we use the term 'potency' to denote the non-null rejection frequency of the test.

This test involves a two-stage process: first detect shifts in the marginal, then use those to detect shifts in the conditional. The properties of the first stage have been considered in Santos and Hendry (2006), so we only note them here, partly to establish notation for the second stage considered in §7.

6 Potency at stage 1

We consider the potency at stage 1 for a mean shift then a variance change, both at time $T_1 = T - k + 1$.

6.1 Detecting a mean shift in the marginal

The potency at the second stage conditional on knowing the break dates in the marginal, and hence correctly retaining every dummy, are easily calculated, but will only be accurate for large magnitude breaks, parameterized below by $\boldsymbol{\lambda}$, when the saturation approach locates all, and only, the relevant impulses. For smaller values of $\boldsymbol{\lambda}$, fewer impulses will be detected in the marginal, and indeed, although the null rejection frequency of the test does not depend on α_1 once $\alpha_1 T > 3$, the potency will in general, suggesting that a relatively non-stringent α_1 should be used. Conversely, that will lead to retaining some 'spurious' impulses in the marginal, albeit fewer than $\alpha_1 T_1$ as shifts lower the remaining null rejection frequency.

Marginal models in their simplest form are:

$$z_{j,t} = \sum_{i \in \mathcal{S}_{\alpha_1}} \tau_{i,j,\alpha_1} 1_{\{t_i\}} + v^*_{2,j,t}, \tag{42}$$

when the marginal process is (43):

$$z_{j,t} = \lambda_j 1_{\{t > T_1\}} + v_{2,j,t} \tag{43}$$

where H_1: $\lambda_j \neq 0 \; \forall j$ holds. The potency to retain each impulse in (42) depends on the probability of rejecting the null for the associated estimated τ_{i,j,α_1}:

$$\widehat{\tau}_{i,j,\alpha_1} = \lambda_j + v^*_{2,j,t_i}.$$

The properties of tests on such impulse indicators are discussed in Hendry and Santos (2005). Let ψ_{λ,α_1} denote the non-centrality, where the subscript α_1 is to denote stage 1, then as $\mathsf{V}[\widehat{\tau}_{i,j,\alpha_1}] = \sigma_{22,j}$:

$$\mathsf{E}\left[\mathsf{t}_{\tau_{i,j,\alpha}=0}\left(\psi_{\lambda,\alpha_1}\right)\right] = \mathsf{E}\left[\frac{\widehat{\tau}_{i,j,\alpha_1}}{\sqrt{\widehat{\sigma}_{22,j}}}\right] \simeq \frac{\lambda_j}{\sqrt{\sigma_{22,j}}} = \psi_{\lambda,\alpha_1} \tag{44}$$

When $v_{2,j,t}$ is normal, the potency could be computed directly from the t-distribution; since most outliers will have been removed, normality should be a reasonable approximation. However, we compute approximate potency functions using an approximation to $\mathsf{t}^2_{\tau_{i,j,\alpha_1}=0}$ by a chi-squared with 1 degree of freedom:

$$\mathsf{t}^2_{\tau_{i,j,\alpha_1}=0}\left(\psi^2_{\lambda,\alpha_1}\right) \sim \chi^2_1\left(\psi^2_{\lambda,\alpha_1}\right). \tag{45}$$

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

$$\chi^2_1\left(\psi^2_{\lambda,\alpha_1}\right) = h\chi^2_m\left(0\right) \tag{46}$$

where:

$$h = \frac{1+2\psi^2_{\lambda,\alpha_1}}{1+\psi^2_{\lambda,\alpha_1}} \text{ and } m = \frac{1+\psi^2_{\lambda,\alpha_1}}{h} \tag{47}$$

then the potency function of the $\chi^2_1\left(\psi^2_{\lambda,\alpha_1}\right)$ test in (45) is approximated by:

$$\mathsf{P}\left[\mathsf{t}^2_{\tau_{i,j,\alpha_1}=0}\left(\psi^2_{\lambda,\alpha_1}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_1\left(\psi^2_{\lambda,\alpha_1}\right) > c_{\alpha_1} \mid \mathsf{H}_1\right] \simeq \mathsf{P}\left[\chi^2_m\left(0\right) > h^{-1}c_{\alpha_1}\right]. \tag{48}$$

For non-integer values of m, a weighted average of the neighbouring integer values is used. For example, when $\psi^2_{\lambda,\alpha_1} = 16$ and $c_{\alpha_1} = 3.84$, then $h \simeq 1.94$ and $m = 8.76$ (taking the nearest integer values as 8 and 9 with weights 0.24 and 0.76) which yields $\mathsf{P}[\mathsf{t}^2_{\tau_{i,j,\alpha_1}=0}(16) > 3.84] \simeq 0.99$, as against the exact t-distribution outcome of 0.975. When $\lambda_j = d\sqrt{\sigma_{22,j}}$, $\psi^2_{\lambda,\alpha_1} = d^2$ and $\mathsf{p}_\lambda = \mathsf{P}[\mathsf{t}^2_{\widehat{\tau}_{i,j,\alpha_1}}\left(d^2\right) > c_{\alpha_1}]$ at $c_{\alpha_1} = 3.84$ rises from 0.17, through 0.50 to 0.86 as d is 1, 2, 3 so the potency is low at $d = 1$ (the t-distribution outcome for $d = 1$ is 0.16), but has risen markedly even by $d = 3$.

In practice, *Autometrics* selects impulses within contiguous blocks with approximately these probabilities, but has somewhat lower probabilities for scattered impulses. For example for the two DGPs:

$$\begin{array}{llll} \text{D1:} & y_{1,t} = d\left(I_{T-19} + \cdots + I_T\right) + u_t, & u_t \sim \mathsf{IN}(0,1) \\ \text{D3:} & y_{3,t} = d\left(I_1 + I_6 + I_{11} + \cdots\right) + u_t, & u_t \sim \mathsf{IN}(0,1) \end{array}$$

where the GUM is just a constant and T dummies for $T = 100$, then while both have 20 relevant indicators, the potency per impulse is:

6.2 Detecting a variance shift in the marginal

We consider a setting where the variance shift $\theta > 1$ occurs when $T_1 > T/2$ so that:

$$z_t = 1 + \left(1_{\{t<T_1\}} + \sqrt{\theta}1_{\{t\geq T_1\}}\right)v_t. \tag{49}$$

D1	$d=0$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
gauge %	1.5	1.2	0.9	0.3	0.7	1.1
potency %	—	4.6	25.6	52.6	86.3	99.0
analytic power %	—	6.1	26.9	65.9	93.7	99.7
D3	$d=0$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
gauge %	1.5	1.0	0.4	0.3	1.0	0.8
potency %	—	3.5	7.9	24.2	67.1	90.2

Table 1: Impulse saturation in *Autometrics* at 1% nominal size, $T = 100$, $M = 1000$

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

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

where $\mathsf{t}_{\tau_t}^2$ has a non-centrality of $\psi_{\theta,\alpha_1}^2 = \theta$. Approximating by $h\chi_m^2(0)$ as in (48), for $\psi_{\theta,\alpha_1}^2 = (2; 5; 10)$ potency will be about (25%, 60%, 90%) respectively at $\alpha_1 = 0.05$. Thus, only large changes in variances will be detected.

Viewing the potencies at stage 1 as the probability p_λ of retaining a relevant impulse from the marginal model, then approximately $\mathsf{p}_\lambda k \leq k$ relevant impulses will be retained for testing in the conditional model, attentuating the non-centrality $\varphi_{\delta,\alpha_1}$of the F-test of $\boldsymbol{\delta} = \mathbf{0}$ in (41) relative to a known break date. Further, retention of irrelevant impulses corresponding to non-break related shocks in the marginal process, will also lower potency relative to knowing the break dates. For the F-test of $\boldsymbol{\delta} = \mathbf{0}$, the increase in its degrees of freedom should only induce a small potency reduction. For a given non-centrality $\varphi_{\delta,\alpha_1}$, however, that effect also differs depending on the magnitude and length of the break in the marginal, since fewer irrelevant impulses will be retained when there is a large, short break.

7 Super exogeneity failure

In this section, we derive the outcome for a super exogeneity failure due a weak exogeneity violation when the marginal process is non-constant, and obtain the non-centrality and approximate potency of the test when there is a location shift in the marginal. Figure 1 showed high constancy-test rejection frequencies for both that setting and even a single impulse. Section 9 reports the simulation outcomes. As seen in §3.1, many causes of failure are possible, including shifts in variances in marginal processes and any cross-links between conditional and marginal parameters, but location shifts due to changes in policy rules are a central scenario.

We use the formulation in §3 for a normally-distributed $n \times 1$ vector $\mathbf{x}_t = (y_t : \mathbf{z}_t')'$ generated by (19), with $\mathsf{E}[y_t|\mathbf{z}_t]$ given by (21), where $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$, $\boldsymbol{\alpha} = \mathbf{0}$ and conditional variance $\omega^2 = \sigma_{11} - \boldsymbol{\sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$. The parameter of interest is $\boldsymbol{\beta}$ in (20), so:

$$y_t = \mu + \boldsymbol{\beta}'\mathbf{z}_t + (\boldsymbol{\gamma} - \boldsymbol{\beta})'\left(\mathbf{z}_t - \boldsymbol{\mu}_{2,t}\right) + \epsilon_t = \mu + \boldsymbol{\beta}'\mathbf{z}_t + (\boldsymbol{\gamma} - \boldsymbol{\beta})'\mathbf{v}_{2,t} + \epsilon_t \tag{50}$$

where $\epsilon_t = y_t - \mathsf{E}[y_t|\mathbf{z}_t]$, so $\mathsf{E}[\epsilon_t|\mathbf{z}_t] = 0$, but $\mathsf{E}[y_t|\mathbf{z}_t] \neq \boldsymbol{\beta}'\mathbf{z}_t$ when $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$, violating weak exogeneity. Instead:

$$\mathsf{E}[y_t \mid \mathbf{z}_t] = \mu + \boldsymbol{\gamma}'\mathbf{z}_t + (\boldsymbol{\beta} - \boldsymbol{\gamma})'\boldsymbol{\mu}_{2,t}$$

which will be non-constant and change as $\boldsymbol{\mu}_{2,t}$ shifts. Such a conditional model is an example of the Lucas (1976) critique where the agents' behavioural rule depends on $\mathsf{E}[\mathbf{z}_t]$ as in (20), whereas the econometric equation uses $\mathbf{z}_t$, leading to (50).

To complete the system, the break in the process for $\{\mathbf{z}_t\}$ which induces the violation in super exogeneity is parametrized as:

$$\mathbf{z}_t = \boldsymbol{\mu}_{2,t} + \mathbf{v}_{2,t} = \boldsymbol{\lambda} 1_{\{t>T_1\}} + \mathbf{v}_{2,t} \tag{51}$$

In practice, there could be multiple breaks in different marginal processes at different times, which may affect one or more $\mathbf{z}_t$s, but little additional insight is gleaned over the one-off break in (51), which is sufficiently general since the proposed test is an F-test on all retained impulses, which does not assume any specific break form at either stage. The advantage of using the explicit alternative in (51) is that analytic calculations are feasible. Since §2 showed that the key shifts are in the long-run mean, we use the Frisch and Waugh (1933) theorem to partial out means, but with a slight abuse of notation, do not alter it. Combining (50) with (51) and letting $\delta = (\boldsymbol{\beta} - \boldsymbol{\gamma})' \boldsymbol{\lambda}$, the DGP is:

$$y_t = \mu + \boldsymbol{\gamma}'\mathbf{z}_t + (\boldsymbol{\beta} - \boldsymbol{\gamma})' \boldsymbol{\lambda} 1_{\{t>T_1\}} + \epsilon_t = \mu + \boldsymbol{\gamma}'\mathbf{z}_t + \delta 1_{\{t>T_1\}} + \epsilon_t \tag{52}$$

Testing for the impulse dummies in the marginal model yields:

$$\mathbf{z}_{t_i} = \sum_{i \in \mathcal{S}} \widehat{\boldsymbol{\rho}}_{i,\alpha_1} 1_{\{i\}} + \mathbf{v}^*_{2,t_i} \tag{53}$$

where $\mathcal{S}$ denotes the set of impulses $\widehat{\boldsymbol{\rho}}_{i,\alpha_1} = \boldsymbol{\lambda} 1_{\{t_i>T_1\}} + \mathbf{v}_{2,t_i}$ where $\mathbf{v}^*_{2,t_i} = \mathbf{0}\ \forall i \in \mathcal{S}$ defined by:

$$\mathsf{t}^2_{\rho_{i,j,\alpha_1}=0} > c_{\alpha_1}. \tag{54}$$

Stacking significant impulses from (54) in $\boldsymbol{\iota}_t$, and adding these to (50), yields the regression:

$$y_t = \tau_0 + \boldsymbol{\tau}_1'\mathbf{z}_t + \boldsymbol{\tau}_2 \boldsymbol{\iota}_t + e_t. \tag{55}$$

The main difficulty in formalizing the analysis is that $\boldsymbol{\iota}_t$ varies between draws in both its length and its contents. Since the test is an F-test, the particular relevant and irrelevant impulses retained do not matter, merely their total numbers from the first stage. Consequently, we distinguish between:
(a) the length of the break, Tr,
(b) the number of relevant retained elements in the index, which on average will be $\mathsf{p}_\lambda Tr$, where p_λ is the probability of retaining any given relevant impulse from §6.1, and
(c) the total number of retained impulses in the model, Ts, usually including some irrelevant ones, where on average $s = (\mathsf{p}_\lambda r + \alpha_1)$ which determines the average degrees of freedom of the test.
The F-test will have Ts numerator degrees-of-freedom and $T(1-s) - n$ denominator (allowing for the constant). The potency of the $\mathsf{F}^{Ts}_{T(1-s)-n}$-test of:

$$\mathsf{H}_0\text{: } \boldsymbol{\tau}_2 = \mathbf{0} \tag{56}$$

in (55) depends on the strengths of the super-exogeneity violations, $(\beta_i - \gamma_i)$; the magnitudes of the breaks, λ_i, both directly and through their detectability, p_λ, in the marginal models, in turn dependent on α_1; the sample size T; the relative number of periods r affected by the break; the number of irrelevant impulses retained, and on α_2. The properties are checked by simulation below, and could be contrasted with the optimal, but generally infeasible, test based on adding the index $1_{\{t>T_1\}}$, instead of the impulses $\boldsymbol{\iota}_t$, equivalent to a Chow (1960) test (see Salkever, 1976).

A formal derivation, could either include $\mathsf{p}_\lambda Tr$ impulses, akin to a mis-specification analysis, or model $\boldsymbol{\iota}_t$ in (55) as containing all Tr relevant impulses, each with probability $\mathsf{p}_\lambda > 0$. The impact of irrelevant retained impulses is merely to reduce the number of available observations, so lowers potency slightly, and can otherwise be neglected. Taking the second route, the full-sample representations are:

$$\begin{array}{rcl} \text{DGP} & : & \mathbf{y}=\mathbf{Z}\boldsymbol{\gamma}+\delta\mathbf{I}^{*}_{Tr}\mathbf{1}_{Tr}+\boldsymbol{\epsilon} \\ \text{Model} & : & \mathbf{y}=\mathbf{Z}\boldsymbol{\tau}_1+\mathbf{J}^{*}_{Tr}\boldsymbol{\tau}_2+\mathbf{e} \\ \text{Exogenous} & : & \mathbf{Z}=\mathbf{I}^{*}_{Tr}\mathbf{1}_{Tr}\boldsymbol{\lambda}'+\mathbf{V}_2 \end{array} \tag{57}$$

where:

$$\mathbf{I}^{*}_{Tr}=\begin{pmatrix} \mathbf{0}_{T(1-r),Tr} \\ \mathbf{I}_{Tr} \end{pmatrix};\ \ \mathbf{J}^{*}_{Tr}=\mathsf{p}_{\lambda}\mathbf{I}^{*}_{Tr} \tag{58}$$

so $\mathbf{1}_{Tr}$ is $Tr\times 1$ with Tr elements of unity etc. Combinations of elements entering $\mathbf{J}^{*}_{Tr}$ also have probability p_{λ}, since it is only their initial chance of selection that matters, and conditional on that, occur with certainty. Then, denoting (e.g.) $\mathbf{I}^{*\prime}_{Tr}\mathbf{Z}=\mathbf{Z}_{Tr}$:

$$\begin{aligned} \begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1-\boldsymbol{\gamma} \\ \widetilde{\boldsymbol{\tau}}_2-\delta\mathbf{1}_{Tr} \end{pmatrix} &= \begin{pmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{J}^{*}_{Tr} \\ \mathbf{J}^{*\prime}_{Tr}\mathbf{Z} & \mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{J}^{*\prime}_{Tr}\mathbf{y} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma} \\ \delta\mathbf{1}_{Tr} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{J}^{*}_{Tr} \\ \mathbf{J}^{*\prime}_{Tr}\mathbf{Z} & \mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr} \end{pmatrix}^{-1} \left(\delta \begin{pmatrix} \mathbf{Z}'\left(\mathbf{I}^{*}_{Tr}-\mathbf{J}^{*}_{Tr}\right)\mathbf{1}_{Tr} \\ \mathbf{J}^{*\prime}_{Tr}\left(\mathbf{I}^{*}_{Tr}-\mathbf{J}^{*}_{Tr}\right)\mathbf{1}_{Tr} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}'\boldsymbol{\epsilon} \\ \mathbf{J}^{*\prime}_{Tr}\boldsymbol{\epsilon} \end{pmatrix}\right) \\ &\simeq -\left(1-\mathsf{p}_{\lambda}\right)\delta \begin{pmatrix} -\mathbf{G}^{-1}\mathbf{Z}'_{Tr} \\ \mathbf{I}_{Tr}+\mathbf{Z}_{Tr}\mathbf{G}^{-1}\mathbf{Z}'_{Tr} \end{pmatrix}\mathbf{1}_{Tr} \\ &\quad + \begin{pmatrix} \mathbf{G}^{-1}\left(\mathbf{Z}'\boldsymbol{\epsilon}-\mathsf{p}_{\lambda}\mathbf{Z}'_{Tr}\boldsymbol{\epsilon}_{Tr}\right) \\ \mathsf{p}_{\lambda}\left(\boldsymbol{\epsilon}_{Tr}-\mathbf{Z}_{Tr}\mathbf{G}^{-1}\left(\mathbf{Z}'\boldsymbol{\epsilon}-\mathsf{p}_{\lambda}\mathbf{Z}'_{Tr}\boldsymbol{\epsilon}_{Tr}\right)\right) \end{pmatrix} \end{aligned} \tag{59}$$

where $\mathbf{G}=\left(\mathbf{Z}'\mathbf{Z}-\mathsf{p}_{\lambda}\mathbf{Z}'_{Tr}\mathbf{Z}_{Tr}\right)$ and the inverse second-moment matrix is:

$$\begin{pmatrix} \mathbf{G}^{-1} & -\mathbf{G}^{-1}\mathbf{Z}'\mathbf{J}^{*}_{Tr}\left(\mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr}\right)^{-1} \\ -\left(\mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr}\right)^{-1}\mathbf{J}^{*\prime}_{Tr}\mathbf{Z}\mathbf{G}^{-1} & \left(\mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr}\right)^{-1}+\left(\mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr}\right)^{-1}\mathbf{J}^{*\prime}_{Tr}\mathbf{Z}\mathbf{G}^{-1}\mathbf{Z}'\mathbf{J}^{*}_{Tr}\left(\mathbf{J}^{*\prime}_{Tr}\mathbf{J}^{*}_{Tr}\right)^{-1} \end{pmatrix}.$$

As $\mathsf{E}\left[\mathbf{Z}_{Tr}\right]=\mathbf{1}_{Tr}\boldsymbol{\lambda}'$ and $\boldsymbol{\lambda}\mathbf{1}'_{Tr}\mathbf{1}_{Tr}=Tr\boldsymbol{\lambda}$, approximating by:

$$\mathsf{E}\left[T\mathbf{G}^{-1}\right]\simeq\left(\mathsf{E}\left[T^{-1}\mathbf{G}\right]\right)^{-1}=\left(\left(1-\mathsf{p}_{\lambda}\right)r\boldsymbol{\lambda}\boldsymbol{\lambda}'+\left(1-\mathsf{p}_{\lambda}r\right)\boldsymbol{\Sigma}_{22}\right)^{-1} \tag{60}$$

then:

$$\mathsf{E}\left[\begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1 \\ \widetilde{\boldsymbol{\tau}}_2 \end{pmatrix}\right]\simeq\begin{pmatrix} \boldsymbol{\gamma} \\ \delta\mathbf{1}_{Tr} \end{pmatrix}-\delta\left(1-\mathsf{p}_{\lambda}\right)\begin{pmatrix} -\mathbf{f}_{\boldsymbol{\lambda}} \\ \left(1+\boldsymbol{\lambda}'\mathbf{f}_{\boldsymbol{\lambda}}\right)\mathbf{1}_{Tr} \end{pmatrix}=\begin{pmatrix} \boldsymbol{\gamma}^{*} \\ \delta^{*}\mathbf{1}_{Tr} \end{pmatrix}$$

where:

$$\mathbf{f}_{\boldsymbol{\lambda}}=r\left(\mathsf{E}\left[T^{-1}\mathbf{G}\right]\right)^{-1}\boldsymbol{\lambda} \tag{61}$$

and:

$$\mathsf{V}\left[\begin{pmatrix} \widetilde{\boldsymbol{\tau}}_1 \\ \widetilde{\boldsymbol{\tau}}_2 \end{pmatrix}\right]\simeq\frac{\sigma^2_e}{T}\begin{pmatrix} \mathbf{G}^{-1} & -\mathsf{p}_{\lambda}\mathbf{G}^{-1}\boldsymbol{\lambda}\mathbf{1}'_{Tr} \\ -\mathsf{p}_{\lambda}\mathbf{1}_{Tr}\boldsymbol{\lambda}'\mathbf{G}^{-1} & \left(T\mathsf{p}^{-1}_{\lambda}\mathbf{I}_{Tr}+\mathsf{p}_{\lambda}\boldsymbol{\lambda}'\mathbf{G}^{-1}\boldsymbol{\lambda}\mathbf{1}_{Tr}\mathbf{1}'_{Tr}\right) \end{pmatrix} \tag{62}$$

where, evaluated at $\boldsymbol{\gamma}^{*}$ and δ^{*}:

$$\sigma^2_e\simeq\sigma^2_{\epsilon}+\delta^2\left(1-\mathsf{p}_{\lambda}\right)^2\mathbf{f}'_{\boldsymbol{\lambda}}\boldsymbol{\Sigma}_{22}\mathbf{f}_{\boldsymbol{\lambda}}+r\delta^2\left(1-\mathsf{p}_{\lambda}\right)^2\left(\mathsf{p}_{\lambda}-\left(1-\mathsf{p}_{\lambda}\right)\boldsymbol{\lambda}'\mathbf{f}_{\boldsymbol{\lambda}}\right)^2. \tag{63}$$

In the special case that $\mathsf{p}_{\lambda}=1$, consistent estimates of $\boldsymbol{\gamma}$ result with $T^{-1}\mathbf{G}=\left(1-r\right)\boldsymbol{\Sigma}_{22}$ and $\sigma^2_e=\sigma^2_{\epsilon}$.

Since:

$$\mathsf{F}^{Ts}_{T(1-s)-n}\left(\boldsymbol{\tau}_2=\mathbf{0}\right)=\left(T\left(1-s\right)-n\right)\frac{\widetilde{\boldsymbol{\tau}}'_2\left(T\mathsf{p}^{-1}_{\lambda}\mathbf{I}_{Tr}+\mathsf{p}_{\lambda}\boldsymbol{\lambda}'\mathbf{G}^{-1}\boldsymbol{\lambda}\mathbf{1}_{Tr}\mathbf{1}'_{Tr}\right)^{-1}\widetilde{\boldsymbol{\tau}}_2}{Ts\widetilde{\sigma}^2_e},$$

using:

$$\mathbf{1}'_{Tr}\left(\mathbf{I}_{Tr}+x\mathbf{1}_{Tr}\mathbf{1}'_{Tr}\right)^{-1}\mathbf{1}_{Tr}=\frac{Tr}{1+Trx}$$

then:

$$T^{-1}\mathsf{p}_{\lambda}\mathbf{1}'_{Tr}\left(\mathbf{I}_{Tr}+T^{-1}\mathsf{p}_{\lambda}^{2}\boldsymbol{\lambda}'\mathbf{G}^{-1}\boldsymbol{\lambda}\mathbf{1}_{Tr}\mathbf{1}'_{Tr}\right)^{-1}\mathbf{1}_{Tr}=\frac{r\mathsf{p}_{\lambda}}{1+r\mathsf{p}_{\lambda}^{2}\boldsymbol{\lambda}'\mathbf{G}^{-1}\boldsymbol{\lambda}} \tag{64}$$

so an approximate explicit expression for the non-centrality of the F-test is:

$$\varphi_{s,F}^{2}\simeq\frac{\left(T\left(1-s\right)-n\right)\mathsf{p}_{\lambda}r\left(\delta^{*}\right)^{2}}{Ts\sigma_{e}^{2}\left(1+r\mathsf{p}_{\lambda}^{2}\boldsymbol{\lambda}'\mathbf{G}^{-1}\boldsymbol{\lambda}\right)}. \tag{65}$$

All the factors affecting the potency of the automatic test are clear in (65). The important mistake is missing relevant impulses: when $\mathsf{p}_{\lambda}<1$ in (59), then $\sigma_{e}^{2}>\sigma_{\epsilon}^{2}$, so φ_{s}^{2} falls rapidly with p_{λ}. Consequently, a relatively loose first-stage significance level seems sensible, e.g., 2.5%. The potency is not monotonic in s since the degrees of freedom of the F-test alter: a given value of δ achieved by a larger s will have lower potency than that from a smaller $s>r$.

For numerical calculations, we allow on average that $\alpha_{1}T$ random extra impulses are retained and $\left(1-\mathsf{p}_{\lambda}\right)rT$ relevant are omitted, and approximate $\mathsf{F}^{Ts}_{T(1-s)-n}\left(\varphi_{s,F}^{2}\right)$ by its numerator as a $\chi_{k}^{2}\left(\varphi_{s}^{2}\right)$ for $\varphi_{s}^{2}=k\varphi_{s,F}^{2}$ using $k=Ts$ by:

$$\mathsf{P}\left[\chi_{k}^{2}\left(k\varphi_{s}^{2}\right)>c_{\alpha_{2}}\mid\mathsf{H}_{1}\right]\simeq\mathsf{P}\left[\chi_{m}^{2}\left(0\right)>h^{-1}c_{\alpha_{2}}\right] \tag{66}$$

where:

$$h=\frac{k+2\varphi_{s}^{2}}{k+\varphi_{s}^{2}}\ \text{ and }\ m=\frac{k+\varphi_{s}^{2}}{h}. \tag{67}$$

Some insight can be gleaned when $n=2$, so $G=\left(\left(1-\mathsf{p}_{\lambda}\right)r\lambda^{2}+\left(1-\mathsf{p}_{\lambda}r\right)\sigma_{22}\right)$ and approximately:

$$\varphi_{s}^{2}\simeq\frac{T\left(1-r\right)r\mathsf{p}_{\lambda}\left(\delta^{*}\right)^{2}}{\sigma_{e}^{2}\left(1+r\mathsf{p}_{\lambda}^{2}G^{-1}\lambda^{2}\right)}\underset{\lambda\to\infty}{\longrightarrow}\frac{T\left(1-r\right)^{2}\left(\beta-\gamma\right)^{2}\sigma_{22}}{\sigma_{\epsilon}^{2}} \tag{68}$$

where the last expression shows the outcome for large λ so $\mathsf{p}_{\lambda}\to1$, which reflects the violation of weak exogeneity, $\left(\beta-\gamma\right)^{2}$, the signal-noise ratio, $\sigma_{22}/\sigma_{\epsilon}^{2}$, the loss from longer break lengths $\left(1-r\right)^{2}$, and the sample size, T. The optimal value of the non-centrality, φ_{r}^{2}, for a known break date and form–so the single variable $1_{\{t>T_{1}\}}$ is added–is:

$$\varphi_{r}^{2}=\frac{Tr\delta^{2}}{\sigma_{\epsilon}^{2}\left(1+r\sigma_{22}^{-1}\lambda^{2}\right)}\underset{\lambda\to\infty}{\longrightarrow}\frac{T\sigma_{22}\left(\beta-\gamma\right)^{2}}{\sigma_{\epsilon}^{2}} \tag{69}$$

Despite the nature of adding Tr separate impulses when nothing is known about the existence or timing of a failure of super exogeneity, so $\varphi_{s}^{2}<\varphi_{r}^{2}$, their powers converge rapidly as the break magnitude λ grows, when r is not too large. The numerical evaluations of 68 in Table 4 below are reasonably accurate.

8 Co-breaking based tests

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

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

$$y_t = \mu_0 + \boldsymbol{\beta}'\mathbf{z}_t + \sum_{j=1}^{s} \phi_j 1_{t_j} + \nu_t. \tag{70}$$

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

$$\mathbf{z}_t = \boldsymbol{\delta}_0 + \sum_{j \in S_{\alpha_1}} \boldsymbol{\delta}_j 1_{t_j} + u_t \tag{71}$$

then the test tries to ascertain whether the timing of the impulses in (70) and (71) overlaps. For example, a perfect match would be strong evidence against super exogeneity, corresponding to the result above that the significance of the marginal-model impulses in the conditional rejects super exogeneity.

9 Simulating the potencies of the automatic super-exogeneity test

We undertook simulation analyses using the bivariate relationship in section 5.1 for a failure of weak exogeneity under non-constancy. Consider violations of super exogeneity due to a failure of weak exogeneity in:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} \beta\mu_{2,t} \\ \mu_{2,t} \end{pmatrix}, \begin{pmatrix} 21 & 10 \\ 10 & 5 \end{pmatrix} \right] \tag{72}$$

so $\gamma = 2$ and $\omega^2 = 1$ but $\beta \neq \gamma$, with a level shift at T_1 in the marginal:

$$\mu_{2,t} = \lambda 1_{\{t>T_1\}} \quad \text{so} \quad \mu_{1,t} = \beta\lambda 1_{\{t>T_1\}}. \tag{73}$$

We vary: $d = \lambda/\sqrt{\sigma_{22}}$ over the values 1, 2, 2.5, 3 and 4; β over 0.75, 1, 1.5 and 1.75, reducing the extent of departure from weak exogeneity; two sample sizes ($T = 100$ and $T = 300$) which have varying break points, T; and the significance levels α_1 and α_2 in the marginal and conditional. A partition of $T/2$ was always used for the impulse saturation in the marginal model, and $M = 10000$ replications.

d/β	0.75	1	1.5	1.75
1.0	0.191	0.153	0.078	0.054
2.0	0.972	0.936	0.529	0.150
2.5	1.000	0.993	0.917	0.339
3.0	1.000	1.000	0.998	0.653
4.0	1.000	1.000	1.000	0.967

Table 2: Level shift at $T_1 = 250$, $T = 300$, $\alpha_1 = \alpha_2 = 0.05$

Table 2 reports the empirical null rejection frequencies of the F-test when $T = 300$ is used with 5% significance levels in both the marginal and conditional models, for a level shift at $T_1 = 250$, so $k = 50$ and $r = 1/6$. The potency of the test increases with the increase in $\beta - \gamma$, as expected, and increases with the magnitude of the level shift d. Even moderate violations of the null are detectable for level shifts of 2.5σ or larger.

Table 3 shows the impact of reducing k to 25 *cet. par.* The empirical potency is never smaller for the shorter break, so the degrees of freedom of the F-test are important.

Using more stringent significance levels of $\alpha_1 = \alpha_2 = 2.5\%$ naturally leads to a less potent test than the 5% in Table 2, although the detection probabilities still rise rapidly with the break magnitude,

d/β	0.75	1	1.5	1.75
1.0	0.377	0.274	0.097	0.060
2.0	1.000	0.997	0.803	0.238
2.5	1.000	1.000	0.990	0.504
3.0	1.000	1.000	1.000	0.797
4.0	1.000	1.000	1.000	0.984

Table 3: Level shift at $T_1 = 275$, $T = 300$, $\alpha_1 = \alpha_2 = 0.05$

d/β	0.75	1	1.5	1.75
1	0.081 *(0.082)*	0.065 *(0.059)*	0.035 *(0.032)*	0.026 *(0.027)*
2	0.717 *(0.990)*	0.612 *(0.915)*	0.220 *(0.220)*	0.062 *(0.053)*
2.5	0.977 *(1.000)*	0.953 *(1.000)*	0.616 *(0.660)*	0.143 *(0.112)*
3	1.000 *(1.000)*	0.999 *(1.000)*	0.953 *(0.970)*	0.372 *(0.241)*
4	1.000 *(1.000)*	1.000 *(1.000)*	1.000 *(1.000)*	0.908 *(0.607)*

Table 4: Level shift at $T_1 = 250$, $T = 300$, $\alpha_1 = \alpha_2 = 0.025$

and even relatively mild departures from weak exogeneity are detected at the break magnitude of $d = 4$. The italic numbers in parentheses report the numerical evaluation of the analytic potency from (68) as a typical example, and the response surface in (74) checks its explanatory ability.

$$\begin{aligned}
\log(\widehat{\mathsf{p}}) &= \underset{(0.04)}{0.84}\ \log(\mathsf{p}^*) + \underset{(0.07)}{0.35}\ \lambda + \underset{(0.04)}{0.21}\ \delta - \underset{(0.13)}{0.50}\ \beta \\
\mathsf{F}_{\mathsf{het}}(8,7) &= 0.35\ \ \mathsf{F}_{\mathsf{reset}}(1,91) = 1,15) = 0.66 \\
\mathsf{R}^2 &= 0.994\ \ \widehat{\sigma}_p = 0.116\ \ \chi^2_{\mathsf{nd}}(2) = 1.06
\end{aligned} \tag{74}$$

Here, R^2 is the squared multiple correlation (when including a constant), $\widehat{\sigma}_p$ is the residual standard deviation, coefficient standard errors are shown in parentheses, the diagnostic tests are of the form $\mathsf{F}_{\mathsf{j}}(k, T-l)$ which denotes an approximate F-test against the alternative hypothesis j for: heteroskedasticity ($\mathsf{F}_{\mathsf{het}}$: see White, 1980); the RESET test ($\mathsf{F}_{\mathsf{reset}}$: see Ramsey, 1969); and $\chi^2_{\mathsf{nd}}(2)$ is a chi-square test for normality (see Doornik and Hansen, 1994); below we also present k^{th}-order serial correlation (F_{ar}: see Godfrey, 1978); k^{th}-order autoregressive conditional heteroskedasticity ($\mathsf{F}_{\mathsf{arch}}$: see Engle, 1982); $\mathsf{F}_{\mathsf{Chow}}$ for parameter constancy over k periods (see Chow, 1960); and *SC* is the Schwarz criterion (see Schwarz, 1978); * and ** denote significant at 5% and 1% respectively. Figure 5 records the response surface fitted and actual values, their cross-plot, the residuals scaled by $\widehat{\sigma}$, and their histogram and density with N[0,1] for comparison.

We now turn to the effect of sample size on potency. Table 5 reports the results for significance levels of 2.5% in both marginal and conditional models when $T = 100$ and $T_1 = 80$, so $k = 20$.

d/β	0.75	1	1.5	1.75
1	0.027	0.027	0.026	0.022
2	0.114	0.098	0.054	0.034
2.5	0.392	0.349	0.159	0.055
3	0.757	0.715	0.434	0.112
4	0.996	0.994	0.949	0.418

Table 5: Level shift at $T_1 = 80$, $T = 100$, $\alpha_1 = \alpha_2 = 0.025$

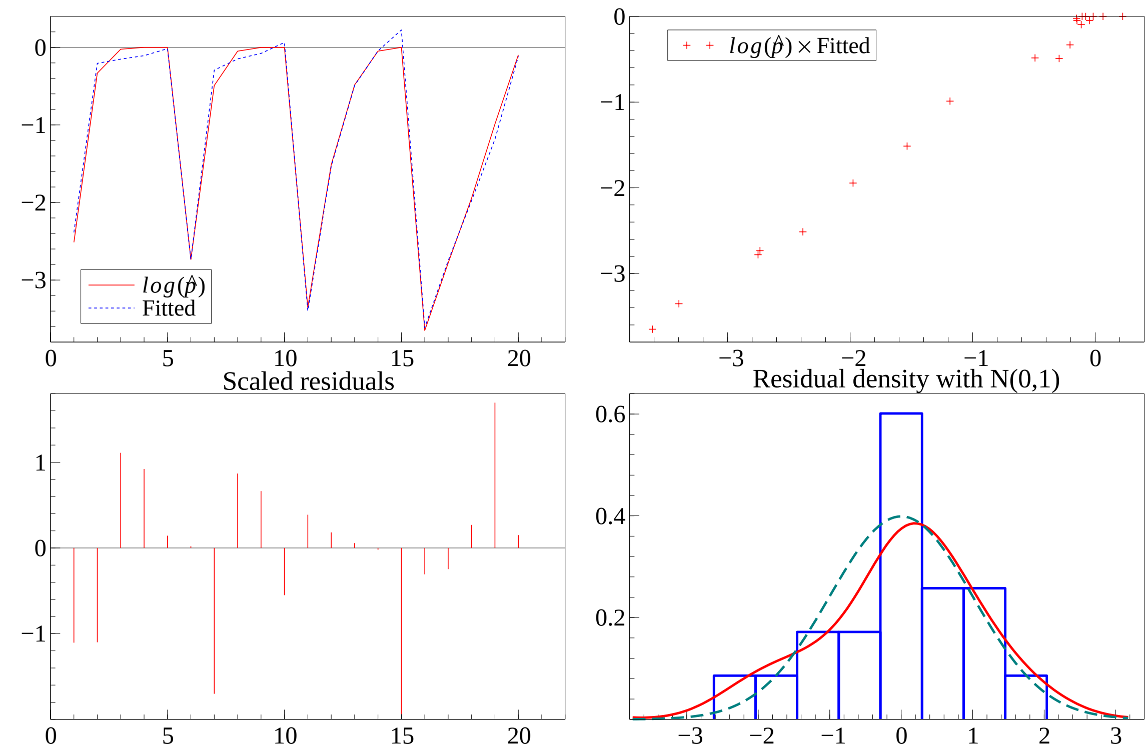

Figure 5: Response surface outcomes for (74)

The test still has reasonable potency for moderate violations of weak exogeneity when breaks are at least 3σ, although there is a loss of potency with the reduction in sample size. The trade off between length of break and potency remains as shown in table 6 for $k = 30$, beginning at observation 71 (small breaks have negligible potency). However, the potency is higher at the larger breaks and smaller weak exogeneity violations, so the impacts of the various determinants are non-monotonic.

d/β	0.75	1	1.5	1.75
2.5	0.260	0.245	0.174	0.118
3.0	0.708	0.680	0.486	0.221
4.0	0.997	0.995	0.967	0.576

Table 6: Level shift at $T_1 = 70$, $T = 100$, $\alpha_1 = \alpha_2 = 0.025$

9.1 Optimal infeasible impulse-based F-test

The optimal infeasible impulse-based F-test with a known break location in the marginal process is computable in simulations. The tables below use $\alpha_2 = 2.5\%$ for testing in the conditional. The empirical rejection frequencies approximate maximum achievable power for this type of test. When $T = 100$, and the break is a mean shift starting at $T_1 = 80$, the correct 20 impulse indicators are always included in the conditional model. Table 7 reports for the failure of super exogeneity.

Relative to the optimal infeasible test, the automatic test based on saturation of the marginal naturally loses considerable potency for breaks of small magnitudes.

Table 8 shows that for a failure of super exogeneity, even when $\beta = 1.75$, the optimal test power increases with k for breaks of $d = 1$ and 2. Thus, the optimal test exhibits power increasing with break length matching (69).

d/β	0.75	1	1.5	1.75
1	1.000	0.994	0.404	0.083
2	1.000	1.000	0.930	0.247
2.5	1.000	1.000	0.973	0.326
3	1.000	1.000	0.985	0.380
4	1.000	1.000	0.988	0.432

Table 7: Level shift at $T_1 = 0.8T = 80$ with known break location and form

d/k	45	40	30	20	15	10	5
1	0.572	0.563	0.515	0.423	0.348	0.259	0.073
2	0.942	0.938	0.920	0.880	0.828	0.720	0.484

Table 8: Super exogeneity failures at T_1 when $T = 100$ with known break location and form

10 Testing super exogeneity in UK money demand

We test super exogeneity in a model of transactions demand for money in the UK using a sample of quarterly observations over 1964(3) to 1989(2), defined by:

M nominal M1
X real total final expenditure (TFE) at 1985 prices
P TFE deflator
R_n net interest rate on retail sight deposits: 3-month local authority interest rate minus own rate

We use the model in Hendry and Doornik (1994) (also see Hendry, 1979, Hendry and Ericsson, 1991, Boswijk, 1992, Hendry and Mizon, 1993, and Boswijk and Doornik, 2004), and express the variables as a vector autoregressive system. Previous cointegration analyses showed two long-run relationships, but confirmed the long-run weak exogeneity of $\{x_t, \Delta p_t, R_{n,t}\}$ in that 4-variable system. The theoretical basis is a model which links demand for real money, $m - p$ (lowercase denoting logs) to (log) income x (transactions motive) and inflation Δp_t, with the interest rate R_n measuring the opportunity cost of holding money. Commencing from the conditional model of $m - p$ on $\{x_t, \Delta p_t, R_{n,t}\}$ with 2 lags of all variables, constant and trend, undertaking selection with impulse saturation on that equation using *Autometrics* at $\alpha_2 = 1\%$ yields:

$$\begin{aligned}(m-p)_t &= \underset{(0.01)}{0.11}\, x_t - \underset{(0.11)}{0.85}\, \Delta p_t - \underset{(0.08)}{0.44}\, R_{n,t} \\ &\quad + \underset{(0.07)}{0.60}\, (m-p)_{t-1} + \underset{(0.07)}{0.30}\, (m-p)_{t-2} - \underset{(0.10)}{0.27}\, R_{n,t-1} \\ &\quad - \underset{(1.1)}{3.5}\, I_{69(2)} + \underset{(1.1)}{4.3}\, I_{71(1)} + \underset{(1.1)}{3.9}\, I_{73(2)} + \underset{(1.1)}{4.2}\, I_{74(4)} - \underset{(1.1)}{2.8}\, I_{83(3)} \qquad (75)\end{aligned}$$

$$\begin{aligned}\mathsf{F}_{\mathsf{ar}}(5,84) &= 1.90 \;\; \mathsf{F}_{\mathsf{arch}}(4,81) = 0.57 \;\; \mathsf{F}_{\mathsf{het}}(22,66) = 0.35 \;\; \mathsf{F}_{\mathsf{reset}}(1,91) = 0.08 \\ \widehat{\sigma}_{(m-p)} &= 0.010 \;\; \chi^2_{\mathsf{nd}}(2) = 0.76 \;\; \mathsf{F}_{\mathsf{Chow}:81(4)}(30,59) = 1.0 \;\; SC = -5.93\end{aligned}$$

The legend is described in §9. The coefficients of the impulses are multiplied by 100 (so are percentage shifts for $(m-p)_t$, x_t and Δp_t).

Despite a large number of previous studies of UK M1, (75) has a major new result: the puzzle of why transactions demand did not depend on the contemporaneous expenditure for which it was held is resolved by finding that it does – once impulse saturation is able to remove the contaminating perturbations. Moreover, the *PcGive* unit-root test is -12.79^{**} strongly rejecting an absence of cointegration; and the derived long-run expenditure elasticity is 1.02 (0.003), so the match with economic theory has been made much closer. Almost all the impulses have historical interpretations: decimalization began in 1969(2) and was completed in 1971(1); 1973(2) saw the introduction of VAT; 1974(4) was the heart of the first Oil crisis; but 1983(3) is unclear.

Next, we selected the significant impulses in congruent marginal models for $\{x_t, \Delta p_t, R_{n,t}\}$ with two lags of every variable, constant and trend, finding:

$$\begin{aligned} x_t &= \underset{(0.32)}{1.24} + \underset{(0.03)}{0.89}\, x_{t-1} - \underset{(0.03)}{0.14}\, R_{n,t-2} + \underset{(0.0002)}{0.0007}\, t \\ &\quad + \underset{(1.0)}{2.9}\, I_{68(1)} + \underset{(1.0)}{3.6}\, I_{72(4)} + \underset{(1.0)}{4.5}\, I_{73(1)} + \underset{(1.0)}{5.7}\, I_{79(2)} \end{aligned} \tag{76}$$

$$\begin{aligned} \mathsf{F}_{\mathsf{ar}}(5,91) &= 1.50 \;\; \mathsf{F}_{\mathsf{arch}}(4,88) = 1.67 \;\; \mathsf{F}_{\mathsf{het}}(13,82) = 1.26 \;\; \mathsf{F}_{\mathsf{reset}}(1,95) = 0.001 \\ \widehat{\sigma}_x &= 0.010 \;\; \chi^2_{\mathsf{nd}}(2) = 0.05 \end{aligned}$$

$$\begin{aligned} \Delta p_t &= -\underset{(0.29)}{1.9} + \underset{(0.07)}{0.43}\, \Delta p_{t-1} + \underset{(0.03)}{0.21}\, x_{t-1} - \underset{(0.01)}{0.03}\, (m-p)_{t-1} - \underset{(0.0002)}{0.0012}\, t \\ &\quad - \underset{(0.68)}{3.1}\, I_{73(2)} + \underset{(0.65)}{2.5}\, I_{74(2)} \end{aligned} \tag{77}$$

$$\begin{aligned} \mathsf{F}_{\mathsf{ar}}(5,92) &= 0.10 \;\; \mathsf{F}_{\mathsf{arch}}(4,89) = 0.84 \;\; \mathsf{F}_{\mathsf{het}}(16,80) = 0.83 \;\; \mathsf{F}_{\mathsf{reset}}(1,96) = 6.5^{*} \\ \widehat{\sigma}_{\Delta p} &= 0.0064 \;\; \chi^2_{\mathsf{nd}}(2) = 0.22 \end{aligned}$$

$$R_{n,t} = \underset{(0.01)}{0.99}\, R_{n,t-1} + \underset{(1.2)}{3.9}\, I_{73(3)} + \underset{(1.2)}{3.5}\, I_{76(4)} - \underset{(1.2)}{3.6}\, I_{77(1)} - \underset{(1.2)}{3.4}\, I_{77(2)} \tag{78}$$

$$\begin{aligned} \mathsf{F}_{\mathsf{ar}}(5,94) &= 1.08 \;\; \mathsf{F}_{\mathsf{arch}}(4,91) = 1.53 \;\; \mathsf{F}_{\mathsf{het}}(6,92) = 1.85 \;\; \mathsf{F}_{\mathsf{reset}}(1,98) = 3.08 \\ \widehat{\sigma}_{R_n} &= 0.012 \;\; \chi^2_{\mathsf{nd}}(2) = 0.09 \end{aligned}$$

Only one mis-specification test is significant at even the 5% level across these three equations, so we judge these marginal models to be congruent. The impulses were selected for using $\alpha_1 = 1\%$ since although the sample size is only $T = 104$, many impulses were already known to matter from the economic turbulence of the 1970s and 1980s in the UK, and indeed 10 are retained across these three models; surprisingly, the 3-day week loss of output in December 1973 did not show up in (76).

Next, we tested the significance of the 10 retained impulses from (76), (77) and (78) in the same unrestricted conditional model of $(m-p)_t$ as used for selecting (75), but without impulse saturation. This yielded $\mathsf{F}_{\mathsf{SE}}(10,81) = 1.28$ so the new test is far from rejecting: the model with impulses had $SC_{22} = -5.11$, whereas the unrestricted model without any impulses had $SC_{12} = -5.41$, both much poorer than (75). The one impulse in common between marginal and conditional models is $I_{73(2)}$, which entered the equation for Δp_t.

Finally, we repeated the super-exogeneity impulse-saturation based test at $\alpha_1 = 2.5\%$, which now led to 37 impulses being retained across the 3 marginal models, and a test statistic of $\mathsf{F}_{\mathsf{SE}}(37,54) = 1.67^{*}$ that just rejects at 5%, which may be partly due to the small remaining degrees of freedom as $SC_{49} = -4.5$, so the conditional model without any impulses has a substantially smaller value of SC. Moreover, the only one of the impulses in (75) selected in any of these marginal models was again $I_{73(2)}$. Thus, we find minimal evidence against the hypothesis that $\{x_t, \Delta p_t, R_{n,t}\}$ are super exogenous for the parameters of the conditional model for $(m-p)_t$ in (75).

Not rejecting the null of super exogeneity implies that agents did not alter their demand for money behaviour despite quite large changes in the processes generating their conditioning variables. In particular, agents could not have been forming expectations based on the marginal models for any of the three variables. This might be because their near unpredictability led to the use of robust forecasting devices of the general forms discussed by Favero and Hendry (1992) and Hendry and Ericsson (1991):

$$\widehat{x}_{t+1} = x_t;\;\; \widehat{\Delta p}_{t+1} = \Delta p_t;\;\; \widehat{R}_{n,t+1} = R_{n,t}.$$

If so, the apparent conditioning variables are actually the basis for robust 1-step ahead forecasting devices used in the face of unanticipated structural breaks as in Hendry (2006). Consequently, the non-rejection of super exogeneity makes sense, and does not contradict the underlying theory of forward-looking money demand behaviour.

11 Conclusion

An automatically-computable test for super exogeneity based on selecting shifts in the marginal process by impulse saturation to test for related shifts in the conditional has been proposed. The test has the correct null rejection frequency in constant conditional models when the nominal test size, α_1, is not too small in the marginal (e.g. 2.5%) even at small sample sizes, for a variety of marginal processes, both constant and with breaks. The approximate power function was derived analytically for regression models, and helps explain the simulation outcomes. These confirm that the test can detect failures of super exogeneity when weak exogeneity fails and the marginal processes change. Although only a single break was considered in detail, the general nature of the test makes it applicable when there are multiple breaks in the marginal processes, perhaps at different times. We have also derived its behaviour for a failure of invariance when the same test is applied despite the altered data generation process, and shown that it has similar properties.

A test rejection outcome indicates a dependence between the conditional model parameters and those of the marginals, warning about potential mistakes from using the conditional model to predict the outcomes of policy changes that alter the marginal processes by location shifts, which is a common policy scenario.

The empirical application to UK M1 delivered new results in a much-studied illustration, and confirmed the feasibility of the test. The status of super exogeneity was not completely clear cut, but suggested at most a small degree of dependence between the parameters.

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

References

Boswijk, H. P. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.

Boswijk, H. P., and Doornik, J. A. (2004). Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica*, **58**, 440–465.

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

Clements, M. P., and Hendry, D. F. (1994). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 9–52. Oxford: Oxford University Press.

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

Doornik, J. A. (1999). *Object-Oriented Matrix Programming using Ox*. London: Timberlake Consultants Press. 3rd edition.

Doornik, J. A. (2007). Autometrics. Working paper, Economics Department, University of Oxford.

Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.

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

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

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

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

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

Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

Hendry, D. F. (1979). Predictive failure and econometric modelling in macro-economics: The transactions demand for money. In Ormerod, P. (ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

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

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

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in *The Economics of Structural Change*, Hagemann, H. Landesman, M. and Scazzieri, R. (eds.), Edward Elgar, Cheltenham, 2002.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426. Special Issue in Honor of Clive Granger.

Hendry, D. F., and Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.

Hendry, D. F., and Doornik, J. A. (1997). The implications for econometric modelling of forecast failure.

Scottish Journal of Political Economy, **44**, 437–461. Special Issue.

Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

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

Hendry, D. F., and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics*, **25**, 33–51.

Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell. Reprinted in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

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. (2007). Automatic index-based tests of super exogeneity. Unpublished paper, Economics Department, 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**, 735–763.

Johansen, S., and Nielsen, B. (2008). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, p. forthcoming. Oxford: Oxford University Press.

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

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

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

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, **31**, 350–371.

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

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

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

Santos, C., and Hendry, D. F. (2006). Saturation in autoregressive models. *Notas Economicas*, **19**, 8–20.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.