



An historical perspective on the forecasting performance of the Treasury Model: Forecasting the growth in UK consumers' expenditure

S Cook

► To cite this version:

S Cook. An historical perspective on the forecasting performance of the Treasury Model: Forecasting the growth in UK consumers' expenditure. Applied Economics, 2011, pp.1. 10.1080/00036846.2010.510465 . hal-00665455

HAL Id: hal-00665455

<https://hal.science/hal-00665455>

Submitted on 2 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**An historical perspective on the forecasting performance of
the Treasury Model: Forecasting the growth in UK
consumers' expenditure**

Journal:	<i>Applied Economics</i>
Manuscript ID:	APE-2009-0677.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	14-Jul-2010
Complete List of Authors:	Cook, S; University of Wales Swansea, Economics
JEL Code:	C53 - Forecasting and Other Model Applications < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, C52 - Model Evaluation and Testing < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods
Keywords:	Forecasting, Forecast evaluation

**An historical perspective on the forecasting performance
of the Treasury Model: Forecasting the growth in UK
consumers' expenditure***

Steve Cook

July 14, 2010

Abstract

Drawing upon *Treasury Official Economic Forecasts Vols. I & II*, a series of Treasury Model forecasts of the percentage growth in real total consumers' expenditure are derived for the period 1967 to 1989. The one-, two- and three-step ahead forecasts examined cover an interesting period which includes major shocks to the UK economy, business cycle effects and changes in economic policy. Whilst a battery of forecast evaluation statistics and tests do not detect any evidence of forecast bias or irrationality over the whole sample, split-sample analysis provides evidence of a switch from overprediction to underprediction around 1977. In addition, the application of 'modified' versions of Holden-Peel (1990) tests provides evidence of the longest horizon forecasts failing to capture the full movement of changes in consumption growth. Using simple regression and a selection of forecast encompassing tests, shorter horizon forecasts are found to dominate longer horizon forecasts, a feature which might be expected logically, but need not occur in practice. Finally, forecast performance is related to changes in model specification and modelling methodology.

Key words: Macroeconomic forecasting; Consumers' expenditure; Forecast evaluation; Forecast bias; Forecast encompassing.

* This paper is a substantially revised version of Cook (1995) which has been cited previously in the forecasting literature (see, *inter alia*, Hendry and Doornik 1997; Hendry 1997; Clements and Hendry 1998; Hendry and Clements 2000; Hendry and Ericsson 2001). I am extremely indebted to David Hendry for his advice and very many helpful suggestions concerning the research undertaken herein. I am grateful also to Mike Clements, Grayham Mizon and Simon Wren-Lewis for additional comments.

1 Introduction

Given its obvious importance for economic policy, the forecasting performance of the Treasury Model (TM) has been the focus of attention of perhaps a relatively limited number of research articles.¹ While Burns (1986) and Mills and Pepper (1999) have considered the accuracy of the TM in terms of forecasting the growth in GDP and inflation in the UK, Church *et al.* (1994) examined the ability of the TM to forecast consumers' expenditure via an analysis of the properties of one-step ahead forecasts of non-durable consumption over the period from 1989 to 1992.² The present paper aims to complement this latter work of Church *et al.* (1994) by providing a more detailed analysis of the forecasting of consumers' expenditure by the TM. Drawing upon *Treasury Official Economic Forecasts Vols. I & II*, a series of one-, two- and three-step ahead TM forecasts of the percentage growth in real total consumers' expenditure have been derived for the period 1967 to 1989. Construction of this data set therefore provides the opportunity to examine forecasts of differing horizons over an interesting sample period capturing changes in Government and economic policy, booms, slumps and major economic shocks, thereby allowing a detailed analysis of the forecasting of consumption by the TM.

To achieve its aims, this paper will proceed as follows. In Section [2] the nature of the data to be examined are discussed. Section [3] provides an initial examination of the properties of the alternative forecasts via the calculation of a battery of forecast evaluation statistics. The issue of forecast bias is examined in section [4]. Initially this is analysed using Holden-Peel (1990) tests of unbiasedness before presenting a split sample analysis of forecast evaluation statistics for the differing horizon forecasts. This section also contains tests of partial rationality and 'modified' Holden-Peel tests to further examine the properties of the forecasts. In section [5] the relative properties of differing horizon forecasts are examined directly via the application of alternative tests of forecast encompassing. Although shorter horizon forecasts might logically be expected to encompass longer horizon forecasts, the recent research of Clements and Hendry (2003, 2008) recognises this might not occur in practice where the idealised conditions of the textbook can be violated by, *inter alia*, data revision and mismeasurement, location shifts, unknown model misspecification and intercept

¹This relative lack of attention is perhaps unsurprising given the asymmetry in information available to those inside and outside of the Treasury. As an illustration of this, Burns (1986, p.105) notes that his research uses "mainly unpublished, internal, forecasts, although some use is made of forecasts published at budget time".

²The TM was one of a number of macroeconomic models considered by Church *et al.* (1994) in their examination of the forecasting of UK consumers' expenditure.

correction. Section [6] provides some concluding remarks.

2 Data

The data considered in this study are actual values (a_t) and forecasts (f_t) of the growth in the real value of total consumers' expenditure over the period 1967 to 1989. More precisely, the figures are annual growth rates which consider both growth from the first half of one year to the first half of the following year (e.g. 1968.1 to 1969.1) and the second half of one year to the second half of the following year (e.g. 1967.2 to 1968.2). Given the timing of the preparation of the forecasts, this results in one- and three-step ahead forecasts (denoted as $f_{t-1,t}$ and $f_{t-3,t}$ respectively) being made for annual growth between the first halves of adjacent years, and two-step ahead forecasts (denoted as $f_{t-2,t}$) for growth between the second halves of adjacent years. Consequently, while direct comparison can be made between $f_{t-1,t}$ and $f_{t-3,t}$ as they provide forecasts for the same period, any comparison with $f_{t-2,t}$ has to acknowledge that this series is producing forecasts for different periods.

Before analysing the forecasting ability of the TM in detail, the behaviour of consumption growth over the period in question can be examined. Looking at the full set of growth rates from 1967.2-1968.2 to 1988.2-1989.2 provided in Figure One below, it can be seen that there are three main peaks in the growth of consumption (1971.2-1972.2, 1978.1-1979.1 and 1987.1-1988.1) each of which is followed almost immediately by a sharp fall. Relatedly, three noteworthy episodes are (i) the slump of the mid to late seventies, (ii) the slump of the early eighties and (iii) the boom of the mid to late eighties. This can all be highlighted with reference to the average growth rate throughout the period of 2.6%, as from 1974.1 to 1977.2 the growth rate averaged -0.4% , from 1980.1 to 1982.1 it averaged 0.3% and from 1986.1 to 1989.1 it averaged 5.4% . However, while volatility may prove problematic when forecasting, there is a distinction between volatility and predictability. In particular, when considering the current sample, the boom of the late 1980s provided a period which saw the growth in consumption outstripping that of income, a marked increase in consumer debt relative to income and an asset-price boom in the stock and housing markets (Muellbauer, 1994). These three factors represented wholly new phenomena which provided difficulties for existing model specifications as the possibility of their occurrence had not been entertained by modellers.

Plots of $\{f_{t-1,t}, f_{t-3,t}\}$ and $\{f_{t-2,t}\}$ against their corresponding actual values of consumption growth (a_t) are provided in Figures Two and Three respectively. In a similar fashion, the one-, two-

and three-step forecast errors ($e_{t-i,t} = y_t - f_{t-i,t}$ for $i = 1, 2, 3$) can be plotted to gauge the extent of errors made when forecasting the growth in consumption. Considering the behaviour of $\{e_{t-1,t}, e_{t-3,t}\}$ depicted in Figure Four, it appears that a tendency exists in both series towards overprediction in the period to 1977 and underprediction thereafter. Figure Four also suggests that in addition to one-step ahead predictions faring better than the three-step ahead predictions, both sets of forecasts manage to miss the major changes in consumption growth. For the three-step forecasts the worst individual errors occur for 1973.1-1974.1, 1977.1-1978.1, 1982.1-1983.1 and 1987.1-1988.1. The worst series of errors for the three-step forecasts occurs in trying to predict consumption growth from 1973.1-1974.1 to 1978.1-1979.1, where an average absolute error of 2.9% is made. This compares to an average absolute actual value of 2.8% for consumption growth over the same period. For the one-step ahead predictions the worst error is -2.3% (-1.1%-3.4%) for 1974.1-1975.1. Figure Five provides a similar depiction for the errors associated with the two-step forecasts suggesting a similar movement from overprediction to underprediction around 1977. It can be noted that these findings concerning over- and underprediction for consumption are line with those presented by Burns (1986) for GDP. Figure Five also shows that the main forecasting errors occur in 1967.2-1968.2, 1974.2-1975.2, 1983.2-1984.2, 1986.2-1987.2 and 1987.2-1988.2.

Figures One to Five about here

While the above figures provide a snapshot of the alternative sets of forecasts and their closeness, or otherwise, to the actual consumption data, a more detailed analysis can be obtained via consideration of various forecast evaluation statistics. This is provided in the following section.

3 Forecast evaluation statistics

To examine the basic properties of the alternative for differing horizons, Table One presents the calculated values of a range of forecast evaluation statistics. The initial rows in Table One present familiar, straightforward evaluation statistics based upon the forecast errors. These statistics are the maximum error (Max E), minimum error (Min E), mean error (ME), mean absolute error (MAE), mean square error (MSE), mean percentage error (MPE) and mean absolute percentage error (MAPE).

The following row contains calculated values Theil's Inequality Statistic, U , which is given as:³

³The U statistic is bounded between 0 and 1, and has a straightforward interpretation with 0 denoting perfect forecasting and 1 denoting the worst possible forecasting performance.

$$U = \frac{\sqrt{\sum_{t=1}^n [(a_t - f_{t-i,t})]^2}}{\sqrt{\sum_{t=1}^n f_{t-i,t}^2} + \sqrt{\sum_{t=1}^n a_t^2}} \quad (1)$$

Considering the results presented in Table One, an immediately striking feature is the consistency between the differing horizon forecasts, with a general worsening of forecasting performance apparent as the forecast horizon is increased. It is also clear that the apparent small values obtained for the unscaled evaluation statistics (e.g. ME, MAE, MSE) need to be interpreted with care as the series being forecast is a percentage growth rate which takes small values. To reinforce this issue, some summary statistics for a_t are provided in the lower portion of Table One. In light of this, the scaled evaluation statistics are of greater interest, and these provide a less pleasing view of the forecasting performance of the TM. For example, the mean absolute percentage errors for $\{f_{t-1,t}, f_{t-2,t}, f_{t-3,t}\}$ take values of 41.36%, 79.88% and 96.01%. A similar pattern is present in the calculated values for U which increase with the forecast horizon.

Table One about here

4 Examining forecasting performance

4.1 Forecasting bias

The previous graphical analysis suggested possible forecasting bias. To examine this more formally, the tests of unbiasedness of Holden and Peel (1990) can be employed. To illustrate these tests, consider the following testing equation:

$$a_t = \alpha + \beta f_{t-i,t} + \varepsilon_t \quad (2)$$

A familiar test for unbiasedness is given by the joint hypothesis, $H_0 : \alpha = 0, \beta = 1$. However, Holden and Peel (1990) show that this is a sufficient but not necessary condition for $f_{t-i,t}$ to be an unbiased forecast of a_t , as the forecast could still be unbiased when $\beta \neq 1$ if $\alpha = (1 - \beta)E[f_{t-i,t}]$. Holden and Peel propose the necessary and sufficient condition, $H_0 : \delta = 0$, in:

$$a_t - f_{t-i,t} = \delta + \nu_t \quad (3)$$

Before applying this test to the alternative forecasts series, the orders of integration of a_t and $f_{t-i,t}$ are examined to ensure they are stationary and hence avoid problems of spurious regression. The results obtained from application of the Elliott *et al.* (1996) GLS-based Dickey-Fuller unit root test and the Kwiatkowski *et al.* (KPSS) (1992) stationarity test are presented in Table Two. The results presented show that while the GLS Dickey-Fuller rejects the unit root null for all series, the null of stationarity is not rejected by the KPSS test at conventionally considered levels of significance. It is therefore concluded that all series can be treated as stationary processes. In Table Three results are presented from application of the Holden-Peel tests to the alternative forecast series. In addition to applying the tests to the full samples available for each set of forecasts, the tests are also applied to sub-samples up to and beyond 1977. This split-sample analysis is motivated by the previously noted possibility of a change in forecasting bias around this period. In all cases the Newey-West (1994) variance-covariance matrix estimator is employed when estimating the Holden-Peel regressions of (4) to overcome potential serial correlation (see Clements 1995).⁴ The results obtained from the Holden-Peel tests are presented in Table Three. Considering these results, it is clear that over the whole sample there is no evidence of significant bias. However, turning to the split-sample results, the previously noted movement from overprediction to underprediction is supported to some extent with the one-step forecasts exhibiting marginally significant overprediction in the earlier sample and the two- and three-step forecasts displaying significant underprediction in the later sample period. Interestingly, movement from two- to three-step forecasts results in increased and more significant evidence of underprediction being detected in the later sample via the estimated value of δ in (4) and its standard error. This reversal of bias reflects the findings of Burns (1986) for the growth in GDP, but contrasts with his findings for inflation where a movement in the opposite direction from underprediction to overprediction was detected.

Tables Two and Three about here

To supplement the above discussion concerning the potential differing properties of forecasts either side of 1977, the forecast evaluation statistics of Table One are presented for the pre- and post-1977 split-samples in Table Four. To summarise the information provided by the range of statistics considered, the movement from overprediction to underprediction around 1977 is present in the calculated

⁴It should be noted that the testing equations for the Holden-Peel tests and all subsequent equations are estimated using the Newey-West covariance matrix estimator.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

mean error statistics for all of the differing horizon forecasts. The results presented also indicate that forecasting performance is worse in the earlier sample with, *inter alia*, higher values of U reported for the earlier sample for all of the sets of forecasts. As noted above, it may have been expected that the latter part of the sample would prove more problematic for forecasting as a result of a number of new and perhaps unexpected phenomena. That the TM should witness an increase in forecasting performance over this period, suggests a marked improvement in underlying forecasting ability.

Table Four about here

4.2 Partial rationality

The above analysis has considered the possibility of bias in the alternative forecasts. However, the presence of unbiasedness is only a necessary condition for partial rationality. To test this latter issue, the work of Pain and Britton (1992) can be drawn upon using the following testing equation:

$$a_t = \alpha + \beta f_{t-i,t} + \gamma a_{t-i} + \varepsilon_t \tag{4}$$

The significance (or otherwise) of $\hat{\beta}$ can be used to compare the forecasts of the TM with forecasts from a simple time-series model based upon the latest value actual value at the time of forecasting. The significance of $\hat{\beta}$ would indicate that the forecast $f_{t-i,t}$ adds information, with $\beta = 1$ being the important hypothesis for unbiasedness. Consideration of the significance of $\hat{\gamma}$ allows testing for omitted factors in the forecast. The results obtained from estimation of (5) for the alternative forecast series are presented in Table Five. From inspection of the results obtained, it can be seen that $\hat{\beta}$ is significantly different from zero for all of the forecast series, while $\hat{\alpha}$ and $\hat{\gamma}$ are insignificant. Testing the hypothesis $H_0 : \beta = 1$ for the differing forecast series via a Wald test leads to p-values of 0.324, 0.494 and 0.371 for the one-, two- and three-step forecasts thus indicating non-rejection of the null in all instances, thereby providing no evidence against partial rationality.

Table Five about here

4.3 ‘Modified’ Holden-Peel tests

The standard Holden-Peel forecasting bias tests above consider the sign and significance of the forecast error. When considering this over the whole sample period, it was found that significant bias was not detected for any of the different horizon forecasts. However, these tests can be modified or extended

to consider the percentage error as a dependent variable. Application of this test leads to the results presented in Table Six. Inspection of these results shows that in contrast to the insignificance of the Holden-Peel tests over the whole sample period, a significant result is obtained for the three-step forecasts. However, given the transformation made to the dependent variable relative to the standard Holden-Peel test, the results derived need to be interpreted carefully. The positive coefficient obtained for the three-step forecast series indicates a tendency for errors and actual values to have the same sign. This then indicates underprediction for positive actual values and overprediction for negative actual values, or alternatively expressed, that the forecasts fail to move fully with movements in the actual series. Interestingly, as noted, this is found only for the longest horizon forecast considered.

Table Six about here

5 Direct comparison of differing horizon forecasts

As noted above, one-step and three-step forecasts are provided for the same data points and as a result their properties can be compared directly. It might be expected that shorter horizon forecasts should be more accurate than longer horizon forecasts. However, following the research of Clements and Hendry (see Clements and Hendry 2008 and references therein for a summary of this work) it is apparent that this may not hold for a number of reasons such as, *inter alia*, the use of intercept corrections (see McNees 1982; Hendry and Clements 1994; Clements and Hendry 1996) which result in forecasts no longer being purely model-based, the prevalence of location shifts in economic data, unknown model misspecification and the impact of data revision.⁵ From inspection of Figure Four it can be seen that $f_{t-3,t}$ is less accurate than $f_{t-1,t}$. As a simple means of examining this, a simple regression can be run using the difference in the absolute values of the forecast errors for the alternative forecast series as a dependent variable and a constant term as a regressor. The results from this regression, again using the Newey-West covariance matrix estimator, are as given below:

$$|e_{t-3,t}| - |e_{t-1,t}| = \underset{[0.001]}{0.976} + \hat{v}_t \quad (5)$$

where the p-value for the significance of the estimated constant term is provided in square brackets.

The difference between the absolute errors is therefore highly significant and also large relative to

⁵For an overview of the nature of data revision in economics and its effects see, *inter alia*, Mankiw *et al.* (1984), Croushore and Stark (2003), Garratt and Vahey (2006) and Cook (2008).

the mean value of 2.6 for the actual series. To further investigate the relative properties of the series, the forecast encompassing tests of Fair and Shiller (1989), Granger and Newbold (1973) and Chong and Hendry (1986) are performed. Each of the above tests is based upon the informational redundancy, or otherwise, of one set of forecasts given the presence of another. As noted by Clements and Harvey (2009), the above forecast encompassing tests can be derived via consideration of the following equation:

$$a_t = \mu + \lambda_1 f_{t-1,t} + \lambda_2 f_{t-3,t} + \zeta_t \quad (6)$$

The simplest of the tests to consider is that of Fair and Shiller (1989) where the forecast encompassing of $f_{t-3,t}$ by $f_{t-1,t}$ (denoted as $f_{t-1,t} \mathcal{FE} f_{t-3,t}$) is examined via the null hypothesis $H_0 : \lambda_2 = 0$ in (7). In a similar fashion, $f_{t-3,t} \mathcal{FE} f_{t-1,t}$ is examined via the null hypothesis $H_0 : \lambda_1 = 0$. To derive the Granger and Newbold (1973) forecasting encompassing test, the restrictions $\mu = 0$ and $\lambda_1 + \lambda_2 = 1$ are imposed in (7) to give:

$$e_{t-1,t} = \gamma (e_{t-1,t} - e_{t-3,t}) + \omega_t \quad (7)$$

with $f_{t-1,t} \mathcal{FE} f_{t-3,t}$ examined via the null hypothesis $H_0 : \gamma = 0$. An analogous test of $f_{t-3,t} \mathcal{FE} f_{t-1,t}$ is provided by swapping the positions of $e_{t-1,t}$ and $e_{t-3,t}$ in (7). The final encompassing test of Chong and Hendry (1986) imposes the restrictions $\mu = 0$ and $\lambda_1 = 0$ in (6) to arrive at the following equation in which $f_{t-1,t} \mathcal{FE} f_{t-3,t}$ is examined via $\delta = 0$:

$$e_{t-1,t} = \delta f_{t-3,t} + \epsilon_t \quad (8)$$

Again encompassing in the opposite direction can be examined with $e_{t-3,t}$ replacing $e_{t-1,t}$ and $f_{t-1,t}$ replacing $f_{t-3,t}$.

Table Seven presents the results obtained from estimation of the relevant forecast encompassing testing equations above using the Newey-West covariance matrix estimator. From inspection of Table Seven it is apparent that very decisive results are obtained. In particular, while it is possible for rejection or non-rejection to occur in both directions thereby leading to inconclusive results, in all instances considered herein, encompassing is unidirectional, with $f_{t-1,t}$ forecast encompassing $f_{t-3,t}$ but the reverse not holding. It can be noted that while the superior informational qualities of shorter horizon forecasts detected here reflects the findings of Burns (1986) for TM forecasts of inflation, it contrasts with findings presented by Burns (1986) for the growth in GDP where it is suggested that

longer horizon forecasts may outperform those over a shorter horizon. However, it should be noted that Burns (1986) does not examine this formally using encompassing tests.

Table Seven about here

6 Concluding remarks

In this paper the Treasury Model's ability to forecast the growth in consumers' expenditure over the period from the second half of 1967 to the second half of 1989 has been examined. The sample examined proved to be an interesting period, capturing several booms and slumps, including the largely unexplained boom of the late 1980s, as well as factors such as the OPEC oil price shocks and the arrival of the Thatcher Government. The work complements existing studies such as those of Burns (1986) and Mills and Pepper (1999) by restricting attention to a single variable, thereby permitting a detailed analysis to be conducted. This analysis included examination of forecasts of different horizons and the consideration of issues such as unbiasedness, rationality and forecast encompassing. The calculation of a battery of forecast evaluation statistics and the application of a variety of tests uncovered a number of interesting results. While inferences of unbiasedness and partial rationality were drawn over the whole sample period, split sample analysis provided evidence of a switch from overprediction to underprediction around 1977. This was detected from both application of Holden-Peel (1990) tests of unbiasedness and forecast evaluation statistics. While further analysis of the properties of the forecasts provided no evidence against rationality, application of 'modified' Holden-Peel tests using the percentage error as a dependent variable uncovered a significant test result for the longest horizon forecast considered. Given the nature of the test employed, this provided evidence of a tendency for underprediction when the actual growth in consumption is positive, and overpredict when it is negative. In short, this indicates a failure to capture the full movement of the actual series when forecasting.⁶ To complete the analysis, the one-step and three-step forecasts for the same data points were compared using a simple regression of absolute forecast errors and a selection of forecast encompassing tests. The results obtained show that while recent research has indicated that longer horizon forecasts need not necessarily encompass shorter horizon forecasts, it was the case for the TM forecasts with the one-step forecasts encompassing the three-step forecasts under all of the tests considered.

⁶This finding could be related to the results of Granger and Newbold (1986, p.131), where it is shown that the variance of a process exceeds the variance of an optimal forecast of it.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A final point which can be noted concerns the changes in model specification and modelling methodology which occurred during the period considered. It is obviously difficult to directly relate any change in forecasting performance to changes in model specification with confidence for a variety of reasons. In particular, it is uncertain to what extent subjective adjustments, or intercept corrections, are applied to model based forecasts. Similarly, the different model specifications are employed over different sample periods and hence the impact of model specification cannot be assessed even in the absence of intercept correction. However, it can be noted that the TM model of consumers' expenditure was heavily influenced by the works of Davidson *et al.* (1978) and Hendry and von Ungern-Sternberg (1980), with the approaches of these papers swiftly adopted. It can be noted also that this period coincides with the increased forecasting accuracy noted in the latter split-sample considered herein. It is therefore tempting to conclude that improved model specification resulted in increased forecasting accuracy.

References

- [1] Burns, T. (1986) 'The interpretation and use of economic predictions', *Proceedings of the Royal Society*, **A407** 103-125.
- [2] Chong, Y. and Hendry, D. (1986) 'Econometric evaluation of macro-economic models', *Review of Economic Studies*, **53**, 671-690.
- [3] Church, K., Smith, P. and Wallis, K. (1994) 'Econometric evaluation of consumers' expenditure equations', *Oxford Review of Economic Policy*, **10(2)**, 71-85.
- [4] Clements, M. (2005) *Evaluating Econometric Forecasts of Economic and Financial Variables*, Basingstoke: MacMillan.
- [5] Clements, M. and Harvey, D. (2009) "Forecast combination and encompassing", Mills, T. and Patterson, K. (eds), *Handbook of Econometrics, Volume 2: Applied Econometrics*, Basingstoke: MacMillan (forthcoming).
- [6] Clements, M. and Hendry, D. (1994) 'Towards a theory of economic forecasting', in Hargreaves, C. (ed) *Non-Stationary Time Series Analyses and Cointegration*, Oxford: Oxford University Press.
- [7] Clements, M. and Hendry, D. (1996) 'Intercept corrections and structural change', *Journal of Applied Econometrics*, **11**, 475-494.
- [8] Clements, M. and Hendry, D. (1998) *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.
- [9] Clements, M. and Hendry, D. (2003) 'Economic forecasting: Some lessons from recent research', *Economic Modelling*, **20**, 301-329.
- [10] Clements, M. and Hendry, D. (2008) 'Economic forecasting in a changing world', *Capitalism and Society*, 3, issue 2, article 1.
- [11] Cook, S. (1995) 'Treasury economic forecasting', *mimeo*, Institute of Economics and Statistics, University of Oxford.
- [12] Cook, S. (2008) 'Cross-data-vintage encompassing', *Oxford Bulletin of Economics and Statistics*, **70**, 849-865.

[13] Croushore, D. and Stark, T. (2003) ‘A real-time data set for macroeconomists: Does the data vintage matter?’, *Review of Economics and Statistics*, **85**, 605-617.

[14] Davidson, J., Hendry, D., Srba, F. and Yeo, S. (1978) ‘Econometric modelling of the aggregate time series relationship between consumers’ expenditure and income in the United Kingdom’, *Economic Journal*, **88**, 661-692.

[15] Elliott, G., Rothenberg, T. and Stock, J. (1996) ‘Efficient tests for an autoregressive unit root’, *Econometrica*, **64**, 813-836.

[16] Fair, R. and Shiller, R. (1989) ‘The informational content of *ex ante* forecasts’, *Review of Economics and Statistics*, **71**, 325-331.

[17] Garratt, A. and Vahey, S. (2006) ‘UK real-time macro data characteristics’, *Economic Journal*, **116**, F119-F135.

[18] Granger, C. and Newbold, P. (1973) ‘Some comments on the evaluation of economic forecasts’, *Applied Economics*, **5**, 35-47.

[19] Granger, C. and Newbold, P. (1986) *Forecasting Economic Time Series* (second edition), San Diego: Academic Press.

[20] Hendry, D. and Clements, M. (1994) ‘On a theory of intercept corrections in macroeconomic forecasting’, in Holly, S. (ed) *Money, Inflation and Employment*, Aldershot: Edward Elgar.

[21] Hendry, D. and Clements, M. (2000) ‘Economic forecasting in the face of structural breaks’, in Holly, S. and Weale, M. (eds), *Econometric Modelling: Techniques and Applications*, Cambridge: Cambridge University Press.

[22] Hendry, D. and von Ungern-Sternberg, T. (1980) ‘Liquidity and inflation effects on consumers’ behaviour’, in Deaton, A.S. (ed), *Essays in the Theory and Measurement of Consumers’ Behaviour*, Cambridge: Cambridge University Press.

[23] Holden, K. and Peel, D. (1990) ‘On testing for unbiasedness and efficiency of forecasts’, *Manchester School*, **58**, 120-127.

- [24] Kwiatkowski D., Phillips P., Schmidt P. and Shin, Y. (1992) 'Testing the null of stationarity against the alternative of a unit root', *Journal of Econometrics*, **54**, 159-178.
- [25] Mankiw, G., Runkle, D. and Shapiro, M. (1984) 'Are preliminary announcements of the money stock rational forecasts?', *Journal of Monetary Economics*, **14**, 15-27.
- [26] McNees, S. (1982) 'The role of macroeconomic models in forecasting and policy analysis in the United States', *Journal of Forecasting*, **1**, 37-48.
- [27] Mills, T. and Pepper, G. (1999) 'Assessing the forecasters: An analysis of the forecasting records of the Treasury, the London Business School and the National Institute', *International Journal of Forecasting*, **15**, 247-257.
- [28] Muellbauer, J. (1994) 'The assessment: Consumer expenditure', *Oxford Review of Economic Policy*, **10(2)**, 1-41.
- [29] Newey, W. and West, K. (1994) 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies*, **61**, 631-653.
- [30] Pain, N. and Britton, A. (1992) 'National Institute Economic Forecasts 1968 to 1991: Some tests of forecast properties', *National Institute Economic Review*, **141**, 81-93.

Table One: Forecast evaluation statistics

	$f_{t-1,t}$	$f_{t-2,t}$	$f_{t-3,t}$
Max E	1.50	3.10	4.40
Min E	-2.30	-2.60	-5.60
ME	-0.09	0.46	0.45
MAE	0.80	1.26	1.78
MSE	0.98	2.41	5.16
MPE	-7.65	26.08	76.95
MAPE	41.36	79.88	96.01
U	0.14	0.24	0.34
Max (a_t)	7.10	6.70	7.10
Min (a_t)	-1.70	-2.30	-1.70
Mean (a_t)	2.62	2.57	2.62
Standard Deviation (a_t)	2.75	2.29	2.75

Notes: The upper portion of the above table contains calculated values of the forecast evaluation statistics of Section [3] for the forecast errors relating to the alternative (differing horizon) forecasts $f_{t-i,t}$. For comparative purposes, the lower portion of the table contains some summary statistics for the actual values (a_t) (annual percentage consumption growth) being forecast.

Table Two: Unit root tests

	<u>1968.1-1969.1 to 1988.1-1989.1</u>			<u>1967.2-1968.2 to 1988.2-1989.2</u>	
	a_t	$f_{t-1,t}$	$f_{t-3,t}$	a_t	$f_{t-2,t}$
τ_{μ}^{GLS}	-3.100**	-3.483**	-2.458*	-3.435**	-2.222*
η_{μ}	0.287	0.192	0.104	0.222	0.096

Notes: The above table contains calculated test statistics for the Elliott *et al.* (1996) (τ_{μ}^{GLS}) GLS-based Dickey-Fuller test and the KPSS (η_{μ}) test. Significance at the 5% and 1% levels of significance are denoted by * and ** respectively. The relevant 10%, 5% and 1% critical values for the η_{μ} test are 0.347, 0.463 and 0.739 respectively.

Table Three: Holden-Peel tests			
	Full sample	Sample I	Sample II
One-step forecasts:	$e_{t-1,t} = -0.086_{[0.730]} + \hat{u}_t$	$e_{t-1,t} = -0.667_{[0.060]} + \hat{u}_t$	$e_{t-1,t} = 0.350_{[0.138]} + \hat{u}_t$
Two-step forecasts:	$e_{t-2,t} = 0.459_{[0.234]} + \hat{v}_t$	$e_{t-2,t} = 0.060_{[0.912]} + \hat{v}_t$	$e_{t-2,t} = 0.892_{[0.049]} + \hat{v}_t$
Three-step forecasts:	$e_{t-3,t} = 0.452_{[0.350]} + \hat{w}_t$	$e_{t-3,t} = -0.778_{[0.232]} + \hat{w}_t$	$e_{t-3,t} = 1.375_{[0.020]} + \hat{w}_t$

Notes: The above table contains estimated regressions for the forecast bias test of Holden and Peel (1990). Figures in square brackets represent p-values for the test of coefficient insignificance. For $f_{t-1,t}$ and $f_{t-3,t}$, Sample I contains actual and forecasted consumption growth for 1968.1-1969.1 to 1976.1-1977.1, while Sample II contains data for 1977.1-1978.1 to 1988.1-1989.1. Samples I and II for $f_{t-2,t}$ cover 1967.2-1968.2 to 1976.2-1977.2 and 1977.2-1978.2 to 1988.2-1989.2.

Table Four: Forecast evaluation statistics

	$f_{t-1,t}$		$f_{t-2,t}$		$f_{t-3,t}$	
	Sample I	Sample II	Sample I	Sample II	Sample I	Sample II
Max E	0.30	1.50	3.10	3.00	1.70	4.40
Min E	-2.30	-1.20	-2.60	-2.10	-5.60	-1.10
ME	-0.67	0.35	-0.06	0.89	-0.78	1.38
MAE	0.84	0.77	1.20	1.31	1.89	1.69
MSE	1.26	0.78	2.31	2.49	5.71	4.75
MPE	-18.42	0.97	17.58	33.16	106.63	53.21
MAPE	67.29	20.62	80.60	79.29	138.99	61.63
U	0.17	0.11	0.25	0.24	0.38	0.32
Max (a_t)	6.10	7.10	6.70	6.00	6.10	7.10
Min (a_t)	-1.70	0.00	-2.30	-0.40	-1.70	0.00
Mean (a_t)	1.50	3.46	2.01	2.96	1.50	3.46
Standard Deviation (a_t)	2.89	2.43	2.40	2.22	2.89	2.43

Notes: For $f_{t-1,t}$ and $f_{t-3,t}$, Sample I contains actual and forecasted consumption growth for 1968.1-1969.1 to 1976.1-1977.1, while Sample II contains data for 1977.1-1978.1 to 1988.1-1989.1. Samples I and II for $f_{t-2,t}$ cover 1967.2-1968.2 to 1976.2-1977.2 and 1977.2-1978.2 to 1988.2-1989.2.

Table Five: Partial rationality tests

$$a_t = \underset{[0.105]}{-0.667} + \underset{[0.000]}{1.097} f_{t-1,t} + \underset{[0.161]}{0.114} a_{t-1} + \hat{u}_t$$

$$a_t = \underset{[0.681]}{0.237} + \underset{[0.000]}{0.915} f_{t-2,t} + \underset{[0.444]}{0.114} a_{t-2} + \hat{v}_t$$

$$a_t = \underset{[0.264]}{1.152} + \underset{[0.001]}{0.810} f_{t-3,t} - \underset{[0.499]}{0.156} a_{t-3} + \hat{w}_t$$

Notes: The above table contains estimated regressions for the partial rationality tests of Pain and Britton (1992). Figures in square brackets represent p-values for the test of coefficient insignificance.

Table Six: ‘Modified’ Holden-Peel tests

$$\text{One-step forecasts:} \quad \frac{e_{t-1,t}}{a_t} = \frac{-7.648}{[0.603]} + \hat{\eta}_t$$

$$\text{Two-step forecasts:} \quad \frac{e_{t-2,t}}{a_t} = \frac{26.079}{[0.321]} + \hat{\epsilon}_t$$

$$\text{Three-step forecasts:} \quad \frac{e_{t-3,t}}{a_t} = \frac{76.952}{[0.033]} + \hat{\xi}_t$$

Notes: The above table contains estimated regressions for ‘modified’ Holden-Peel tests using percentage forecast errors, rather than forecast errors, as a dependent variable. Figures in square brackets represent p-values for the test of coefficient insignificance.

Table Seven: Forecast encompassing tests

Fair-Shiller (1989)

$$a_t = -0.406 + 1.077 f_{t-1,t} + 0.052 f_{t-3,t} + \hat{u}_t$$

[0.288] [0.000] [0.746]

Granger-Newbold (1973)

$$e_{t-1,t} = 0.020 (e_{t-1,t} - e_{t-3,t}) + \hat{v}_t$$

[0.868]

$$e_{t-3,t} = 1.020 (e_{t-3,t} - e_{t-1,t}) + \hat{w}_t$$

[0.000]

Chong-Hendry (1986)

$$e_{t-1,t} = 0.026 f_{t-3,t} + \hat{\eta}_t$$

[0.715]

$$e_{t-3,t} = 0.341 f_{t-1,t} + \hat{\epsilon}_t$$

[0.005]

Notes: The above table contains estimated regressions for the forecast encompassing tests of Fair and Shiller (1989), Granger and Newbold (1973) and Chong and Hendry (1986). Figures in square brackets represent p-values for the test of coefficient insignificance.

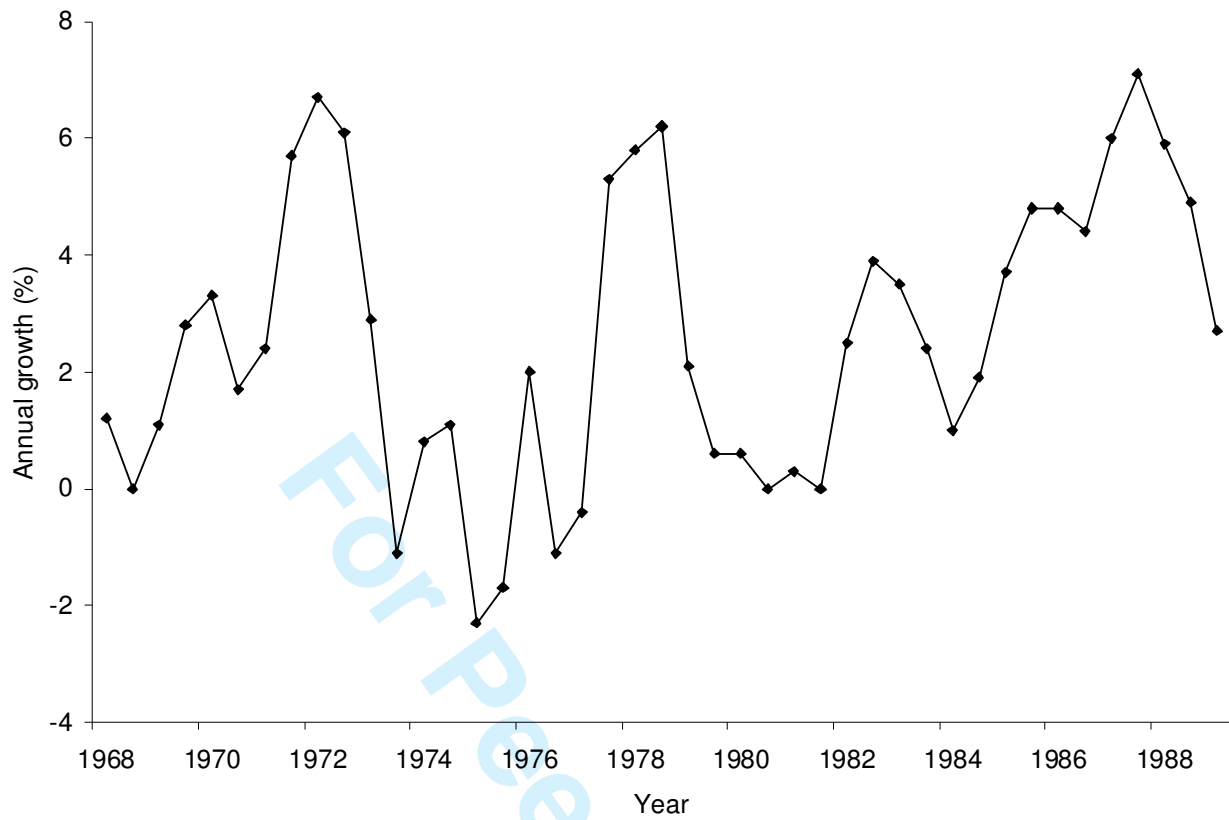


Figure One: Annual percentage growth in consumers' expenditure

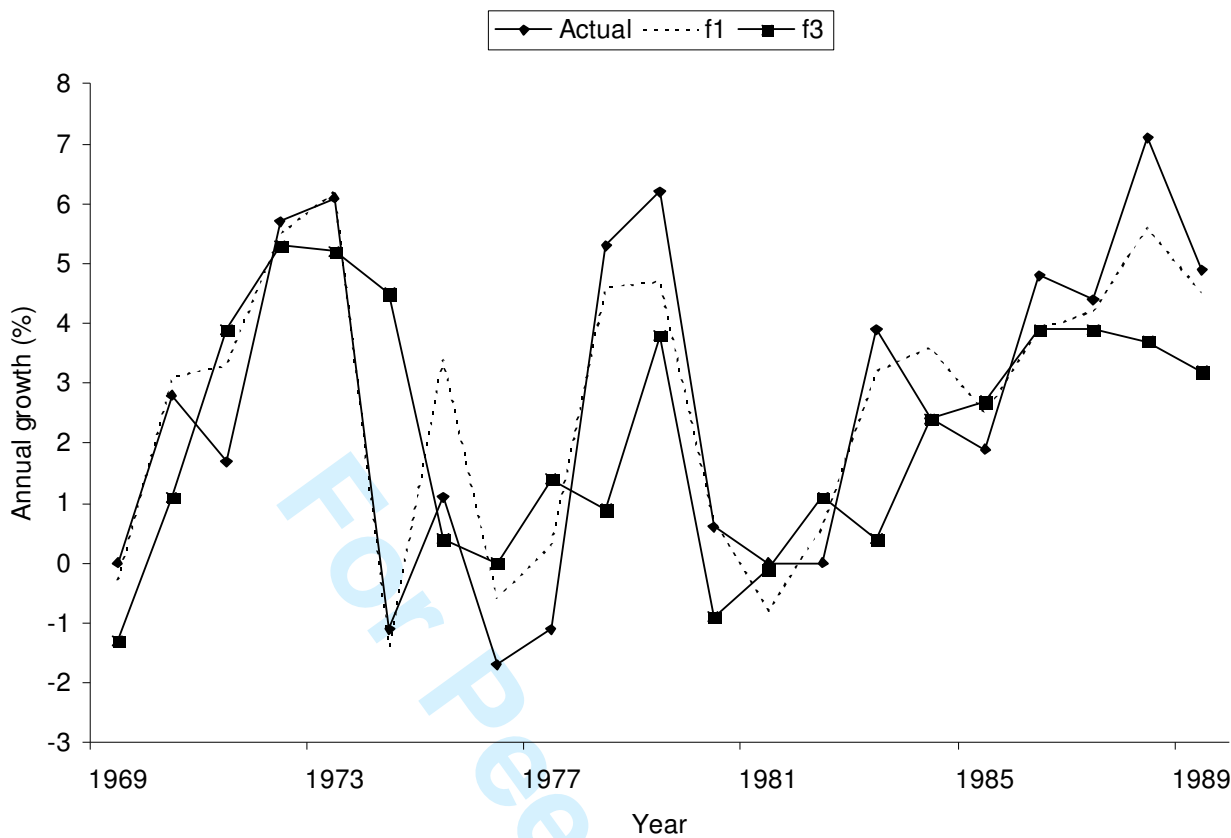


Figure Two: One-step and three-step forecasts of consumption growth

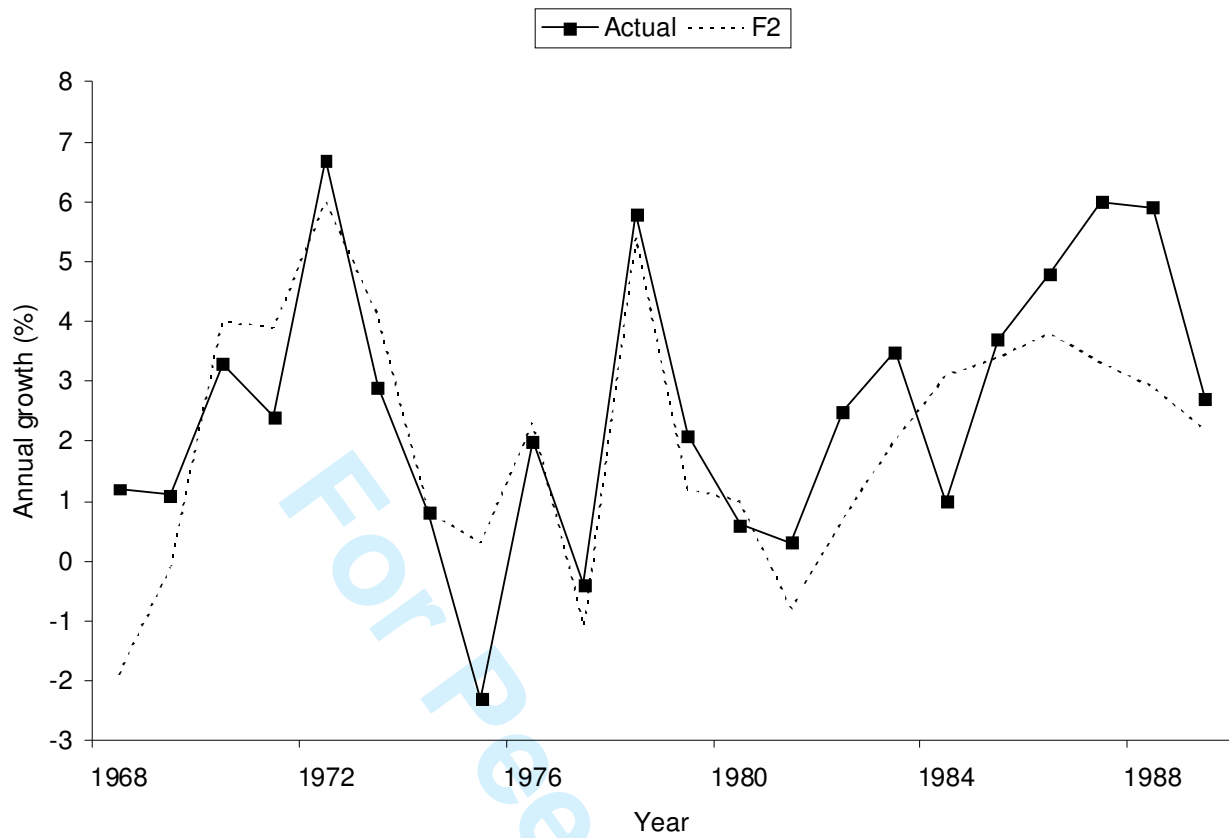


Figure Three: Two-step forecasts of consumption growth

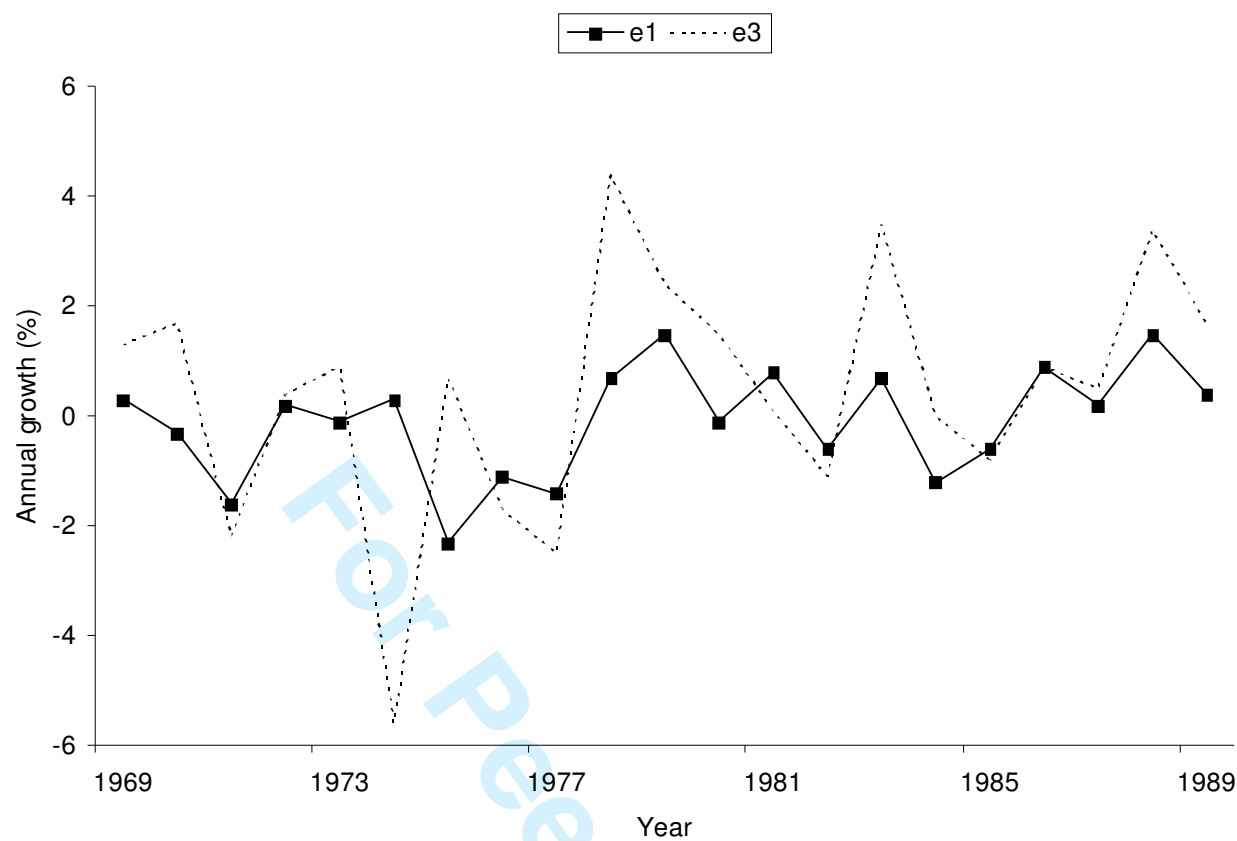


Figure Four: One-step and three-step forecast errors

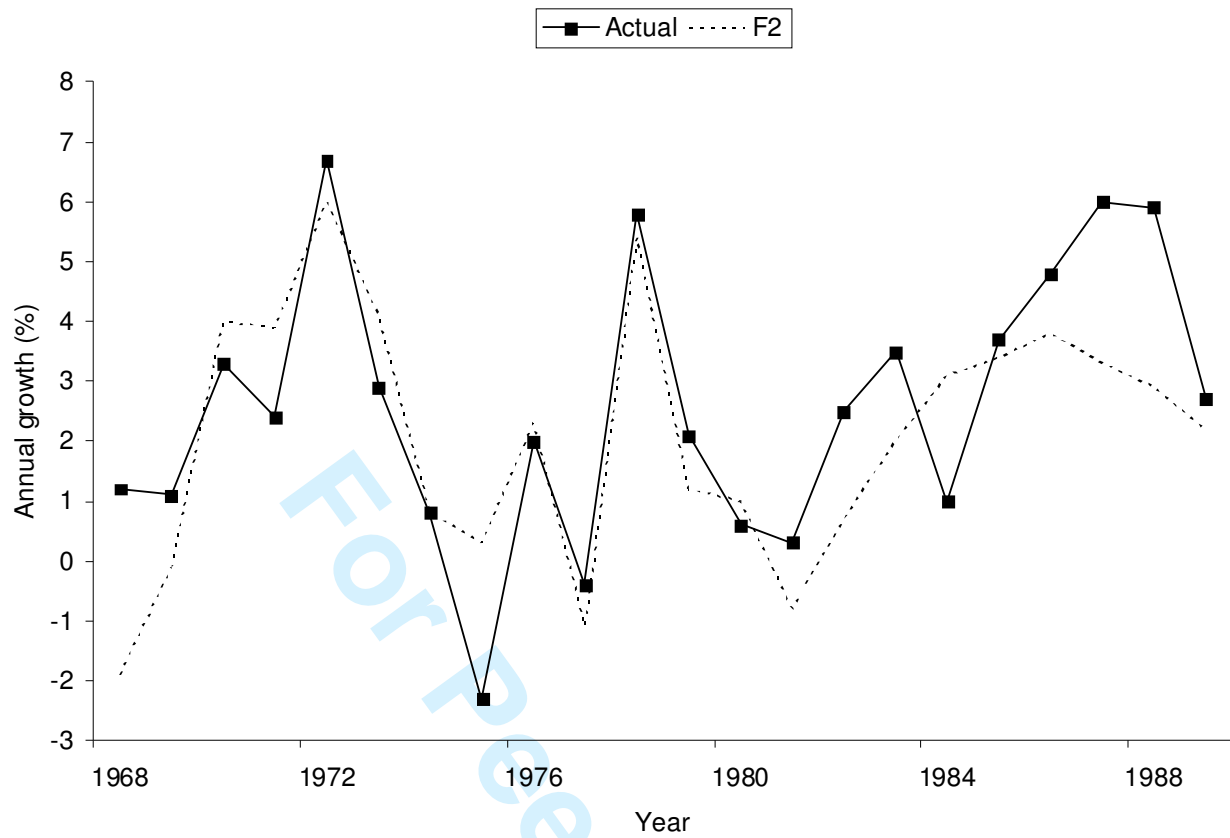


Figure Five: Two-step forecast errors