

Regression Models with Data-based Indicator Variables

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

November 3, 2003

Abstract

OLS estimation of an impulse-indicator coefficient is inconsistent, but its variance can be consistently estimated. Although the ratio of the inconsistent estimator to its standard error has a t -distribution, that test is inconsistent: one solution is to form an index of indicators. We provide Monte Carlo evidence that including a plethora of indicators need not distort model selection, permitting the use of many dummies in a general-to-specific framework. Although White's (1980) heteroskedasticity test is incorrectly sized in that context, we suggest an improvement. Finally, a possible modification to impulse 'intercept corrections' is considered.

JEL classifications: C51, C22.

Contents

1	Introduction	2
2	OLS estimation in an unrestricted dummies model	2
	2.1 Properties of \mathbf{D}	3
	2.2 Properties of $\hat{\gamma}$	3
	2.3 Properties of $V[\hat{\gamma}]$	4
	2.4 Properties of inference on γ	4
	2.5 t -test inconsistency	5
	2.6 Fisher information	5
3	The properties of linear combinations of indicators	6
	3.1 OLS estimation of an index parameter	6
	3.2 Mis-specified weights	8
4	Model selection in the UDM	9
	4.1 Null distribution	9
	4.2 Mixture of distributions	10
5	The behaviour of White's heteroskedasticity test	11
	5.1 Large sample sizes	13
	5.2 An index representation	14
	5.3 Power of normality test	14
6	Last sample-observation indicators	14
7	Conclusion	16
	References	17

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

1 Introduction

Indicator variables, also called impulse dummies, and combinations thereof, regularly occur in estimated time-series relationships, eliminating residuals that would otherwise be outliers (usually in excess of 2 standard errors in absolute value). Any location shifts or other changes in the coefficients of deterministic variables can induce such outliers, (usually called innovation outliers), as can data measurement or recording errors (additive outliers). These are equivalent in static regression models with strongly exogenous variables, but have different consequences in dynamic models.

Despite the prevalence of indicators in models, several aspects of their use do not seem to have been fully investigated, and here we address some of the more pertinent of these. First, although OLS estimation is unbiased for the coefficient of an indicator, it is inconsistent—yet its variance can be consistently estimated. Nonetheless, the ratio of the inconsistent estimator to its consistently estimated standard error has the usual t -distribution when the errors are normally distributed, but the power of the t -test does not rise as the sample size increases. The inconsistency results from the lack of divergence in the Fisher information, as there is only a single observation on the indicator, so we consider overcoming this by forming linear combinations of indicators. We establish necessary conditions for consistent estimation of the parameter of an index of indicators, showing that consistency can occur even with mis-specified weights. We also report Monte Carlo evidence that inclusion of a plethora of indicators does not distort model selection, permitting the use of many dummies in a general-to-specific (*Gets*) framework. However, when large numbers of indicators are used, inferences from mis-specification tests might be distorted, and we note the low power of normality tests in this setting. The heteroskedasticity test proposed by White (1980) in fact has the wrong size in an unrestricted dummies model (also see Messer and White, 1984), so we suggest an improvement to the test in that context. The use of an index could also alleviate such a distributional problem. Lastly, an indicator for the final observation in a sample is considered in relation to intercept corrections.

The aim of the paper is to investigate the impact of selecting data-based indicators on other parameter estimates and on mis-specification tests. To do so, we first consider the properties of estimators of coefficients of indicators and tests thereon in section 2, when no selection is involved. Then the impact of combining indicators into an index based on sample evidence is considered in section 3. In section 4, the effects of adding many dummies are discussed under both null (section 4.1, when no dummies matter) and alternative (when the error comes from a mixture of distributions, one of which generates outliers: section 4.2). Potential distortions to the distribution of the heteroskedasticity test are then examined in section 5, and possible corrections considered. Finally, in section 6, we note the role of an indicator added for a discrepant final sample observation.¹ Section 7 concludes.

2 OLS estimation in an unrestricted dummies model

We consider models of the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\gamma} + \mathbf{v} \quad (1)$$

where $\mathbf{v} \sim \text{IN}_T[\mathbf{0}, \sigma_v^2 \mathbf{I}]$ is a $T \times 1$ random vector when T is the sample size, and \mathbf{X} ($T \times k_1$) and \mathbf{D} ($T \times k_2$) are matrices of strongly exogenous regressors with $k = k_1 + k_2$, so k is the total number of regressors including a constant. The columns of \mathbf{D} are zero-one observation-specific indicators, denoted $1_{\{t=t_b\}}$ when unity at observation t_b and zero otherwise, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $k_1 \times 1$ and $k_2 \times 1$ vectors of

¹All of the computations were undertaken using *PcGets* and *Ox* (see Hendry and Krolzig, 2001, and Doornik, 2001).

constant parameters. We assume that:

$$\lim \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right) = \mathbf{Q}, \quad (2)$$

where \mathbf{Q} is a finite positive-definite matrix. Equation (1) is referred to as the unrestricted dummies model (UDM).

2.1 Properties of \mathbf{D}

The matrix \mathbf{D} acts to select elements in cross products, so that when (e.g.) $k_2 = 1$ there is a single indicator $d_t = 1_{\{t=t_b\}}$, then:

$$\mathbf{D}'\mathbf{D} = \mathbf{d}'\mathbf{d} = 1, \quad \mathbf{X}'\mathbf{d} = \mathbf{x}_{t_b} \quad \text{and} \quad \mathbf{d}'\mathbf{y} = y_{t_b},$$

where \mathbf{x}_{t_b} is a $k_1 \times 1$ vector. Let $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ so $\mathbf{M}\mathbf{X} = \mathbf{0}_T$ then:

$$\mathbf{d}'\mathbf{M}\mathbf{d} = 1 - \mathbf{x}'_{t_b}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{t_b}.$$

Thus:

$$\lim \mathbf{d}'\mathbf{M}\mathbf{d} = \lim \left(1 - \frac{\mathbf{d}'\mathbf{X}}{T} \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \mathbf{X}'\mathbf{d} \right) = (1 - \mathbf{0}'_{1 \times k_1} \mathbf{Q}^{-1} \mathbf{x}_{t_b}) = 1, \quad (3)$$

so:

$$\lim T^{-1} \mathbf{d}'\mathbf{M}\mathbf{d} = 0, \quad (4)$$

in contrast to (2).

2.2 Properties of $\hat{\gamma}$

For a single indicator, from the Frisch and Waugh (1933) theorem, the OLS estimator of the parameter γ is given by:

$$\hat{\gamma} = (\mathbf{d}'\mathbf{M}\mathbf{x})^{-1} \mathbf{x}'\mathbf{M}\mathbf{y}. \quad (5)$$

Substituting (1) into (5) and simplifying:

$$\hat{\gamma} - \gamma = (\mathbf{d}'\mathbf{M}\mathbf{x})^{-1} \mathbf{d}'\mathbf{v} - (\mathbf{d}'\mathbf{M}\mathbf{x})^{-1} \mathbf{d}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}. \quad (6)$$

That the OLS estimator of γ is unbiased follows immediately from applying expectations to (6) as both $\mathbf{E}[\mathbf{d}'\mathbf{v}] = 0$ and $\mathbf{E}[\mathbf{X}'\mathbf{v}] = \mathbf{0}$.

However, in the UDM, the OLS estimator of γ is inconsistent. As $T \rightarrow \infty$, the last term in (6) vanishes since the probability limit of a non-stochastic sequence is equal to its non-probabilistic limit, and:

$$\lim \left[\frac{\mathbf{d}'\mathbf{X}}{T} \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \right] = \mathbf{0}'_{1 \times k_1} \mathbf{Q}^{-1} = \mathbf{0}'_{1 \times k_1}$$

with:

$$\text{plim}_{T \rightarrow \infty} \frac{\mathbf{X}'\mathbf{v}}{T} = \mathbf{0}_{k_1 \times 1}, \quad (7)$$

using Slutsky's theorem (see Sargan, 1988). Hence using (3):

$$\text{plim}_{T \rightarrow \infty} \hat{\gamma} = \gamma + \text{plim}_{T \rightarrow \infty} \mathbf{d}'\mathbf{v} = \gamma + \text{plim}_{T \rightarrow \infty} v_{t_b}. \quad (8)$$

Since $\text{plim } v_{t_b} = v_{t_b} \neq 0$, as v_{t_b} has a non-degenerate limiting distribution:

$$\text{plim}_{T \rightarrow \infty} \hat{\gamma} \neq \gamma. \quad (9)$$

The non-degeneracy of $V[\hat{\gamma}]$ shown in (16) below confirms (9).

2.3 Properties of $V[\hat{\gamma}]$

In spite of the inconsistency of $\hat{\gamma}$, the estimator of its variance is unbiased and consistent. Consider for simplicity the case $k_1 = 0$ and $k_2 = 1$, that is the DGP:

$$y_t = \gamma d_t + v_t \quad (10)$$

where $d_t = 1_{\{t=t_b\}}$. The OLS estimator of γ is simply:

$$\hat{\gamma} = \frac{\sum_{t=1}^T d_t y_t}{\sum_{t=1}^T d_t^2} = \gamma + \frac{\sum_{t=1}^T d_t v_t}{\sum_{t=1}^T d_t^2} = \gamma + v_{t_b} = y_{t_b} \quad (11)$$

implying $\hat{v}_{t_b} = 0$, and that $V[\hat{\gamma}] = E[v_{t_b}^2] = \sigma_v^2$ so:

$$V[\hat{\gamma}] = \hat{\sigma}_v^2 = \frac{\sum_{t=1}^T (y_t - \hat{\gamma} d_t)^2}{T-1} = \frac{\sum_{t=1}^{t_b-1} v_t^2 + \sum_{t=t_b+1}^T v_t^2}{T-1} + \frac{(y_{t_b} - \hat{\gamma})^2}{T-1}. \quad (12)$$

As $\hat{\gamma} = y_{t_b}$, from (12):

$$E[\hat{\sigma}_v^2] = E\left[\frac{\sum_{t=1}^{t_b-1} v_t^2 + \sum_{t=t_b+1}^T v_t^2}{T-1}\right] = \sigma_v^2, \quad (13)$$

confirming that the estimator of the residual variance is unbiased. Further, since $(v_t/\sigma_v)^2 \sim \chi_{(1)}^2$, and as the $\{v_t\}$ are independent:

$$\sum_{t=1}^{t_b-1} \left(\frac{v_t}{\sigma_v}\right)^2 \sim \chi_{(t_b-1)}^2,$$

hence:

$$\left(\sigma_v^2 \sum_{t=1}^{t_b-1} \left(\frac{v_t}{\sigma_v}\right)^2 + \sigma_v^2 \sum_{t=t_b+1}^T \left(\frac{v_t}{\sigma_v}\right)^2\right) \sim \chi_{(T-1)}^2. \quad (14)$$

Thus, $V[\hat{\sigma}_v^2]$ converges to zero as the sample size increases

$$V[\hat{\sigma}_v^2] = V\left[\frac{\chi_{(T-1)}^2}{T-1}\right] = \frac{2}{T-1} \rightarrow 0. \quad (15)$$

From (13) and (15), the estimator of the variance of the OLS estimator of the indicator variable parameter is mean-square convergent to the true variance, so:

$$\text{plim}_{T \rightarrow \infty} V[\hat{\gamma}] = V[\hat{\gamma}] = \sigma_v^2, \quad (16)$$

also confirming the inconsistency of $\hat{\gamma}$. Such a result contrasts with what would occur for $\hat{\beta}$ (say) when $k_1 = 1$ and $k_2 = 0$ as $\text{plim}_{T \rightarrow \infty} V[\hat{\beta}] = 0$.

2.4 Properties of inference on γ

Surprisingly, despite the inconsistency of $\hat{\gamma}$ in the UDM, under the assumption that $v_{t_b} \sim N[0, \sigma_v^2]$:

$$\frac{\hat{\gamma} - \gamma}{\sqrt{V[\hat{\gamma}]}} \sim t_{(T-1)}. \quad (17)$$

To show this, for simplicity again consider the model in (10), where:

$$\hat{\gamma} \sim \mathbf{N}[\gamma, \sigma_v^2], \quad (18)$$

with $\hat{\sigma}_v^2 \sim \chi_{(T-1)}^2$, and $(\hat{\gamma} - \gamma)$ and $\hat{\sigma}_v^2$ are independently distributed, so the results in (e.g.) Hendry (1995, section A2.9.4), imply:

$$\frac{\hat{\gamma} - \gamma}{\sqrt{\widehat{\mathbf{V}}[\hat{\gamma}]}} = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_v} \sim \mathbf{t}_{(T-1)}. \quad (19)$$

Thus, tests on γ can be conducted when v_{t_b} is normal with mean zero and variance σ_v^2 .

2.5 t-test inconsistency

Let W be a test statistic. The test is consistent if, for any fixed significance level α , and for any fixed alternative H_1 :

$$\Pr(|W| > c_\alpha; H_1) \rightarrow 1 \quad (20)$$

as $T \rightarrow \infty$ (see Cox and Hinkley, 1974), where c_α is the critical value.

Consider the UDM, assuming for simplicity that $k_2 = 1$ and $k_1 = 0$. Then, when testing the null hypothesis:

$$H_0 : \gamma = \gamma^* \quad (21)$$

versus the alternative:

$$H_1 : \gamma \neq \gamma^* \quad (22)$$

the power of the t-test is given by:

$$\Pr\left(\left|\frac{\hat{\gamma} - \gamma^*}{\hat{\sigma}_\gamma}\right| > c_\alpha\right) \quad (23)$$

Since $\hat{\sigma}_\gamma \rightarrow \sigma_\gamma \neq 0$, there are significance levels and values of γ^* for which (23) does not converge to unity. Therefore the test on an impulse is not consistent.

2.6 Fisher information

The Fisher information for an observation-specific indicator variable is asymptotically negligible. Assume for simplicity that $k_2 = 1$. Given that $v_t \sim \text{IN}[0, \sigma_v^2] \forall t$, the log-likelihood for one observation is given by:

$$\ln \mathcal{L}(y_t, \mathbf{x}_t, d_t; \boldsymbol{\beta}, \gamma, \sigma_v^2) = -\ln\left(\frac{1}{\sigma_v \sqrt{2\pi}}\right) - \frac{1}{2\sigma_v^2} \left(y_t - \sum_{i=1}^{k_1} \beta_i x_{i,t} - \gamma d_t\right)^2. \quad (24)$$

If that observation is $t = t_b$, then:

$$\frac{\partial \ln \mathcal{L}}{\partial \beta_2} = \frac{1}{\sigma_v^2} \left(y_{t_b} - \sum_{i=1}^{k_1} \beta_i x_{i,t_b} - \gamma\right). \quad (25)$$

Hence, the information equality implies:

$$\mathbb{E} \left[-\frac{\partial^2 \ln \mathcal{L}(y_t, \mathbf{x}_t, d_t; \boldsymbol{\beta}, \gamma, \sigma_v^2)}{\partial \gamma^2} \Big|_{\gamma=\gamma^0} \right] = \frac{1}{\sigma_v^2} \quad (26)$$

where γ^0 denotes the true parameter value. The Fisher information about the parameter γ^0 is zero for all other observations. Since the v_t are independent, the sample Fisher information equals the sum of

the information for each random variable, so the sample information about γ^0 is still σ_v^{-2} . Hence, as $T \rightarrow \infty$, the sample information about the indicator variable parameter is negligible:

$$\lim \frac{1}{T\sigma_v^2} = 0. \quad (27)$$

As a corollary, the OLS estimator of the observation-specific dummy variable parameter estimator is efficient: the Cramér–Rao lower bound for $V[\hat{\gamma}]$ in the model defined by (10) is σ_v^2 , and we established above that $V[\hat{\gamma}] = \sigma_v^2$.

Given these properties of estimation and inference about indicator coefficients, we consider selecting them from data evidence, and first discuss forming a data-based index of indicators.

3 The properties of linear combinations of indicators

The idea of forming an index to substitute for the original dummies was used by Hendry (1999) in analyzing US Food Expenditure, 1931–1989, and Hendry (2001) in an empirical study of UK inflation from 1865 to 1991. Replacing the indicators by a linear combination was, in the context of these papers, motivated by the excessive number of dummies initially needed in each model. In the first paper, the indicators were almost contiguous over 1931–1945 so were an application of the forecast (actually, backcast) approach in Salkever (1976), but were then simplified to two indexes. Earlier research on the second topic had also revealed an abundance of outliers, perhaps unsurprising in a sample that comprises two world wars and two oil crises, the breakdown of the Bretton Woods system, and a many legislative and technological changes. Twenty-two indicators remained in the final model using data-based restrictions which were acceptable at a 5% significance level, inducing 22 zero residuals. Three groups of dummies were formed, roughly matching ‘very big (12%)’, ‘large (6%)’ and ‘medium (4%)’ outliers, then each group was assigned a weight, where 4% mapped to unity, to form an index. After this reduction to a single index, the model remained congruent.

Neither the theoretical properties nor the small-sample behaviour of test statistics and estimators are known when both large residuals and the ensuing zero clusters have been eliminated. Section 3.1 establishes overly strong sufficient conditions for consistent estimation of the index parameter, and section 3.2 allows for mis-specified weights. Both sections take the ‘significant’ indicators as known, but allow for omitting some of the shifts that actually occurred.

3.1 OLS estimation of an index parameter

We postulate the following simplified DGP:

$$y_t = \sum_{i=1}^{K_T} \phi_i d_{i,t} + v_t \quad (28)$$

where $v_t \sim \text{IN}[0, \sigma_v^2]$. This is an unrestricted dummies DGP, where for simplicity, $k_1 = 0$ and $k_2 = K_T$. The $d_{i,t}$ are observation-specific indicators, zero except for $1_{\{t=t_i\}}$ for some set of K_T time occurrences in a set $\mathcal{S}_{K_T} = \{t_1 \dots t_{K_T}\}$. Hence, the DGP for y_t is a white-noise stochastic process perturbed at some points in time by transient location shifts. We assume that as $T \rightarrow \infty$, $K_T \rightarrow \infty$, but $K_T < T$ such that $1 > K_T/T \rightarrow c_K > 0$, where c_K denotes the ‘average arrival rate of shifts’. Based on realistic historical foundations, outliers are assumed to keep occurring in the future. We do not assume a specific arrival process for such shocks, which would generate a meta-stationary process (e.g., a Poisson process): rather shocks are assumed to keep on happening, but not every period.

In the absence of data revisions, either a dummy has a significant effect immediately or never, since information on the individual indicators does not accrue. Thus, even with these assumptions, dummies will usually only be included in an econometric model of $\{y_t\}$ for ‘significant’ shocks. Several criteria could be used to assess that need, such as $|\phi_i + v_{t_i}|/\sigma_v > 2.0$ (say).² Let k_T be the number of significant shocks in a set $\mathcal{S}_{k_T} \subseteq \mathcal{S}_{K_T}$, and hence the number of indicators in the econometric model. Then we also assume that as $T \rightarrow \infty$, then $k_T \rightarrow \infty$ with $k_T < K_T$ and $k_T/T_T \rightarrow c_k > 0$. This assumption ensures that as the sample size increases, significant shocks also keep occurring, so would be included in the UDM according to the given criterion.

Re-write (28) as:

$$y_t = \sum_{j \in \mathcal{S}_{k_T}} \phi_j d_{j,t} + \omega_t = \boldsymbol{\phi}' \mathbf{d}_t + \omega_t \quad (29)$$

where there are $K_T - k_T$ omitted indicators, with:

$$\omega_t = \sum_{j \in \mathcal{S}_{K_T} - \mathcal{S}_{k_T}} \phi_j d_{j,t} + v_t,$$

so:

$$\mathbb{E}[\omega_t] = \begin{cases} \phi_j & \text{for } j \in \mathcal{S}_{K_T} - \mathcal{S}_{k_T} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbb{E}[\omega_t^2] = \begin{cases} \phi_j^2 + \sigma_v^2 & \text{for } j \in \mathcal{S}_{K_T} - \mathcal{S}_{k_T} \\ \sigma_v^2 & \text{otherwise} \end{cases}.$$

We take the omitted $\{\phi_j\}$ to have an average of zero, which would arise if (e.g.) the smaller outliers were randomly drawn from a symmetric distribution. Thus, $\omega_t \sim \text{ID}[0, \sigma_{\omega_t}^2]$, and is assumed independent of the retained $\{d_{i,t}\}$. In effect, smaller shocks are subsumed in the equation’s error process, making $\{\omega_t\}$ heteroskedastic with an average variance greater than σ_v^2 . Let $\sigma_{\omega_t}^2 < B \in \mathbb{R}^+ \forall t$, so the variance of the combined error never exceeds an upper bound B , even asymptotically, noting that large ϕ_j are removed by indicator dummies.

Given the unrestricted dummies DGP as in (29), consider the econometric model:

$$y_t = \delta I_t + u_t \quad (30)$$

with $\delta \neq 0$ where:

$$I_t = \sum_{j \in \mathcal{S}_{k_T}} w_j d_{j,t} = \mathbf{w}' \mathbf{d}_t. \quad (31)$$

We first assume that the weights $\{w_j\}$ are correctly specified for the ‘significant outliers’, so that:

$$w_j = \frac{\phi_j}{\delta}, \quad (32)$$

and hence $\boldsymbol{\phi} = \delta \mathbf{w}$ (section 3.2 considers a class of mis-specified weights). Then, $\widehat{\delta}$ is a weakly consistent estimator of δ . The proof requires the error term to be the same in (29) and (30), so the two representations are isomorphic provided (32) holds. Then as $\boldsymbol{\phi}' \mathbf{d}_t = \delta \mathbf{w}' \mathbf{d}_t = \delta I_t$:

$$\widehat{\delta} = \frac{\sum_{t=1}^T I_t y_t}{\sum_{t=1}^T I_t^2} = \frac{\sum_{t=1}^T I_t (\boldsymbol{\phi}' \mathbf{d}_t + \omega_t)}{\sum_{t=1}^T I_t^2} = \delta + \frac{\sum_{t=1}^T I_t \omega_t}{\sum_{t=1}^T I_t^2}. \quad (33)$$

²As σ_v will be unknown, there is the potential problem of detecting outliers using $\widehat{\sigma}_v$ which initially reflects the omitted outliers; this is one reason we allow for ‘smaller’ transient shifts to be omitted.

Taking expectations of both sides of (33) shows that $\widehat{\delta}$ is unbiased when $E[\sum_{t=1}^T I_t \omega_t] = 0$:

$$E[\widehat{\delta}] = \delta + \frac{E\left[\sum_{t=1}^T I_t \omega_t\right]}{\sum_{t=1}^T I_t^2} = \delta. \quad (34)$$

Further, as:

$$E\left[\left(\sum_{t=1}^T I_t \omega_t\right)^2\right] = E\left[\sum_{t=1}^T \sum_{s=1}^T I_t I_s \omega_t \omega_s\right] = \sum_{t=1}^T I_t^2 \sigma_{\omega_t}^2,$$

then:

$$V[\widehat{\delta}] = E\left[\left(\widehat{\delta} - \delta\right)^2\right] = E\left[\left(\frac{\sum_{t=1}^T I_t \omega_t}{\sum_{t=1}^T I_t^2}\right)^2\right] = \frac{\sum_{t=1}^T I_t^2 \sigma_{\omega_t}^2}{\left(\sum_{t=1}^T I_t^2\right)^2}. \quad (35)$$

Next (see e.g., White, 1984):

$$\frac{1}{k_T} \sum_{t=1}^T I_t^2 = \frac{1}{k_T} \sum_{t=1}^T \sum_{j \in \mathcal{S}_{k_T}} w_j^2 d_{j,t} = \frac{1}{k_T} \sum_{j \in \mathcal{S}_{k_T}} w_j^2 \sum_{t=1}^T d_{j,t} = \frac{1}{k_T} \sum_{j \in \mathcal{S}_{k_T}} w_j^2 \rightarrow \overline{w^2} > 0.$$

Further:

$$\frac{1}{k_T} \sum_{t=1}^T I_t^2 \sigma_{\omega_t}^2 < B \frac{1}{k_T} \sum_{t=1}^T I_t^2 \rightarrow B \overline{w^2} > 0,$$

so that:

$$\lim_{T \rightarrow \infty} V[\widehat{\delta}] = \lim_{T \rightarrow \infty} \frac{1}{k_T} \frac{\frac{1}{k_T} \sum_{t=1}^T I_t^2 \sigma_{\omega_t}^2}{\left(\frac{1}{k_T} \sum_{t=1}^T I_t^2\right)^2} = \lim_{T \rightarrow \infty} \frac{1}{k_T} \frac{B \overline{w^2}}{\left(\overline{w^2}\right)^2} = 0. \quad (36)$$

Sufficient conditions for mean-square convergence of $\widehat{\delta}$ to δ are, therefore, verified, so that:

$$\text{plim}_{T \rightarrow \infty} \widehat{\delta} = \delta. \quad (37)$$

Indeed:

$$\sqrt{T}(\widehat{\delta} - \delta) = \sqrt{k_T} \frac{\frac{\sqrt{k_T}}{\sqrt{T}} \sum_{t=1}^T I_t \omega_t}{\frac{k_T}{T} \sum_{t=1}^T I_t^2} \rightarrow \frac{1}{\sqrt{c_k}} \left(\sqrt{k_T} \sum_{i \in \mathcal{S}_{k_T}} \omega_{t_i} \right) \underset{\sim}{\sim} \text{N} \left[0, \frac{\sigma_{\omega}^2}{c_k} \right]. \quad (38)$$

3.2 Mis-specified weights

Unfortunately, this result is of little practical value, as the index model will not in general be isomorphic to the UDM, since δ is unknown when defining the weights. In any empirical application, the index weights are bound to be mis-specified, so we establish sufficient conditions for consistent estimation of δ even though the index weights are mis-specified. The analysis is close to that mapping (28) to (29).

Let:

$$\phi_i = \delta w_i + \mu_i \quad \text{for } i \in \mathcal{S}_{k_T},$$

so that:

$$\sum_{i \in \mathcal{S}_{k_T}} \phi_i d_{i,t} = \delta \sum_{i \in \mathcal{S}_{k_T}} w_i d_{i,t} + \varepsilon_t = \delta I_t + \varepsilon_t \quad (39)$$

where:

$$\varepsilon_t = \sum_{i \in \mathcal{S}_{k_T}} \mu_i d_{i,t},$$

so:

$$y_t = \delta I_t + v_t + \varepsilon_t = \delta I_t + \eta_t \quad (40)$$

where:

$$\mathbb{E}[\eta_t] = \begin{cases} \mu_j & \text{for } j \in \mathcal{S}_{k_T} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbb{E}[\eta_t^2] = \begin{cases} \mu_j^2 + \sigma_v^2 & \text{for } j \in \mathcal{S}_{k_T} \\ \sigma_v^2 & \text{otherwise} \end{cases}.$$

For the mis-specification not to affect consistency, the omitted components must continue to act like random errors. As before, we assume $\sigma_{\eta,t}^2 < L \in \mathbb{R}^+ \forall t$ and $\mathbb{C}[\eta_t, \eta_{t-s}] = 0, \forall s \neq t$, and require that the $\{\mu_j\}$ average to zero. Also, $\mathbb{E}[I_t v_t] = 0$ so the key is that $\mathbb{E}[I_t \varepsilon_t] = 0$ and we have:

$$\mathbb{E}[I_t \varepsilon_t] = \mathbb{E} \left[\sum_{i \in \mathcal{S}_{k_T}} w_i d_{i,t} \sum_{j \in \mathcal{S}_{k_T}} \mu_j d_{j,t} \right] = \sum_{i \in \mathcal{S}_{k_T}} w_i \mu_i.$$

Thus, mistakes in assigning weights must on average be ‘uncorrelated’ with the weight assigned. If so, then (40) satisfies the assumptions of section 3.1, and hence the OLS estimator of δ is weakly consistent.

4 Model selection in the UDM

The Monte Carlo study developed in this section addresses the issue of whether or not adding dummies that do not actually matter will distort model selection. We considered a rather extreme scenario where the number of dummies represents 62.5% of the sample size. Here, $T = 40$, and apart from the 25 zero-one observation-specific dummy variables, only one other regressor was considered: for each replication these were drawings from the uniform distribution with support in the unit interval. $M = 10000$ replications were conducted. We first note the null distribution for ‘near-saturated’ regressions, then consider the alternative when the errors come from a mixture of distributions, one of which generates outliers.

4.1 Null distribution

The DGP is given by the following UDM:

$$y_t = \beta x_t + v_t \quad (41)$$

where $v_t \sim \text{IN}[0, \sigma_v^2]$. Thus, no dummies were included in the DGP (41) to generate outliers in the data, although $\Pr(v_t^2/\sigma_v^2 > 4) \simeq 0.05$.

As a baseline, first consider adding one unnecessary impulse dummy to (41). There is no bias, but an efficiency loss of $\mathcal{O}(T^{-1})$. Next, consider a model where $x_t = 1$ and $T - 2$ impulse dummies $d_{j,t}$, $j = 1, \dots, T - 2$ are added, leaving just one degree of freedom:

$$y_t = \beta + \sum_{j=1}^{T-2} \gamma_j d_{j,t} + v_t. \quad (42)$$

Although it is difficult to saturate the problem any more than (42) without a perfect fit, nevertheless the dummies merely reduce the sample size to 2, with $\hat{\beta} = \frac{1}{2}(y_{T-1} + y_T)$ so that $\hat{\gamma}_j = v_j + \beta - \hat{\beta}$. Providing the last 2 observations are representative (so v_{T-1} and v_T are neither outliers nor very small), then $0.05T$ dummies will be significant by chance on average, but with considerable variability. Moreover, the number of significant values will be altered only marginally if a selection routine eliminates the

insignificant $\hat{\gamma}_j$. While simplistic, this reasoning seems to characterize why there need not be a major problem under the null from adding many dummies.

In the Monte Carlo, $\sigma_v^2 = 1$ and $\beta = 1$ in 41) without loss of generality. However, the econometric model contains dummies that are ‘randomly’ added, in the sense that there is no *a priori* reason to think they correspond to outliers. The observations for which the dummies are introduced were selected by simulating a Bernoulli distribution with parameter 0.6, and were the same across the 10000 replications.

Individual significance tests on the indicators should have a t-distribution with 14 degrees of freedom, so the nominal critical values used are ± 1.76 . The resulting empirical critical values for the indicators were close to the theoretical ones, and average rejection frequencies were virtually identical to the postulated significance levels. Furthermore, the inclusion of 25 dummy variables did not affect the bias of the estimator of the coefficient β on x , nor its significance. The number of irrelevant dummies retained at each replication followed a binomial distribution with parameters $p = 0.01$ (when a rule close to $2.5\hat{\sigma}$ was used) and $N = 25$.³ Hence, on average, only 0.25 irrelevant dummies were retained in each regression despite nearly ‘saturating’ the model with indicators. In practice, therefore, almost no irrelevant dummy variable will be retained, revealing low costs of inference and search in this context (see Hendry, 2000).

4.2 Mixture of distributions

An alternative DGP is one where the errors come from a mixture of distributions, one of which generates relatively rare outliers relative to the other. We worked with a version of the unrestricted dummies model where \mathbf{X} contained a constant and a uniformly distributed regressor with support in the unit interval: the parameter values were set to $\beta_0 = 0.25$ and $\beta_1 = 0.45$ respectively. The sample size was $T = 200$, and $M = 1000$ replications were conducted. Three key features are worth noticing in the design of this Monte Carlo experiment:

1. The vector \mathbf{v} was generated from a mixture of a standard normal distribution and a member of Student’s t family of distributions.⁴ The probability that a drawing from the standard normal would be generated was chosen to be equal to:

$$\Pr(Z = 0) \tag{43}$$

where

$$Z \sim \text{Poisson}(\lambda = 0.2) \tag{44}$$

2. The alternative distribution in the mixture varied across experiments. We conducted simulations for drawings from $t_{(4)}$, $t_{(3)}$ and $t_{(2)}$. These choices reflect that the lower the degrees of freedom of the t distribution, the heavier its tails, and hence the more likely it is that many outliers will be generated.

3. Outliers were defined as observations with associated OLS residuals greater than $2.5\hat{\sigma}$ in absolute value.

After generating \mathbf{v} as the mixture just described, the model without dummies was estimated. The times of residuals greater than $2.5\hat{\sigma}$ in absolute value were used to create the matrix of dummy variables \mathbf{D} of dimension $T \times n$ (for n outliers). Finally, that UDM was estimated by OLS at every replication. The mean estimates and t-values, and those from the regression without dummies, are reported in tables 1 and 2, together with the rejection frequencies of the t-tests using one-sided 5% critical values of 1.645.

In both cases, the parameters are unbiasedly estimated. However, not including dummies to account for the outliers induces a loss of power for rejecting the nulls for β_0 and β_1 . The parameter values

³The complete results of this Monte Carlo experiment are available in Santos (2003).

⁴This is a common method of simulating aberrant observations for Monte Carlo studies (see Abraham and Chuang, 1989) which does not entail that this is the true economic mechanism generating the v_t .

	$t_{(4)}$	$t_{(3)}$	$t_{(2)}$
β_0	0.25	0.24	0.24
β_1	0.44	0.46	0.45
t_{β_0}	1.85	1.75	1.68
t_{β_1}	1.84	1.90	1.80
RF_{β_0}	0.57	0.54	0.51
RF_{β_1}	0.56	0.58	0.56

Table 1 Model with Outlier Generated Dummies.

	$t_{(4)}$	$t_{(3)}$	$t_{(2)}$
β_0	0.25	0.25	0.25
β_1	0.45	0.46	0.42
t_{β_0}	1.58	1.44	1.17
t_{β_1}	1.6	1.52	1.21
RF_{β_0}	0.46	0.42	0.32
RF_{β_1}	0.48	0.47	0.35

Table 2 Model without dummies.

in the DGP imply relatively low non-centralities of the t-tests, so for the $t_{(2)}$ simulation, the rejection frequency of $\beta_1 = 0$ when outliers are ignored is 62% of the rejection frequency when the dummies are included. Thus, including the dummies in the model, when the data suggests doing so, seems beneficial, relative to not keeping the dummies when they matter.

5 The behaviour of White's heteroskedasticity test

The inclusion of a large number of dummy variables, relative to the sample size, generates many zero residuals which might give rise to misleading inference when using residual-based mis-specification tests. Following this intuition, we developed a Monte Carlo experiment to assess how close were nominal and empirical critical values in the heteroskedasticity test suggested by White (1980). The DGP is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \sum_{i=1}^{k_2} \gamma_i d_{i,t} + v_t \quad (45)$$

so (45) is a UDM, and we focus on the size of White's test.

$x_{1,t}$ and $x_{2,t}$ are strongly exogenous regressors, so, $k_1 = 3$. $d_{i,t}$ is an observation-specific indicator, and k_2 is allowed to vary across experiments. $x_{1,t}$ was generated as a set of drawings from a uniform distribution in the unit interval, scaled up by a factor of 100, and $x_{2,t}$ was generated as a set of drawings from $N[0, 4]$. The same drawings for $x_{1,t}$ and $x_{2,t}$ were used in every replication. The sample size T was first allowed to vary across experiments using 50, 60, 70, 80 (for local variation; below we also consider more 'asymptotic samples' of 800, 2000 and 10000).

We chose the parameter values reported in table 3 for the simulations, so $k_2 = 9$.

Given the asymptotic nature of White's test statistic, we first conducted an experiment with no dummies in either the DGP or econometric model (so $k_2 = 0$) to assess the closeness of the theoretical and empirical quantiles. $M = 10000$ replications were used throughout. Table 4 reports our results for

β_0	β_1	β_2	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
4	3.3	5	14	22	35	19	24	19	27	12	25

Table 3 DGP parameter values - 9 dummies.

$T \setminus \alpha$	1%	5%	10%
80	13.42	9.29	7.53
70	13.02	9.14	7.58
60	13.03	9.15	7.54
50	13.06	9.19	7.58
$\chi^2_{(4)}$	13.28	9.49	7.78

Table 4 Baseline model.

White's test without cross products. α is the significance level.

We conclude from table 4 that, even for the small sample sizes we are considering, the limiting distribution $\chi^2_{(q)}$, where q is the number of regressors in the auxiliary regression, is a good approximation to the empirical distribution. This provides the baseline for assessing the impact on the test of including dummy variables in the DGP and econometric model. The number of dummies represents approximately 18%, 15%, 12% and 11% of the sample sizes. The outliers in the DGP were introduced at observations 9, 13, 19, 21, 22, 33, 36, 38 and 41, held constant across replications

Table 5 reports the empirical critical values of White's heteroskedasticity test without cross products, for each sample size, and when the DGP and the econometric model have 9 observation-specific dummy variables. The nominal critical values are also reported (the test statistic asymptotically has a $\chi^2_{(13)}$).

Comparing tables 4 and 5 reveals a striking difference, at all relevant quantiles, between the nominal and the empirical critical values.

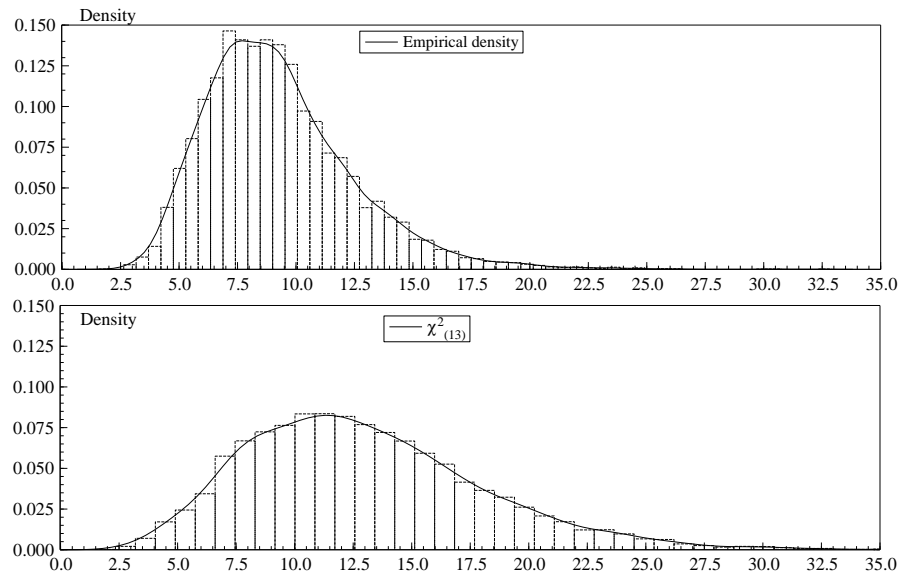


Figure 1 Empirical and Nominal Sampling Distributions.

In figure 1, the first graph reports the empirical distribution of White's test statistic, for $T = 50$ and $k_2 = 9$. The second reports the actual $\chi^2_{(13)}$ density. The empirical critical values are always lower than the nominal critical values, implying that the use of the theoretical distribution would lead to substantial

$T \setminus \alpha$	1%	5%	10%
80	19.07	14.78	12.99
70	18.84	14.95	13.04
60	18.97	14.87	13.15
50	19.16	15.09	13.44
$\chi^2_{(13)}$	27.69	22.36	19.81

Table 5 No cross products, 9 dummies.

$T \setminus \alpha$	1%	5%	10%
800	17.88	14.06	12.27
2000	17.68	13.78	12.17
10000	17.55	14.05	12.33
$\chi^2_{(13)}$	27.69	22.36	19.81

Table 6 Large samples, no cross products, 9 dummies.

under-rejection of the null hypothesis of homoskedasticity.

5.1 Large sample sizes

We repeated the experiment for much larger sample sizes. Table 6 highlights that the problem remains in large samples even for $k_2 = 9$.

In table 7, we report the Monte Carlo results for a modified version of White's test, whereby dummies are included in the DGP and in the baseline econometric model, but are excluded from the auxiliary regressions used to calculate the test:

$$e_t^2 = \mu + \eta_1 x_{1,t} + \eta_2 x_{2,t} + \eta_3 x_{1,t}^2 + \eta_4 x_{2,t}^2 + \xi_t. \quad (46)$$

This version does not appear to be suggested in the literature, although Messer and White (1984) consider estimation of a heteroskedasticity-consistent covariance matrix when there is a singularity due to zero residuals, and suggest dropping those residuals and the dummies from estimation of the covariance matrix. In the test computed here, only the dummy variables are dropped from the auxiliary regression, but the zero residuals are not. Defaults from the previous subsection apply, and table 7 refers to the modified White test without cross products.

As can be seen, finite-sample distortions induce only small differences between nominal and empirical critical values in this modified test. Although such a modified heteroskedasticity test need not be the optimal solution to the problem of proliferating indicators, it shows that heteroskedasticity testing in the UDM need not be problematic.

$T \setminus \alpha$	1%	5%	10%
80	13.81	9.18	7.40
70	12.25	9.01	7.38
60	11.52	7.97	6.67
50	11.11	8.24	7.08
$\chi^2_{(4)}$	13.28	9.49	7.78

Table 7 Modified test, no cross products, 9 dummies.

(47)				(48)			
$T \setminus \alpha$	1%	5%	10%	$T \setminus \alpha$	1%	5%	10%
50	14.82	11.69	10.3	50	11.9	9.6	8.5
60	14.59	11.43	10.1	60	12	9.2	8.3
70	14.45	11.21	9.85	70	11.6	9.4	8.1
80	14.37	11.22	9.8	80	11.7	9.6	8
$\chi^2_{(11)}$	24.7	19.7	17.3	$\chi^2_{(4)}$	13.3	9.5	7.78

Table 8 White's Test No Cross Products.

5.2 An index representation

Next, to assess whether conventional nominal critical values provide a useful guide for White's test when the index model is estimated as in section 3.1, we considered the following UDM as the DGP:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \sum_{i=1}^{k_2} \gamma_i d_{i,t} + v_t. \quad (47)$$

The econometric models under study are therefore the same as (47) and:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \delta I_t + u_t \quad (48)$$

where I_t is defined by (31) and (32), and the numerical weights deliberately allowed some misspecification. $M = 10000$ replications were used (details are provided in Santos, 2003). The unrestricted model includes 9 observation-specific dummy variables, and hence the index has 9 positive weights. Table 8 reports the results for both models when White's test is conducted without cross products.

The asymptotic distribution of White's heteroskedasticity test is close to the empirical when the test is conducted for this index model, in spite of the mis-specified weights. The intuition for such a result, and for the noticeable difference relative to (47), is that the index no longer generates zero residuals, unlike unrestricted dummies.

5.3 Power of normality test

We conducted a Monte Carlo study of the Bowman and Shenton (1975) test for non-normality, to evaluate its power in the UDM setting. Such tests are primarily designed to detect leptokurtosis rather than the mesokurtosis which will result in the present setting. Thus, the alternative was a mixture of a standard Gaussian distribution with zeroes, to mimic the outcome of OLS estimation of the UDM when $\sigma_v^2 = 1$. $M = 10000$ replications were used. For each experiment, the power of the test was estimated by the mean rejection rate of normality at a 5% significance level. For a model with 5 zeroes, the average power of the test was 8% for sample sizes ranging from 20 to 40. This could be contrasted with a mean rejection frequency of 64% when the alternative is a $t_{(2)}$. The problem becomes less relevant as the sample size increases, although at each sample size, the power is smaller than against a $t_{(2)}$. Nevertheless, failure to reject normality in the UDM should be viewed cautiously.

6 Last sample-observation indicators

Intercept corrections of the form of setting a model 'back on track' prior to forecasting are widely used in practice. It is well known that an indicator entered for the final observation in a sample and continued

at unity into the forecast period doubles the forecast error variance (see e.g., Clements and Hendry, 1998). We re-establish that result to consider situations under which it would nevertheless be beneficial to correct for a discrepant final observation.

Again the simplest regression model suffices as an illustration:

$$y_t = \beta x_t + v_t \text{ where } v_t \sim \text{IN} [0, \sigma_v^2],$$

for forecasting y_{T+1} using:

$$\hat{y}_{T+1} = \hat{\beta} x_{T+1},$$

where x_{T+1} is known and:

$$\hat{\beta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} = \beta + \frac{\sum_{t=1}^T x_t v_t}{\sum_{t=1}^T x_t^2}. \quad (49)$$

Under an unchanged process, the forecast error is:

$$\hat{v}_{T+1} = y_{T+1} - \hat{y}_{T+1} = (\beta - \hat{\beta}) x_{T+1} + v_{T+1},$$

with mean-square forecast error (MSFE):

$$\text{E} [\hat{v}_{T+1}^2] = \text{V} [\hat{\beta}] x_{T+1}^2 + \text{E} [v_{T+1}^2].$$

In a stationary environment, $\text{E} [v_{T+1}^2] = \sigma_v^2$ in which case as above, $\text{V}[\hat{\beta}] \simeq \sigma_v^2 / (T\sigma_x^2)$ (where the approximation is of $T^{-1} \sum x_t^2$ by σ_x^2), leading to the well-known result:

$$\text{E} [\hat{v}_{T+1}^2] \simeq \sigma_v^2 \left(1 + \frac{x_{T+1}^2}{T\sigma_x^2} \right). \quad (50)$$

If instead, an indicator is added for the final, and future, observation the model becomes:

$$y_t = \beta x_t + \delta 1_{\{t \geq T\}} + v_t, \quad (51)$$

with:

$$\tilde{\beta} = \frac{\sum_{t=1}^{T-1} x_t y_t}{\sum_{t=1}^{T-1} x_t^2},$$

which is equivalent in this static context to ignoring the final data point, where $\tilde{\delta} = \tilde{y}_T - x_T \tilde{\beta} = \tilde{v}_T$. Continuing the value of the indicator at unity for $T + 1$ leads to:

$$\bar{y}_{T+1} = x_{T+1} \tilde{\beta} + \tilde{\delta},$$

so that:

$$\bar{y}_{T+1} = \tilde{y}_T + \Delta x_{T+1} \tilde{\beta},$$

and hence:

$$\bar{v}_{T+1} = y_{T+1} - \bar{y}_{T+1} = v_{T+1} + x_{T+1} (\beta - \tilde{\beta}) - \tilde{v}_T. \quad (52)$$

Thus, treating the terms in (52) as statistically independent:

$$\text{E} [\bar{v}_{T+1}^2] = x_{T+1}^2 \text{V} [\tilde{\beta}] + \text{E} [v_{T+1}^2] + \text{E} [\tilde{v}_T^2] \simeq 2\sigma_v^2 \left(1 + \frac{x_{T+1}^2}{(T-1)\sigma_x^2} \right). \quad (53)$$

Compared to (50), the error variance is doubled.

However, if the indicator is just added for the final observation and not extrapolated, namely $\delta 1_{\{t=T\}}$, then:

$$\tilde{y}_{T+1} = \tilde{\beta} x_{T+1}$$

so:

$$\mathbb{E} [\tilde{v}_{T+1}^2] = x_{T+1}^2 \mathbb{V} [\tilde{\beta}] + \mathbb{E} [v_{T+1}^2] \simeq \sigma_v^2 \left(1 + \frac{x_{T+1}^2}{(T-1)\sigma_x^2} \right). \quad (54)$$

Consequently, relative to (50) and (53), it is not the ‘setting back on track’ *per se* that doubles the error variance, but the assumption that the location shift persists into the forecast period as an intercept correction. There is an MSFE loss of (54) over (50) of:

$$\mathbb{E} [\tilde{v}_{T+1}^2] - \mathbb{E} [\hat{v}_{T+1}^2] \simeq \frac{\sigma_v^2 x_{T+1}^2}{T(T-1)\sigma_x^2},$$

which is of order $\mathcal{O}(T^{-2})$, so only a small cost ensues.

In practice, an indicator is often added to correct an outlier in the final observation, which is probably measured less accurately than earlier ones, deriving from the DGP being the same as (51):

$$y_t = \beta x_t + \delta 1_{\{t=T\}} + v_t. \quad (55)$$

The alternative of not including the indicator entails that (49) would become:

$$\hat{\beta}_\delta = \hat{\beta} + \frac{\delta x_T}{\sum_{t=1}^T x_t^2},$$

with forecast MSFE from $\hat{y}_{\delta, T+1} = \hat{\beta}_\delta x_{T+1}$ of:

$$\mathbb{E} [\hat{v}_{\delta, T+1}^2] = x_{T+1}^2 \mathbb{E} \left[\left(\beta - \hat{\beta}_\delta \right)^2 \right] + \mathbb{E} [v_{T+1}^2] \simeq \sigma_v^2 \left(1 + \frac{x_{T+1}^2}{T\sigma_x^2} \right). \quad (56)$$

Thus:

$$\begin{aligned} \mathbb{E} [\hat{v}_{\delta, T+1}^2] - \mathbb{E} [\tilde{v}_{T+1}^2] &= \frac{\sigma_v^2 x_{T+1}^2}{T\sigma_x^2} - \frac{\sigma_v^2 x_{T+1}^2}{(T-1)\sigma_x^2} + \frac{\sigma_v^2 x_{T+1}^2}{T\sigma_x^2} \frac{\delta^2 x_T^2}{T\sigma_x^2} \\ &= \frac{\sigma_v^2 x_{T+1}^2}{T\sigma_x^2} \left(\frac{\delta^2 x_T^2}{T\sigma_x^2} - \frac{1}{(T-1)} \right) \\ &\simeq \frac{\sigma_v^2 x_{T+1}^2}{T\sigma_x^2} \left(\frac{T(\delta^2 - 1) - \delta^2}{(T-1)T} \right), \end{aligned}$$

so that for a reasonable size of sample, $\mathbb{E}[\hat{v}_{\delta, T+1}^2] \geq \mathbb{E}[\tilde{v}_{T+1}^2]$ when $\delta^2 > 1$. Thus, a relatively small outlier justifies setting the model back on track before forecasting, separately from the decision to extrapolate the indicator into the future.

7 Conclusion

We have considered the addition of impulse indicators in static regressions, their combination in an index, and their data-based selection, both when needed to correct outliers, and when unnecessary. The implications seem remarkably benign. The coefficients of such dummies are unbiased but inconsistent; their standard errors are consistent; the ratio of the first to the second has a t-distribution for normally-distributed errors, but provides an inconsistent test in general. Even nearly saturating a model with

impulse dummies need not induce ‘spurious’ results, hence selecting the ‘most significant’ of these is not problematic either. Although too many dummies can distort some mis-specification tests, solutions exist, either by forming an index, or modifying the test. An index can be consistently estimated when not ‘too mis-specified’ for the correct weights. Including dummies in a model, when the data suggests doing so, seems beneficial, relative to not keeping the dummies when they matter; including dummies when they don’t matter seems relatively harmless, although there is a small efficiency loss.

The baseline case of a static regression plays a useful pedagogic role, but it is well known that results on dummies in such models do not generalize easily to either dynamic models or integrated data processes: see e.g., Doornik, Hendry and Nielsen (1998) and Nielsen (2003). Nevertheless, we believe the above results serve to mitigate some of the fears we have encountered from referees on the role of dummies in econometric modelling.

References

- Abraham, B., and Chuang, A. (1989). Outlier detection and time series modeling. *Technometrics*, **31**, 241–248.
- Bowman, K. O., and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, **62**, 243–250.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics*: Chapman and Hall.
- Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox* 4th edn. London: Timberlake Consultants Press.
- Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, **12**, 533–572.
- Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1999). An econometric analysis of US food expenditure, 1931–1989. In Magnus, J. R., and Morgan, M. S. (eds.), *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, pp. 341–361. Chichester: John Wiley and Sons.
- Hendry, D. F. (2000). Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or Science?*, pp. 467–490. Oxford: Oxford University Press. New Edition.
- Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, **16**, 255–275.
- Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
- Messer, K., and White, H. (1984). A note on computing the heteroscedasticity consistent covariance matrix using instrumental variable techniques. *Oxford Bulletin of Economics and Statistics*, **46**, 181–184.
- Nielsen, H. B. (2003). Cointegration analysis in the presence of outliers. Discussion paper 03-05, Institute of Economics, University of Copenhagen.

- Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, **4**, 393–397.
- Santos, C. (2003). Statistical inference in econometric models with indicator variables. Unpublished MPhil thesis, Department of Economics, University of Oxford.
- Sargan, J. D. (1988). *Lectures on Advanced Econometric Theory*. Oxford: Basil Blackwell.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.
- White, H. (1984). *Asymptotic Theory for Econometricians*. London: Academic Press.