ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

MODEL SELECTION IN UNDER-SPECIFIED EQUATIONS FACING BREAKS

Jennifer L. Castle and David F. Hendry

Number 509
October 2010

Manor Road Building, Oxford OX1 3UQ

Model Selection in Under-specified Equations Facing Breaks

Jennifer L. Castle† and David F. Hendry⋆ *

† Institute for Economic Modelling and Magdalen College, University of Oxford, UK

⋆ Institute for Economic Modelling and Department of Economics, University of Oxford, UK

October 21, 2010

Abstract

Although a general unrestricted model may under-specify the data generation process, especially when breaks occur, model selection can still improve over estimating a prior specification. Impulse-indicator saturation (IIS) can 'correct' non-constant intercepts induced by location shifts in *omitted* variables, which surprisingly leave slope parameters unaltered even when correlated with included variables. However, location shifts in *included* variables do induce changes in slopes when there are correlated omitted variables. IIS acts as a 'robust method' when models are mis-specified, and helps mitigate the adverse impacts of induced location shifts on non-constant intercepts and equation standard errors.

JEL classifications: C51, C22.
KEYWORDS: Model selection, mis-specification, location shifts, impulse-indicator saturation, costs of search, costs of inference, *Autometrics*.

1 Introduction

Omitted variables are a common problem in empirical econometrics, resulting in biased, inconsistent, or non-constant parameter estimates.[1] While often expressed as omitted-variables bias, the key issue is that a different parameter vector is induced by omissions. The theory of reduction describes the operations implicitly applied to the data generating process (DGP) to obtain the local data generating process (LDGP, the generating process in the space of variables under analysis: see e.g., Hendry, 2009). Choosing the set of variables, $\mathbf{x}_t = (\mathbf{y}_t, \mathbf{z}_t)$, for analysis determines the properties of the LDGP, and hence of any models thereof. Omitting from the LDGP any variables that matter in the DGP (i.e., a relevant set $\mathbf{w}_t$) defines a less useful LDGP, denoted LDGP_1, than the 'correct specification', denoted

*This research was supported in part by grants from the Open Society Foundation and the Oxford-Martin School. Contact details: jennifer.castle@magd.ox.ac.uk and david.hendry@nuffield.ox.ac.uk.

[1] Our paper is dedicated to Clive Granger with fond memories of insightful discussions. It builds on the issues of model specification and evaluation, which were the focus of Granger (1999), leading to the exchange about automatic model selection in Granger and Hendry (2005).

LDGP_0. The included conditioning variables $\mathbf{z}_t$ then 'pick up' the correlated parts of the excluded variables $\mathbf{w}_t$. Estimation, in effect, minimizes the costs of omission by 'spreading' what can be captured across the included variables. Then, if $\boldsymbol{\mu}$ is the parameter vector for LDGP_0 when including all relevant conditioning variables, $(\mathbf{z}_t, \mathbf{w}_t)$, and $\boldsymbol{\gamma}$ is the parameter vector for LDGP_1 when the relevant variables, $\mathbf{w}_t$, are omitted, the model approximating LDGP_1 will estimate $\boldsymbol{\gamma}$, which may mistakenly be interpreted as $\boldsymbol{\mu}$. In addition, any model of either LDGP may be under-specified by omitting relevant functions such as lags, $\mathbf{z}_{t-s}$, breaks, or non-linear transformations, $f(\mathbf{w}_t)$, even when the complete set of determining variables is being analyzed.

A 'general-to-specific' (Gets) approach to model selection seeks to mimic the theory of reduction by commencing from a large set of potentially relevant variables, then reducing that set by undertaking multi-path searches to eliminate irrelevant variables. There are two requirements for a successful model selection procedure to result in a close approximation to the LDGP. First, the LDGP must be a 'well specified' process which satisfies properties such as constant parameters and innovation errors at a reasonable lag length, so locating that LDGP will be worthwhile, and inferences are well based during selection. Secondly, the initial general unrestricted model (GUM) that is specified by the econometrician must nest the LDGP. The primary difficulty is satisfying the first criterion, namely that the LDGP is not under-specified for the DGP, as otherwise, this will entail a model that approximates LDGP_1 as opposed to LDGP_0, resulting in mis-specification relative to LDGP_0. The second criterion is now less of a difficulty for empirical modeling, as it has the solution of starting from a more general model that does nest the LDGP. Here, we use an automatic model selection approach, *Autometrics* (see e.g., Doornik, 2009, Hendry and Doornik, 2009), to address a key aspect of that second difficulty, namely induced location shifts. The *Autometrics* algorithm seeks to locate the LDGP for the chosen set of variables, building on Hoover and Perez (1999) and Hendry and Krolzig (2005), by embedding the anticipated LDGP in a much larger model that allows for long lags, non-linearities, and multiple location shifts. Its properties are documented in Castle, Doornik and Hendry (2009, 2010a), Castle and Hendry (2010), and Hendry and Mizon (2010) *inter alia*. The main point of this paper is that by addressing the second difficulty, we can also mitigate some of the worst impacts of the first.

Analytical results for 'omitted-variables bias' in constant DGPs are well-known, but the empirically-relevant setting is one where breaks occur. Structural breaks abound in economic data, and we distinguish between internal breaks, which are shifts in the relationship being modeled, and external breaks,

which affect the DGPs of the unmodeled variables, but leave the model under analysis constant (see Castle, Fawcett and Hendry, 2010b). The latter is the case of interest here in that a constant, congruent model could be developed when the LDGP is correctly formulated, but that might not happen if an inadvertently omitted variable shifts. Our objective is to mitigate the impacts of mis-specifications due to omitted variables that are subject to breaks. We focus on location shifts where the means of variables change abruptly, but the methodology applies when any aspect of the distribution, or indeed the entire distribution, shifts. Since the timing of such breaks is usually unknown for unknown omitted variables, a 'portmanteau' approach is required that eliminates potential location shifts at any point in the sample, so we use the location shift elimination procedure of impulse-indicator saturation (IIS) to detect and partial out breaks when the number, timing and magnitude of shifts in both the included and omitted variables are unknown.

IIS includes an impulse indicator for every observation in the set of candidate regressors, so adds T variables when there are T observations. The properties of IIS in an IID setting are analyzed in Hendry, Johansen and Santos (2008), and extended by Johansen and Nielsen (2009) to both stationary and unit-root autoregressions. Castle *et al.* (2009) consider its ability to detect multiple location shifts, and Hendry and Santos (2010) apply IIS to develop an automatic test for super exogeneity (see Engle, Hendry and Richard, 1983, and Engle and Hendry, 1993). Johansen and Nielsen (2009) also relate IIS to robust estimation, and show that it is a highly efficient method: under the null of no breaks, outliers or data contamination, the cost of applying IIS is the loss of αT of the sample, where α is the significance level, so at $\alpha = 0.01$, is 99% efficient for $T = 100$ despite including 100 'irrelevant' impulse indicators in the search set. While IIS entails more candidate variables than observations, $N + T > T$, when N is the number of conditioning variables initially considered, this case is feasible as *Autometrics* undertakes expanding as well as contracting searches.

Here we are concerned with the behavior of IIS under the alternative in mis-specified equations where the breaks and outliers are induced by omitting relevant variables that shift. If IIS is not needed, the costs are almost negligible (as just explained); and if it is needed, we show that the most pernicious effects of induced location shifts on non-constant intercepts and equation standard errors are corrected. Both static and dynamic equations are investigated.

The structure of the paper is as follows. Section 2 provides the theory derivations when a static model is mis-specified by omitting variables that have location shifts, interacting with a break in the

included variables. Section 3 discusses the implications of these mis-specifications. Section 4 then considers the possible role of impulse-indicator saturation in removing the effects of such location shifts. Section 5 provides a Monte Carlo study matching the theory to evaluate the magnitudes of the effects when the LDGP equation is constant and only the included and omitted variables change. Section 6 then investigates mis-specified dynamic models where both included and omitted variables have location shifts, with simulation findings in §7. Section 8 briefly discusses the distinction between omitting DGP variables when creating the LDGP, and mis-specifying the GUM for that LDGP. Section 9 concludes. Appendix §10 provides some of the mathematical calculations.

2 Breaks in included and excluded variables

To highlight the impacts of changes affecting mis-specified models in a setting where analytic results can be obtained, we first consider a linear, static, constant-parameter, conditional LDGP equation with white-noise errors:

$$y_t = \boldsymbol{\beta}_1' \mathbf{z}_t + \boldsymbol{\beta}_2' \mathbf{w}_t + \epsilon_t \tag{1}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ with:

$$\begin{pmatrix} \mathbf{z}_t \\ \mathbf{w}_t \end{pmatrix} \sim \mathsf{IN}_{k_1+k_2}\left[\begin{pmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\delta}_t \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right] \tag{2}$$

where the regression model is mis-specified as:

$$y_t = \gamma_0 + \boldsymbol{\gamma}_1' \mathbf{z}_t + e_t \tag{3}$$

including an intercept since $\mathsf{E}\left[\mathbf{w}_t\right] \neq \mathbf{0}$. Thus, $\mathbf{w}_t$ is unknowingly omitted in (3), and both sets of variables have one-off location shifts.

To keep the analysis tractable, the breaks occur just once at times $1 < T^0 < T$ and $1 < T^* < T$:

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\mu}_1 & t < T^0 \\ \boldsymbol{\mu}_2 & t \geq T^0 \end{cases} \quad \text{and} \quad \boldsymbol{\delta}_t = \begin{cases} \boldsymbol{\delta}_1 & t < T^* \\ \boldsymbol{\delta}_2 & t \geq T^* \end{cases} \tag{4}$$

The key to understanding the induced non-constancy is what happens to the relationship between the

omitted and included variables. From (2):

$$\mathbf{w}_t = (\boldsymbol{\delta}_t - \boldsymbol{\Psi}\boldsymbol{\mu}_t) + \boldsymbol{\Psi}\mathbf{z}_t + \mathbf{u}_t \tag{5}$$

where $\mathsf{E}[\mathbf{z}_t\mathbf{u}_t'] = \mathbf{0}$ and $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. Thus from (3) and (5), the reduced LDGP (LDGP$_1$) is:

$$\begin{aligned} y_t &= \boldsymbol{\beta}_1'\mathbf{z}_t + \boldsymbol{\beta}_2'\mathbf{w}_t + \epsilon_t = \boldsymbol{\beta}_1'\mathbf{z}_t + \boldsymbol{\beta}_2'\left((\boldsymbol{\delta}_t - \boldsymbol{\Psi}\boldsymbol{\mu}_t) + \boldsymbol{\Psi}\mathbf{z}_t + \mathbf{u}_t\right) + \epsilon_t \\ &= \boldsymbol{\beta}_2'\left(\boldsymbol{\delta}_t - \boldsymbol{\Psi}\boldsymbol{\mu}_t\right) + \left(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2'\boldsymbol{\Psi}\right)\mathbf{z}_t + \boldsymbol{\beta}_2'\mathbf{u}_t + \epsilon_t. \end{aligned} \tag{6}$$

Analysis of the full-sample estimators of (6) requires sub-sample derivations due to the multiple breaks in $\pi_t = \boldsymbol{\beta}_2'\left(\boldsymbol{\delta}_t - \boldsymbol{\Psi}\boldsymbol{\mu}_t\right)$. We take $T^* < T^0$ for explicit calculations, and report the relevant calculations in the appendix (§10). Full-sample estimation of (3) then yields:

$$\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\boldsymbol{\gamma}}_1 \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^T \mathbf{z}_t' \\ \sum_{t=1}^T \mathbf{z}_t & \sum_{t=1}^T \mathbf{z}_t\mathbf{z}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T \mathbf{z}_t y_t \end{pmatrix}.$$

Letting $\lambda = \left(T^0 - 1\right)/T$ and $\kappa = \left(T^* - 1\right)/T$, $\mathbf{r} = \left(\lambda\boldsymbol{\mu}_1 + (1-\lambda)\,\boldsymbol{\mu}_2\right)$, $\mathbf{s} = \left(\kappa\boldsymbol{\delta}_1 + (1-\kappa)\,\boldsymbol{\delta}_2\right)$ and $\mathbf{M} = \left(\lambda\boldsymbol{\mu}_1\boldsymbol{\delta}_1' + (\kappa-\lambda)\,\boldsymbol{\mu}_2\boldsymbol{\delta}_1' + (1-\kappa)\,\boldsymbol{\mu}_2\boldsymbol{\delta}_2'\right)$:

$$\begin{aligned} \mathsf{E}\left[\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\boldsymbol{\gamma}}_1 \end{pmatrix}\right] &\simeq \begin{pmatrix} 1 & \mathbf{r}' \\ \mathbf{r} & \left(\mathbf{H} - \mathbf{r}\mathbf{r}'\right) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\beta}_1'\mathbf{r} + \boldsymbol{\beta}_2'\mathbf{s} \\ \left(\mathbf{H} - \mathbf{r}\mathbf{r}'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \mathbf{M}\right)\boldsymbol{\beta}_2 \end{pmatrix} \\ &= \begin{pmatrix} \left(\mathbf{s}' - \mathbf{r}'\mathbf{H}^{-1}\left(\boldsymbol{\Sigma}_{12} + \lambda\left(1-\kappa\right)\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\left(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\right)'\right)\right)\boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_1 + \mathbf{H}^{-1}\left(\boldsymbol{\Sigma}_{12} + \lambda\left(1-\kappa\right)\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\left(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\right)'\right)\boldsymbol{\beta}_2 \end{pmatrix} \\ &= \begin{pmatrix} \gamma_{0,p} \\ \boldsymbol{\gamma}_{1,p} \end{pmatrix} \end{aligned} \tag{7}$$

as:

$$\mathbf{M} - \mathbf{r}\mathbf{s}' = \lambda\left(1-\kappa\right)\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\left(\boldsymbol{\delta}_1' - \boldsymbol{\delta}_2'\right)$$

and:

$$\begin{aligned}\mathbf{H} &= \boldsymbol{\Sigma}_{11} + \lambda\boldsymbol{\mu}_1\boldsymbol{\mu}_1' + (1-\lambda)\,\boldsymbol{\mu}_2\boldsymbol{\mu}_2' - \mathbf{r}\mathbf{r}' \\ &= \boldsymbol{\Sigma}_{11} + \lambda\,(1-\lambda)\,(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\,(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\,.\end{aligned}$$

Thus, in the special case when $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{r} = \boldsymbol{\mu}$, then $\mathbf{H} = \boldsymbol{\Sigma}_{11}$ and hence:

$$\left(\begin{array}{c}\gamma_{0,p}\\ \boldsymbol{\gamma}_{1,p}\end{array}\right)_{\boldsymbol{\mu}_1=\boldsymbol{\mu}_2} = \left(\begin{array}{c}\left((\kappa\boldsymbol{\delta}_1 + (1-\kappa)\,\boldsymbol{\delta}_2)' - \boldsymbol{\mu}'\boldsymbol{\Psi}'\right)\boldsymbol{\beta}_2\\ (\boldsymbol{\beta}_1 + \boldsymbol{\Psi}'\boldsymbol{\beta}_2)\end{array}\right). \tag{8}$$

Conversely, when $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{s} = \boldsymbol{\delta}$, then $\mathbf{H} \neq \boldsymbol{\Sigma}_{11}$, and still depends on $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, leading to:

$$\left(\begin{array}{c}\gamma_{0,p}\\ \boldsymbol{\gamma}_{1,p}\end{array}\right)_{\boldsymbol{\delta}_1=\boldsymbol{\delta}_2} = \left(\begin{array}{c}\left(\boldsymbol{\delta}' - \mathbf{r}'\mathbf{H}^{-1}\boldsymbol{\Sigma}_{12}\right)\boldsymbol{\beta}_2\\ \boldsymbol{\beta}_1 + \mathbf{H}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}_2\end{array}\right). \tag{9}$$

We now consider the consequences of the formulae in (7), (8), and (9).

3 Implications

As (9) shows, the slope parameter $\boldsymbol{\gamma}_{1,p}$ shifts, changing the bias, if the DGP of the *included* variables, $\mathbf{z}_t$, changes when $\mathbf{w}_t$ is omitted, irrespective of changes in the parameters of the latter's DGP. Surprisingly, as (8) shows, $\boldsymbol{\gamma}_{1,p}$ does not shift if $\mathbf{w}_t$ alone changes when the $\{\mathbf{z}_t\}$ process is constant. Thus, a break in the included variables' DGP alters both slopes and intercepts, as the biases in their estimated coefficients lead to an induced non-constancy in the estimated slope parameters. Consequently, when there is any mis-specification due to an omitted variable that is correlated with included variables, breaks in any *included* variable will cause the model's parameters to change. Essentially, therefore, if breaks occur intermittently, estimated models will be constant only if all substantive variables are included. Section 4 discusses the extent to which this rather disastrous implication can be offset by impulse-indicator saturation, but it also argues strongly for commencing from very general models as proposed in Castle *et al.* (2010a).

Next, changes in the intercept terms associated with $\boldsymbol{\beta}_1$ cancel in all three cases. However, even

when $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $\gamma_{0,p}$ shifts as seen in (8) to:

$$\left(\gamma_{0,p}\right)_{\boldsymbol{\mu}_1=\boldsymbol{\mu}_2} = \left(\kappa\boldsymbol{\delta}_1' + (1-\kappa)\,\boldsymbol{\delta}_2' - \boldsymbol{\mu}'\boldsymbol{\Psi}'\right)\boldsymbol{\beta}_2.$$

Thus, there are locations shifts in (3) whenever the omitted variables shift. Conversely, when $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2$, from (9):

$$\left(\gamma_{0,p}\right)_{\boldsymbol{\delta}_1=\boldsymbol{\delta}_2} = \left(\boldsymbol{\delta}' - (\lambda\boldsymbol{\mu}_1 + (1-\lambda)\,\boldsymbol{\mu}_2)'\,\mathbf{H}^{-1}\boldsymbol{\Sigma}_{12}\right)\boldsymbol{\beta}_2$$

so the model's intercept experiences a location shift when $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.

Finally, in the special case $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, and using population values, an analysis of the estimated equation error variance is possible. First, let $\widehat{e}_t = y_t - \widehat{y}_t$, then for $t = 1, \ldots, T^* - 1$, $\{\widehat{e}_t\}_{\boldsymbol{\mu}_1=\boldsymbol{\mu}_2}$ is given by:

$$\begin{aligned}
\{\widehat{e}_t\}_{\boldsymbol{\mu}_1=\boldsymbol{\mu}_2} &= \boldsymbol{\beta}_1'\mathbf{z}_t + \boldsymbol{\beta}_2'\mathbf{w}_t + \epsilon_t - \gamma_{0,p} - \boldsymbol{\gamma}_{1,p}'\mathbf{z}_t \\
&= \left(\boldsymbol{\delta}_1' - \boldsymbol{\mu}'\boldsymbol{\Psi}'\right)\boldsymbol{\beta}_2 + \mathbf{z}_t'\left(\boldsymbol{\beta}_1 + \boldsymbol{\Psi}\boldsymbol{\beta}_2\right) + \mathbf{u}_t'\boldsymbol{\beta}_2 + \epsilon_t \\
&\quad - \left(\left(\kappa\boldsymbol{\delta}_1' + (1-\kappa)\,\boldsymbol{\delta}_2' - \boldsymbol{\mu}'\boldsymbol{\Psi}'\right)\boldsymbol{\beta}_2 + \mathbf{z}_t'\left(\boldsymbol{\beta}_1 + \boldsymbol{\Psi}\boldsymbol{\beta}_2\right)\right) \\
&= (1-\kappa)\left(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\right)\boldsymbol{\beta}_2 + \mathbf{u}_t'\boldsymbol{\beta}_2 + \epsilon_t
\end{aligned}$$

and for $t = T^*, \ldots, T$:

$$\{\widehat{e}_t\}_{\boldsymbol{\mu}_1=\boldsymbol{\mu}_2} = \kappa\left(\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1\right)\boldsymbol{\beta}_2 + \mathbf{u}_t'\boldsymbol{\beta}_2 + \epsilon_t$$

so overall the residual variance is inflated by the induced non-constancy over $\sigma_\epsilon^2 + \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2$ by:

$$\kappa\,(1-\kappa)\,\boldsymbol{\beta}_2'\left(\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1\right)\left(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\right)'\boldsymbol{\beta}_2 \tag{10}$$

thereby lowering the precision of estimation, possibly considerably, as well as creating a heteroskedastic residual at the unknown break point of the omitted variables. Non-constancy in $\boldsymbol{\gamma}_{1,p}$ would exacerbate that problem.

The aim of applying IIS to such settings is to remove these location-shift induced non-constancies in intercepts and equation standard errors, as we now discuss.

4 Impulse-indicator saturation

The theory of IIS is derived under the null of no outliers or location shifts. In the simplest analysis (the 'split-half' approach), a regression initially only includes the first $T/2$ of these indicators together with the relevant regressors, when $(N + T/2) < T$. By dummying out the first half of the observations, estimates are based on the remaining data, so any observations in the first half that are discrepant will result in significant indicators–in essence, that lies behind the approach in Salkever (1976) for testing parameter constancy using indicators. The location of the significant indicators is recorded, then the first $T/2$ are replaced by the second half and the procedure repeated. The two sets of significant indicators are finally added to the model with the N regressors for selection of the indicators that remain significant (possibly also selecting over the non-dummy variables). Then Hendry *et al.* (2008) and Johansen and Nielsen (2009) show that αT impulse indicators will be retained on average at significance level α. Setting $\alpha \leq r/T$ maintains the average false null retention at r outliers, equivalent to 'losing' r observations, which is a small efficiency loss. Under the alternative, IIS can detect outliers and multiple location shifts, including breaks close to the start and end of the sample (see Castle *et al.*, 2009, for Monte Carlo evidence on the detectability of internal breaks), and hence in our setting, IIS can potentially remove the intercept non-constancy, as well as mitigate the poor, and non-constant, fit.

From (1) and (2), using (5), the reduced LDGP is given in (6) as:

$$y_t = \boldsymbol{\beta}_2' \left(\boldsymbol{\delta}_t - \boldsymbol{\Psi} \boldsymbol{\mu}_t \right) + \left(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2' \boldsymbol{\Psi} \right) \mathbf{z}_t + \boldsymbol{\beta}_2' \mathbf{u}_t + \epsilon_t$$

The inconsistent coefficient of $\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2' \boldsymbol{\Psi}$ on $\mathbf{z}_t$ is a 'classical' omitted-variables bias problem. IIS will not alter this, but can correct the non-constancy of the intercept, and hence of changes in the *estimated* slope and the goodness of fit. The 'optimal' solution to the intercept shift and its attendant changes would be to include step dummies that changed at T^* and T^0, which is an infeasible knowledge level when it is not even known that $\mathbf{w}_t$ is relevant. Thus, we will examine the extent to which the addition of a complete set of impulse indicators to the candidate regressors can mimic that effect here for induced shifts.

5 Static equation simulation results

Monte Carlo experiments are used to assess the implications of under-specification when there are non-constant parameters as outlined in §3. The DGP is given by (1) and (2) for scalar z_t and w_t. The baseline parameter values are: $\beta_1 = 1$, $\beta_2 = 1$, $\sigma_\epsilon^2 = 1$, $\Sigma_{12} = \Psi = 0.5$, $\Sigma_{11} = \Sigma_{22} = 1$ and $T = 100$. $M = 1000$ replications are undertaken. Since $\Psi = 0.5$, the implied location shift in (6) from a break in z_t is half as large as that induced by the same change in w_t.

5.1 Under the null of correct specification

First, we assess estimation and selection of the correctly specified model to establish properties under the null when the model matches the DGP, so there are no induced shifts, but IIS is nevertheless applied. Two specifications are considered including a break in one regressor and a break in both regressors at different dates. For the first specification, we assume a break in w_t with parameters given by $T^0 = 0$, $\mu_1 = \mu_2 = 0$ and $T^* = 81$, $\delta_1 = 0$, $\delta_2 = 5$. The parameters for the specification in which there is a break in both variables are given by $T^0 = 91$, $\mu_1 = 0$, $\mu_2 = -5$, $T^* = 81$, $\delta_1 = 0$ and $\delta_2 = 5$. The outcomes should not depend on whether or not $T^* \lessgtr T^0$.

The model augments the DGP in (1) with an intercept:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 w_t + \nu_t \tag{11}$$

The step-shift dummies are given by:

$$D_{z,t} = \begin{cases} 0 & t < T^0 \\ 1 & t \geq T^0 \end{cases} \quad \text{and} \quad D_{w,t} = \begin{cases} 0 & t < T^* \\ 1 & t \geq T^* \end{cases} \tag{12}$$

with parameters $\beta_{s,z}$ and $\beta_{s,w}$ respectively. As break dates are unlikely to be known, particularly if the break occurs in an omitted variable, impulse-indicator saturation is a robust estimation method in this setting. Results are reported in table 1. The columns refer to:

(a) estimation of the correct LDGP equation (11);

(b) selecting from (11);

(c) selection from (11) augmented with step-shift dummies at the break dates;

(d) selection from (11) also applying IIS.

Selection is undertaken using *Autometrics* at the $\alpha = 1\%$ significance level, and ι refers to the number of impulse indicators retained on average per replication in (d). In case (c), the step-shift dummies are unnecessary, so their retention rate checks the size.

	(a)	(b)	(c)	(d)
Constant z_t, break in w_t				
$\widehat{\beta}_0$	$0.001\,[0.107]$ $(0.109)[(0.008)]$	$0.000\,[0.029]$ $(0.001)[(0.010)]$	$0.000\,[0.029]$ $(0.001)[(0.010)]$	$-0.001\,[0.030]$ $(0.001)[(0.010)]$
$\widehat{\beta}_1$	$1.001\,[0.102]$ $(0.101)[(0.007)]$	$1.001\,[0.101]$ $(0.100)[(0.007)]$	$1.001\,[0.102]$ $(0.100)[(0.007)]$	$1.002\,[0.105]$ $(0.096)[(0.008)]$
$\widehat{\beta}_2$	$1.002\,[0.046]$ $(0.044)[(0.003)]$	$1.002\,[0.041]$ $(0.041)[(0.003)]$	$1.001\,[0.055]$ $(0.042)[(0.009)]$	$1.000\,[0.043]$ $(0.039)[(0.003)]$
$\widehat{\beta}_{s,w}$			$0.006\,[0.204]$ $(0.008)[(0.071)]$	
$\widehat{\sigma}_\epsilon$	$0.998\,[0.071]$	$0.998\,[0.071]$	$0.997\,[0.071]$	$0.957\,[0.078]$
ι	-	-	-	1.12
Break in z_t and w_t				
$\widehat{\beta}_0$	$0.001\,[0.108]$ $(0.109)[(0.008)]$	$0.000\,[0.030]$ $(0.001)[(0.010)]$	$0.000\,[0.032]$ $(0.001)[(0.011)]$	$-0.001\,[0.034]$ $(0.001)[(0.011)]$
$\widehat{\beta}_1$	$1.000\,[0.063]$ $(0.060)[(0.004)]$	$1.000\,[0.062]$ $(0.060)[(0.004)]$	$0.999\,[0.068]$ $(0.060)[(0.006)]$	$1.001\,[0.067]$ $(0.057)[(0.005)]$
$\widehat{\beta}_2$	$1.002\,[0.047]$ $(0.046)[(0.003)]$	$1.002\,[0.045]$ $(0.044)[(0.003)]$	$1.002\,[0.058]$ $(0.045)[(0.008)]$	$1.002\,[0.047]$ $(0.042)[(0.004)]$
$\widehat{\beta}_{s,z}$			$-0.007\,[0.186]$ $(0.007)[(0.066)]$	
$\widehat{\beta}_{s,w}$			$0.000\,[0.233]$ $(0.009)[(0.080)]$	
$\widehat{\sigma}_\epsilon$	$0.998\,[0.071]$	$0.997\,[0.071]$	$0.997\,[0.071]$	$0.945\,[0.084]$
ι	-	-	-	1.49

Table 1: Monte Carlo estimates for the correctly specified model. Standard errors reported in parentheses and MCSDs reported in brackets.

As can be seen, the coefficient estimates and the equation standard error are close to their DGP values. The outcomes are almost identical with and without selection, and the Monte Carlo standard deviation (MCSD) for the intercept reflects that selection will frequently exclude that insignificant variable. As the model is correctly specified, the step-shift dummies are redundant and are rarely retained. Applying impulse-indicator saturation reduces the equation standard error by removing approximately 1.5 observations on average when there is a break in both regressors (slightly above the 1 indicator implied by $\alpha = 1\%$), but there is no effect on coefficient estimates.[2] Hence, the correct model is estimated and selected as expected and IIS has low cost under the null hypothesis of the correct model specification.

[2] Johansen and Nielsen (2009) note a feasible bias correction for $\widehat{\sigma}_\epsilon$ under the null.

5.2 The alternative of an under-specified model

We next consider the alternative hypothesis of a mis-specified model given by (3). Two cases are considered; a break of small magnitude ($2\sigma_\epsilon$) and a larger break ($5\sigma_\epsilon$). The specifications considered are reported in table 2 along with the theoretical coefficients derived in §2. Tables 3–6 record the results. In every table, the columns refer to:

(a) estimation of (3);

(b) estimation of (3) with step-shift dummies at break dates;

(c) selection from (3);

(d) selection applying IIS.

Selection is undertaken using *Autometrics* at the $\alpha = 1\%$ significance level, and the intercept is forced to enter the final specification (i.e., it is not selected over).

H_a	Parameters	Break date	γ_0	γ_1	σ_e
(i) No break	$\mu_1 = \mu_2 = \delta_1 = \delta_2 = 0$	$T^0 = T^* = 0$	0	1.5	1.41
(ii) Break in w_t	$\mu_1 = \mu_2 = \delta_1 = 0$, $\delta_2 = 2$	$T^0 = 0$, $T^* = 81$	0.4	1.5	1.62
(iii) Break in w_t	$\mu_1 = \mu_2 = \delta_1 = 0$, $\delta_2 = 5$	$T^0 = 0$, $T^* = 81$	1	1.5	2.45
(iv) Break in z_t	$\mu_1 = \delta_1 = \delta_2 = 0$, $\mu_2 = -2$	$T^0 = 81$, $T^* = 0$	0.12	1.30	
(v) Break in z_t	$\mu_1 = \delta_1 = \delta_2 = 0$, $\mu_2 = -5$	$T^0 = 81$, $T^* = 0$	0.1	1.1	
(vi) Break in z_t and w_t	$\mu_1 = \delta_1 = 0$, $\mu_2 = -2$, $\delta_2 = 2$	$T^0 = 91$, $T^* = 81$	0.37	0.84	
(vii) Break in z_t and w_t	$\mu_1 = \delta_1 = 0$, $\mu_2 = -5$, $\delta_2 = 5$	$T^0 = 91$, $T^* = 81$	0.38	-0.23	

Table 2: Parameter specification under the alternative with theoretically derived coefficients.

We first consider (i), where the equation is mis-specified but there are no breaks, reported in Table 3. When both the included and omitted variable are constant, the coefficient estimates correspond to their theory counterparts, although the equation standard error is slightly too small (but not significantly different from $\sqrt{2}$), reduced by IIS retaining approximately 1.4 indicators on average, as before. IIS has no impact on the coefficient estimates, so is not costly; the only efficiency loss is a reduced effective sample size of $\widetilde{T} = 98.6$ instead of $T = 100$.

5.3 The excluded variable shifts

When the omitted variable breaks, but the included variable remains constant, (cases (ii) & (iii)), estimates are close to their theoretical counterparts in table 2. Unconditional estimates are reported after selection, which differ from the conditional estimates when the break is small because the intercept is

	(a)	(c)	(d)
Constant z_t and w_t			
$\widehat{\gamma}_0$	$-0.015\,[0.096]$	$-0.001\,[0.016]$	$-0.019\,[0.106]$
	$(0.128)[(0.008)]$	$(0.000)[(0.005)]$	$(0.123)[(0.010)]$
$\widehat{\gamma}_1$	$1.471\,[0.098]$	$1.471\,[0.098]$	$1.480\,[0.112]$
	$(0.126)[(0.008)]$	$(0.126)[(0.008)]$	$(0.121)[(0.010)]$
$\widehat{\sigma}_\epsilon$	$1.284\,[0.085]$	$1.281\,[0.084]$	$1.222\,[0.103]$
ι	-	-	1.40

Table 3: Monte Carlo estimates for the mis-specified static model when w_t is omitted, case (i). Standard errors reported in parentheses, and Monte Carlo standard deviations reported in brackets. ι is number of impulse-indicators retained on average.

often not retained, but there is virtually no difference in conditional and unconditional estimates when the break is large. Table 4 records the outcomes.

	(a)	(b)	(c)	(d)
Constant z_t, break in w_t, $\delta_2 = 5$				
$\widehat{\gamma}_0$	$0.985\,[0.096]$	$-0.070\,[0.110]$	$0.984\,[0.100]$	$0.018\,[0.175]$
	$(0.249)[(0.010)]$	$(0.143)[(0.010)]$	$(0.248)[(0.013)]$	$(0.146)[(0.019)]$
$\widehat{\gamma}_1$	$1.499\,[0.098]$	$1.469\,[0.098]$	$1.499\,[0.098]$	$1.536\,[0.147]$
	$(0.244)[(0.010)]$	$(0.126)[(0.008)]$	$(0.244)[(0.010)]$	$(0.149)[(0.019)]$
$\widehat{\gamma}_{s,w}$		$5.274\,[0.256]$		
		$(0.320)[(0.021)]$		
$\widehat{\sigma}_\epsilon$	$2.485\,[0.100]$	$1.282\,[0.085]$	$2.485\,[0.100]$	$1.327\,[0.192]$
ι	-	-	-	18.20

Table 4: Monte Carlo results for the mis-specified static model omitting w_t, case (iii).

Figure 1 records the recursive parameter estimates and recursive break-point Chow (1960) tests for:
(a) in the first row, estimation of (3)
(b) in the second row, estimation with the step dummy providing the optimal infeasible benchmark, and
(d) selection with IIS in the final row.

When the break occurs in the omitted variable, the intercept shifts (top-left panel), but the slope parameter remains constant, albeit with increasing standard errors (next panel in first row), as goodness of fit deteriorates sharply, and the Chow test rejects 100% of the time till after the break date. Including the step-shift dummy at the date of the break in the omitted variable picks up the non-constancy, resulting in a constant intercept (middle-left panel) as well as a constant slope parameter and a constant estimated equation standard error so the Chow test rejects at its nominal size. Thus, knowing when there are breaks in omitted variables does deliver a constant-parameter model, but in practice is infeasible, as the variable would have been included otherwise. IIS helps overcome this problem, as it accounts for the unknown

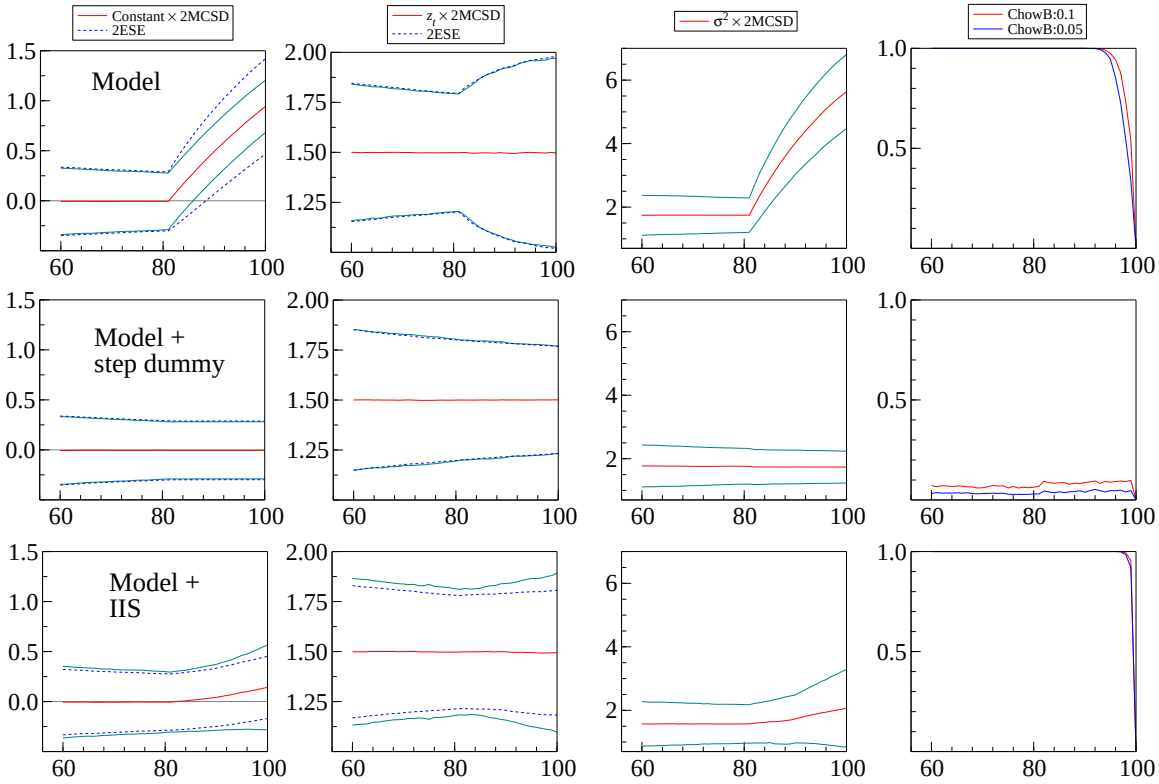

Figure 1: Incorrect specification with non-constant w_t, $\delta_1 = 0$, $\delta_2 = 5$: case (iii).

break using selected indicators. For a break of 5 standard deviations, IIS retains 18 indicators on average (for a break of 20 observations). An empirical investigator would surely notice that the indicators all had similar magnitude, same sign coefficients and combine these into a step-shift dummy, with almost no cost (see Hendry and Santos, 2005).

The estimate of the intercept increases slightly relative to the step-dummy case, but most of the parameter non-constancy evident in (a) is removed (third row). Increasing the significance level (to $\alpha = 2.5\%$) results in more indicators being retained, and the non-constancy of the intercept is almost fully removed, at a cost of retaining more indicators under the null. The equation standard error increases slightly, but is less than half that of the estimated mis-specified model (a). A Chow test clearly detects the break, as would be expected with 18 indicators retained. If the break is small ($2\sigma_\epsilon$), fewer indicators are retained with the consequence that the intercept is biased upwards as the break is not fully accounted for, but the slope coefficient is constant.

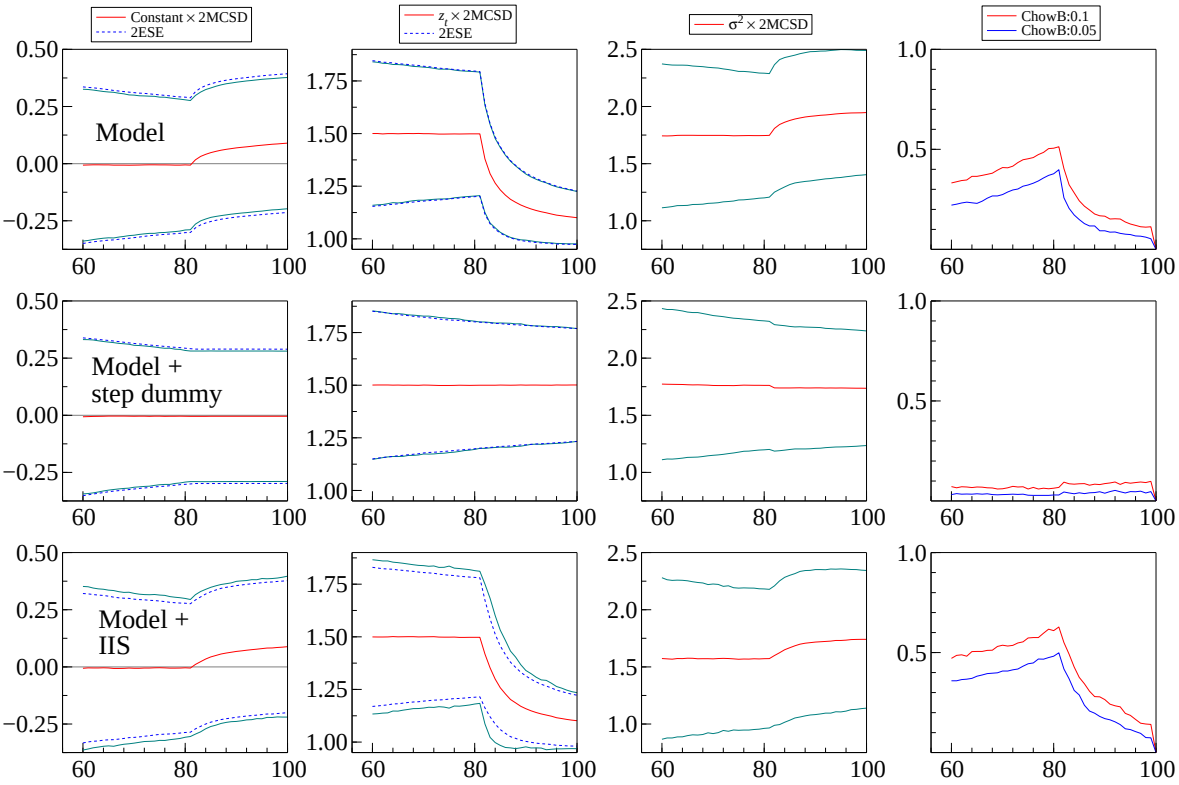

Figure 2: Incorrect specification with non-constant z_t, $\mu_1 = 0$, $\mu_2 = -5$: case (v).

5.4 The included variable shifts

When the included variable breaks, but the omitted variable is constant, both the intercept and slope parameters exhibit non-constancy.

Figure 2 records recursive parameter estimates for direct estimation, (a), inclusion of a step dummy, (b), and selection with IIS, (d), in rows. As expected, the shift in the intercept is of smaller magnitude than for a break in δ_t, and the Chow test shows that detecting such a break is more difficult. This is a function of the degree of correlation between the included and omitted regressors: with $\Psi = 0.5$, the break in z_t is only half that of the shift in the omitted w_t. The step-shift dummy is significant, reflecting the correlation, and although it is estimated quite imprecisely, it does help correct for the non-constancy in both the intercept and slope parameter. IIS picks up fewer indicators as the break is not as evident, and roughly the same number of indicators are retained as under the constant parameter model. However, the non-constancy of the parameters is substantially mitigated by the few large indicators retained.

The ability of IIS to proxy the step-shift dummy that removes the non-constancy is diminished for the break in the included variable unless it is highly correlated with the omitted variable. If $\Psi = 0.8$, IIS

	(a)	(b)	(c)	(d)
Break in z_t, constant w_t, $\mu_2 = -5$				
$\widehat{\gamma}_0$	$0.027\,[0.109]$	$-0.070\,[0.110]$	$0.010\,[0.063]$	$0.023\,[0.119]$
	$(0.150)[(0.010)]$	$(0.143)[(0.010)]$	$(0.031)[(0.061)]$	$(0.143)[(0.011)]$
$\widehat{\gamma}_1$	$1.051\,[0.047]$	$1.469\,[0.098]$	$1.048\,[0.043]$	$1.050\,[0.051]$
	$(0.061)[(0.004)]$	$(0.126)[(0.008)]$	$(0.057)[(0.004)]$	$(0.058)[(0.005)]$
$\widehat{\gamma}_{s,z}$		$2.619\,[0.535]$		
		$(0.702)[(0.047)]$		
$\widehat{\sigma}_\epsilon$	$1.367\,[0.088]$	$1.282\,[0.085]$	$1.364\,[0.087]$	$1.293\,[0.108]$
ι	-	-	-	1.62

Table 5: Monte Carlo results for the mis-specified static model omitting w_t, case (v).

retains 16 indicators on average and results are similar to the case above with a break in the omitted variable, with IIS delivering parameter constancy. This highlights an apparent catch-22: breaks that induce significant parameter non-constancy are damaging to the model specification but are easily detected and addressed by IIS. Breaks that induce smaller non-constancies are difficult to detect, and so potentially more damaging for policy. However, in practice breaks are likely to occur in both included and omitted regressors, and correlations are likely to be high, a situation in which IIS excels.

5.5 Both included and excluded variables shift

When breaks occur in both included and omitted variables (cases (vi) & (vii)), both the intercept and slope parameters shift, with the intercept shifting from the point at which a break in z_t occurs. Figure 3 records the recursive results for a break of $5\sigma_\epsilon$ in both variables. The step-shift dummies mop up all of the non-constancy. IIS retains all 20 indicators on average, proxying the optimal step shift dummy. As a result, parameters are close to constant and the equation standard error only increases slightly.

	(a)	(b)	(c)	(d)
Break in z_t and w_t ($\mu_2 = -5$, $\delta_2 = 5$)				
$\widehat{\gamma}_0$	$0.732\,[0.102]$	$-0.070\,[0.110]$	$0.588\,[0.336]$	$-0.050\,[0.130]$
	$(0.250)[(0.010)]$	$(0.143)[(0.010)]$	$(0.189)[(0.106)]$	$(0.141)[(0.015)]$
$\widehat{\gamma}_1$	$0.535\,[0.059]$	$1.476\,[0.098]$	$0.512\,[0.082]$	$1.503\,[0.198]$
	$(0.135)[(0.006)]$	$(0.125)[(0.008)]$	$(0.134)[(0.009)]$	$(0.141)[(0.014)]$
$\widehat{\gamma}_{s,z}$		$1.719\,[0.630]$		
		$(0.827)[(0.055)]$		
$\widehat{\gamma}_{s,w}$		$5.605\,[0.344]$		
		$(0.428)[(0.029)]$		
$\widehat{\sigma}_\epsilon$	$2.396\,[0.099]$	$1.275\,[0.085]$	$2.410\,[0.108]$	$1.257\,[0.142]$
ι	-	-	-	19.99

Table 6: Monte Carlo results for the mis-specified static model omitting w_t, case (vii).

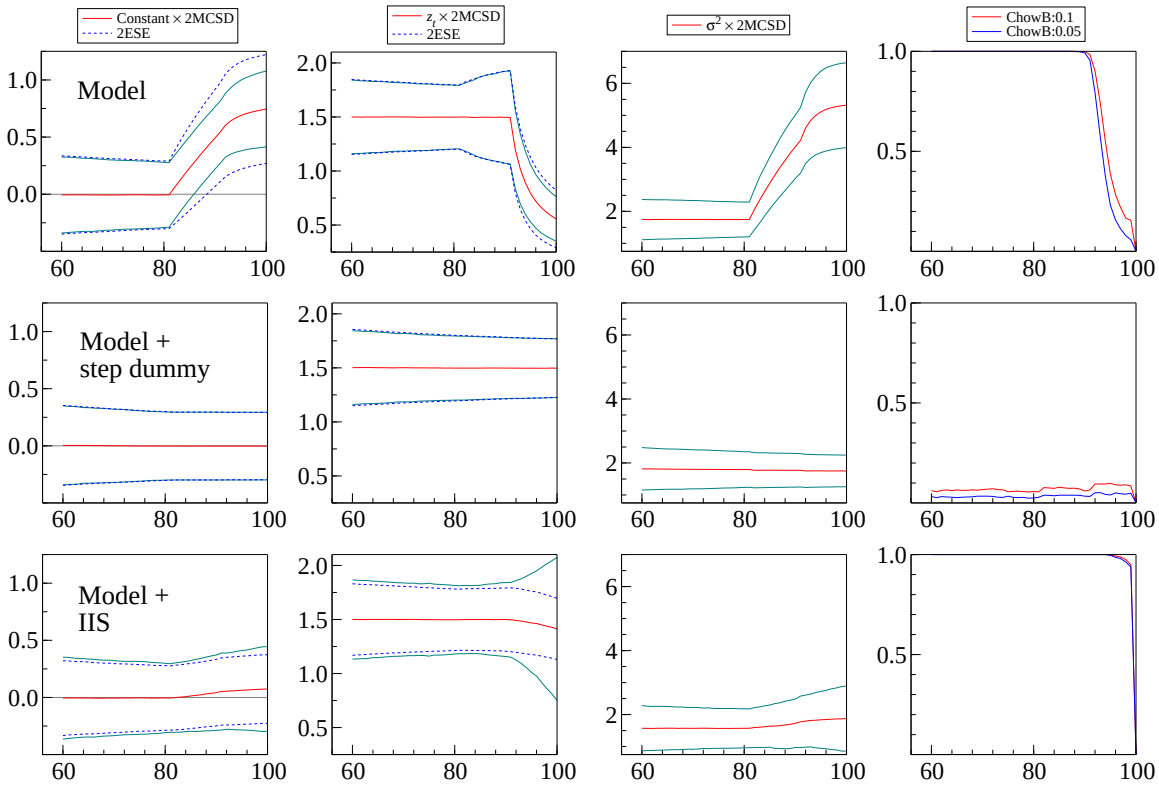

Figure 3: Mis-specified static model with non-constant z_t and w_t, $\delta_1 = 0$, $\delta_2 = 5$, $\mu_1 = 0$, $\mu_2 = -5$.

5.6 Multiple breaks

The breaks considered so far are cases in which a Chow test has power to detect the break and a step shift dummy excels if the break is known. In practice multiple breaks are common, so we examine a case where there is an intermittent break in the intercept of 5 observations in length. The DGP and model are given by (1) and (3) but the included and omitted variables are given by:

$$\begin{aligned} z_t &= \mu_1 + \mu_2 \left[I_{20} + \cdots + I_{25} + I_{40} + \cdots + I_{45} + I_{60} + \cdots + I_{65} + I_{80} + \cdots + I_{85} \right] + u_{z,t} \\ w_t &= \delta_1 + \delta_2 \left[I_{20} + \cdots + I_{25} + I_{40} + \cdots + I_{45} + I_{60} + \cdots + I_{65} + I_{80} + \cdots + I_{85} \right] + u_{w,t} \end{aligned}$$

where

$$\begin{pmatrix} u_{z,t} \\ u_{w,t} \end{pmatrix} \sim \mathsf{IN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right] \tag{13}$$

The baseline parameter values are: $\beta_1 = 1$, $\beta_2 = 1$, $\sigma_\epsilon^2 = 1$, $\Sigma_{12} = \Psi = 0.8$, $\Sigma_{11} = \Sigma_{22} = 1$ and $T = 100$ with recursive estimation commencing from $t = 6$. $M = 1000$ replications are undertaken.

Table 7 reports the results for multiple breaks of $5\sigma_\epsilon$ in the included and omitted variables, with recursive graphics recorded in figure 4. When the break occurs in the excluded variable, the intercept shifts at the point of the break. IIS mitigates this non-constancy by retaining approximately 17 indicators on average, removing most of the non-constancy in the intercept. The coefficient of the included variable should be constant, as is observed, although the recursive figures commence at $t = 6$ so there is some estimation uncertainty evident due to the small sample. The equation standard error with IIS is almost half that of the mis-specified model, and the power to detect the breaks is high. When there are intermittent breaks in the included variable, the non-constancy is not as evident. Although IIS only retains approximately 6 indicators on average, the resulting equation standard error is close to that for the LDGP in column (b), so most of the mis-specification is corrected. In neither case does selection for the mis-specified equation make any difference (columns (a) and (c) are the same).

	(a)	(b)	(c)	(d)
Constant z_t, multiple breaks in w_t ($\delta_2 = 5$)				
$\widehat{\gamma}_0$	0.993 [0.115] (0.233)[(0.011)]	0.001 [0.131] (0.130)[(0.009)]	0.993 [0.115] (0.233)[(0.011)]	0.177 [0.246] (0.146)[(0.027)]
$\widehat{\gamma}_1$	1.800 [0.235] (0.235)[(0.020)]	1.801 [0.118] (0.118)[(0.012)]	1.800 [0.235] (0.235)[(0.020)]	1.799 [0.211] (0.147)[(0.029)]
$\widehat{\gamma}_{s,w}$		4.987 [0.289] (0.291)[(0.021)]		
$\widehat{\sigma}_\epsilon$	2.323 [0.519]	1.161 [0.196]	2.323 [0.519]	1.361 [0.803]
ι	-	-	-	16.53
Multiple breaks in z_t, constant w_t ($\mu_2 = -5$)				
$\widehat{\gamma}_0$	0.152 [0.142] (0.150)[(0.011)]	0.001 [0.131] (0.130)[(0.009)]	0.152 [0.142] (0.150)[(0.011)]	0.145 [0.170] (0.131)[(0.016)]
$\widehat{\gamma}_1$	1.157 [0.059] (0.061)[(0.005)]	1.801 [0.118] (0.118)[(0.012)]	1.157 [0.059] (0.061)[(0.005)]	1.157 [0.071] (0.054)[(0.007)]
$\widehat{\gamma}_{s,z}$		3.992 [0.653] (0.656)[(0.067)]		
$\widehat{\sigma}_\epsilon$	1.368 [0.260]	1.161 [0.196]	1.368 [0.260]	1.170 [0.352]
ι	-	-	-	6.12

Table 7: Monte Carlo results for the mis-specified static model with multiple breaks. Legend as Table 3.

5.7 Diagnostic tracking

Diagnostic testing is an integral part of the *Gets* procedure, ensuring that candidate models are congruent. *Autometrics* undertakes diagnostic testing when a terminal model is reached to ensure the reduction is valid, backtracking to find a valid reduction if the mis-specification tests fail. When diagnostic tracking is switched on, irrelevant variables can proxy part of a chance departure from the null of one of the

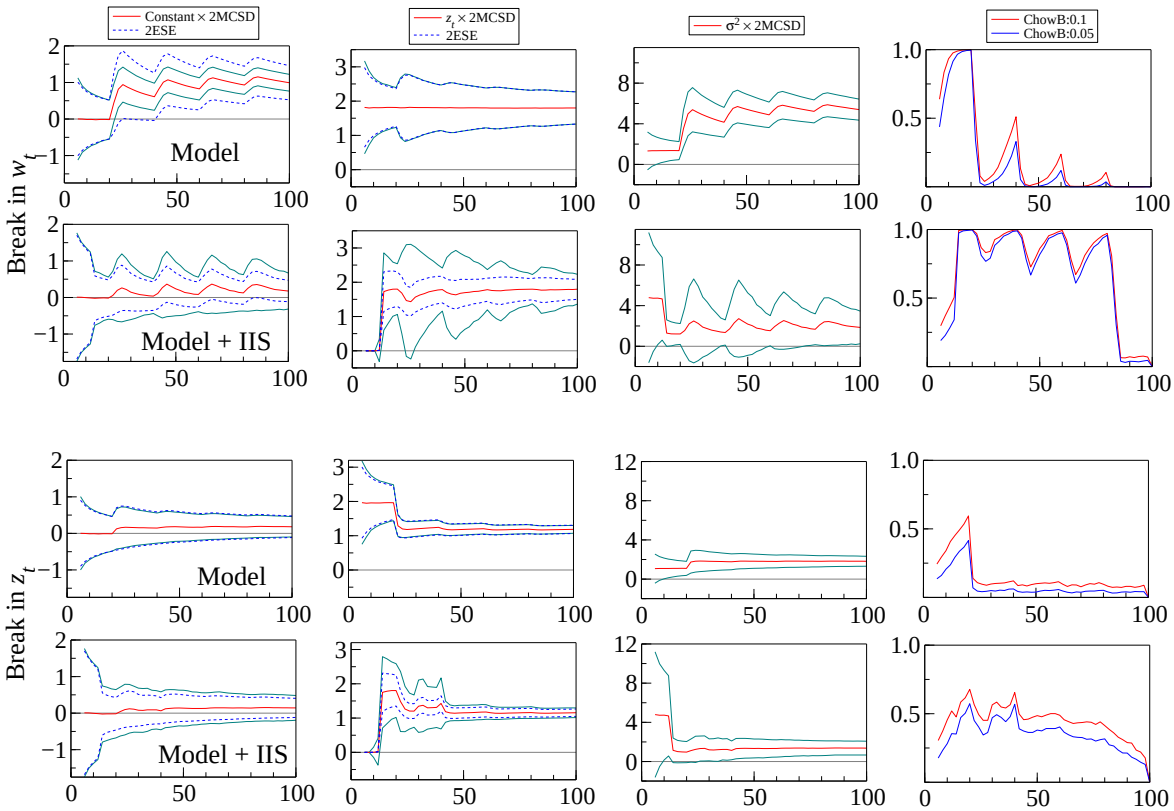

Figure 4: Incorrect specification with multiple breaks in w_t (top 2 panels) and z_t (bottom 2 panels), $\delta_1 = 0$, $\delta_2 = 5$, $\mu_1 = 0$, $\mu_2 = -5$.

mis-specification tests or the encompassing check, and then be retained despite insignificance.

The simulations were also conducted with diagnostic tracking switched on. The number of indicators retained on average increased slightly, with ι between 0.1 and 0.2 larger than those reported in tables 4–6. There was no substantive effect on the parameter estimates.

5.8 Forcing variables

Often the econometrician will have some prior information based on theoretical or institutional knowledge or past evidence. This can be incorporated into the selection procedure by 'forcing' variables in the final specification, such that selection is only applied to the variables that freely enter the search space. Forcing variables enables inclusion of theoretical priors, so if there was a strong theoretical case for including z_t, this could be a forced variable, but that obviously does not guarantee its significance, nor even the 'correct' sign if in fact the theory is incorrect (see e.g., Hendry and Johansen, 2010).

The simulations were conducted when both the intercept and z_t were forced to enter, so only the indicators were selected. Forcing the included variable had little impact on the results, but forcing the

intercept to be retained was important when the mean of either the included or omitted variable shifted.[3]

6 Mis-specifications and breaks in a dynamic equation

We now extend the analysis to investigate the properties for a dynamic specification. The constant-parameter conditional LDGP is given by:

$$y_t = \beta_y y_{t-1} + \beta_z z_t + \beta_w w_t + \epsilon_t \tag{14}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$, and:

$$\begin{aligned} z_t &= \mu_t + \theta_z z_{t-1} + u_{z,t} \\ w_t &= \delta_t + \theta_w w_{t-1} + u_{w,t} \end{aligned} \tag{15}$$

where:

$$\begin{pmatrix} u_{z,t} \\ u_{w,t} \end{pmatrix} \sim \mathsf{IN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right] \tag{16}$$

and the intercepts both have one-off location shifts given by (4).

From (15):

$$\begin{aligned} w_t &= \delta_t + \theta_w w_{t-1} + \Psi u_{z,t} \\ &= \delta_t + \theta_w w_{t-1} + \Psi\left(z_t - \mu_t - \theta_z z_{t-1}\right) + u_t \\ &= \left(\delta_t - \Psi\mu_t\right) + \theta_w w_{t-1} + \Psi z_t - \Psi\theta_z z_{t-1} + u_t \end{aligned} \tag{17}$$

where $\mathsf{E}[z_t u_t] = 0$. Solving to eliminate both current and lagged w in (17) from the LDGP using (14) lagged:

$$\theta_w y_{t-1} = \beta_z \theta_w z_{t-1} + \beta_y \theta_w y_{t-2} + \beta_w \theta_w w_{t-1} + \theta_w \epsilon_{t-1} \tag{18}$$

[3]Detailed results with diagnostic tracking switched on, and forced variables, are available on request.

then:

$$\begin{aligned}\beta_w w_t &= \beta_w\left(\delta_t - \Psi\mu_t\right) + \beta_w\theta_w w_{t-1} + \beta_w\Psi z_t - \beta_w\Psi\theta_z z_{t-1} + \beta_w u_t \\ &= \beta_w\left(\delta_t - \Psi\mu_t\right) + \beta_w\Psi z_t + \theta_w y_{t-1} - \left(\beta_z\theta_w + \beta_w\Psi\theta_z\right) z_{t-1} - \beta_y\theta_w y_{t-2} + \beta_w u_t - \theta_w\epsilon_{t-1}\end{aligned}$$

so:

$$\begin{aligned}y_t = \beta_w\left(\delta_t - \Psi\mu_t\right) + \left(\beta_z + \beta_w\Psi\right) z_t - \left(\beta_w\Psi\theta_z + \beta_z\theta_w\right) z_{t-1} + \left(\beta_y + \theta_w(1-\psi)\right) y_{t-1} \\ - \beta_y\theta_w y_{t-2} + \epsilon_t + \beta_w u_t - \theta_w\eta_{t-1}\end{aligned} \tag{19}$$

after letting:

$$\epsilon_{t-1} = \psi y_{t-1} + \eta_{t-1}$$

where $\psi = \sigma_\epsilon^2/\sigma_y^2$, to approximately orthogonalize y_{t-1} with ϵ_{t-1}.

There are a variety of possible models that are mis-specified due to excluding w_t and its lags. The most general model considered could be an ADL model:

$$y_t = \gamma_0 + \sum_{i=1}^{p}\gamma_i y_{t-i} + \sum_{j=0}^{q}\pi_i z_{t-j} + v_t \tag{20}$$

with sufficiently general lags (p, q), whereas the most mis-specified model omitting all dynamics is:

$$y_t = \gamma_0 + \pi_1 z_t + e_t. \tag{21}$$

The inclusion of both z_{t-1} and y_{t-2} in the reduced form model provides further support for a *Gets* approach. Mis-specified dynamic models entail more complex dynamics to capture the mis-specification, as well as IIS to account for parameter non-constancy, suggesting commencing with a sufficiently general lag length (p, q).

The analytics for the above dynamic case with mis-specification and location shifts is close to intractable as the step shifts alter the expectations at every observation after their breaks. Hence, we rely on simulations to analyze the more complex case, but note that the static case highlights the key issues, with the simulations in §7 providing more realistic evidence for time-series.

	no IIS	IIS	no IIS	IIS	no IIS	IIS	no IIS	IIS	no IIS	IIS
Constant z_t, break in w_t, $\delta_2 = 5$										
$\widehat{\gamma}_0$	7.088 (1.626)	0.035 (0.301)	0.426 (0.281)	0.173 (0.180)	0.339 (0.245)	0.059 (0.183)	0.246 (0.219)	0.062 (0.181)	0.247 (0.219)	0.067 (0.180)
$\widehat{\gamma}_1$			1.010 (0.017)	0.768 (0.033) [1.0]	1.022 (0.015)	0.800 (0.033) [1.0]	1.425 (0.078)	0.995 (0.058) [1.0]	1.400 (0.090)	0.996 (0.051) [1.0]
$\widehat{\gamma}_2$							−0.419 (0.080)	−0.347 (0.071) [0.499]	−0.391 (0.093)	−0.344 (0.072) [0.424]
$\widehat{\pi}_0$	2.407 (1.005)	2.431 (0.186) [0.998]	0.603 (0.167)	1.226 (0.132) [0.716]	1.536 (0.217)	1.465 (0.155) [1.0]	1.517 (0.192)	1.467 (0.156) [1.0]	1.516 (0.193)	1.465 (0.152) [1.0]
$\widehat{\pi}_1$					−1.268 (0.220)	−1.082 (0.190) [0.586]	−1.476 (0.199)	−1.261 (0.180) [0.642]	−1.363 (0.278)	−1.176 (0.187) [0.632]
$\widehat{\pi}_2$									−0.133 (0.226)	−0.734 (0.160) [0.208]
$\widehat{\sigma}$	15.56	2.511	2.477	1.744	2.135	1.528	1.871	1.504	1.868	1.497
ι	-	25.72	-	15.55	-	14.06	-	11.00	-	11.60

Table 8: Monte Carlo estimates for the mis-specified dynamic model omitting w_t. Mean coefficient estimates conditional on selection, with retention rates for selection reported in brackets.

7 Simulating dynamic mis-specification

The baseline parameter values are $\beta_y = 0.5$, $\beta_z = 1$, $\beta_w = 1$, $\sigma_\epsilon^2 = 1$, $\rho\Sigma_{11} = \Psi = 0.5$, $\Sigma_{11} = \Sigma_{22} = 1$, $\theta_z = \theta_w = 0.8$ and $T = 100$, with $y_0 = 0$ and 20 initial observations discarded. $M = 1000$ replications are undertaken. The break dates and parameters are the same as in the static case, recorded in table 2. Tables 8 and 9 record the results for a range of possible GUM specifications based on (20) with varying lag length (p, q). Each pair of columns reports estimation of the GUM specification under 'no IIS' and selection from the GUM with IIS, forcing the intercept to enter the final specification.[4] Selection is undertaken using *Autometrics* at the $\alpha = 1\%$ significance level. Mean coefficient estimates are reported conditional on being retained in the final selection, with retention rates for selection reported in brackets.

In dynamic models, an unmodeled break results in an estimate of the sum of the lagged dependent variables close to unity, as imposing a unit root is the only way to 'pick up' the shift in mean, thereby adapting to the location shift. IIS mitigates this effect, resulting in stationary estimates of the lagged dependent variables' coefficients. The most significant impact of IIS is on the intercept, as the impulse indicators correctly estimate the mean shift. When $\theta_z = \theta_w$, then the coefficient of z_{t-1} is θ_z times that

[4]Results for the break of $2\sigma_\epsilon$ are omitted for brevity but are available on request.

	no IIS	IIS	no IIS	IIS	no IIS	IIS	no IIS	IIS	no IIS	IIS
Constant w_t, break in z_t, $\mu_2 = -5$										
$\widehat{\gamma}_0$	0.217 (0.354)	0.124 (0.301)	0.026 (0.198)	0.016 (0.174)	0.046 (0.186)	0.037 (0.173)	0.051 (0.175)	0.046 (0.166)	0.054 (0.176)	0.048 (0.168)
$\widehat{\gamma}_1$			0.596 (0.040)	0.614 (0.037) [1.0]	0.756 (0.053)	0.736 (0.047) [1.0]	1.004 (0.089)	0.942 (0.071) [1.0]	0.969 (0.097)	0.894 (0.063) [1.0]
$\widehat{\gamma}_2$							−0.217 (0.063)	−0.246 (0.059) [0.69]	−0.160 (0.089)	−0.254 (0.059) [0.43]
$\widehat{\pi}_0$	1.928 (0.039)	2.130 (0.058) [1.0]	0.841 (0.076)	0.967 (0.078) [1.0]	1.296 (0.127)	1.290 (0.115) [1.0]	1.275 (0.121)	1.275 (0.112) [1.0]	1.266 (0.121)	1.232 (0.102) [1.0]
$\widehat{\pi}_1$					−0.781 (0.184)	−0.813 (0.174) [0.82]	−0.843 (0.175)	−0.867 (0.166) [0.90]	−0.702 (0.232)	−0.845 (0.170) [0.67]
$\widehat{\pi}_2$									−0.174 (0.191)	−0.577 (0.119) [0.35]
$\widehat{\sigma}$	3.262	2.663	1.760	1.489	1.608	1.485	1.512	1.427	1.506	1.439
ι	-	11.30	-	9.90	-	4.16	-	2.47	-	2.15

Table 9: Monte Carlo estimates for the mis-specified dynamic model omitting w_t.

of z_t with opposite sign. The estimates of $\widehat{\pi}_1$ are close to $-0.8\widehat{\pi}_0$ and IIS brings the estimate closer to the theory. The lagged exogenous variable is retained roughly two thirds of the time, and the retention rate of z_{t-2} is significantly greater than the 1% significance level.

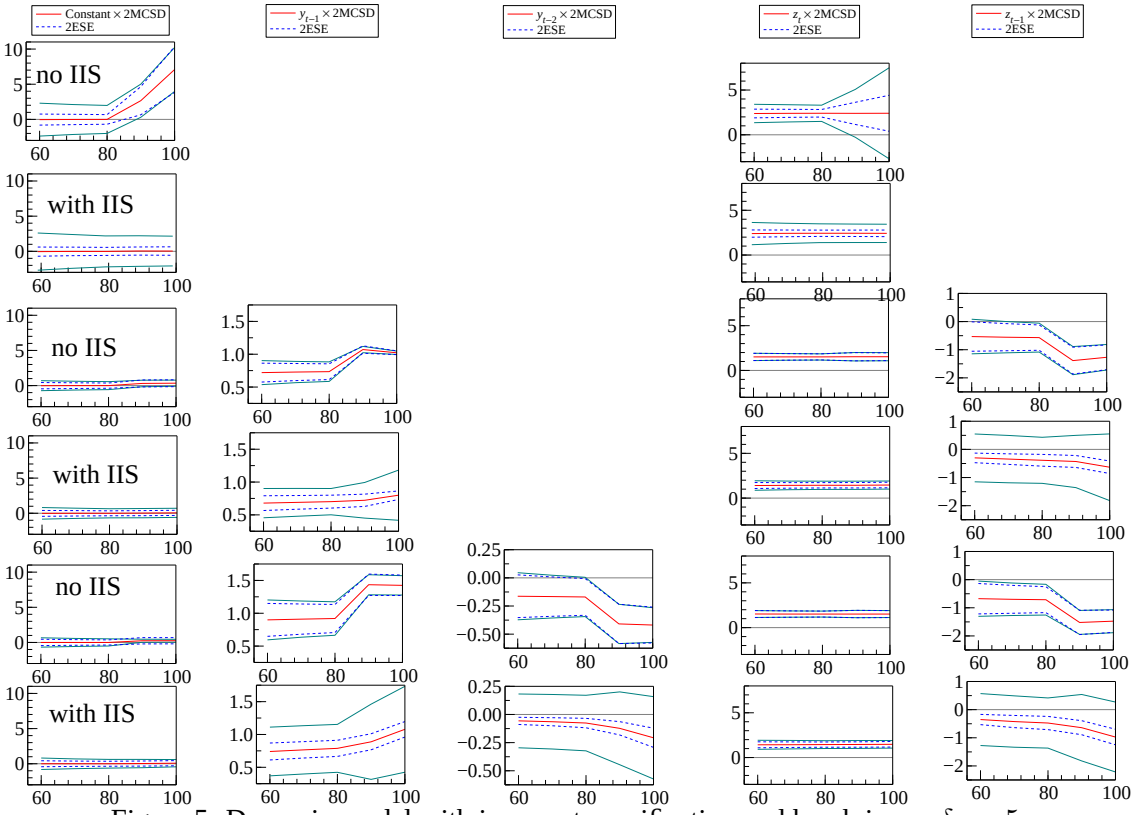

Figure 5: Dynamic model with incorrect specification and break in w_t, $\delta_2 = 5$.

Figures 5 and 6 record the recursive coefficient estimates for both estimation of the GUM and selection with IIS for three cases; $(p = 0, q = 0)$, $(p = 1, q = 1)$ and $(p = 2, q = 1)$. The static model exhibits severe non-constancy. The non-constancy in the dynamic model is not smooth, flattening off after 10 observations. The unit root adapts slowly, but then retains the correct mean. IIS smooths the shift to the new mean but as fewer indicators are retained compared to the static model the degree of persistence is larger due to more unmodeled breaks. As in the static case the coefficient on z_t should remain constant, but the lagged exogenous variable, z_{t-1}, does shift. When there is a break in the included variable and the correlations are low the break is difficult to detect, as in the static case.

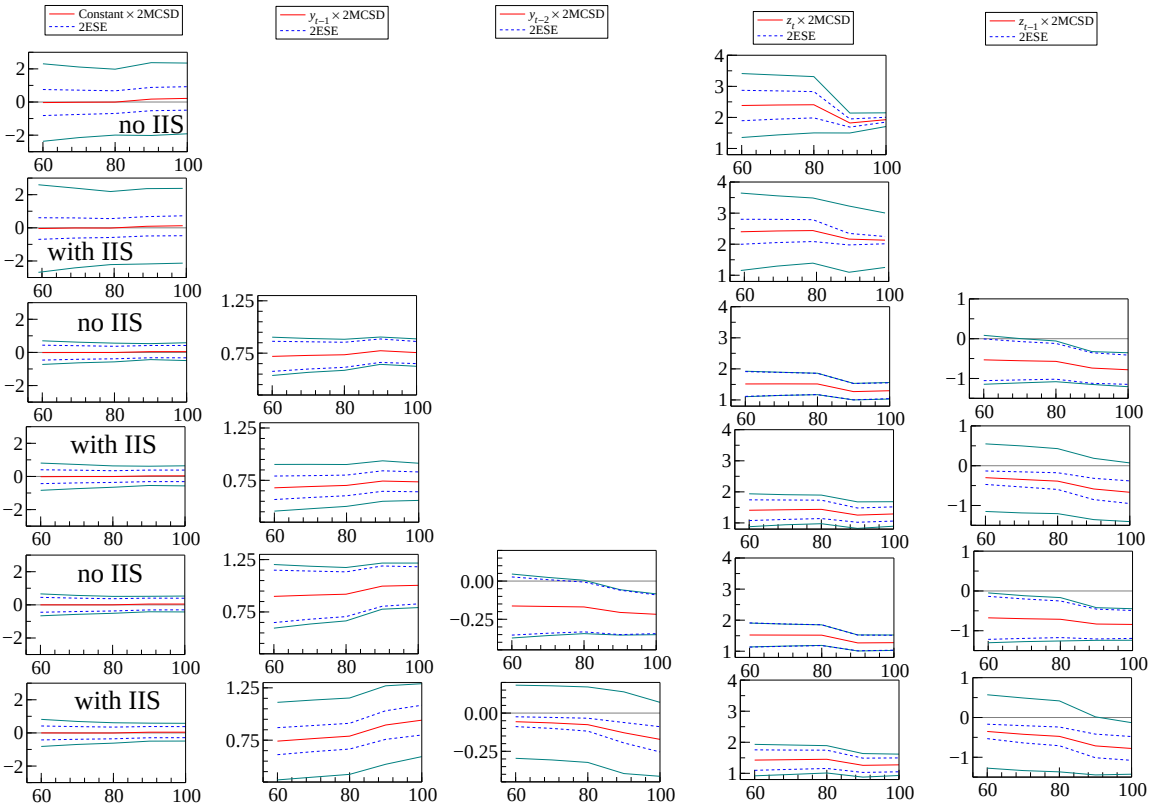

Figure 6: Dynamic model with incorrect specification and break in z_t, $\mu_2 = -5$.

Extensions to alternative break types including seasonal shifts, intermittent outliers, trend breaks, and multiple breaks are all feasible, including both stationary and unit-root processes. Castle *et al.* (2009) provide evidence on the ability of IIS to detect such breaks. If the break(s) occur in the omitted variables, IIS will deliver close to constant parameter estimates for the model approximating the LDGP defined by the included variables. If the break(s) occur in the included variables, IIS performs well if the variables are highly correlated with the omitted variables. However, if the correlation is low, the extent of parameter non-constancy will anyway be small.

8 Omitted variables versus other mis-specifications

The reduced LDGP_1 is different in concept from a model where the set of variables is correct, but the dynamic specification (say) is wrong. The former is a well-defined target for selection, albeit not the most useful one due to the mis-specification. However, if an omission is unknown, little can be done to correct the problem. On the contrary, the latter is relatively easily corrected by commencing from a suitably more general GUM that does nest the corresponding LDGP.

In practice, both difficulties are likely: the LDGP does not include all the relevant variables, and the GUM does not nest it. Since *Autometrics* can handle large numbers of variables, lags and functional-form transformations, as well as IIS, very general initial specifications are feasible, including more variables than observations as noted above. As the empirical modeling example in Hendry and Mizon (2010) illustrates, IIS can help retrieve a theory-consistent specification that fails on direct estimation, but is correctly recovered when embedded in a sufficiently general GUM, which mitigates any otherwise unmodeled non-constancies.

9 Conclusion

There are many critiques of model selection, but almost all of these are applied to nearly correct models with constant parameters showing that simply fitting the chosen specification usually dominates over selecting. This is not a realistic characterization of the situation confronting empirical investigators. Data processes are complicated, evolving, and subject to intermittent unanticipated location shifts; and models derived from theory provide only a guide to some of the main determinants, rarely addressing breaks, outliers, or data contamination. Non-stationarities vitiate any *ceteris paribus* assumptions, and many features of models are not uniquely derivable *a priori*. Model selection is inevitable.

Consequently, it is fortunate that automatic methods based on general initial models can perform well in a variety of settings and can jointly tackle selection, functional form specification, breaks, outliers and data contamination, while having low costs when those problems are in fact absent, yet high efficiency to handle them when needed. Here, we have shown that the benefits of model selection extend to under-specified settings when variables have location shifts and relevant variables are omitted. Impulse-indicator saturation appropriately tackles multiple breaks at unknown sample points, and we have shown that it 'robustifies' estimated models against non-constancy induced by omitted variables.

The analysis also reveals that breaks in only the omitted variables do not contaminate the slope parameter, whereas when there are any correlated omitted variables, breaks in the included variables lead to non-constant slopes as well as shifting intercepts despite that aspect having been correctly specified. In all cases, the equation standard error becomes non-constant. The omission *per se* cannot be corrected by IIS or model selection, so policy derivatives will be incorrect until the correct specification is discovered, but the model-selection, IIS-based parameters will be more constant, and forecasts based on such a model will face one less difficulty. A sufficiently general initial model may, of course, succeed in including all the substantively relevant variables.

10 Appendix calculations

First, from §2:

$$\sum_{t=1}^{T} \mathbf{w}_t \mathbf{z}_t' = \sum_{t=1}^{T^0-1} \mathbf{w}_t \mathbf{z}_t' + \sum_{t=T^0}^{T^*-1} \mathbf{w}_t \mathbf{z}_t' + \sum_{t=T^*}^{T} \mathbf{w}_t \mathbf{z}_t'$$

where, from (2):

$$\begin{aligned}
\mathsf{E}\left[\sum_{t=1}^{T^0-1} \mathbf{w}_t \mathbf{z}_t'\right] &= T\lambda\left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\delta}_1 \boldsymbol{\mu}_1'\right) \\
\mathsf{E}\left[\sum_{t=T^0}^{T^*-1} \mathbf{w}_t \mathbf{z}_t'\right] &= T\left(\kappa - \lambda\right)\left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\delta}_1 \boldsymbol{\mu}_2'\right) \\
\mathsf{E}\left[\sum_{t=T^*}^{T} \mathbf{w}_t \mathbf{z}_t'\right] &= T\left(1 - \kappa\right)\left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\delta}_2 \boldsymbol{\mu}_2'\right).
\end{aligned}$$

with $\mathsf{E}\left[\sum_{t=1}^{T} \mathbf{w}_t\right] = T\left(\kappa \boldsymbol{\delta}_1 + \left(1 - \kappa\right)\boldsymbol{\delta}_2\right)$. Next:

$$\begin{aligned}
\mathsf{E}\left[\sum_{t=1}^{T} \mathbf{z}_t\right] &= T\left(\lambda \boldsymbol{\mu}_1 + \left(1 - \lambda\right)\boldsymbol{\mu}_2\right) \\
\mathsf{E}\left[\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right] &= T\left(\boldsymbol{\Sigma}_{11} + \lambda \boldsymbol{\mu}_1 \boldsymbol{\mu}_1' + \left(1 - \lambda\right)\boldsymbol{\mu}_2 \boldsymbol{\mu}_2'\right)
\end{aligned}$$

Also:

$$\mathsf{E}\left[y_t\right] = \boldsymbol{\beta}_1' \mathsf{E}\left[\mathbf{z}_t\right] + \boldsymbol{\beta}_2' \mathsf{E}\left[\mathbf{w}_t\right] = \boldsymbol{\beta}_1' \boldsymbol{\mu}_t + \boldsymbol{\beta}_2' \boldsymbol{\delta}_t$$

so:

$$\begin{aligned}\mathsf{E}\left[\sum_{t=1}^{T} y_t\right] &= \mathsf{E}\left[\sum_{t=1}^{T^0-1} y_t\right] + \mathsf{E}\left[\sum_{t=T^0}^{T^*-1} y_t\right] + \mathsf{E}\left[\sum_{t=T^*}^{T} y_t\right] \\ &= T\lambda\left(\boldsymbol{\beta}_1'\boldsymbol{\mu}_1 + \boldsymbol{\beta}_2'\boldsymbol{\delta}_1\right) + T\left(\kappa-\lambda\right)\left(\boldsymbol{\beta}_1'\boldsymbol{\mu}_2 + \boldsymbol{\beta}_2'\boldsymbol{\delta}_1\right) + T\left(1-\kappa\right)\left(\boldsymbol{\beta}_1'\boldsymbol{\mu}_2 + \boldsymbol{\beta}_2'\boldsymbol{\delta}_2\right)\end{aligned}$$

hence:

$$\mathsf{E}\left[\sum_{t=1}^{T} y_t\right] = T\boldsymbol{\beta}_1'\left(\lambda\boldsymbol{\mu}_1 + \left(1-\lambda\right)\boldsymbol{\mu}_2\right) + T\boldsymbol{\beta}_2'\left(\kappa\boldsymbol{\delta}_1 + \left(1-\kappa\right)\boldsymbol{\delta}_2\right)$$

$$\begin{aligned}\mathsf{E}\left[\sum_{t=1}^{T} \mathbf{z}_t y_t\right] &= \mathsf{E}\left[\sum_{t=1}^{T^0-1} \mathbf{z}_t y_t\right] + \mathsf{E}\left[\sum_{t=T^0}^{T^*-1} \mathbf{z}_t y_t\right] + \mathsf{E}\left[\sum_{t=T^*}^{T} \mathbf{z}_t y_t\right] \\ &= T\lambda\left[\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_1\boldsymbol{\delta}_1'\right)\right]\boldsymbol{\beta}_2 \\ &+ T\left(\kappa-\lambda\right)\left[\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_2\boldsymbol{\mu}_2'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_2\boldsymbol{\delta}_1'\right)\boldsymbol{\beta}_2\right] \\ &+ T\left(1-\kappa\right)\left[\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_2\boldsymbol{\mu}_2'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_2\boldsymbol{\delta}_2'\right)\boldsymbol{\beta}_2\right] \\ &= T\left[\boldsymbol{\Sigma}_{11} + \left(\lambda\boldsymbol{\mu}_1\boldsymbol{\mu}_1' + \left(1-\lambda\right)\boldsymbol{\mu}_2\boldsymbol{\mu}_2'\right)\right]\boldsymbol{\beta}_1 \\ &+ T\left[\boldsymbol{\Sigma}_{12} + \lambda\boldsymbol{\mu}_1\boldsymbol{\delta}_1' + \left(\kappa-\lambda\right)\boldsymbol{\mu}_2\boldsymbol{\delta}_1' + \left(1-\kappa\right)\boldsymbol{\mu}_2\boldsymbol{\delta}_2'\right]\boldsymbol{\beta}_2\end{aligned}$$

as:

$$\begin{aligned}\mathsf{E}\left[\sum_{t=1}^{T^0-1} \mathbf{z}_t y_t\right] &= \mathsf{E}\left[\sum_{t=1}^{T^0-1} \mathbf{z}_t \mathbf{z}_t'\right]\boldsymbol{\beta}_1 + \mathsf{E}\left[\sum_{t=1}^{T^0-1} \mathbf{z}_t \mathbf{w}_t'\right]\boldsymbol{\beta}_2 \\ &= T\lambda\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1'\right)\boldsymbol{\beta}_1 + T\lambda\left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_1\boldsymbol{\delta}_1'\right)\boldsymbol{\beta}_2\end{aligned}$$

and:

$$\begin{aligned}\mathsf{E}\left[\sum_{t=T^0}^{T^*-1} \mathbf{z}_t y_t\right] &= \mathsf{E}\left[\sum_{t=T^0}^{T^*-1} \mathbf{z}_t \mathbf{z}_t'\right]\boldsymbol{\beta}_1 + \mathsf{E}\left[\sum_{t=T^0}^{T^*-1} \mathbf{z}_t \mathbf{w}_t'\right]\boldsymbol{\beta}_2 \\ &= T\left(\kappa-\lambda\right)\left[\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_2\boldsymbol{\mu}_2'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_2\boldsymbol{\delta}_1'\right)\boldsymbol{\beta}_2\right]\end{aligned}$$

$$\mathsf{E}\left[\sum_{t=T^*}^{T} z_t y_t\right] = \mathsf{E}\left[\sum_{t=T^*}^{T} \mathbf{z}_t \mathbf{z}_t'\right] \boldsymbol{\beta}_1 + \mathsf{E}\left[\sum_{t=T^*}^{T} \mathbf{z}_t \mathbf{w}_t'\right] \boldsymbol{\beta}_2$$
$$= T\left(1-\kappa\right)\left[\left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_2\boldsymbol{\mu}_2'\right)\boldsymbol{\beta}_1 + \left(\boldsymbol{\Sigma}_{12} + \boldsymbol{\mu}_2\boldsymbol{\delta}_2'\right)\boldsymbol{\beta}_2\right]$$

References

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2009). Model selection when there are multiple breaks. Working paper 472, Economics Department, University of Oxford.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2010a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, forthcoming.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2010b). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **158**, 25–36.

Castle, J. L., and Hendry, D. F. (2010). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T.(eds.), *System Identification, Environmental Modelling and Control*, New York: Springer, forthcoming.

Castle, J. L., and Shephard, N.(eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

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

Doornik, J. A. (2009). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.

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

Granger, C. W. J. (1999). *Empirical Modeling in Economics: Specification and Evaluation*. Cambridge: Cambridge University Press.

Granger, C. W. J., and Hendry, D. F. (2005). A dialogue concerning a new instrument for econometric modeling. *Econometric Theory*, **21**, 278–297.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson, K. D.(eds.), *Palgrave Handbook of*

Econometrics, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F., and Doornik, J. A. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Hendry, D. F., and Johansen, S. (2010). Model selection when forcing retention of theory variables. Unpublished paper, Economics Department, University of Oxford.

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. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.

Hendry, D. F., and Mizon, G. E. (2010). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, forthcoming.

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

Hendry, D. F., and Santos, C. (2010). An automatic test of super exogeneity. In Watson, M. W., Bollerslev, T., and Russell, J.(eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–191.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

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