Model Selection in Under-specified Equations Facing Breaks

Jennifer L. Castle and David F. Hendry*

Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

Abstract

When a model under-specifies the data generation process, model selection can improve over estimating a prior specification, especially if location shifts occur. Impulse-indicator saturation (IIS) can 'correct' non-constant intercepts induced by location shifts in *omitted* variables, which leave slope parameters unaltered even when correlated with included variables. Location shifts in *included* variables induce changes in slopes when there are correlated omitted variables. IIS helps mitigate the adverse impacts of induced location shifts on non-constant intercepts and estimated standard errors, and can provide an automatic intercept correction to improve forecasts following location shifts.

JEL classifications: C51, C22.

KEYWORDS: Model selection, mis-specification, breaks, impulse-indicator saturation, *Autometrics*.

1 Introduction

Omitted variables are a ubiquitous problem in empirical econometrics.[1] Analytical results for stationary DGPs are well-known, including general limiting distributions (see e.g., Hendry, 1979). The theory of reduction characterizes the operations implicitly applied to the data generating process (DGP) to obtain the local DGP (LDGP, the generating process in the space of the variables under analysis: see e.g., Hendry, 2009). Choosing the variables $\mathbf{x}_t = (\mathbf{y}_t, \mathbf{z}_t)$ for analysis determines the properties of the LDGP, and hence of models thereof. Omitting relevant variables ($\mathbf{w}_t$) that are correlated with the included $\mathbf{z}_t$ creates an LDGP with parameters that do not coincide with those that determine outcomes. Nevertheless, LDGPs can be written after sequential factorization with mean-innovation errors.

*The research was supported in part by grants from the Open Society Foundations and the Oxford Martin School. Comments from the editors and 3 anonymous referees are acknowledged, as are helpful discussions with Neil R. Ericsson, Roger Hammersland and Bent Nielsen. 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 many insightful discussions spanning four decades. It builds on the issues of model specification and evaluation that were the focus of Granger (1999), leading to the exchange about automatic model selection in Granger and Hendry (2005).

However, the empirically-relevant setting is one where structural breaks occur. We focus on location shifts, where the previous unconditional means of variables shift. A linear static model highlights the impacts of such shifts, as their full effect is immediate. An LDGP can be written with constant parameters using indicators to capture such location shifts. If unmodeled, however, location shifts can induce non-constant estimated parameters in models of the LDGP, so deliver incorrect policy implications. When it is not known what variables $\mathbf{w}_t$ are omitted, their correlations with included variables, or the numbers, timings and magnitudes of location shifts, a portmanteau approach is required to detect breaks at any point. We show that the impacts of shifts in unknowingly omitted variables can be mitigated to re-establish near parameter constancy in models of the LDGP using impulse-indicator saturation (denoted IIS: see Hendry, Johansen and Santos, 2008, and Johansen and Nielsen, 2009).

Although a model of the LDGP could also be under-specified by omitting relevant lags $\{\mathbf{z}_{t-s}\}$ or non-linear transformations $\mathbf{f}(\mathbf{z}_t)$, such difficulties can be avoided by commencing with a sufficiently general set of variables to nest the LDGP. We use the multi-path automatic modeling approach *Autometrics* to select a minimal undominated model (see Doornik, 2009a). As the LDGP is the generating process in the space of the variables under analysis and can always be written with constant parameters and mean-innovation errors, our approach searches for constant-parameter congruent models thereof (see Hendry, 1995, and Bontemps and Mizon, 2008). Castle and Hendry (2011) and Castle, Doornik and Hendry (2011, 2012) respectively consider selection in non-linear and in dynamic models, and the last show that IIS can also take account of non-normality in the form of fat tails.

The structure of the paper is as follows. Section 2 analyzes a static model mis-specified by omitting variables that have location shifts, interacting with mean shifts in included variables. Section 3 uses IIS to detect location shifts. Section 4 provides a Monte Carlo study of the effects when the included and omitted variables' DGPs change. Section 5 considers forecasting and policy. Section 6 concludes.

2 Location shifts in included and excluded variables

Consider a linear, static, constant-parameter, DGP with independent white-noise errors:

$$y_t = \boldsymbol{\beta}_1'\mathbf{z}_t + \boldsymbol{\beta}_2'\mathbf{w}_t + \epsilon_t \text{ where } \epsilon_t|\mathbf{z}_t, \mathbf{w}_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1}$$

when $\mathsf{IN}\left[\theta, \sigma^2\right]$ denotes an independent normal distribution with mean θ and variance σ^2, with:

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

Both $\mathbf{w}_t$ and $\mathbf{z}_t$ have one-off location shifts at times $1 < T^0 < T$ and $1 < T^* < T$, with $\nabla\boldsymbol{\mu} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ and $\nabla\boldsymbol{\delta} = (\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1)$ where:

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\mu}_1 & t < T^0 \\ \boldsymbol{\mu}_2 & t \geq T^0 \end{cases} \quad \text{and} \quad \boldsymbol{\delta}_t = \begin{cases} \boldsymbol{\delta}_1 & t < T^* \\ \boldsymbol{\delta}_2 & t \geq T^* \end{cases} \tag{3}$$

so $\boldsymbol{\mu}_t = \boldsymbol{\mu}_1 + \nabla\boldsymbol{\mu}1_{\{t \geq T^0\}}$ and $\boldsymbol{\delta}_t = \boldsymbol{\delta}_1 + \nabla\boldsymbol{\delta}1_{\{t \geq T^*\}}$ where $1_{\{\cdot\}}$ denotes an indicator function (unity when the argument is true, zero otherwise). A conditional regression model is inadvertently mis-specified as:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \gamma_0 + \boldsymbol{\gamma}_1'\mathbf{z}_t \tag{4}$$

including an intercept as $\mathsf{E}\left[\mathbf{w}_t\right] \neq \mathbf{0}$. We would like to investigate the consequences of estimating this mis-specified model and how any possibly adverse outcomes can be remedied by IIS.

There is an induced non-constancy in the parameters of (4) from the relationship between the omitted and included variables. Conditioning $\mathbf{w}_t$ on $\mathbf{z}_t$ in (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 $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. From (1) and (5), the LDGP for y_t is:

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

The intercept is non-constant, but the slope parameter is constant, though estimates of model (4) need not reflect those implications. To analyze those full-sample estimators given the LDGP (6):

$$\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\boldsymbol{\gamma}}_1 \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T}\mathbf{z}_t' \\ \sum_{t=1}^{T}\mathbf{z}_t & \sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T}\mathbf{z}_t y_t \end{pmatrix} \tag{7}$$

requires sub-sample derivations due to the breaks in $\boldsymbol{\beta}_2'\left(\boldsymbol{\delta}_t - \boldsymbol{\Psi}\boldsymbol{\mu}_t\right)$. We take $T^* < T^0$ for explicit calculations, letting $\lambda_{T^0} = \left(T^0 - 1\right)/T$ and $\kappa_{T^*} = \left(T^* - 1\right)/T$, with $\mathbf{r} = \left(\boldsymbol{\mu}_1 + (1 - \lambda_{T^0})\nabla\boldsymbol{\mu}\right)$, $\mathbf{s} = \left(\boldsymbol{\delta}_1 + (1 - \kappa_{T^*})\nabla\boldsymbol{\delta}\right)$ and $\mathbf{M} = \left(\lambda_{T^0}\boldsymbol{\mu}_1\boldsymbol{\delta}_1' + (\kappa_{T^*} - \lambda_{T^0})\boldsymbol{\mu}_2\boldsymbol{\delta}_1' + (1 - \kappa_{T^*})\boldsymbol{\mu}_2\boldsymbol{\delta}_2'\right)$ so $\mathbf{M} - \mathbf{r}\mathbf{s}' = \lambda_{T^0}\left(1 - \kappa_{T^*}\right)\nabla\boldsymbol{\mu}\nabla\boldsymbol{\delta}'$. As the relevant second moments are finite positive definite, expectations of (7) yield the following Nagar (1959) approximations where the subscript $_{(T)}$ denotes a full-sample estimate:

$$\mathsf{E}\left[\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\boldsymbol{\gamma}}_1 \end{pmatrix}\right] \approx \begin{pmatrix} \left(\mathbf{s}' - \mathbf{r}'\mathbf{H}^{-1}\left(\boldsymbol{\Sigma}_{12} + \lambda_{T^0}\left(1 - \kappa_{T^*}\right)\nabla\boldsymbol{\mu}\nabla\boldsymbol{\delta}'\right)\right)\boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_1 + \mathbf{H}^{-1}\left(\boldsymbol{\Sigma}_{12} + \lambda_{T^0}\left(1 - \kappa_{T^*}\right)\nabla\boldsymbol{\mu}\nabla\boldsymbol{\delta}'\right)\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \gamma_{0(T)} \\ \boldsymbol{\gamma}_{1(T)} \end{pmatrix} \tag{8}$$

where $\mathbf{H} = \boldsymbol{\Sigma}_{11} + \lambda_{T^0}\left(1 - \lambda_{T^0}\right)\nabla\boldsymbol{\mu}\nabla\boldsymbol{\mu}'$. Sub-sample results are obtained by setting λ_{T^0} and/or κ_{T^*} to zero and unity respectively. Fitting separately to the sub-samples would reveal the shifted estimates, but requires knowledge of when the breaks occurred; recursive estimation could also reflect that problem if the initial sample was less than $T^* < T^0$, with estimates shifting to those in (8). The impacts on the intercept from terms involving $\boldsymbol{\beta}_1$ always cancel exactly in (8): the estimated intercept is unaffected by the relevance or otherwise of $\mathbf{z}_t$ in the LDGP, and is non-constant when either $\nabla\boldsymbol{\mu}$ or $\nabla\boldsymbol{\delta}$ are non-zero.

(8) shows that the postulated model (4) does not have constant parameters, so should be written as:

$$y_t = \gamma_{0(T)} + \boldsymbol{\gamma}'_{1(T)}\mathbf{z}_t + v_t \tag{9}$$

Let $\eta_t = \boldsymbol{\beta}'_2\mathbf{u}_t + \epsilon_t$ and $\varphi_t = (\boldsymbol{\beta}'_2\boldsymbol{\delta}_t - \gamma_{0(T)} + (\boldsymbol{\beta}'_1 - \boldsymbol{\gamma}'_{1(T)})\boldsymbol{\mu}_t)$, then:

$$v_t = \varphi_t + \left(\left(\boldsymbol{\beta}'_1 + \boldsymbol{\beta}'_2\boldsymbol{\Psi}\right) - \boldsymbol{\gamma}'_{1(T)}\right)(\mathbf{z}_t - \boldsymbol{\mu}_t) + \eta_t \tag{10}$$

Despite $\mathsf{E}[\mathbf{z}_t - \boldsymbol{\mu}_t] = \mathbf{0}$, $\mathsf{V}[\mathbf{z}_t - \boldsymbol{\mu}_t] = \boldsymbol{\Sigma}_{11}\ \forall t$, and $\mathsf{E}[T^{-1}\sum_{t=1}^{T} v_t] = 0$, using (8) and $\boldsymbol{\beta}'_2\nabla\boldsymbol{\delta} = \tau$:

$$\varphi_t = (\kappa_{T^*} - 1_{\{t<T^*\}})\tau + \left(\lambda_{T^0} - 1_{\{t<T^0\}}\right)\nabla\boldsymbol{\mu}'\mathbf{H}^{-1}\left(\boldsymbol{\Sigma}_{12} + \lambda_{T^0}\left(1 - \kappa_{T^*}\right)\nabla\boldsymbol{\mu}\nabla\boldsymbol{\delta}'\right)\boldsymbol{\beta}_2 \tag{11}$$

Consequently, v_t in (10) will be heteroskedastic and autocorrelated if any of the shifts $\boldsymbol{\delta}_t$ or $\boldsymbol{\mu}_t$ are step functions. In congruent models, inference and decisions during model selection can be based on conventional procedures, such as t-statistics. Here heteroskedastic-consistent standard errors (HCSEs: White, 1980), and autocorrelation-consistent (HACSEs) generalizations thereof (see e.g., Andrews, 1991) appear to be needed. However, the LDGP errors η_t are neither heteroskedastic nor autocorrelated, so HACSEs would not correctly attribute such problems to unmodeled breaks. Although φ_t vanishes in expectation given the definitions of κ_{T^*} and λ_{T^0}, the presence of the indicators in (11) entails that shifts affect all sub-sample estimates unless $t < T^* < T^0$. Shifts in intercepts can be removed by including the appropriate indicator variables $1_{\{t<T^*\}}$ and $1_{\{t<T^0\}}$, but those are unknown, and although their location can be consistently estimated, in practice breaks may be more frequent than just one at a common date in each variable set, so section 3 considers using IIS to approximate the indicators. We now analyze the two cases where one of $\nabla\boldsymbol{\mu}$ or $\nabla\boldsymbol{\delta}$ is zero.

2.1 A mean shift occurs only in the excluded variables

When $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{r} = \boldsymbol{\mu}$, then $\mathbf{H} = \boldsymbol{\Sigma}_{11}$ so $\boldsymbol{\gamma}_{1(T)} = \boldsymbol{\gamma}_1$ with the constant-parameter LDGP:

$$y_t = \left(\boldsymbol{\beta}_2'\boldsymbol{\delta}_1 + \boldsymbol{\beta}_1'\boldsymbol{\mu}\right) + \left(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2'\boldsymbol{\Psi}\right)(\mathbf{z}_t - \boldsymbol{\mu}) + \tau 1_{\{t \geq T^*\}} + \eta_t \tag{12}$$

Hence, once the correct indicator is included, a congruent model matching the LDGP (12) is:

$$y_t = \gamma_0 + \boldsymbol{\gamma}_1'\mathbf{z}_t + \gamma_2 1_{\{t \geq T^*\}} + \eta_t \tag{13}$$

Taking deviations from means for $\kappa_{T^*} < 1$ when (1) and (2) hold, full-sample estimation of (13) yields:

$$\sqrt{T}\begin{pmatrix} \widehat{\gamma}_1 - (\boldsymbol{\beta}_1 + \boldsymbol{\Psi}'\boldsymbol{\beta}_2) \\ \widehat{\gamma}_2 - \boldsymbol{\beta}_2'(\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1) \end{pmatrix} \xrightarrow{D} \mathsf{N}_{k_1+1}\left[\mathbf{0}, \sigma_\eta^2 \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0}' & (\kappa_{T^*}(1 - \kappa_{T^*}))^{-1} \end{pmatrix}\right] \tag{14}$$

where $\sigma_\eta^2 = \sigma_\epsilon^2 + \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2$. Thus, adding the indicator $1_{\{t \geq T^*\}}$ to a regression of y_t on an intercept and $\mathbf{z}_t$ delivers a constant parameter congruent model of the LDGP (12), with estimators having a normal distribution, and a homoskedastic error, so conventional inference will be valid, albeit unknowingly about $(\boldsymbol{\beta}_1 + \boldsymbol{\Psi}'\boldsymbol{\beta}_2)$ rather than $\boldsymbol{\beta}_1$. Forecast errors will have their anticipated properties of normal, mean zero and constant variance, provided no further breaks occur and it is known to extend the break dummy. To the extent that model selection with IIS captures $1_{\{t \geq T^*\}}$, its results will approximate (13) and (14).

When no location shift indicator is included, from (11) the unexplained component of the model becomes $v_t = \left(\kappa_{T^*} - 1_{\{t < T^*\}}\right)\tau + \eta_t$ so:

$$\mathsf{E}\left[v_t^2\right] \approx \begin{pmatrix} \sigma_\eta^2 + (1 - \kappa_{T^*})^2\tau^2 & \text{for } t < T^* \\ \sigma_\eta^2 + \kappa_{T^*}^2\tau^2 & \text{for } t \geq T^* \end{pmatrix}$$

and hence is heteroskedastic. For a sufficiently large sample before and after the break, with fixed κ_{T^*}:

$$\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left[v_t^2\right] = \sigma_\epsilon^2 + \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2 + \kappa_{T^*}(1 - \kappa_{T^*})\tau^2 = \varpi^2 \tag{15}$$

so the full-sample residual variance is inflated by $\kappa_{T^*}(1 - \kappa_{T^*})\tau^2$ over σ_η^2. Despite the LDGP errors η_t being IID, the full-sample v_t are autocorrelated (and also manifest ARCH, Engle, 1982):

$$\mathsf{E}\left[v_t v_{t-k}\right] = \rho_k \approx \frac{\tau^2\kappa_{T^*}\left(1 - \kappa_{T^*} - \frac{k}{T}\right)}{\sigma_\eta^2 + \tau^2\kappa_{T^*}(1 - \kappa_{T^*})} \quad k = 1, \ldots, T - T^* + 1 \tag{16}$$

so:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t \xrightarrow{D} \mathsf{D}\left[0, \varpi^2 \boldsymbol{\iota}'\boldsymbol{\Omega}\boldsymbol{\iota}\right]$$

where $\boldsymbol{\iota}$ is a vector of 1s and $\boldsymbol{\Omega}$ is a symmetric band matrix with 1s on the diagonal, and ρ_k on the k^{th} off-diagonal then zeroes for $k > T - T^*$. As the fourth cumulant of a normal is zero, using the approach in Hendry (1979), since none of the three triples of four terms is non-zero, and $\mathsf{E}[\sum_{t=1}^{T}(\mathbf{z}_t - \boldsymbol{\mu})\, v_t] = \mathbf{0}$ by construction, then:

$$\mathsf{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbf{z}_t v_t v_s \mathbf{z}_s'\right] = \mathbf{0} \tag{17}$$

so:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} (\mathbf{z}_t - \boldsymbol{\mu})\, v_t \xrightarrow{D} \mathsf{D}\left[\mathbf{0}, \varpi^2 \boldsymbol{\Sigma}_{11}\right].$$

However, inference on the intercept is distorted by the residual autocorrelation:

$$\mathsf{E}\left[\left(\widehat{\gamma}_0 - \gamma_{0(T)}\right)^2\right] = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathsf{E}\left[v_t v_s\right] = \frac{\varpi^2}{T}\boldsymbol{\iota}'\boldsymbol{\Omega}\boldsymbol{\iota}$$

so its conventionally estimated standard errors will differ from sampling standard deviations. Combining:

$$\sqrt{T}\left(\begin{array}{c} \widehat{\gamma}_0 - \gamma_{0(T)} \\ \widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{array}\right)_{\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2} \xrightarrow{D} \mathsf{D}_{k_1+1}\left[\mathbf{0}, \varpi^2 \left(\begin{array}{cc} \boldsymbol{\iota}'\boldsymbol{\Omega}\boldsymbol{\iota} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{11}^{-1} \end{array}\right)\right] \tag{18}$$

2.2 A mean shift occurs only in the included variables

In a correctly specified model, shifts in the means of included variables would have the beneficial effect of increasing information, yet induce shifts in estimated slopes as well as intercepts when there are correlated excluded variables. When $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{s} = \boldsymbol{\delta}$, then $\mathbf{H} \neq \boldsymbol{\Sigma}_{11}$ as $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, so estimates of both slope and intercept will be non-constant. Because the latter is continually changing with λ_{T^0}, conditional expectations are non-constant, and 1-step ahead predictors are neither unbiased nor minimum mean-square error, so distributions of estimators vary accordingly.

2.3 Implications

Both $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2$ are needed to avoid the intercept shifting, but the slope parameter $\boldsymbol{\gamma}_1$ does not shift when either excluded or included variables have mean shifts. However, the estimated slope parameter $\boldsymbol{\gamma}_1$ shifts if the DGP of the *included* variables, $\mathbf{z}_t$, changes when $\mathbf{w}_t$ is omitted. Since $\varpi^2 > \sigma_\eta^2$ by $\kappa_{T^*}(1 - \kappa_{T^*})\,\tau^2$, when τ^2 is large, comparing (14) and (18) shows that |t|-statistics could be smaller than the chosen critical value c_α when the indicator is excluded, yet larger with, exacerbating the omitted variables problem. (14) shows that adding an indicator that matches the shift removes all non-constancies, so could avoid false deletions of apparently insignificant but relevant variables; residual autocorrelation and heteroskedasticity are also removed. Section 3 discusses the extent to which such implications can be matched by IIS removing the location-shift induced non-constancies in estimated intercepts and equation standard errors. Finally, there are two types of risk with a mis-specified model, using a quadratic loss function over all the parameters for a single location shift. First, losses for $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ relative to $\gamma_{0(T)}$ and $\gamma_{1(T)}$ are small in large samples. However, for any estimator that does not handle the break, equation (8) shows that the risks associated with $\gamma_{0(T)}$ must be unbounded compared to the intercept in the LDGP (6) or the DGP (1) as the break magnitude increases. Conversely, based on the analysis in Hendry and Santos (2010), the potency of IIS will then tend to unity, leading to bounded risk. Section 5.1 explains why the intercept should be included in risk assessment.

3 Impulse-indicator saturation

IIS includes in the set of candidate regressors an impulse indicator for every observation, so adds T variables when there are T observations. The distributional properties of IIS under the null of no outliers or location shifts are analyzed in Hendry *et al.* (2008). In their analysis (the 'split-half' approach, which can be generalized to multiple splits of unequal size), the regression includes the first $T/2$ indicators and the regressors (N), provided $(N + T/2) < T$. By dummying out the first half of the observations, estimates are based on the remaining data, so any first-half observations that are discrepant will have significant indicators: see Salkever (1976). Their locations are recorded, the first $T/2$ indicators are replaced by the second half, and the procedure repeated. The recorded indicators are combined for selection of those that remain significant. Johansen and Nielsen (2009) extend the analysis to both stationary and unit-root autoregressions and relate IIS to robust estimation. They show that under the null of no location shifts or outliers, the cost of applying IIS is the loss of αT observations, where α is the significance level. Thus, at $\alpha = 0.01$, IIS is 99% efficient for $T = 100$ despite including 100 'irrelevant' impulse indicators in the search set and checking for location shifts and outliers at every data point. IIS entails more candidate variables than observations as $N + T > T$, but selection is feasible using expanding and contracting block searches.

If the intercept shifts were known the relevant step dummies that changed at T^* and T^0 could be included, but this is an infeasible knowledge level when it is not known that $\mathbf{w}_t$ is relevant. We will examine the extent to which IIS can mimic that effect. Castle *et al.* (2012) established the ability of IIS to detect multiple internal location shifts and outliers, including breaks close to the start and end of the sample. For a single break $\boldsymbol{\delta}_2 \neq \boldsymbol{\delta}_1$, Hendry and Santos (2010) show the detection power is determined by the magnitude of the shift τ; sample size T; the length of the break interval $T - T^*$; the equation error variance σ_η^2; and the significance level, α, using a normal-distribution critical value c_α. Let ψ_t denote the non-centrality for a single indicator $\widehat{\tau}_t$ conditional on removing all others, with $\mathsf{V}[\widehat{\tau}_t] = \omega_t$, then:

$$\psi_t = \mathsf{E}\left[\mathsf{t}_{\tau_t=0}\right] = \mathsf{E}\left[\frac{\widehat{\tau}_t}{\sqrt{\widehat{\omega}_t}}\right] \approx \frac{\tau}{\sqrt{\sigma_\epsilon^2 + \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2}} \tag{19}$$

with non-null rejection frequency for $\tau > 0$ of approximately:

$$\mathsf{P}\left[|\mathsf{t}|_{\tau_t=0} > c_\alpha \mid \psi_t\right] \approx \mathsf{P}\left[|\mathsf{t}|_{\tau_t=0} > (c_\alpha - \psi_t)\right] \tag{20}$$

When $\alpha = 0.01$, so $c_\alpha \approx 2.6$ and $\psi_t = 3$, the rejection frequency will be about 70%. A failure to detect other indicators will lower this power, whereas larger shifts will have higher power. Testing the significance of the block of impulse indicators that precisely correspond to a known break period would be equivalent to a Chow (1960) $\mathsf{F}^{T^*}_{T-T^*-n}$-test, with a similar non-centrality, whereas testing the correct step indicator by a t-test would have a non-centrality that is larger by $(1-r)^2$ where $r = (T - T^*)/T$ (see Hendry and Santos, 2010). Castle *et al.* (2012) compare an IIS-based test with that proposed by Bai and Perron (1998) and find similar power when all location shifts lie within the central sample, but improved detection when there are large breaks and outliers close to the start or end of the sample.

4 Simulation results

Monte Carlo experiments are used to assess the magnitude implications of under-specification when there are non-constant parameters as outlined in §2.3. The DGP is given by (1) and (2) for scalar z_t and w_t.[2] 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 $T = 100$. $M = 10,000$ replications are undertaken and regressors are stochastic. 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. The simulations were conducted without diagnostic testing and with a forced intercept which means that selection is not applied to the intercept; it is always retained.[3]

[2]Alternative DGP specifications including dynamic models have been considered but are omitted due to space constraints. 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. Results are available on request.

[3]All results and graphs were computed using Ox (Doornik, 2009b). *Autometrics* can undertake diagnostic testing when a terminal model is reached, backtracking to find a valid reduction if any mis-specification tests fail. Irrelevant variables can then proxy a chance departure from the null of a mis-specification or encompassing test, and be retained despite insignificance.

4.1 IIS applied to a correctly-specified model

First, we check the properties of applying IIS when the model matches the DGP, so there are no induced shifts due to omitted variables. We consider a break in both variables with $T^0 = 91$, $\mu_1 = 0$, $\mu_2 = -5$, $T^* = 81$, $\delta_1 = 0$ and $\delta_2 = 5$. The outcomes do not depend on whether or not $T^* \lesseqgtr T^0$. The model augments the DGP in (1) with an intercept:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 w_t + \nu_t \tag{21}$$

The optimal step-shift dummies to match the breaks are given by:

$$D_{z,t} = \begin{cases} 0 & t < T^0 \\ 1 & t \geq T^0 \end{cases} \quad \text{and} \quad D_{w,t} = \begin{cases} 0 & t < T^* \\ 1 & t \geq T^* \end{cases} \tag{22}$$

with parameters $\beta_{s,z}$ and $\beta_{s,w}$ respectively. Results are reported in Table 1, where 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. As the step-shift dummies are unnecessary, their retention rate checks the null retention frequency.

Let $\widetilde{\beta}_{i,k}$ denote the outcome on the k^{th} trial, then $\overline{\boldsymbol{\beta}}_i = \frac{1}{M}\sum_{k=1}^{M}\widetilde{\boldsymbol{\beta}}_{i,k}$, with Monte Carlo standard deviation, MCSD= $\sqrt{(\frac{1}{M-1}\sum_{k=1}^{M}(\widetilde{\boldsymbol{\beta}}_{i,k} - \overline{\boldsymbol{\beta}}_i)^2)}$ and the estimated standard error, ESE= $\frac{1}{M}\sum_{k=1}^{M}\mathsf{SE}[\widetilde{\boldsymbol{\beta}}_{i,k}]$, with HACSEs computed using Andrews (1991). Similarly for other simulated estimators like $\widetilde{\gamma}_i$ and $\widetilde{\sigma}$.

In Table 1, mean coefficient estimates and the mean equation standard error are close to their DGP values. The outcomes are almost identical with and without selection. As the model is correctly specified, the step-shift dummies are redundant and are rarely retained. Applying IIS reduces the equation standard error by removing approximately 1.4 observations on average when there is a break in both regressors, but there is no effect on coefficient estimates. so IIS has low costs for the correct model specification. HACSEs are close to their uncorrected counterparts for $\overline{\beta}_2$ but there is some divergence for the intercept and $\overline{\beta}_1$ despite no residual heteroskedasticity or autocorrelation.

4.2 IIS applied to a mis-specified model

Table 2 reports the parameter specifications for the experiments considered, with theoretical coefficient values derived in §2, based on (8) with $T = 100$, for breaks of magnitude $5\sigma_\epsilon$. Tables 3–5 record the results, where selection uses *Autometrics* at $\alpha = 1\%$.

Simulations were also conducted with diagnostic tracking, and the number of retained indicators increased slightly, being between 0.1 and 0.2 larger than in tables 4 and 5 below, but without substantive effects on parameter estimates. Detailed results with diagnostic tracking switched on and forced regressors, where selection is not applied to z_t, are available on request.

	Estimating the model	Selection from model with known break dummies	Selection from model with IIS
Breaks in z_t and w_t			
$\overline{\beta}_0$	$0.000\,[0.110]$ $(0.110)\,[(0.008)]$ $(0.107)\,[(0.016)]$	$0.000\,[0.110]$ $(0.110)\,[(0.008)]$ $(0.056)\,[(0.083)]$	$0.000\,[0.117]$ $(0.106)\,[(0.009)]$ $(0.119)\,[(0.077)]$
$\overline{\beta}_1$	$1.000\,[0.060]$ $(0.060)\,[(0.006)]$ $(0.054)\,[(0.011)]$	$1.000\,[0.066]$ $(0.061)\,[(0.007)]$ $(0.107)\,[(0.016)]$	$0.999\,[0.064]$ $(0.058)\,[(0.006)]$ $(0.103)\,[(0.016)]$
$\overline{\beta}_2$	$1.000\,[0.048]$ $(0.048)\,[(0.004)]$ $(0.044)\,[(0.009)]$	$1.000\,[0.057]$ $(0.049)\,[(0.007)]$ $(0.054)\,[(0.012)]$	$1.000\,[0.051]$ $(0.046)\,[(0.005)]$ $(0.052)\,[(0.010)]$
$\overline{\beta}_{s,z}$		$-0.001\,[0.208]$ $(0.007)\,[(0.070)]$ $(0.001)\,[(0.005)]$	
$\overline{\beta}_{s,w}$		$0.002\,[0.199]$ $(0.007)\,[(0.067)]$ $(0.001)\,[(0.013)]$	
$\overline{\sigma}_\epsilon$	$0.998\,[0.071]$	$0.997\,[0.071]$	$0.947\,[0.085]$
ι	-	-	1.417

Table 1: Monte Carlo results for the correctly specified model with breaks in z_t and w_t. ESEs reported in parentheses and MCSDs in brackets, with uncorrected standard errors given below parameter estimates and HACSEs below the uncorrected standard errors.

Case (i), where the equation is mis-specified but there are no breaks, is reported in Table 3.[4] When there are no mean shifts in both the included and omitted variable, the mean coefficient estimates correspond to their theory counterparts. $\overline{\sigma}_v$ is slightly less than σ_v, as IIS retains approximately 1.5 indicators on average, leading to an effective sample size of $\widetilde{T} = 98.5$ instead of $T = 100$.

[4] A break of magnitude $2\sigma_\epsilon$ was also considered: note that the effective shift is smaller as $\sigma_e \simeq 1.4\sigma_\epsilon$. Thus, a t-test of a single indicator at 1% for the $2\sigma_\epsilon$ shift in w_t would have power of about 11%, and in z_t, as $\beta_2\Psi(\mu_1 - \mu_2)\sigma_e \simeq 0.7$, power would be about 3%. The simulations respectively delivered potencies of 4.1% and 1.4%. Corresponding powers for a single $5\sigma_\epsilon$ shift would be approximately 83% and 20%, compared to the simulation potencies of 75% and 8% in those cases. Further results are available on request.

H_a	Parameters	Break date	$\gamma_{0(T)}$	$\gamma_{1(T)}$	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	$\mu_1 = \mu_2 = \delta_1 = 0$, $\delta_2 = 5$	$T^0 = 0$, $T^* = 81$	1	1.5	2.45
(iii) Break in z_t	$\mu_1 = \delta_1 = \delta_2 = 0$, $\mu_2 = -5$	$T^0 = 81$, $T^* = 0$	0.1	1.1	
(iv) 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 values under mis-specifications, with theoretically derived coefficients.

	Estimating the model	Selection from model with IIS
Constant z_t and w_t		
$\overline{\gamma}_0$	$0.002\,[0.133]$ $(0.133)\,[(0.010)]$ $(0.129)\,[(0.017)]$	$0.001\,[0.141]$ $(0.127)\,[(0.011)]$ $(0.154)\,[(0.052)]$
$\overline{\gamma}_1$	$1.498\,[0.134]$ $(0.134)\,[(0.014)]$ $(0.128)\,[(0.022)]$	$1.497\,[0.142]$ $(0.128)\,[(0.014)]$ $(0.124)\,[(0.017)]$
$\overline{\sigma}_v$	$1.320\,[0.094]$	$1.252\,[0.112]$
ι	-	1.474

Table 3: Monte Carlo results for the mis-specified model when w_t is omitted, case (i): legend as Table 1.

4.3 The excluded variable shifts

When the omitted variable breaks, but the included variable remains constant, (case (ii)), mean estimates are close to their theoretical counterparts: see Table 4. If the break is not accounted for, despite the LDGP innovation error inference is affected by the heteroskedastic and autocorrelated residuals, as shown by the large $\overline{\sigma}_v$ and the difference between mean ESEs (or mean HACSEs) and MCSDs for the intercept. The constant is then usually significant, although some theory models (e.g., of consumption and inflation) entail zero intercepts (here $\gamma_0 = 0$).

	Estimating the model	Selection from model with known break dummy	Selection from model with IIS
Constant z_t, break in w_t, $\delta_2 = 5$			
$\overline{\gamma}_0$	$1.002\,[0.134]$ $(0.242)\,[(0.013)]$ $(0.448)\,[(0.030)]$	$0.001\,[0.147]$ $(0.148)\,[(0.011)]$ $(0.294)\,[(0.065)]$	$0.199\,[0.249]$ $(0.162)\,[(0.027)]$ $(0.246)\,[(0.086)]$
$\overline{\gamma}_1$	$1.500\,[0.245]$ $(0.243)\,[(0.022)]$ $(0.230)\,[(0.044)]$	$1.498\,[0.135]$ $(0.134)\,[(0.014)]$ $(0.144)\,[(0.021)]$	$1.431\,[0.385]$ $(0.151)\,[(0.043)]$ $(0.181)\,[(0.063)]$
$\overline{\gamma}_{s,w}$		$5.003\,[0.334]$ $(0.332)\,[(0.024)]$ $(0.129)\,[(0.023)]$	
$\overline{\sigma}_v$	$2.404\,[0.123]$	$1.320\,[0.095]$	$1.484\,[0.281]$
ι	-	-	15.50

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

Figure 1 records the simulation-mean recursive parameter estimates with $\pm$2ESE and $\pm$2MCSD. When the break occurs in the omitted variable, the intercept shifts (top-left panel), but the slope param-

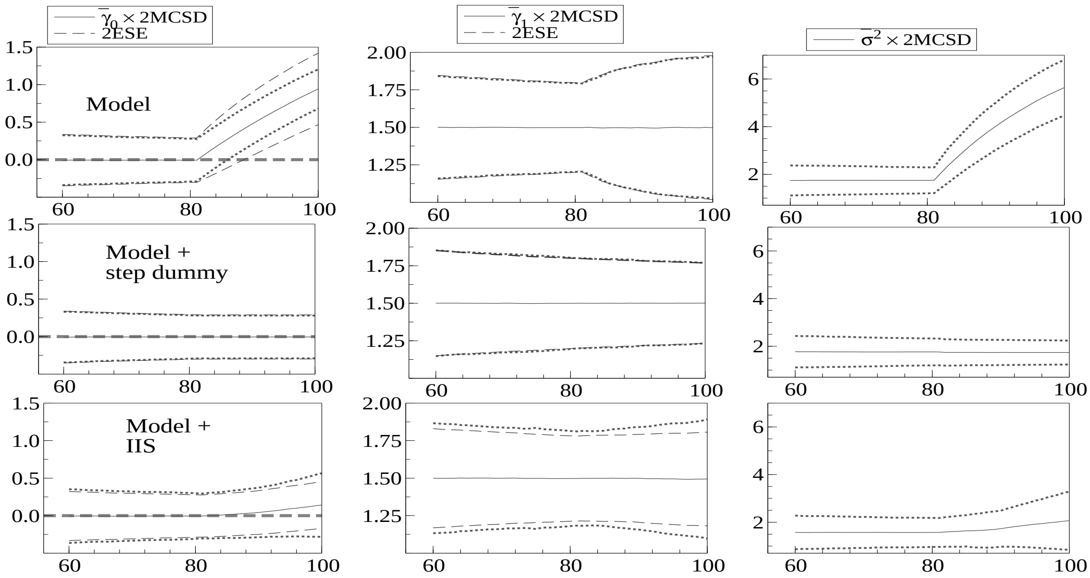

Figure 1: Incorrect specification with non-constant w_t, $\delta_1 = 0$, $\delta_2 = 5$: case (ii).

eter remains constant, albeit with increasing standard errors (next panel in first row), as goodness of fit deteriorates sharply. Including the correct step-shift dummy results in a constant intercept (middle-row left panel) as well as a constant slope parameter and a constant estimated equation standard error. Thus, knowing when there are breaks in omitted variables delivers a constant-parameter model, but is infeasible. IIS helps account for the unknown break using selected indicators. For a break of 5 standard deviations, IIS retains 15.5 indicators on average (for a break of 20 observations). As the indicators all have similar magnitudes and same signs, they could be combined into a step-shift dummy. The estimated intercept increases slightly after the break, but most of the estimated parameter non-constancy in the top row is removed (third row). Increasing α to 2.5% retains more indicators, so the non-constancy of the intercept is almost fully removed, at a cost of retaining more indicators under the null. $\overline{\sigma}_v$ is slightly larger than if a known step-shift dummy were included, but less than half that of just estimating the model.

4.4 The included variable shifts

When the included variable alone breaks, both estimated parameters are non-constant, see Table 5. Section 5.1 shows the intercept will therefore shift with policy changes in z_t. The shift in the intercept is smaller than for a break in δ_t: with $\Psi = 0.5$, the equivalent shift in z_t is only half that of w_t. The step-shift dummy is significant, and although it is estimated imprecisely, it corrects the non-constancy in both the intercept and slope parameter. IIS picks up fewer indicators, close to the number retained in the constant parameter model. When the correlation with the omitted variable is higher, e.g., $\Psi = 0.8$, IIS retains 16 indicators on average so results are similar to a break in the omitted variable, with near estimated parameter constancy.

	Estimating the model	Selection from model with known break dummy	Selection from model with IIS
Break in z_t, constant w_t, $\mu_2 = -5$			
$\overline{\gamma}_0$	$0.099\,[0.150]$ $(0.153)\,[(0.011)]$ $(0.147)\,[(0.022)]$	$0.016\,[0.155]$ $(0.148)\,[(0.011)]$ $(0.541)\,[(0.297)]$	$0.097\,[0.157]$ $(0.147)\,[(0.012)]$ $(0.153)\,[(0.093)]$
$\overline{\gamma}_1$	$1.097\,[0.062]$ $(0.063)\,[(0.005)]$ $(0.062)\,[(0.011)]$	$1.435\,[0.218]$ $(0.115)\,[(0.033)]$ $(0.114)\,[(0.021)]$	$1.098\,[0.066]$ $(0.060)\,[(0.006)]$ $(0.143)\,[(0.021)]$
$\overline{\gamma}_{s,z}$		$2.104\,[1.305]$ $(0.555)\,[(0.325)]$ $(0.096)\,[(0.058)]$	
$\overline{\sigma}_v$	$1.393\,[0.099]$	$1.325\,[0.097]$	$1.329\,[0.113]$
ι	-	-	1.63

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

4.5 Both included and excluded variables shift

When breaks occur in both included and omitted variables (case (iv)), the LDGP without indicators is non-constant, and estimates of both the intercept and slope parameters accordingly shift, with the intercept changing from the point at which a break in z_t occurs. Figure 2 shows the simulation-mean recursive parameter estimates with $\pm$2ESE and $\pm$2MCSD for a break of $5\sigma_\epsilon$ in both variables. The step-shift dummies mop up all of the intercept non-constancy. IIS retains all 20 indicators on average, proxying the step-shift dummy. As a result, parameter estimates are close to their LDGP values, are constant, and the equation standard error only increases slightly, in sharp contrast to the uncorrected case with substantial shifts in estimated parameters and large increases in standard errors.

4.6 Multiple breaks in included and excluded variables

For the breaks considered so far, the correct single step-shift dummy excels. In practice, multiple breaks occur, so we examine intermittent mean shifts of 5 observations. The DGP and model are given by (1) and (4), with the included and omitted variables generated by (2) where:

$$\mu_t = \mu_1 + \mu_2\left[I_{20} + \cdots + I_{25} + I_{40} + \cdots + I_{45} + I_{60} + \cdots + I_{65} + I_{80} + \cdots + I_{85}\right]$$

$$\delta_t = \delta_1 + \delta_2\left[I_{20} + \cdots + I_{25} + I_{40} + \cdots + I_{45} + I_{60} + \cdots + I_{65} + I_{80} + \cdots + I_{85}\right]$$

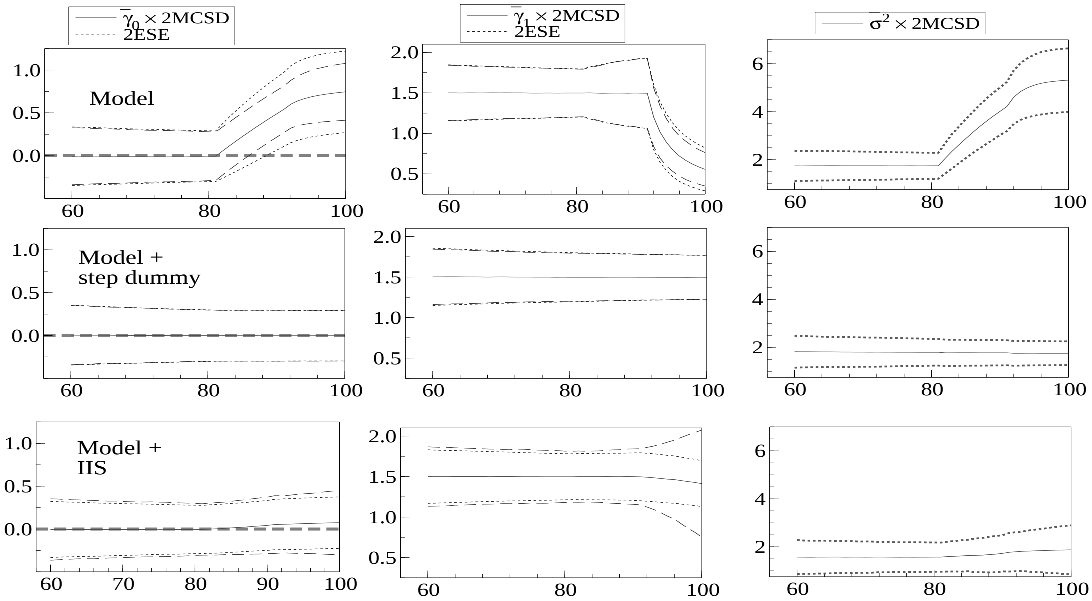

Figure 2: Mis-specified model with non-constant z_t and w_t, $\delta_1 = 0$, $\delta_2 = 5$, $\mu_1 = 0$, $\mu_2 = -5$.

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 $t = 6$. $M = 10,000$ replications are undertaken with stochastic regressors and a forced intercept. Table 6 reports the results for multiple breaks of $5\sigma_\epsilon$ in each of included and omitted variables.

For the mean shift in w_t, the intercept shifts at the point of the break, but IIS mitigates most of this non-constancy by retaining approximately 16 indicators on average. With IIS, $\overline{\sigma}_v$ is almost half that in the mis-specified model, and the power to detect the breaks is high. For intermittent mean shifts in z_t, the non-constancy is not as evident as the break magnitude is smaller and hence more difficult to detect.

5 1-step ahead forecasts

We now consider the implications for 1-step ahead forecasting for known z_{T+1} when there is a break in the omitted variable. Table 7 records the root mean square forecast errors (RMSFEs) for a 5σ shift in the excluded variable, using the same parameters as above. Breaks at $T^* = 81$ and $T^* = 99$ are considered. In the second case, break detection tests such as Bai and Perron (1998) have no power as the break lies outside the break detection window. Conversely, IIS does have power to detect breaks at the beginning and end of the sample: see Castle *et al.* (2012).

We consider either a break that continues into the forecast period, or ends at the last in-sample observation so the mean reverts to δ_1 in the forecast period. In the first case, IIS is proxying a step-shift dummy which should be extended into the forecast period. We do this by generating intercept-corrected

	Estimating the model	Selection from model with known break dummies	Selection from model with IIS
Constant z_t, multiple breaks in w_t ($\delta_2 = 5$)			
$\overline{\gamma}_0$	$1.201\,[0.119]$ $(0.245)\,[(0.011)]$ $(0.401)\,[(0.024)]$	$0.001\,[0.134]$ $(0.134)\,[(0.010)]$ $(0.254)\,[(0.048)]$	$0.392\,[0.375]$ $(0.170)\,[(0.038)]$ $(0.265)\,[(0.103)]$
$\overline{\gamma}_1$	$1.801\,[0.247]$ $(0.247)\,[(0.021)]$ $(0.235)\,[(0.042)]$	$1.798\,[0.119]$ $(0.118)\,[(0.012)]$ $(0.131)\,[(0.019)]$	$1.788\,[0.277]$ $(0.170)\,[(0.041)]$ $(0.221)\,[(0.077)]$
$\overline{\gamma}_{s,w}$		$5.002\,[0.272]$ $(0.274)\,[(0.020)]$ $(0.113)\,[(0.020)]$	
$\overline{\sigma}_v$	$2.441\,[0.111]$	$1.163\,[0.084]$	$1.563\,[0.412]$
ι	-	-	15.73
Multiple breaks in z_t, constant w_t ($\mu_2 = -5$)			
$\overline{\gamma}_0$	$0.171\,[0.149]$ $(0.154)\,[(0.011)]$ $(0.147)\,[(0.021)]$	$0.001\,[0.134]$ $(0.134)\,[(0.010)]$ $(0.621)\,[(0.110)]$	$0.167\,[0.158]$ $(0.147)\,[(0.013)]$ $(0.140)\,[(0.073)]$
$\overline{\gamma}_1$	$1.140\,[0.056]$ $(0.058)\,[(0.005)]$ $(0.059)\,[(0.010)]$	$1.798\,[0.120]$ $(0.118)\,[(0.012)]$ $(0.131)\,[(0.019)]$	$1.139\,[0.059]$ $(0.056)\,[(0.005)]$ $(0.142)\,[(0.021)]$
$\overline{\gamma}_{s,w}$		$3.993\,[0.661]$ $(0.651)\,[(0.068)]$ $(0.113)\,[(0.020)]$	
$\overline{\sigma}_v$	$1.370\,[0.098]$	$1.163\,[0.084]$	$1.302\,[0.116]$
ι	-	-	1.45

Table 6: Monte Carlo results for the mis-specified model with multiple breaks. Legend as table 1.

forecasts. The final column of Table 7 refers to an intercept correction when an impulse indicator is retained for the last in-sample observation, $T = 100$, where the intercept correction augments the forecast by the coefficient of the final in-sample indicator $\widehat{y}^{IC}_{T+1} = \widehat{y}_{T+1} + \widehat{\tau}_T 1_T$. In the second case, IIS should not be extended into the forecast period, so the third column, selection with IIS which picks up outliers in the in-sample period only, is appropriate.

If the break extends into the forecast horizon, the uncorrected forecasts are biased, as can be seen in figure 3 which records the distribution of forecast errors for the estimated mis-specified model. If the break was known, a step-shift dummy would be included which acts as an intercept correction (second column), which results in unbiased forecasts. This is an infeasible knowledge level as it would require knowing the timing and persistence of the shifts, so IIS is again used to proxy these. If IIS is applied in-sample but the break is not extrapolated into the forecast period the forecasts will still be biased,

	Estimating the model	Selection with break dummies	Selection with IIS	Selection with IIS and IC
Break continues into forecast period				
Break at $T^{*}=81$	4.237	1.369	5.177	2.808
Break at $T^{*}=99$	5.092	1.633	5.187	2.455
Break ends at $T=100$				
Break at $T^{*}=81$	1.683	1.347	1.371	5.138
Break at $T^{*}=99$	1.350	1.340	1.341	2.482

Table 7: 1-step ahead RMSFEs for a 5σ shift in the omitted variable w_t.

and could be marginally worse than estimating the mis-specified model. However, when an impulse indicator is retained for the final in-sample observation and used as an intercept correction, the forecasts are significantly improved. Here, the forecast is simply adjusted by the magnitude of the last indicator, but more sophisticated adjustments that average across the last few observations could also be used which would reduce the bimodality.

In contrast, if the break does not extend into the forecast horizon then intercept correction is costly. The second column includes a break dummy that exactly matches the shift so provides the infeasible benchmark; selection with IIS is close to this benchmark. If the break is short, the impact of not modeling the break on 1-step forecasts is small, in the latter case there are just 2 outliers at $T = 99$ and 100.

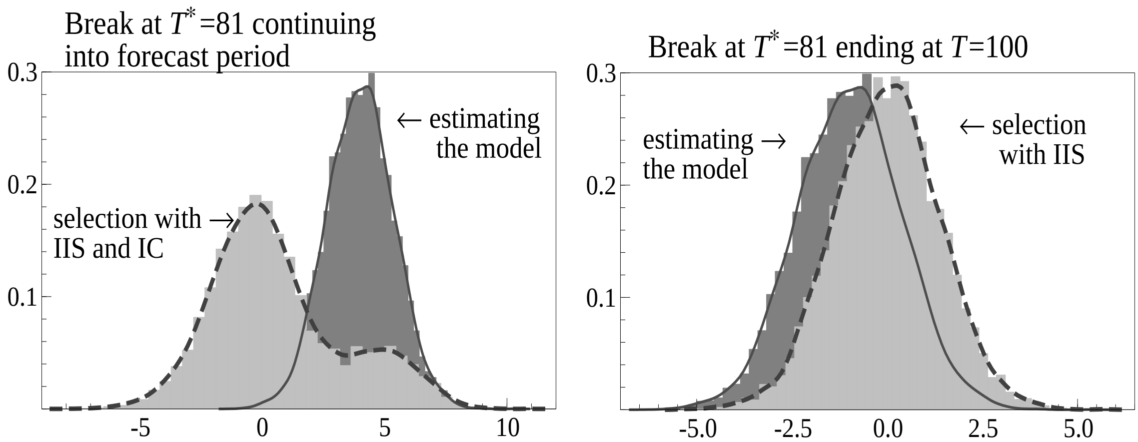

Figure 3: 1-step ahead forecast error densities.

Figure 3 shows that the intercept-corrected forecast-error distribution is centered on zero. There is some bimodality due to the last in-sample indicator not being retained for every replication, so not all forecasts are adjusted.[5] Even when the shift occurs at the end of the sample, IIS still has power to detect it, so intercept corrections can be applied. When the break is not extended into the forecast period, estimating the mis-specified model leads to biased forecasts, but IIS corrects this bias.

[5]When the shift occurs at $T^{*} = 99$, an indicator is retained 88% of the time at $T = 100$, so 12% of forecasts are unadjusted.

5.1 Policy implications

From (1), a policy that changed $\mathbf{z}_t$ to $\mathbf{z}_t + \nabla \mathbf{z}_t$ would shift y_t by $\boldsymbol{\beta}_1' \nabla \mathbf{z}_t$, whereas (4) entails a predicted shift of $\gamma_1' \nabla \mathbf{z}_t$. An intercept shift of $-\boldsymbol{\beta}_2' \boldsymbol{\Psi} \nabla \mathbf{z}_t$ must occur in the model, so policy would not have the expected outcome, an indirect failure of super exogeneity (see Engle and Hendry, 1993). When previous policy changes have occurred, the IIS-test in Hendry and Santos (2010) ascertains their timing in $\{\mathbf{z}_t\}$, and checks whether they coincided with shifts in the model $y_t | \mathbf{z}_t$. If so, that would reject super exogeneity. Locating indicators in $\{\mathbf{z}_t\}$ directly, rather than their effect mediated through $\boldsymbol{\Psi}$, is more powerful and finds the source of shifts, warning of a super exogeneity failure before a policy is changed.

6 Conclusion

We consider model selection in under-specified settings when variables have location shifts and relevant variables are omitted. Mean shifts in the omitted variables alone induce intercept shifts in the local data generation process (LDGP), but do not contaminate the slope parameters, even when correlated with included variables. Mean shifts in included variables would be beneficial under correct specification, but with mis-specification lead to shifting intercepts in the LDGP, and non-constant estimated slopes. In both cases, the equation and estimated standard errors become non-constant.

Impulse-indicator saturation (IIS) tackles multiple shifts at unknown sample points, and 'robustifies' estimated models against non-constant intercepts even when induced by omitted variables. If IIS is not needed, the costs of adding to the set of candidate regressors an impulse indicator for every observation are relatively small; and if IIS is needed, the most pernicious effects of induced location shifts on non-constant intercepts and standard errors are corrected. Simulations in Castle *et al.* (2012) show that similar findings hold for breaks in correctly-specified dynamic models, including some integrated data processes. As their location can be consistently estimated, the generalization of IIS to step-indicator saturation could allow a more parsimonious representation of location shifts (see Ericsson, 2012, for an illustration). Based on the results presented in this paper, the model-selection IIS-based parameters in static equations would appear to be more constant, and inferences more reliable such that additional incorrect omissions of relevant variables are less likely, and forecasts based on such models using intercept corrections will face one less difficulty. As a sufficiently general initial model is less likely to omit substantively relevant variables, there are, we believe, strong arguments for such an approach.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.

Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.

Bontemps, C., and Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, **70**, 721–750.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1097.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., and Hendry, D. F. (2011). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T. (eds.), *System Identification, Environmental Modelling and Control*, pp. 229–250. New York: Springer.

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. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). *Object-Oriented Matrix Programming using Ox* 7th edn. London: Timberlake Consultants Press.

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

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

Ericsson, N. R. (2012). Detecting parameter nonconstancy and changes in regime. Working paper, Federal Reserve Board of Governors, Washington, D.C.

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. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics*, **9**, 295–314.

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

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., 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 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.

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.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, **27**, 575–595.

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

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