

# Econometric Modelling of Time Series with Outlying Observations

David F. Hendry

Department of Economics, University of Oxford, UK

Grayham E. Mizon

Division of Economics, School of Social Sciences, University of Southampton, UK.\*

## Abstract

Economies are buffeted by natural shocks, wars, policy changes, and other unanticipated events. Observed data can be subject to substantial revisions. Consequently, a ‘correct’ theory can manifest serious mis-specification if just fitted to data ignoring its time-series characteristics. Modelling US expenditure on food, the simplest theory implementation fails to describe the evidence. Embedding that theory in a general framework with dynamics, outliers and structural breaks, using impulse-indicator saturation, the selected model performs well, despite commencing with more variables than observations (see Doornik, 2009b), producing useful robust forecasts. Although this illustration involves a simple theory, the implications are generic and apply to sophisticated theories.

*JEL classifications:* C51, C22.

**KEYWORDS:** Econometric modelling; Food Expenditure; Outliers; Impulse-indicator saturation; Robust forecasting; *Autometrics*.

## 1 Introduction

The recent financial crisis and resulting global recession are stark reminders that economies can experience unexpected changes with substantive effects on their state and operation. Such sudden changes—or more precisely, structural breaks—lead to difficulties in economic forecasting, and later empirical modelling, as well as raising doubts about the value of economic models and the theories underlying them. Although the arguments in Engsted (2009) are attractive, the fact that some large changes are unexpected entails that economists must now assess the applicability of the economic theory, and the performance of the modelling methodology used to develop their models, as well as the appropriateness and quality of the empirical observations (especially given the frequency and size of some data revisions). For models to be relevant, reliable and robust they must exploit as fully as possible all the available information from economic theory, empirical observations, computer technology (hardware and software), historical and institutional knowledge, as well as being able to allow for, or adapt to, shifts in the economy. This paper proposes an approach to doing so.

Although the ideal for reliable analysis may be the development of a structural model *ab initio*, it is more realistic to seek an ability to quickly: (i) identify a new regime’s characteristics, and (ii) develop a model of that regime. It is unclear precisely how this can be done within a framework of theory-led models developed using a simple-to-general strategy that leads to a tightly-specified empirical implementation of that theory, and only introduces modifications of a limited nature, such as *ad hoc* stickiness to deal with mis-specified dynamics. Such simple-to-general modelling is fraught with difficulties as has

---

\*We are grateful to Jennifer L. Castle, two anonymous referees and the editors for helpful comments, and to James J. Reade for updating the data. All the calculations reported below were based on *PcGive* (see Hendry and Doornik, 2009).

been explained by numerous authors (for recent contributions see *inter alia*, Hendry, 1995, Johansen, 2006, Juselius and Johansen, 2006, Mizon, 1995, and Spanos, 1995). The modelling strategy adopted in this paper is designed to function in a world of intermittent unanticipated location shifts. By modelling from general-to-simple and focusing on selecting variables rather than models, recent developments in the automation of model selection have produced remarkable results. For discussions of this achievement, see Castle, Doornik and Hendry (2009), Castle *et al.* (2009) and Doornik and Hendry (2009). The results of this large body of research are embodied in the software package *Autometrics* (see Doornik, 2008, 2009a, Hendry and Doornik, 2009), which we use below.

To illustrate that even though a theory is essentially ‘correct’, it can exhibit serious mis-specification if it is just fitted without paying attention to the observed characteristics of the data, especially substantive external events such as wars, major recessions, and policy regime changes, we model expenditure on food in the USA. Just adding dynamics, or just tackling outliers and breaks, is insufficient. However, when the same theory is embedded in the more general framework embracing dynamics and the possibility of intermittent structural changes, then that original theory can perform well even over an extended data period, as shown here applying impulse-indicator saturation (see Hendry, Johansen and Santos, 2008 and Johansen and Nielsen, 2009). The rapid evolution of economic theory makes it unwise to impose today’s theory on data, as tomorrow’s theory may lead to such ‘evidence’ being discarded. To achieve the right balance of also not neglecting valuable analytic insights, available theory can be embedded in the modelling exercise, to be retained when it is complete and correct, rejected if invalid, yet possibly rescued or modified when relevant additional aspects are included in the general model. Doing so also avoids reliance on doubtful assumptions about the sources of problems like residual autocorrelation—perhaps a symptom of breaks—so that ‘corrections’ fail to achieve valid inference. Although our particular illustration involves a straightforward theory, the point being made is generic, and applies no matter how sophisticated the theory: in wide-sense non-stationary time series, all substantively relevant factors must be included if useful empirical models are to result.

The paper is organized as follows. Some previous analyses of food expenditure in the USA from shorter data sets are summarized in Section 2, and the need to deal with structural breaks and other unanticipated changes is emphasized. Section 3 outlines the theory behind impulse-indicator saturation (denoted IIS). Estimates of the static theory model are presented in Section 4, then that model is re-estimated using IIS. Next, embedding the imposed theory in a dynamic specification is explored in Section 5, but without IIS, cointegration fails. Like many time series, the sample experienced a number of major shifts, here from food programs, World War II, and the Korean war. It is the combination of including dynamics and IIS that dramatically improves the results. Now selecting the theory variables to eliminate insignificant coefficients barely changes the findings, and delivers a cointegrating vector corresponding to the initial theory. Section 5.1 shows that without the indicator variables, cointegration is not found. The conditional forecasting performance of the model, both from 1952 and from 2003, is evaluated in Section 6. Two solutions to the poor forecasting performance over the final period are explored in Section 6.1, and conclusions are presented in Section 7.

## 2 Modelling food demand

An early time-series analysis of USA food expenditure was Tobin (1950), followed by a number of studies reported in Magnus and Morgan (1999). Building on Hendry (2009), we re-examine US food expenditure over 1931–2002, based on the update of Tobin (1950) by Magnus and Morgan (1999) to 1989 and extended as described in Reade (2008).

The data variables are defined as follows (lower case denotes logs):  
 $e_f$ : constant-price per capita expenditure on food;

$p_f - p$ : real price of food;  
 $e$ : constant-price per capita total expenditure;  
 $p$ : deflator of total expenditure;  
 $y$ : constant-price per capita income;  
 $s = y - e$ : an approximation to the savings ratio;  
 $a$ : average family size;  
 $n$ : total population of the USA.

Here the basic theory is that:

$$e_f = f(e, p_f - p, s, a) \quad (1)$$

The earlier study of a cointegrated VAR for the system in Hendry (2009) established that  $e$ ,  $s$ , and  $p_f - p$  were weakly exogenous in the food demand equation, justifying the analysis of models conditional on these variables. Correction to per capita should mean that  $n$  is irrelevant, and demographic effects are captured by  $a$ . According to conventional theory, the anticipated signs are:

$$\frac{\partial e_f}{\partial e} > 0, \quad \frac{\partial e_f}{\partial (p_f - p)} < 0, \quad \frac{\partial e_f}{\partial s} > 0, \quad \frac{\partial e_f}{\partial a} < 0, \quad \frac{\partial e_f}{\partial n} = 0 \quad (2)$$

Because his estimates of quantities of food supply suggested these had not changed greatly over the period 1912–1948, Tobin (1950) ‘inverted’ (1) to use food price as the dependent variable in his time-series model (he also estimated a cross-section model). Here, we take food prices as being determined in the world market, so model per capita expenditure. Like Tobin, we take  $f(\cdot)$  to be multiplicative, and hence linear after logs.

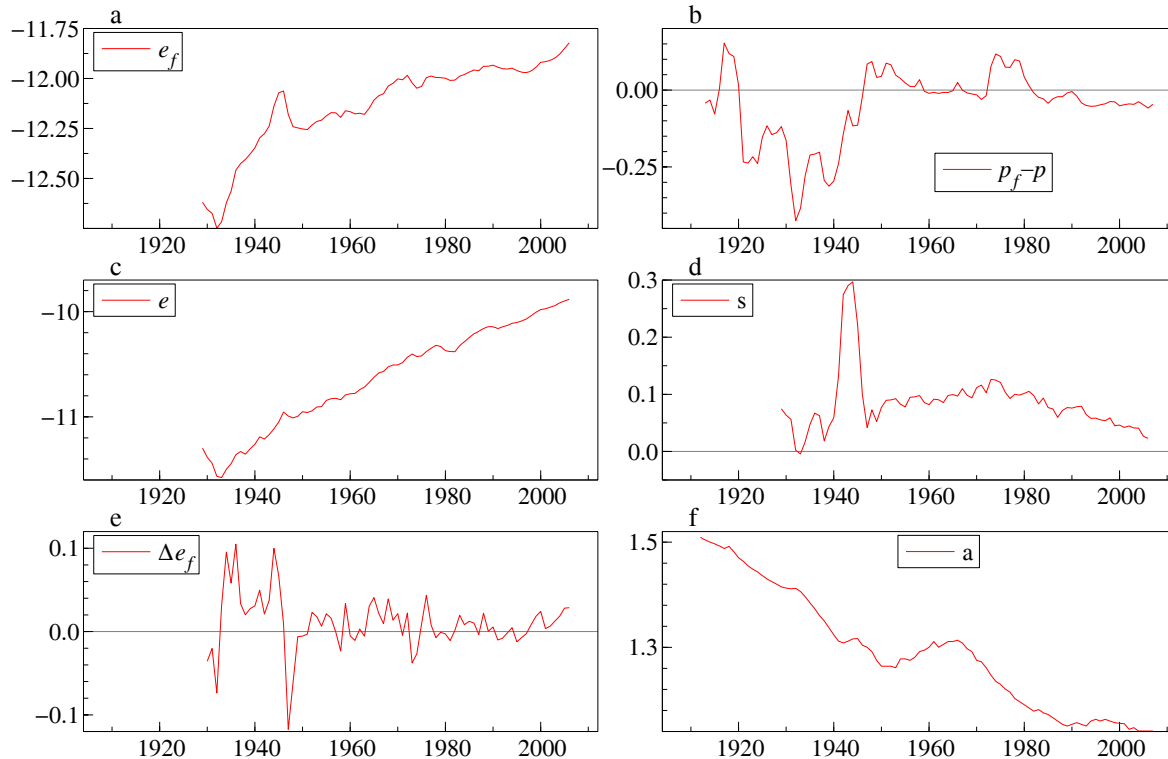


Figure 1: Food expenditure and related time series

Figure 1 shows the time series, which reveal considerable changes over the period. After falling sharply at the commencement of the Great Depression, both  $e_f$  and  $e$  rise substantially till World War II, fall after, then resume a gentle rise (panels a and c), so  $\Delta e_f$  is much more volatile pre-war (panel e:

$\Delta e_t$  has a similar but less pronounced pattern). Next,  $p_f - p$  is quite volatile till after the war, then is relatively stable (panel b), whereas the dramatic rise in  $s$  from ‘forced saving’ during the war is manifest (panel d). Average family size (panel f) has fallen considerably, partly reflecting changes in social mores.

Section 4 illustrates that even if an economic theory is fundamentally correct, it need not be coherent with the data. Here, the estimated static model has wrong coefficient signs and is seriously mis-specified, either of which could lead to the (false) rejection of the theory. Most contributors to Magnus and Morgan (1999) found their dynamic models of (1) were so non-constant over the combined inter-war and post-war samples that they did not model the data before 1950. However, Hendry (1999) found that adding impulse indicators for the sample till 1950 detected some very large outliers (later identified by archival research as being for a food program and post-war de-rationing), which allowed a constant equation to be developed over the (then) whole sample 1931–1989. Therefore, we now briefly describe how impulse-indicator saturation (IIS) can successfully deal jointly with outliers, structural breaks and data revisions even though the resulting model involves more variables than observations.

### 3 Modelling with data revisions and breaks

Economic theory is rarely explicit about outliers and structural breaks, but if not modelled, these can have a pernicious effect on a model’s fit, parameter estimates, and constancy even when the underlying theory is ‘valid’. Breaks matter greatly for food demand in a world subject to unanticipated shifts due to health scares, major wars, and the introduction of policy programs like food relief. Indeed not modelling breaks can result in estimated models exhibiting unit roots when none are there (see *inter alia*, Perron, 1989 and Hendry and Neale, 1991). Outliers arise both from sudden behavioural shifts (scares over the safety of apples from alar; mercury in swordfish, salmonella in eggs, etc.), and from data measurement errors: see Doornik, Hendry and Nielsen (1998) and Nielsen (2004) for analyses of the impacts of outliers in cointegrated systems. Since few theory models allow for such shifts, these are dealt with here by including indicators that remove both location shifts and outliers, possibly from data mistakes.

The inclusion of deterministic time trends, seasonal dummy variables, and event-specific dummy variables is a well established practice in empirical econometric modelling, but the recent developments in allowing an impulse indicator for every observation among the candidate variables (so there are more potential explanatory variables than the sample size) are less well known.<sup>1</sup> Since there are likely to be multiple breaks, impulse-indicator saturation (IIS) offers an attractive approach. Indeed, the concept originated in an earlier re-analysis of the model in Tobin (1950), by including indicators for the inter-war period to investigate the heterogeneity that had led other investigators to eschew that sample as noted above. Hendry *et al.* (2008) and Johansen and Nielsen (2009) analyze the theory of IIS under the null that there are no breaks or outliers, and the latter relate it to robust statistical methods which seek to eliminate possible data contamination. Since IIS is a form of robust estimation, it can jointly tackle data contamination and ‘fat-tailed’ distributions (by removing extreme observations), as well as location shifts and innovation outliers, at the same time as selecting over potential regressors. We now briefly describe the IIS procedure.

First generate impulse indicators for every observation,  $1_{\{t=s\}} \forall s$ , still including an intercept. The indicators are divided into  $J$  subsets (e.g. 2 of equal size in the split-sample approach), which form a set of initial general unrestricted models (GUMs) each of which is then estimated. The significant indicators from each subset are then selected in terminal models and recorded. The joint model is formulated as the union of these  $J$  terminal models, estimated and the indicators re-selected. Under the null that there are no outliers,  $\alpha T$  indicators will be retained in a sample of  $T$  observations on average at a significance

---

<sup>1</sup>There are in fact many ways to estimate models with more candidate variables than observations, including stepwise forward searches, (see e.g., Kiiveri, 2006 and Stodden, 2008), but Doornik (2009b) shows that they are not all equally effective.

level  $\alpha$ . Johansen and Nielsen (2009) show that the cost of testing for the significance of  $T$  indicators under the null is extraordinarily low for small  $\alpha$ : for example, when  $\alpha = 1/T$ , only one observation is ‘dummied out’ on average. Thus, for  $T = 100$  and  $\alpha = 0.01$ , IIS is 99% efficient despite including an additional 100 candidate variables. Castle *et al.* (2009) show that IIS also alleviates fat-tailed draws, and allows near-normal inference, important both during search and for the post-selection bias corrections which assume normality. Although the analysis in Johansen and Nielsen (2009) does not jointly select over regressors as well as indicators, as we do below, simulation studies in Castle *et al.* (2009) show that does not affect their conclusions.

Given the trivial cost of IIS under the null, the next issue is its ability to detect outliers, breaks, and data contamination. This is difficult to ascertain theoretically for multiple sources of discrepancies, but Hendry and Santos (2010) and Castle *et al.* (2009) provide analyses and simulations under the alternatives of location shifts and outliers, and the latter demonstrate that it performs favourably compared to procedures like Bai and Perron (1998). Moreover, since only  $\alpha T$  indicators would be retained under the null on average, when many are found the procedure must have detected substantive departures from the null. Since there will always be more candidate variables,  $N$ , than observations,  $T$ , when IIS is applied to a model, and other splits than  $J = 2$  need to be considered, the use of an automatic selection procedure is essential.

Applying IIS in econometric modelling thus not only assesses the adequacy of a model, allows for external events that have significant effects on the phenomena being analyzed, and removes any adverse impacts from large data errors, it also complements the standard use of dummies for known shifts, both checking that their timing is correct and that all outlying events have been included.

## 4 Static models

In this section, the simple theoretical framework of equation (1) with no lagged values of variables is estimated, and following its poor performance, is re-estimated with impulse-indicator saturation (IIS), to investigate whether ‘removing’ major breaks and outliers is sufficient by itself.

The results obtained for the imposed static theory model are:

$$e_{f,t} = \begin{matrix} 5.30 \\ (4.02) \end{matrix} + \begin{matrix} 0.77 \\ (0.14) \end{matrix} e_t + \begin{matrix} 0.11 \\ (0.08) \end{matrix} (p_f - p)_t + \begin{matrix} 0.72 \\ (0.14) \end{matrix} s_t - \begin{matrix} 0.36 \\ (0.23) \end{matrix} a_t - \begin{matrix} 0.73 \\ (0.22) \end{matrix} n_t \quad (3)$$

$$R^2 = 0.94 \quad \hat{\sigma} = 0.055 \quad \chi_{nd}^2(2) = 19.5^{**} \quad F_{ar}(2, 66) = 44.3^{**} \\ F_{arch}(1, 72) = 216.8^{**} \quad F_{reset}(2, 66) = 18.1^{**} \quad F_{het}(10, 63) = 23.2^{**}$$

In (3),  $R^2$  is the squared multiple correlation, and  $\hat{\sigma}$  is the residual standard deviation, with coefficient standard errors shown in parentheses. The diagnostic tests are of the form  $F_j(k, T - l)$  which denotes an approximate F-test against the alternative hypothesis  $j$  for:  $k^{th}$ -order serial correlation ( $F_{ar}$ : see Godfrey, 1978),  $k^{th}$ -order autoregressive conditional heteroscedasticity ( $F_{arch}$ : see Engle, 1982), heteroscedasticity ( $F_{het}$ : see White, 1980); the RESET test ( $F_{reset}$ : see Ramsey, 1969); and a chi-square test for normality ( $\chi_{nd}^2(2)$ : see Doornik and Hansen, 2008). In the later sections of the paper we also report a parameter constancy test ( $F_{Chow}$ : see Chow, 1960) over  $k$  periods.

The static economic-theory model has a very poor fit, and does not capture well the behaviour of the observed data. The price elasticity has the ‘wrong sign’, contradicting (2), but is insignificant. Also, although it is theoretically irrelevant, population  $n_t$  is significant. Finally, every mis-specification test strongly rejects. Figure 2 shows that the estimated model fails to describe the 1930s.

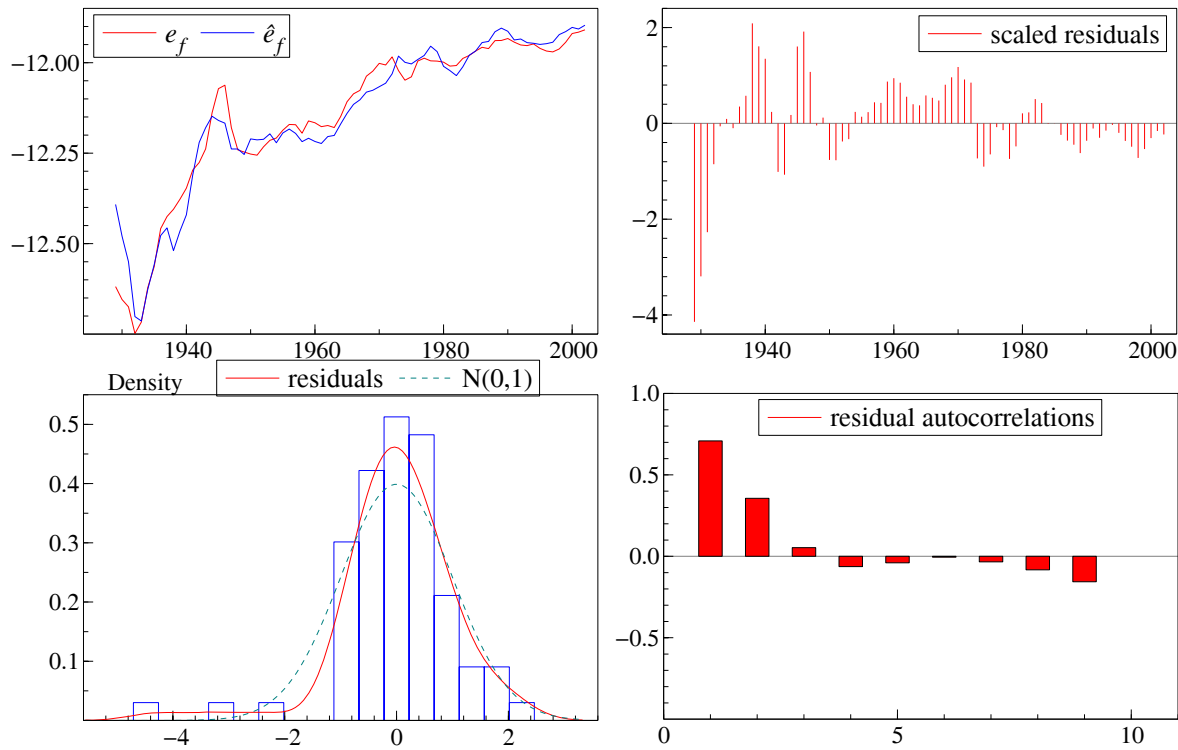


Figure 2: Static unselected ‘theory’ equation

#### 4.1 Static ‘theory’ equation with IIS

Given its poor performance, and the manifestly different behaviour of the pre and post war time series, IIS is next introduced into the specification.

The theory of IIS is derived under the null that no outliers or breaks matter. The distributional results in Johansen and Nielsen (2009) apply to well-specified dynamic equations, albeit possibly with unit roots in the generating processes, and show that null retention rates are close to nominal, and the distributions of the regression model parameter estimates are close to those derived when IIS is not used. Here, it is clear that the static model is invalid even if there were no breaks, and as yet it has not been established that the set of variables cointegrates. Thus, this exercise serves to illustrate the problems of not including all the relevant influences, and also shows that IIS does not fix that by itself. One cannot infer from the rejection of a null (here no breaks) that it is false *per se*, but can infer that the simpler model is wrong.

While the results in Johansen and Nielsen (2009) apply to static equations in  $I(1)$  variables, such relations are assumed to be correctly specified so that conventional inference is valid given Sims, Stock and Watson (1990). We undertook a simulation study to establish the null retention rate of impulse indicators when the DGP is a constant, dynamic, cointegrated bivariate system, but a static regression is estimated with IIS: a] when the static  $I(1)$  relation was correctly specified with white-noise errors; and b] the dynamic model was mis-specified as a static relation. In the former case, impulse indicators were retained 1.4% of the time at a nominal significance level of 1%, as anticipated, In the latter case, impulse indicators were retained 7.5% of the time, possibly ‘picking up’ some of the omitted lagged dependent variable. Thus, IIS correctly reveals that the static representation is incorrect, and although a naive investigator might interpret the retained impulse indicators as showing ‘breaks or outliers’, that also highlights the dangers of seeking to conduct inference in non-congruent models.

The resulting selected model is:

$$\begin{aligned}
e_{f,t} = & \underset{(0.06)}{0.64} e_t - \underset{(0.03)}{0.15} (p_f - p)_t + \underset{(0.1)}{1.0} s_t - \underset{(0.08)}{0.01} a_t - \underset{(0.10)}{0.47} n_t \\
& + \underset{(1.8)}{0.24} - \underset{(0.02)}{0.30} I_{29} - \underset{(0.02)}{0.27} I_{30} - \underset{(0.02)}{0.26} I_{31} - \underset{(0.02)}{0.21} I_{32} \\
& - \underset{(0.02)}{0.16} I_{33} - \underset{(0.02)}{0.12} I_{34} - \underset{(0.02)}{0.11} I_{35} - \underset{(0.02)}{0.08} I_{36} - \underset{(0.02)}{0.06} I_{37} \\
& - \underset{(0.02)}{0.08} I_{41} - \underset{(0.02)}{0.17} I_{42} - \underset{(0.02)}{0.16} I_{43} - \underset{(0.02)}{0.10} I_{44} + \underset{(0.02)}{0.10} I_{46} \\
& + \underset{(0.02)}{0.09} I_{47} + \underset{(0.02)}{0.03} I_{70} + \underset{(0.02)}{0.03} I_{72} - \underset{(0.02)}{0.034} I_{73} - \underset{(0.02)}{0.03} I_{98} \tag{4}
\end{aligned}$$

$$\begin{aligned}
R^2 &= 0.997 \quad \hat{\sigma} = 0.015 \quad F_{\text{ar}}(2, 47) = 4.9 \quad \chi_{\text{nd}}^2(2) = 2.3 \\
F_{\text{arch}}(1, 72) &= 6.7^* \quad F_{\text{reset}}(2, 47) = 3.1 \quad F_{\text{het}}(10, 44) = 2.3
\end{aligned}$$

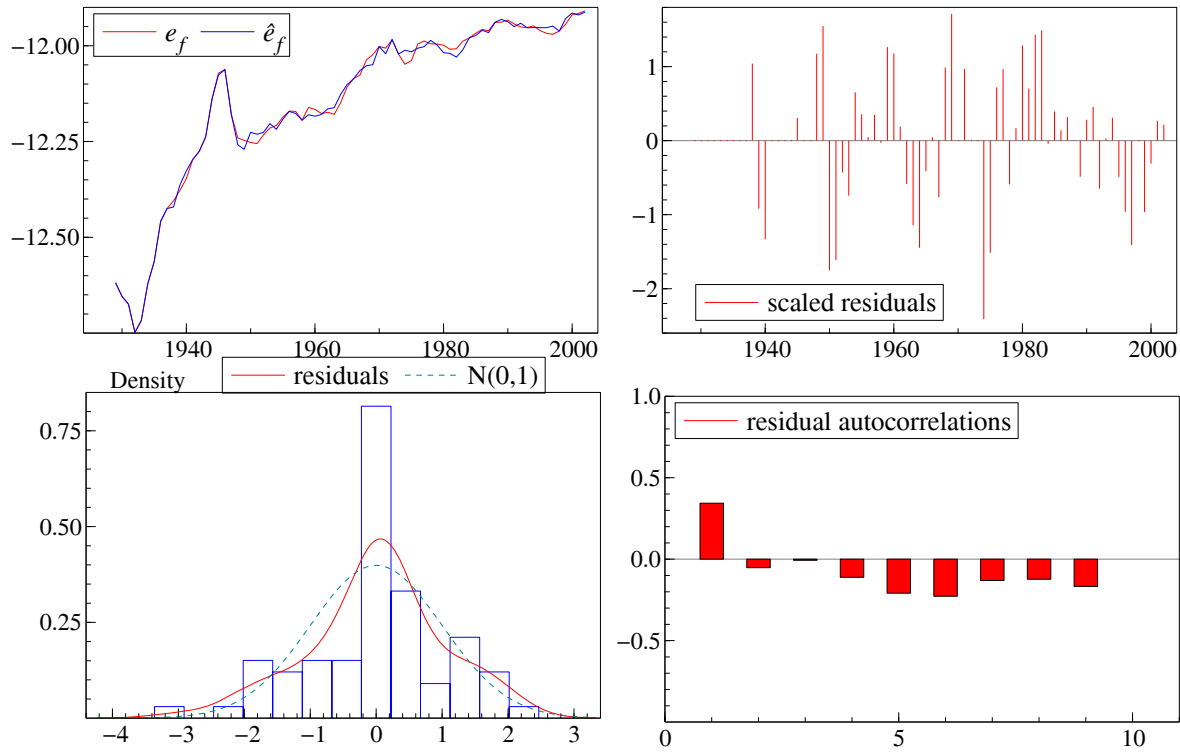


Figure 3: Static unselected ‘theory’ equation with IIS

The impulse indicators dummy out almost all interwar and war data, and as a result, few of the misspecification test statistics are significant. In effect, IIS has produced by model search a Salkever (1976) dummy-variable implementation of a Chow (1960) test of post-war versus earlier, as first described by Hendry (1999). Further, the estimated model is much closer to (1), though  $n$  remains significant, a feature of all subsequent results. It is tempting to conclude that the theory in (1) is valid—but only after the war for some reason, although Tobin originally estimated his time-series model over the very period now excluded. We will see that it is features of the data ‘outside’ of the theory which induce that failure (mainly location shifts from policy interventions and wars), and the final model we obtain satisfies most of the theory for most of the sample.

## 5 Dynamic models

When modelling aspects of sophisticated modern economies subject to adjustment (rather than instantaneous responses), habit, expectations formation, and external surprises, then dynamics clearly matter. However, even when a theory is explicit about which variables are relevant, it is rarely exact about dynamics, since the unit of time for agents' decisions has never been derived theoretically. Here, the data are annual, and so two lagged values of each variable are introduced to check whether the absence of a dynamic specification is the reason for the previous 'rejection' of the theory. In fact, for the demand for food two lags will transpire to matter empirically.

The dynamic model is first estimated without selection for all variables (including  $n$  as a check) and 2 lags:

$$\begin{aligned}
 e_{f,t} = & \quad 0.71 e_t - 0.20 (p_f - p)_t + 0.26 s_t + 0.05 a_t + 0.13 n_t \\
 & \quad (0.08) \quad (0.06) \quad (0.08) \quad (0.39) \quad (1.5) \\
 & + 0.98 e_{f,t-1} - 0.14 e_{f,t-2} - 0.73 e_{t-1} + 0.01 e_{t-2} \\
 & \quad (0.14) \quad (0.12) \quad (0.13) \quad (0.12) \\
 & + 0.25 (p_f - p)_{t-1} - 0.04 (p_f - p)_{t-2} + 0.01 s_{t-1} - 0.04 s_{t-2} \\
 & \quad (0.09) \quad (0.06) \quad (0.13) \quad (0.09) \\
 & - 0.10 a_{t-1} - 1.45 n_{t-1} + 1.41 n_{t-2} - 3.10 \\
 & \quad (0.38) \quad (2.72) \quad (1.41) \quad (1.80)
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 R^2 &= 0.996 \quad \hat{\sigma} = 0.015 \quad F_{ar}(2, 53) = 0.56 \\
 \chi_{nd}^2(2) &= 0.14 \quad F_{arch}(1, 70) = 2.60 \quad F_{reset}(2, 53) = 1.79 \quad F_{het}(32, 39) = 3.78^{**}
 \end{aligned}$$

The *PcGive* unit-root test does not reject the null of no cointegration ( $t_{ur} = -2.59$ : see Banerjee and Hendry, 1992, and Ericsson and MacKinnon, 2002) and delivers the pseudo long-run solution:

$$c_1 = e_f + 0.06e - 0.09(p_f - p) - 1.42s + 0.31a - 0.59n + 19.6$$

Thus, although the fit is hugely improved (see Figure 4), adding lags does not solve the mis-match between the theory and evidence: both key long-run price and expenditure elasticities are 'wrong signed'.

Applying *Autometrics* with the theory variables forced to be retained, but selecting the dynamic reactions, has almost no impact on the match of the theory and evidence, the fit of the model, the coefficient magnitudes and their standard errors, or the lack of cointegration. Selection simply eliminates insignificant lagged variables:

$$\begin{aligned}
 e_{f,t} = & \quad 0.74 e_t - 0.17 (p_f - p)_t + 0.27 s_t - 0.05 a_t - 1.21 n_t \\
 & \quad (0.07) \quad (0.05) \quad (0.04) \quad (0.07) \quad (0.72) \\
 & \quad 0.84 e_{f,t-1} - 0.75 e_{t-1} + 0.17 (p_f - p)_{t-1} + 1.30 n_{t-1} - 3.01 \\
 & \quad (0.05) \quad (0.06) \quad (0.05) \quad (0.74) \quad (1.49)
 \end{aligned} \tag{6}$$

$$\begin{aligned}
 R^2 &= 0.995 \quad \hat{\sigma} = 0.015 \quad F_{ar}(2, 60) = 0.80 \\
 \chi_{nd}^2(2) &= 0.90 \quad F_{arch}(1, 70) = 0.01 \quad F_{reset}(2, 60) = 1.60 \quad F_{het}(18, 53) = 2.24^*
 \end{aligned}$$

where:

$$c_2 = e_f + 0.045e - 0.011(p_f - p) - 0.56n + 19.1 - 1.72s + 0.33a \quad t_{ur} = -3.32$$

The finding of no cointegration with 'wrong signs', and an estimated lagged dependent variable coefficient close to unity, are consistent with unit roots 'capturing' breaks that are not dealt with directly (see *inter alia* Perron, 1989, Hendry and Neale, 1991, and Banerjee, Lumsdaine and Stock, 1992).



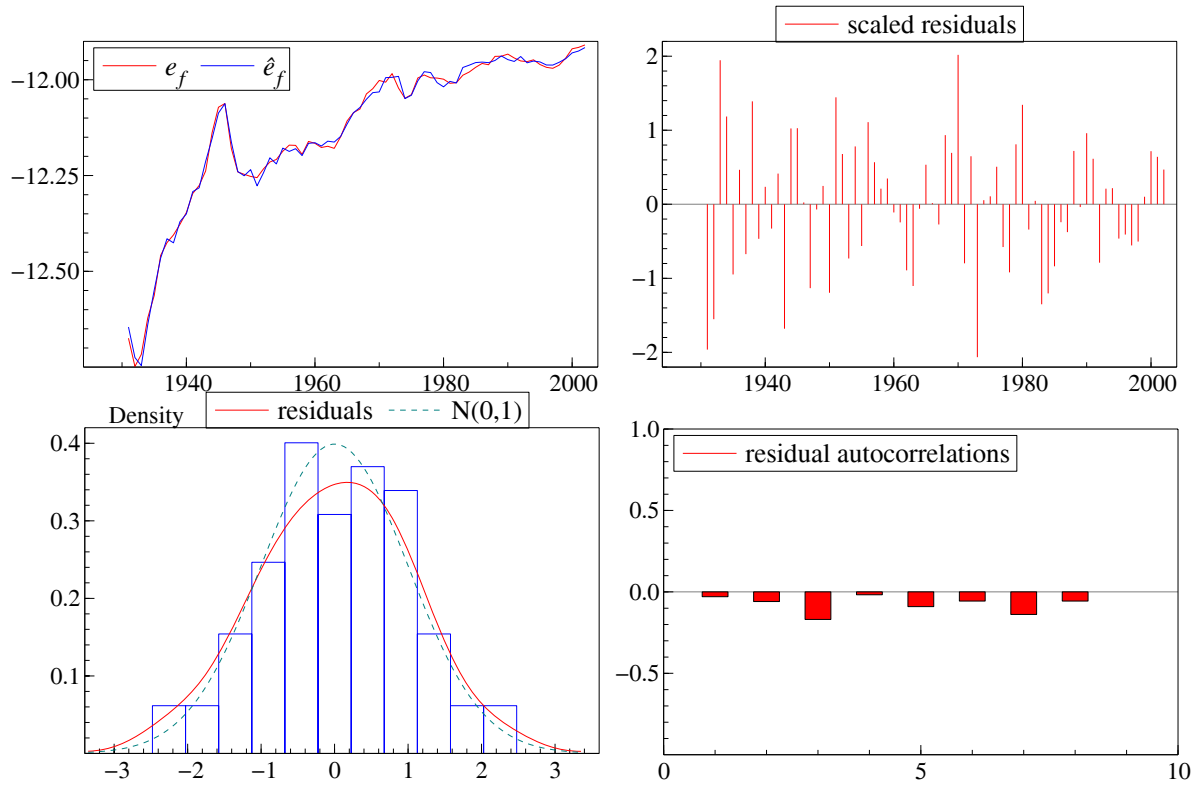


Figure 4: Estimated theory model with dynamics

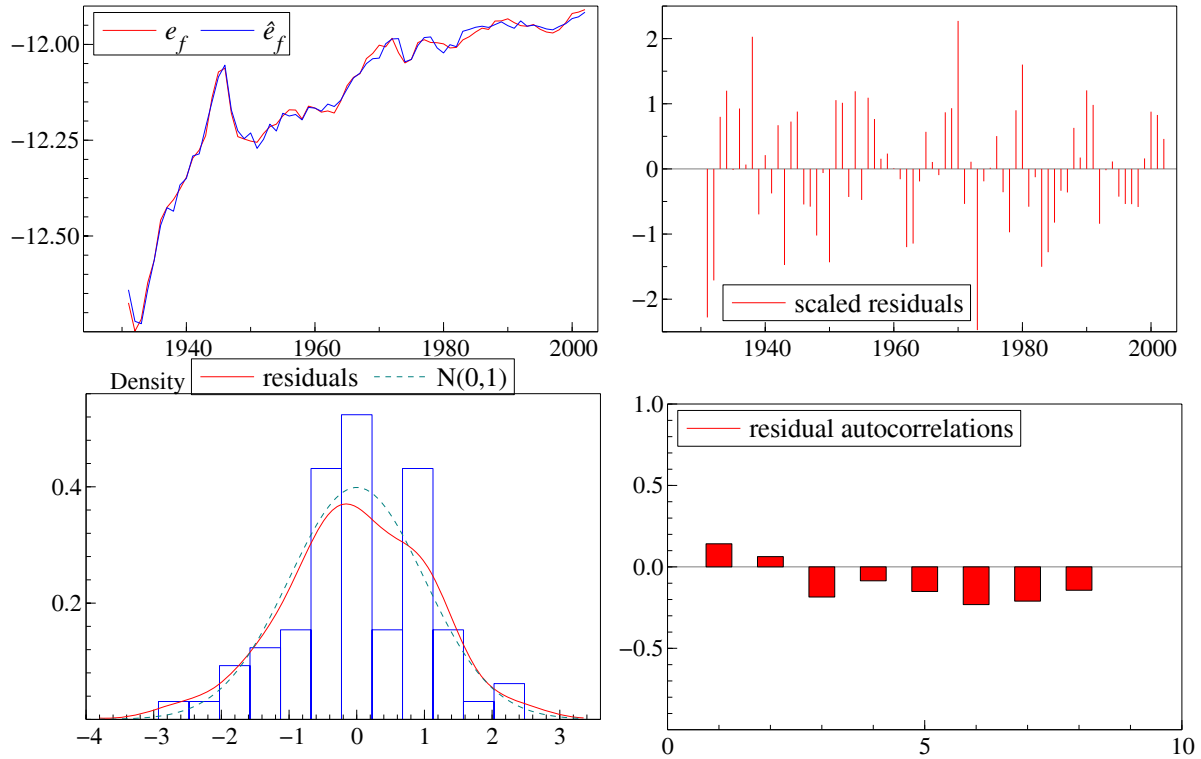


Figure 5: Dynamic model with imposed 'theory' but selected dynamics

Hence, we next estimated a dynamic model including IIS, selecting the dynamics and impulse indi-



with no substantive effect on the theory-model relationship:

$$\begin{aligned}
e_{f,t} = & \begin{matrix} 0.59 & -0.32 & 0.23 & -0.16 & 0.77 \\ (0.04) & (0.03) & (0.04) & (0.02) & (0.05) \end{matrix} e_t - (p_f - p)_t + s_t - n_t + e_{f,t-1} \\
& - \begin{matrix} 0.13 & 0.36 & 0.27 & 0.17 \\ (0.04) & (0.05) & (0.03) & (0.05) \end{matrix} e_{f,t-2} - e_{t-1} + (p_f - p)_{t-1} + s_{t-1} \\
& - \begin{matrix} 0.11 & 0.11 & 0.03 & -0.03 & 0.03 \\ (0.01) & (0.01) & (0.01) & (0.01) & (0.01) \end{matrix} I_{31} - I_{32} + I_{34} - I_{43} + I_{70} \quad (8)
\end{aligned}$$

$$\begin{aligned}
(R^*)^2 &= 0.999 \quad \hat{\sigma} = 0.0085 \quad F_{ar}(2, 56) = 0.69 \\
\chi_{nd}^2(2) &= 1.69 \quad F_{arch}(1, 70) = 0.19 \quad F_{reset}(2, 56) = 1.16 \quad F_{het}(18, 48) = 1.02
\end{aligned}$$

with the resulting solved cointegrating relation:

$$c_4 = e_f - \begin{matrix} 0.63 \\ (0.01) \end{matrix} e + \begin{matrix} 0.13 \\ (0.04) \end{matrix} (p_f - p) - \begin{matrix} 1.12 \\ (0.08) \end{matrix} s + \begin{matrix} 0.45 \\ (0.01) \end{matrix} n \quad \text{where } t_{ur} = -12.1^{**}$$

That there are no substantive changes from selecting over all variables, merely the elimination of insignificant regressors, is our experience of such methods when applied to general initial models. Figure 8 plots the resulting  $c_{4,t}$ , which shows a steady return from the disequilibrium following 1929, a further dip at the onset of World War II, peaking around the time of the Korean War, then benign behaviour thereafter (dummies are excluded from the calculation of  $c_{4,t}$ ).

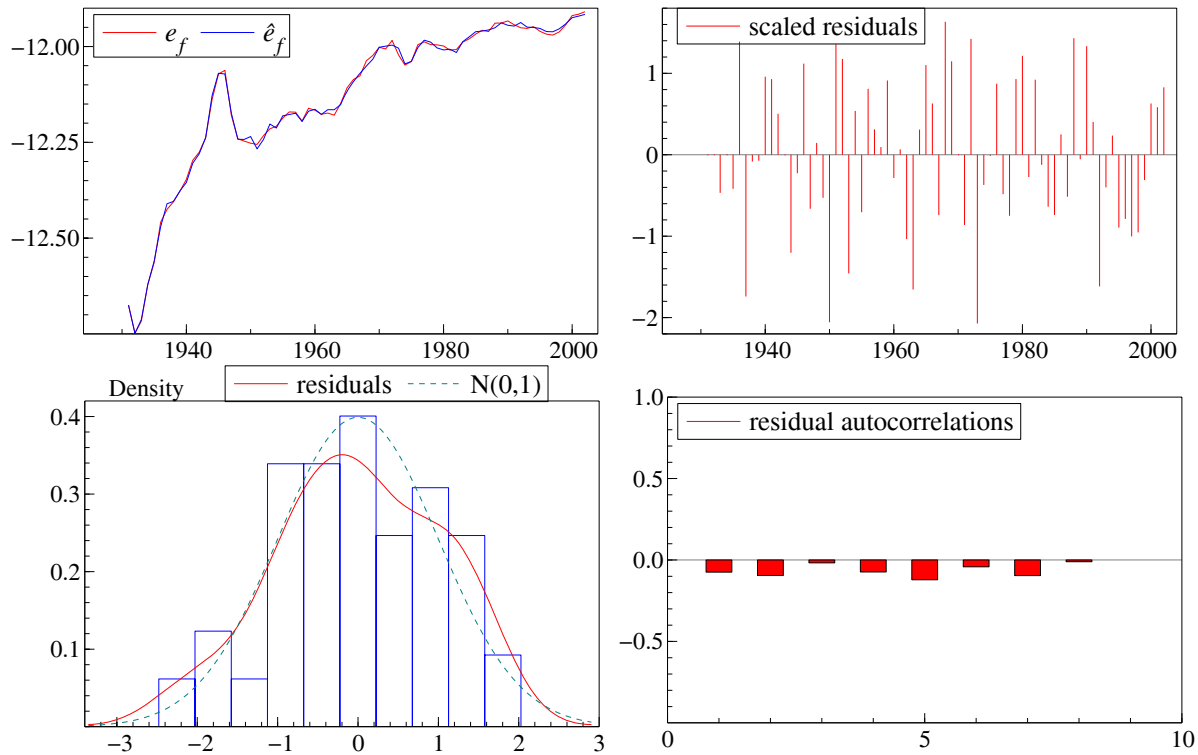


Figure 7: Selecting across all variables, dynamics and IIS

## 5.1 Cointegration

Since an EqCM has a much clearer economic interpretation than a levels autoregression (see e.g., Hendry, 1995, 2008), and the previous results produced a strong and economically meaningful cointegration

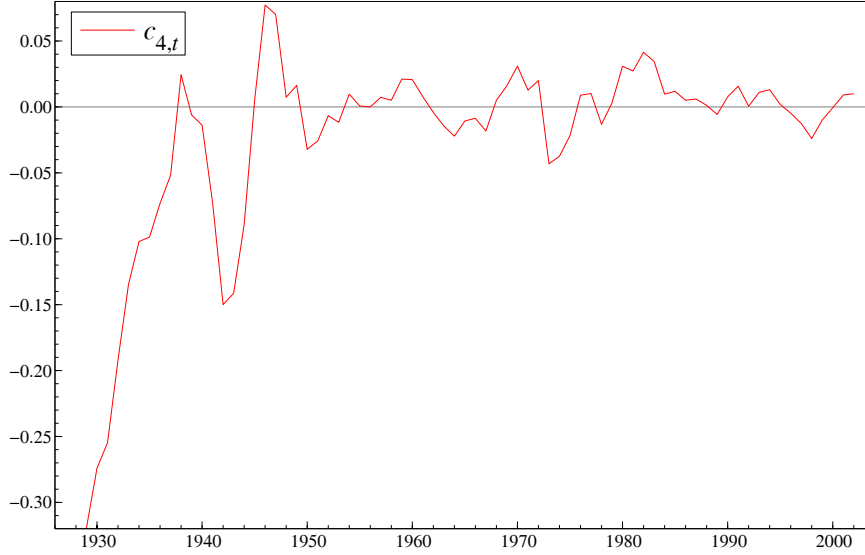


Figure 8: Solved EqCM,  $c_{4,t}$ , from (8)

relationship, we next estimated this re-parameterization. Nevertheless, using the solved cointegrating relation  $c_{4,t}$  as the feedback in an EqCM, but without IIS, again results in the failure of several misspecification tests ( $(R^*)^2$  is computed including an intercept):

$$\Delta e_{f,t} = \begin{matrix} 0.85 \\ (0.06) \end{matrix} \Delta e_t - \begin{matrix} 0.20 \\ (0.05) \end{matrix} \Delta(p_f - p)_t - \begin{matrix} 0.60 \\ (0.17) \end{matrix} \Delta p_{t-1} + \begin{matrix} 0.28 \\ (0.06) \end{matrix} \Delta s_t - \begin{matrix} 0.22 \\ (0.03) \end{matrix} c_{4,t-1}$$

$$(R^*)^2 = 0.82 \quad \hat{\sigma} = 0.015 \quad F_{ar}(2, 65) = 0.50$$

$$\chi_{nd}^2(2) = 3.08 \quad F_{arch}(1, 70) = 0.28 \quad F_{reset}(2, 65) = 5.46^{**} \quad F_{het}(10, 61) = 8.09^{**}$$

However, introducing IIS into the EqCM with the cointegrating relation  $c_4$  delivers an appropriate  $l(0)$  form:

$$\begin{aligned} \Delta e_{f,t} = & \begin{matrix} 0.13 \\ (0.04) \end{matrix} \Delta e_{f,t-1} + \begin{matrix} 0.58 \\ (0.03) \end{matrix} \Delta e_t - \begin{matrix} 0.32 \\ (0.03) \end{matrix} \Delta(p_f - p)_t + \begin{matrix} 0.23 \\ (0.04) \end{matrix} \Delta s_t - \begin{matrix} 0.36 \\ (0.02) \end{matrix} c_{4,t-1} \\ & - \begin{matrix} 0.11 \\ (0.01) \end{matrix} I_{31} - \begin{matrix} 0.11 \\ (0.01) \end{matrix} I_{32} + \begin{matrix} 0.03 \\ (0.01) \end{matrix} I_{34} - \begin{matrix} 0.03 \\ (0.01) \end{matrix} I_{43} + \begin{matrix} 0.03 \\ (0.01) \end{matrix} I_{70} \end{aligned} \quad (9)$$

$$(R^*)^2 = 0.95 \quad \hat{\sigma} = 0.0083 \quad F_{ar}(2, 60) = 0.60$$

$$\chi_{nd}^2(2) = 1.51 \quad F_{arch}(1, 70) = 0.06 \quad F_{reset}(2, 59) = 0.44 \quad F_{het}(10, 56) = 1.08$$

where  $F_{Chow}(21, 41) = 0.62$  for a break after 1981. This model satisfies all the theory and statistical criteria—and would have been the outcome from a one-off general-to-simple selection with IIS after an  $l(0)$  reduction.

## 6 Conditional forecasting

Given that a model with good economic and econometric properties has been estimated, we now assess its ex post ‘forecasting’ ability. Using data up to 1952 to estimate the EqCM with IIS, forecasts of  $\Delta e_f$  were generated up to 2002, conditional on observed values of the explanatory variables. The estimated

sub-sample model was:

$$\begin{aligned} \Delta e_{f,t} = & \underset{(0.06)}{0.11} \Delta e_{f,t-1} + \underset{(0.06)}{0.62} \Delta e_t - \underset{(0.05)}{0.28} \Delta(p_f - p)_t + \underset{(0.04)}{0.24} \Delta s_t \\ & - \underset{(0.03)}{0.35} c_{4,t-1} - \underset{(0.01)}{0.11} I_{31} - \underset{(0.02)}{0.109} I_{32} + \underset{(0.01)}{0.02} I_{34} - \underset{(0.01)}{0.03} I_{43} \quad (10) \\ (R^*)^2 = & 0.98 \quad \hat{\sigma} = 0.0098 \quad F_{ar}(2, 11) = 0.37 \\ \chi_{nd}^2(2) = & 2.42 \quad F_{arch}(1, 20) = 0.24 \quad F_{reset}(2, 10) = 0.70 \quad F_{het}(10, 7) = 0.22 \end{aligned}$$

Apart from two observations in the early 1970s (reflected in the retention of an indicator in (9)), parameter constancy is clear—see Figure 9—albeit taking  $c_4$  as given. The 1-step (ex post) ‘forecasts’ over the period 1953–2002 perform well, as indicated by the corresponding Chow forecast statistic  $F_{Chow}(50, 13) = 0.81$ . This result is particularly impressive given that many other investigators of these data had to omit the inter-war period as being discrepant, although its manifestly much larger data variation is invaluable in the precision of parameter estimates.

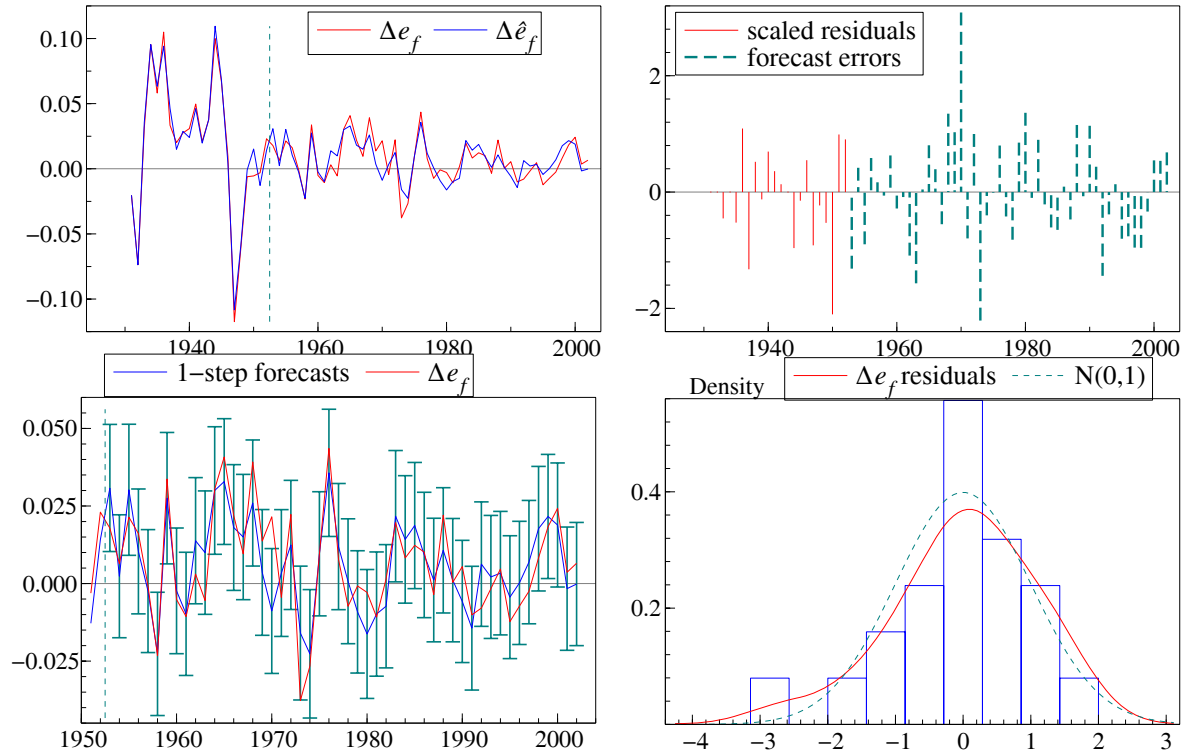


Figure 9: Conditional ‘forecasts’ over 1953–2002 from (10)

### 6.1 ‘Forecasting’ 2003–2006

However, the model is less impressive in generating conditional forecasts from 2002 of the 4 observations 2003–2006, as is evidenced by Figure 10, and the large Chow statistic  $F_{Chow}(4, 62) = 10.95^{**}$ .

This inability of the model to produce reasonable forecasts of the data for 2003–2006 illustrates that change is ever with us, so apparently good models can break down at any time. The fact that in the presence of unanticipated changes, the pre-existing conditional expectation is no longer the minimum mean squared error predictor is proved in Hendry and Mizon (2009). This not only entails it is imperative that modellers quickly identify a new regime’s presence and characteristics, and develop a model thereof,

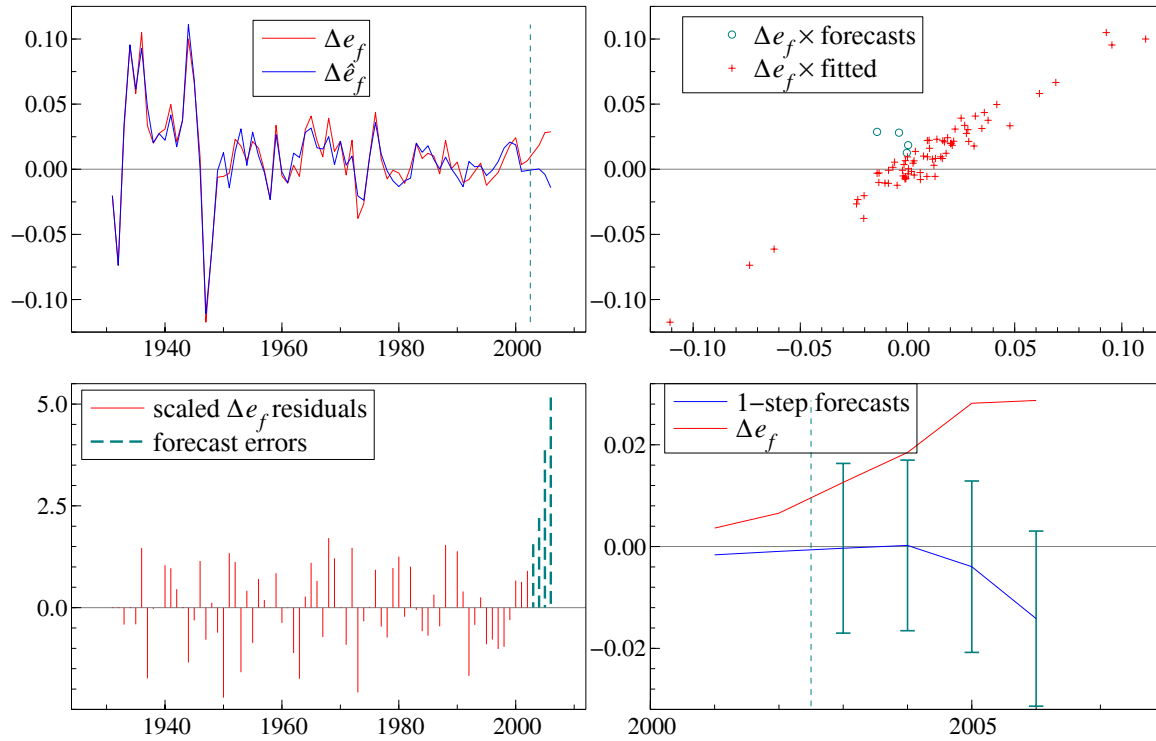


Figure 10: Forecasts over 2003–2006 from (9)

but also that rational expectations theory-led models cannot be invariant when the underlying process changes.

That economies are subject to unanticipated change, and that this can make forecasting difficult, is well documented. Recent research has explained the ability of robust forecasting methods to out-perform econometric models, and hence can be used to improve the latter's performance (see e.g., Clements and Hendry, 1998, 1999, Hendry, 2006, and Castle and Hendry, 2008). Drawing on this literature, we generate robust forecasts by using the coefficient values from the congruent parsimonious model (9), expressed in the differences of all variables:

$$\begin{aligned} \Delta e_{f,t} &= \Delta e_{f,t-1} + 0.13\Delta^2 e_{f,t-1} + 0.59\Delta^2 e_t + 0.23\Delta^2 s_t - 0.36\Delta c_{4,t-1} - 0.32\Delta^2(p_f - p)_t \\ F_{\text{Chow}}(4, 71) &= 0.24 \end{aligned} \quad (11)$$

The results recorded in Figure 11 show no evidence of forecast failure.

Differencing reduces step shifts to impulses, and hence offsets systematic mis-forecasting following a location shift. However, note that the forecasts, based on identical parameter values but applied to the differences of the original variables, now show a rise, whereas previously (e.g., Figure 10), they showed a fall. Hendry (2006) explains why differencing is so successful, not only here but in many other settings (see e.g., Hendry, 2004). After unmodelled parameter changes, an estimated EqCM will still use previous parameter values, and so mis-forecast; but the data must reflect the parameter shifts, and hence so must the change in the dependent variable, making it an excellent robust forecasting device, albeit one period later. In algebraic terms, for an  $n$ -dimensional vector of variables  $\mathbf{x}_t$  determined by a VEQCM:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\lambda}(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t$$

when the parameters shift over a forecast period from  $T$  to (say):

$$\Delta \mathbf{x}_{T+1} = \boldsymbol{\gamma}^* + \boldsymbol{\lambda}^*((\boldsymbol{\beta}^*)'\mathbf{x}_T - \boldsymbol{\mu}^*) + \boldsymbol{\epsilon}_{T+1}$$

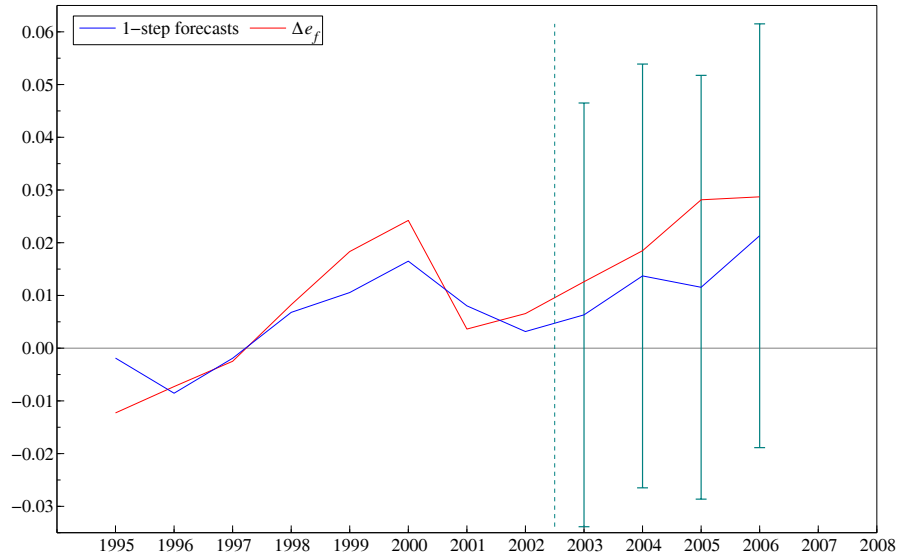


Figure 11: Robust forecasts over 2003–2006 from (11)

but stay constant at these new values over  $T + 2$ :

$$\Delta \mathbf{x}_{T+2} = \gamma^* + \lambda^* ((\beta^*)' \mathbf{x}_{T+1} - \mu^*) + \epsilon_{T+2} = \Delta \mathbf{x}_{T+1} + [\lambda^* ((\beta^*)' \Delta \mathbf{x}_{T+1}) + \Delta \epsilon_{T+2}] \quad (12)$$

then  $\Delta \mathbf{x}_{T+1}$  ‘contains’ all the information about the shifts—without any modelling or estimation—other than  $I(-1)$  noise, shown in brackets. This analysis suggests that (11) should also be able to ‘forecast’ the entire post-war period, and that is the case, as Figure 12 shows: indeed, the forecast errors are smaller than the in-sample residuals.

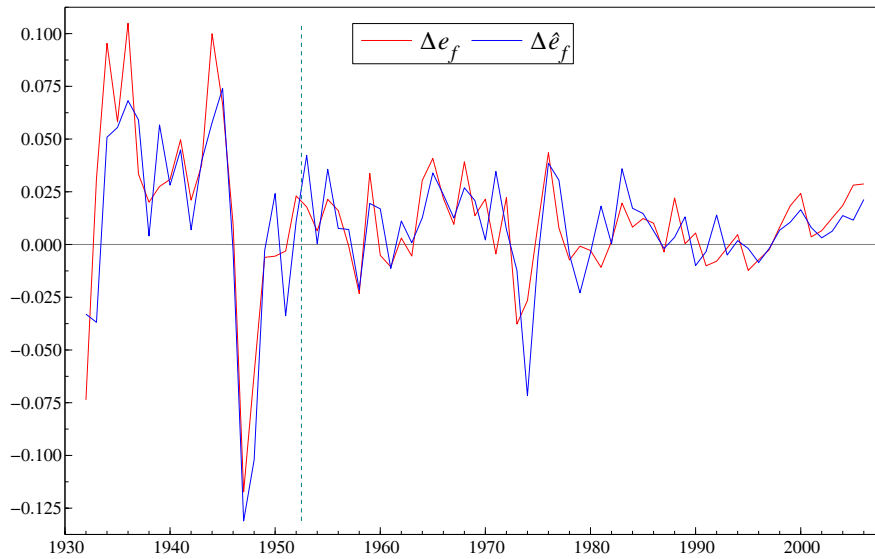


Figure 12: Robust forecasts over 1953–2006 from (11)

As an alternative, we also selected the model with IIS over the whole sample 1931–2006:

$$\begin{aligned}
\Delta e_{f,t} = & \frac{0.14}{(0.04)} \Delta e_{f,t-1} + \frac{0.59}{(0.04)} \Delta e_t - \frac{0.32}{(0.03)} \Delta(p_f - p)_t + \frac{0.23}{(0.04)} \Delta s_t \\
& - \frac{0.35}{(0.02)} c_{4,t-1} - \frac{0.11}{(0.01)} I_{31} - \frac{0.11}{(0.01)} I_{32} + \frac{0.03}{(0.01)} I_{34} \\
& - \frac{0.03}{(0.01)} I_{43} + \frac{0.03}{(0.01)} I_{70} + \frac{0.03}{(0.01)} I_{05} + \frac{0.04}{(0.01)} I_{06} \tag{13}
\end{aligned}$$

$$\begin{aligned}
(R^*)^2 &= 0.94 \quad \hat{\sigma} = 0.0086 \quad F_{\text{ar}}(2, 62) = 0.05 \\
\chi_{\text{nd}}^2(2) &= 0.64 \quad F_{\text{arch}}(1, 74) = 0.21 \quad F_{\text{reset}}(2, 62) = 0.43 \quad F_{\text{het}}(10, 58) = 0.96
\end{aligned}$$

Hence, the results are almost the same as the estimates in (10), but with two extra large indicators in 2005 and 2006 relative to that model. These results show that the final few observations require careful analysis in order to understand the changes that have taken place. Possible explanations include:

- (a) increased measurement errors at the forecast origin—some of the data analyzed here differ considerably from the earlier measures (see Hendry, 2009);
- (b) increased income inequality makes per capita measures less satisfactory, especially as the poor spend a larger proportion of their income on food;
- (c) increased credit facilities allowed expenditure to exceed normal levels, especially for poorer families previously unable to access it.

The data for 2008–2009 will allow the last to be tested, but are not available as yet.

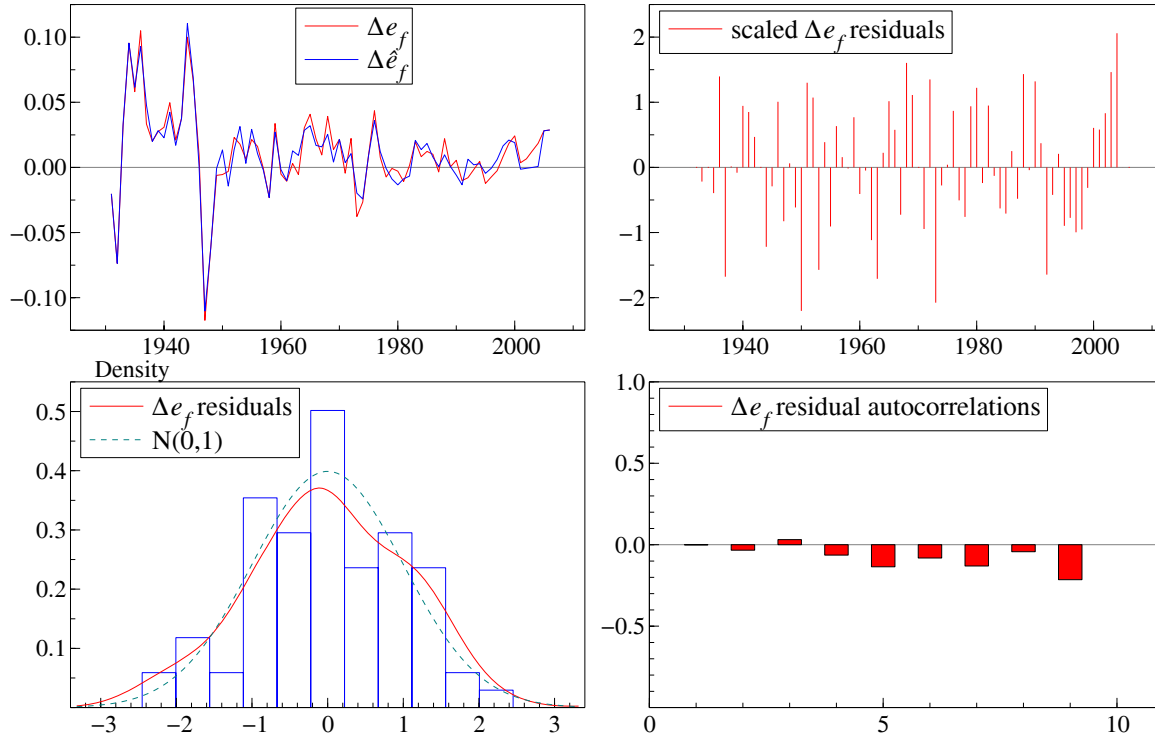


Figure 13: Fit of (13) over 2003–2006



## 7 Conclusion

The focus of this paper has been the difficulties for economic modelling that arise when the underlying economic generation processes change. The recent financial crisis and resulting global recession are vivid reminders that unexpected changes can have significant effects on both economies and econometric models thereof. Unanticipated changes cause problems for forecasting, and result in estimated models that cast serious doubt on the economic theories on which they are based, when in fact this doubt might be removed by appropriate treatment of discrepant observations. It is also possible that an economic analysis might have to be modified in the light of important new information.

We have revisited a model of expenditure on food in the USA using an extended dataset covering the period 1931–2002, and illustrated the fact that even though a theory is essentially ‘correct’, it can exhibit serious signs of mis-specification if it is just fitted to the data without paying attention to the observed time-series characteristics of the data and significant external events such as wars and policy interventions. However, when the theory is embedded in a more general framework embracing dynamics, outliers and structural change, using automatic model selection, then the original theory may perform well even over an extended data period during which there has been substantial change. Generally, a theory will be retained when valid, but can be rejected if invalid; rescued when a direct fit would have rejected, yet extended if initially incomplete: little is lost and much is gained by embedding theory models in general formulations. Although our particular illustration involves a simple theory, the point being made is generic and applies no matter how sophisticated the theory.

We also conclude that model selection is beneficial in detecting and removing structural shifts and outliers, yet innocuous in eliminating insignificant variables, such that here the theory implications were always improved by commencing from the general model with impulse-indicator saturation. Thus, this approach allows all available theory information to be incorporated, evaluated, and retained where it is helpful empirically, while revealing what aspects need modification or even elimination. Despite some claims that there is an inherent ‘trade-off’ between economic theory consistency and data congruence (see e.g., Pagan, 2003), we believe there need be no such conflict once a general modelling methodology is adopted, within which the best available theory is embedded as the above analysis illustrated.

Finally, robust forecasts can function well even in the face of further unexplained location shifts, although the causes of such shifts merit deeper investigation.

## References

- Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.
- Banerjee, A., and Hendry, D. F. (1992). Testing integration and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, **54**, 225–255.
- Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992). Recursive and sequential tests of the unit root and trend break hypothesis. *Journal of Business and Economic Statistics*, **10**, 271–288.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2009). Model selection when there are multiple breaks. Working paper 472, Economics Department, University of Oxford.
- Castle, J. L., and Hendry, D. F. (2008). Forecasting UK inflation: Empirical evidence on robust forecasting devices. In Rapach, D. E., and Wohar, M. E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty: Frontiers of Economics and Globalization Volume 3*, pp. 41–92. Bingley, UK: Emerald Group.
- 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.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- Colander, D. (ed.) (2006). *Post Walrasian Macroeconomics*. Cambridge: Cambridge University Press.
- Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, **70**, 915–925.
- Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.
- Doornik, J. A. (2009b). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.
- Doornik, J. A., and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, **70**, 927–939.
- Doornik, J. A., and Hendry, D. F. (2009). Empirical model discovery. Monograph, Economics Department, Oxford University.
- Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, **12**, 533–572.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- Engsted, T. (2009). Statistical vs. economic significance in economics and econometrics: Further comments on McCloskey and Ziliak. Research paper 2009-17, CREATES, Aarhus University.
- Ericsson, N. R., and MacKinnon, J. G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal*, **5**, 285–318.
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1999). An econometric analysis of US food expenditure, 1931–1989. In Magnus, and Morgan (1999), pp. 341–361.
- Hendry, D. F. (2004). Forecasting long-run TV advertising expenditure in the UK. Commissioned report, Ofcom, London. <http://www.ofcom.org.uk/research/tv/reports/tvadvmarket.pdf>.
- Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.
- Hendry, D. F. (2008). Equilibrium-correction models. In Durlauf, S., and Blume, L. (eds.), *New Palgrave Dictionary of Economics, 2nd Edition*, p. doi:10.1057/9780230226203.0496. London: Palgrave Macmillan.
- Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson, K. D. (eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.
- Hendry, D. F., and Doornik, J. A. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.
- Hendry, D. F., 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 Mizon, G. E. (2009). Economic policy modelling in a rapidly changing world.

- Revised version of paper given at Conference in Honour of Svend Hylleberg, August 2009. Mimeo, Division of Economics , University of Southampton.
- Hendry, D. F., and Neale, A. J. (1991). A Monte Carlo study of the effects of structural breaks on tests for unit roots. In Hackl, P., and Westlund, A. H. (eds.), *Economic Structural Change, Analysis and Forecasting*, pp. 95–119. Berlin: Springer-Verlag.
- 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. (2006). Confronting the economic model with the data. In Colander (2006), pp. 287–300.
- 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.
- Juselius, K., and Johansen, S. (2006). Extracting information from the data: A European view on empirical macro. In Colander (2006), pp. 301–334.
- Kiiveri, H. (2006). Fitting statistical models to data with many more variables than observations. Working paper, CSIRO CMIS, Australian National University, Canberra.
- Magnus, J. R., and Morgan, M. S. (eds.)(1999). *Methodology and Tacit Knowledge: Two Experiments in Econometrics*. Chichester: John Wiley and Sons.
- Mizon, G. E. (1995). Progressive modelling of macroeconomic time series: the LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, pp. 107–169. Dordrecht: Kluwer Academic Press.
- Nielsen, H. B. (2004). Cointegration analysis in the presence of outliers. *Econometrics Journal*, **7**, 249–271.
- Pagan, A. R. (2003). Report on modelling and forecasting at the Bank of England. *Bank of England Quarterly Bulletin*, Spring.
- Perron, P. (1989). The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, **31**, 350–371.
- Reade, J. J. (2008). Updating Tobin’s food expenditure time series data. Working paper, Department of Economics, University of Oxford.
- Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, **4**, 393–397.
- Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, **58**, 113–144.
- Spanos, A. (1995). On theory testing in econometric modelling with non-experimental data. *Journal of Econometrics*, **67**, 189–226.
- Stodden, V. (2008). Model selection with many more variables than observations. Working paper, Microsoft Research Asia, Stanford University.
- Tobin, J. (1950). A statistical demand function for food in the U.S.A.. *Journal of the Royal Statistical Society, A*, **113**(2), 113–141.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.