

Forecasting Seasonal UK Consumption Components

Michael P. Clements and Jeremy Smith*

Department of Economics,

University of Warwick.

June 12, 1997

Abstract

Periodic models for seasonal data allow the parameters of the model to vary across the different seasons. This paper uses the components of UK consumption to see whether the periodic autoregressive (PAR) model yields more accurate forecasts than non-periodic models, such as the airline model of Box and Jenkins (1970), and autoregressive models that pre-test for (seasonal) unit roots. We analyse possible explanations for the relatively poor forecast performance of the periodic models that we find, notwithstanding the apparent support such models receive from the data in-sample.

Keywords: periodic models, mean shifts, forecasting.

*The first author acknowledges financial support under ESRC grant L116251015. We are grateful to workshop participants at Warwick for helpful comments.

1 Introduction

Traditionally, the class of seasonal autoregressive-moving average (SARMA) models developed by Box and Jenkins (1970) have been popular amongst time-series analysts for modelling seasonal time series. More recently, increasing attention has been paid to periodic models of seasonality, where the parameters of the model are allowed to vary with the season. Such models are highly parameterised compared to the parsimonious description of the data accorded by the SARMA models. Important papers on periodic models include Tiao and Grupe (1980) and Osborn (1988), and recently, Philip Hans Franses and his co-authors have developed modelling approaches for periodic models, see, e.g., Franses and Paap (1994) and Franses (1996). Such models arise quite naturally when consumer tastes change over the year: see Osborn (1988).

However, whether the added complexity of periodic models offers a much improved forecast performance relative to simpler, non-periodic models (such as the SARMA class referred to above) is still a relatively open question. Osborn and Smith (1989) (henceforth OS) was an early contribution on the practical usefulness of periodic models for forecasting seasonal series. They found some improvement in forecast accuracy for the components of seasonal UK consumption at horizons less than a year.

In this paper we also use the components of UK consumers' expenditure to assess the usefulness of periodic models for forecasting. We are able to extend OS's study in a number of directions. Firstly, we now have at our disposal an additional 11 years of quarterly observations, for the period 1984:1 – 1994:4. So, by following OS and using the period from 1979:1 onwards for post-sample forecasting, we obtain 55 1, 2, ..., 10 -step ahead forecasts, compared to OS's 20 1-steps, 19 2-steps ahead, etc. The larger sample of multi-step forecasts, particularly at medium and longer horizons, allows conclusions concerning the relative merits of the models to be drawn with greater confidence, and we are also able to test whether apparent differences in forecast accuracy are statistically significant, using the test developed by Diebold and Mariano (1995).

Secondly, there have been a number of theoretical advances in periodic modelling, such as the recog-

dition that processes with periodically varying parameters may be periodically intergrated. We assess the forecast gains from imposing periodic integration when this phenomenon is not rejected by the data.

Thirdly, it is now recognised that a satisfactory in-sample fit is no guarantee of out-of-sample forecast performance (see, for example, Clements and Hendry, 1996, 1997b, Fildes and Makridakis, 1995), and we wish to explore the implications of structural breaks for models of seasonal processes.

Our results generally indicate that the short-run gains to periodic models found by OS do not extend to the larger sample of forecasts that we are able to calculate, and the remainder of the paper addresses the question of why we are unable to exploit the periodicity that appears to be a feature of the data to forecast more accurately.

The plan of the paper is as follows. Section 2 describes the data on the components of seasonally unadjusted UK non-durable consumption used in the analysis. Section 3 describes the models that we consider, and the estimates we obtain from fitting these models to the data. Section 4 reports the results of the forecast comparison exercise, and section 5 offers an explanation for the pattern of our results based on a simulation study. Finally, section 6 reviews our main results and offers some concluding remarks.

2 Description of the consumption data

The data is quarterly seasonally unadjusted nondurable consumption for the UK over the period 1955:1 to 1994:4, taken from Economic Trends Annual Supplement 1996.2. Consumers' expenditure divides into seven categories: food; alcohol and tobacco (referred to as 'alcohol'); clothing and footwear ('clothing'); energy products ('energy'); other goods; rents, rates and water charges; and other services ('services'). For reasons outlined in Osborn and Smith (1989), rents, rates and water charges are excluded from the analysis and total nondurable consumption ('total') is then defined as the sum of the remaining six component categories.

Summary statistics on these series are presented in table 1. Plots of the seven series (six components plus the total) are given in figures 1 and 2 for each of the 4 quarters separately. The seasonal patterns

of each series are essentially distinct, although alcohol, clothing and other goods all exhibit increasing consumption over the four quarters of the year to peak in quarter four.

3 Models of seasonal economic time series

3.1 AR models

The HEGY (Hylleberg, Engle, Granger and Yoo (1990)) testing procedure is based on the regression equation:

$$\alpha(L)\Delta_4 x_t = \mu_t + \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{3,t-1} + \xi_t \quad (1)$$

where:

$$\mu_t = \mu_1 + \mu_2 t + \sum_{i=1}^3 \mu_{2+i} Q_{it} \quad (2)$$

and:

$$z_{1t} = (1 + L + L^2 + L^3 + L^4)x_t, \quad z_{2t} = -(1 - L)(1 + L^2)x_t, \quad z_{3t} = -(1 - L^2)x_t.$$

$\alpha(L)$ is a p^{th} order lag polynomial, where p is chosen to whiten the error term (as judged by the usual diagnostic tests for serial correlation, heteroscedastic and normality). Taylor (1997) discusses some of the problems with the HEGY testing procedure – in particular, the choice of p , and generalising (2) to include *seasonal* deterministic trends.

The tests for roots at the zero, bi-annual and annual frequencies are based on whether $\pi_1 = 0$, $\pi_2 = 0$ and $\pi_3 = \pi_4 = 0$, respectively.

We adapt the HEGY procedure to generate models for forecasting seasonal processes, whereby after testing for unit roots, the regression is re-run omitting the z_i which are not significantly different from zero. We also omit the linear trend term because the null hypothesis that $\pi_1 = 0$ is not rejected for any of the series.¹ A zero-frequency root in conjunction with a linear trend would imply the rate of change

¹Strictly, the null is rejected for total non-durable consumption using the HEGY procedure, but standard unit root tests on seasonally-adjusted data suggest the aggregate is $I(1)$.

of the variable follows a linear trend, which appears unreasonable, albeit that it may not matter much for short horizon forecasts.

Table 2 reports the outcomes of the HEGY-testing procedure based on a regression of the form of (1) over 1958:1-1994:4 for each of the consumption components.

Notice that all roots are imposed when the null hypotheses that $\pi_1 = 0$, $\pi_2 = 0$ and $\pi_3 = \pi_4 = 0$ are not rejected. Separately to pre-testing for unit roots, we estimate models that have this property, since Clements and Hendry (1997a) found such models forecast reasonably well. The models were specified as:

$$\varphi(L) \Delta_4 x_t = \eta + \nu_t$$

where $\varphi(L)$ is at most a sixth-order lag polynomial. The model implies that the expected annual rate of growth ($E[\Delta_4 x_t]$) is a constant ($\varphi(1)^{-1} \eta$), and it seems unreasonable to include higher deterministic terms (such as a trend).

3.2 SARMA models

For quarterly data, the general class of seasonal autoregressive-moving average (SARMA) models developed by Box and Jenkins (1970) can be written as:

$$\phi(L) (1 - L) (1 - L^4) x_t = \mu + (1 - \theta_1 L) (1 - \theta_4 L^4) \epsilon_t \quad (3)$$

where $\epsilon_t \sim \text{IN}(0, \sigma_\epsilon^2)$, $|\theta_1| < 1$, $|\theta_4| < 1$ and $L^k x_t = x_{t-k}$. $\phi(L)$ is a polynomial in L with all its roots outside the unit circle, which we restrict to the form $\phi(L) = (1 - \phi_1 L)(1 - \phi_4 L^4)$.

Franses (1996) p.42–46 references empirical studies that fit SARMA models and provides examples. When $\phi(L) = 1$ this model is sometimes known as the ‘airline’ model. The filter $\Delta_4 = (1 - L^4)$ captures the tendency for the value of the series in a particular season to be highly correlated with the value in the same season a year earlier. The filter $\Delta = (1 - L)$ relates to the non-seasonal part of the model and specifies a stochastic trend in the level of the series (with drift when $\mu \neq 0$). The model suggests that

the expected quarterly change in the annual rate of growth depends on μ (equals μ when $\phi(L) = 1$), so not surprisingly the empirical estimates of μ are close to zero.

Expanding $\Delta\Delta_4$ we obtain:

$$(1 - L)(1 - L^4) = (1 - L)[(1 - L)(1 + L)(1 - iL)(1 + iL)]$$

suggesting two zero-frequency roots, as well as roots at the bi-annual and seasonal frequencies (taking the two complex roots as $(1 - iL)(1 + iL) = (1 + L^2)$). The results of the HEGY procedure (see table 2) suggest that only for food, alcohol, and other goods do we find a *single* zero-frequency root and roots at all the seasonal frequencies. There is no evidence of a second zero-frequency root for any of the series. This tension concerning the number of roots has been documented by a number of authors, e.g, Osborn (1990), Hylleberg, Jørgensen and Sørensen (1993). Clements and Hendry (1997a) review a number of possible explanations, such as the neglect of the MA component in the HEGY testing procedure leading to over-sized tests (see, for example, Franses and Koehler, 1994), and the possibility that ‘over-differencing’ in (3) converts level shifts in seasonal means into ‘blips’ which may be mistaken for outliers. Conversely, Smith and Otero (1997) show that HEGY will have low power to reject seasonal roots if there are shifts in the seasonal pattern, so that such testing procedures may overstate the number of roots.

Table (3) gives the results from estimating models of the form of (3) over the sample 1955:1-1994:4 (less periods lost by taking lags).

3.3 Periodic Autoregressive Models

The Periodic Autoregressive (PAR) model can be written as:

$$y_t = \mu_s + \phi_{1s}y_{t-1} + \dots + \phi_{ps}y_{t-p} + \epsilon_t, \quad t = 1, 2 \dots n \quad (4)$$

or more succinctly as:

$$\phi^s(L)(y_t - \mu_s^*) = \epsilon_t$$

where the intercepts (μ_s, μ_s^*) and the autoregressive parameters $(\phi_{1s}, \dots, \phi_{ps})$ may vary with the season, s , where for quarterly data $s = 1, \dots, 4$, and there are $n = 4N$ quarterly observations, where N is the number of annual observations. The disturbance term is assumed to be independently distributed with zero mean and variance σ^2 . Tiao and Grupe (1980) extend this basic model by allowing for periodic Moving Average (MA) terms.

Franses and Paap (1994) recommend a strategy for selecting the lag order p of the PAR model based on minimizing an information criterion, such as the Schwarz Criterion (SC), subject to an F -test of the restriction $H_0 : \phi_{p+1,s} = 0, s = 1, \dots, 4$ failing to reject. However, we found that a strict application of this strategy sometimes resulted in models with serially correlated and heteroscedastic disturbances, or in highly parameterised models compared to e.g., Franses (1996), Table 7.10, p. 112.

Consequently, for each series we selected a model with no holes in the lag distribution (the PAR model), taking in to account the properties of the equation disturbance term, as well as a restricted PAR model (RPAR) which sets insignificant lags to zero. For non-durable consumption, we obtain $p = 1$ following the mixed strategy (as in Franses, 1996, Table 7.10, p. 112 and Proietti, 1996, Table 1) and beginning with a maximum lag of 5, since $p = 1$ minimises SC and $F_{\phi_{2,s}=0}$ does not reject. However, lag 5 is highly significant: $F_{\phi_{5,s}=0} = 10.15$ (which is $F_{4,126}$ under the null). Hence for total consumption we estimated a fifth-order model (the PAR model), and a restricted PAR model (RPAR) with lags 1 and 5 only, since intermediate lags were not significant. For food, the mixed strategy led to a fourth-order model, as lower order models suffered from serial correlation and ARCH. Intermediate lags were not clearly insignificant, so the RPAR and PAR models are one and the same. For alcohol, the mixed strategy gave $p = 4$. On the basis of the individual t -statistics on the second and third lags, we also estimated a RPAR model with lags 1 and 4 only. For clothing, the mixed strategy led to $p = 4$. The SC suggested a lower order model, but the F -test for deleting lag 2 rejected, and similarly for testing lag 3 in a third-order model, and lag 4 in a fourth-order model. For energy, the mixed strategy led to $p = 4$.

For other goods, the mixed strategy suggested $p = 1$. However, all models of order lower than 5

rejected the null of serially uncorrelated errors with p -values numerically indistinguishable from zero. Hence the PAR model order was set at 5, and we also considered an RPAR with lags at 1, 4 and 5 only. The story for services was the same as that for other goods, and we ended up with a fifth-order model and an RPAR with lags at 1, 4 and 5 only.

Table 4 summarizes our results on model selection. It reports the order p of the PAR model for each of the components, the SC, the F -test for the null hypothesis $H_0 : \phi_{p,s} = 0$ of whether the order of the model can be further reduced, and the usual diagnostic statistics of model adequacy. The final column records the lags that comprise the RPAR model.²

Conditional on p , for the (unrestricted) PAR models we run tests of whether the slope parameters exhibit periodic variation. This is the F_{PAR} test which tests (4) against the nested non-periodic (in the slopes) AR(p) model:

$$y_t = \mu_s + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t. \quad (5)$$

The test is that $\phi_{js} = \phi_j$, $s = 1, \dots, 4$, $j = 1, \dots, p$, which has an $F_{3p, n-4p-4}$ distribution under the null of no periodic variation. We find a clear rejection of the null for each of the component series at the 5% level.

Franses (1996) suggests including a periodic trend for testing UK non-durable consumption, and we follow this advice for each of the components. In that case, the test of periodic variation is whether the AR parameters are equal across seasons in:

²The PAR, PIAR and NPIAR models include the following dummy variables to address the problem of outliers: total non-durable consumption included two dummy variables, one for Q1 of 1974 – 1977, and the other for 1980:2; Alcohol & Tobacco had two dummy variables, for 1973:2 and 1980:2; clothing & footwear two dummy variables for 1973:1 and 1978:3; and for services a single dummy variable for 1976:1.

For the HEGY and 5th-order fourth-differenced model we included the following dummies: alcohol, 1980:2; energy, 1963:3; other goods, 1968:2; services, 1978:1; total consumption, 1974:1 and 1978:1.

$$y_t = \mu_s + \gamma_s t + \phi_{1s} y_{t-1} + \dots + \phi_{ps} y_{t-p} + \epsilon_t, \quad (6)$$

denoted by $F(t)_{PAR}$. For all the components series and the total, except other goods, we are able to reject the null hypothesis at the 5% significance level – see table 4.

Equation (6) can be reparameterised as:

$$(1 - \alpha_s L) y_t = \mu_s + \gamma_s t + \sum_{j=1}^{p-1} \beta_{js} (1 - \alpha_{s-j} L) y_{t-j} + \epsilon_t \quad (7)$$

where $\alpha_{s-4i} = \alpha_s, i = 1, 2, \dots$. This parameterisation motivates the notion of periodic integration (PI). A series is said to be periodically integrated if $\prod_{s=1}^4 \alpha_s = 1$, in which case the periodic filter $(1 - \alpha_s L)$ removes the stochastic trend. Such models are known as PIAR models. A consequence is that such series can not be decomposed into seasonal and stochastic trend components: see Franses (1996) ch. 8.

Table 5 presents two tests of the hypothesis of PI due to Boswijk and Franses (1996): the $LR(t)$ statistic tests the restriction on the α_i 's based on (7), and the LR tests the same restriction but with the periodic trend term $(\gamma_s t)$ absent. For all the series the LR test fails to reject the null of periodic integration. Similarly for the $LR(t)$ test, with the exception of services.

Finally, the $F_{(1-B)}$ and $F(t)_{(1-B)}$ statistics test whether, assuming periodic integration, we can impose the further restriction that $\alpha_s = 1$. In that case the model is non-periodically integrated – NPIAR. The first statistic tests this restriction against (7) with periodic integration imposed and the seasonal trend term absent, and the second allows the seasonal trend term to be present. When the restriction is not rejected the PAR model simplifies to:

$$\Delta y_t = \mu_s + \gamma_s t + \sum_{j=1}^{p-1} \beta_{js} \Delta y_{t-j} + \epsilon_t. \quad (8)$$

$F_{(1-B)}$ rejects the simplification to (8) for all series at the 5% significance level, with the exception of other goods. Using $F(t)_{(1-B)}$ we are unable to reject the unit root at the 10% level for alcohol, other goods and services.

Following, e.g. Franses (1996) p.151, we can show that the PIAR model with seasonal intercepts suggests that the annual growth rate will differ across the seasons, suggesting increasing seasonal variation over time. This is most easily seen using a PIAR(1) model:

$$y_t = \mu_s + \alpha_s y_{t-1} + \epsilon_t, \quad \alpha_1 \alpha_2 \alpha_3 \alpha_4 = 1$$

so that by backward substitution:

$$\Delta_4 y_t = \delta_s + \epsilon_t + \alpha_s \epsilon_{t-1} + \alpha_s \alpha_{s-1} \epsilon_{t-2} + \alpha_s \alpha_{s-1} \alpha_{s-2} \epsilon_{t-3}$$

where δ_s depends on the α 's and μ 's. Hence $E[\Delta_4 y_t] = \delta_s$. Even when $\mu_s = \mu$, the δ_s will differ unless $\alpha_s = 1$, i.e., we have a NPIAR process. Hence, to rule out increasing seasonal variation in the PIAR case we need to set $\mu_s = 0$, whereas in the NPIAR we can allow $\mu_s = \mu$.

4 Empirical forecast comparison exercise

Each of the models discussed in the preceding section was used to generate 'rolling forecasts' of the UK consumption components. Following Osborn and Smith (1989) the models were specified on the full sample (in our case 1955–94), and then initially estimated on data up to 1978:4. Forecasts for the subsequent 10 periods (1979:1 – 1981:2) were calculated, yielding a single sequence of 1 to 10-step ahead forecasts. The estimation period is then increased by one observation to include 1979:1, and another sequence of 1 to 10-step ahead forecasts is calculated, and so on up to an estimation period that includes 1992:3. This rolling method of forecasting yields a sample of 55 1 to 10-step ahead forecasts. Table 6 reports (for selected horizons) the Mean Squared Forecast Error (MSFE) measure of forecast accuracy:

$$MSFE_h = \frac{1}{55} \sum_{j=1}^{55} (y_j - \hat{y}_{j,h})^2$$

where $\hat{y}_{j,h}$ is the forecast of y_j made at time $j - h$.

In table 6, BIC_{μ_s} refers to the PAR model of OS, chosen according to the BIC criterion by considering subsets of lagged regressors (see Osborn and Smith, 1989 for details). PAR_{μ_s} refers to the model given by

(4) which includes seasonal intercepts, $\text{PIAR}_{\mu=0}$ is the PAR model with PI imposed (and no intercepts), and NPIAR_{μ} to the non-periodically integrated PAR model with a non-seasonal intercept. AR_{Δ_4} denotes the linear model in the fourth-differences of the variable. For some of these models we have considered a number of variants with different (seasonal) deterministic terms. The reported models are those which forecast the best.

Firstly, we compare the results for the alternative periodic model specifications. The PAR_{μ_s} models generally dominate the BIC_{μ_s} models indicating that the subsets of regressors chosen by OS can be improved upon on the current vintage of data and sample period.

Table 5 provides evidence that the variables are periodically integrated. Imposing this restriction $\text{PIAR}_{\mu=0}$ generally leads to more accurate forecasts (an exception is total consumption). There is little evidence that further restricting the model to have a unit root, NPIAR_{μ} , improves forecast accuracy, and for alcohol, other goods and services, for which the restriction appears to be reasonable (when we allow for seasonal trends), the performance 10-steps ahead is worse.

Comparing the AR_{Δ_4} and HEGY models, the AR_{Δ_4} model is better for alcohol, otherwise the two models' forecasts are of similar accuracy. HEGY imposes roots at all frequencies for alcohol, so in this respect the models are equivalent, and the gains to the AR_{Δ_4} model result from $p = 5$ rather than the augmentation suggested by the HEGY testing procedure (1 for alcohol). The largest 1-step ahead gain for HEGY over AR_{Δ_4} is for total consumption, when it is of the order of 5%. Again, both models impose the same roots (all) for this variable.

The results also include tests of the statistical significance of differences in forecast accuracy, expressed as pairwise comparisons of models against the SARMA. Notice that the 1-step forecasts of food, clothing, other goods and services from the $\text{PIAR}_{\mu=0}$ model are all significantly less accurate than those from the SARMA model at the 10% level, while at this significance level we are unable to reject the null of equal forecast accuracy of the AR_{Δ_4} and HEGY model, relative to the SARMA, for any of the series. While the SARMA forecasts of total consumption are not statistically superior to those from the

PIAR $_{\mu=0}$, they are numerically more accurate on MSFE at $h = 1, 2$.

At longer horizons (e.g., $h = 10$) the autoregressive models (PIAR $_{\mu=0}$, AR $_{\Delta_4}$ and HEGY) tend to have numerically smaller MSFEs than the SARMA, the exception being the non-periodic AR models for alcohol.

Concern over the large number of parameters of the PAR models, which were either of order 4 or 5 (including the PIAR and NPIAR models), led us to include restricted PAR models (RPAR) in the comparisons for alcohol, other goods, services and total consumption. The RPAR forecasts were generally only a little better, suggesting that our suspicion that the over-parameterisation of the periodic models may account for their relatively poor performance is ungrounded.

5 Post forecast comparison analysis: shifts in seasonal means

While the formal testing procedures clearly reject the null of no periodic variation in the slope parameters for all the components, allowing for periodic variation yields no significant gains in forecast accuracy relative to the non-periodic models (SARMA, AR $_{\Delta_4}$ and HEGY), and in a number of cases the forecasts are significantly less accurate. Only for alcohol are the PIAR $_{\mu=0}$ model forecasts numerically more accurate than those of the AR or HEGY model forecasts, and even then they are matched at short horizons by the SARMA forecasts.

In this section we explore a potential explanation for the poor performance of the periodic models framed in terms of shifts in deterministic seasonal components.

Un-modelled structural breaks that show up in mean shifts are known to bias tests of zero frequency roots toward non-rejection (Perron, 1989, 1990), and seasonal mean shifts have a similar effect on tests for seasonal roots. Franses and Vogelsang (1995) and Smith and Otero (1997) discuss testing for roots in seasonal time series with mean shifts. Clements and Hendry (1997a) argue that more accurate out-of-sample forecasts may result from imposing extraneous roots when there are shifts in seasonal means over the forecast period. The imposition of unit roots can partially robustify sequences of rolling fore-

casts against (untreated) shifts in the deterministic seasonal components of the series, yielding improved forecast accuracy.

Equally, un-modelled seasonal mean shifts may also show up as seasonal variation in slope parameters, so that periodic models are found (see Franses and McAleer, 1995). Thus, while periodic models may appear to fit the data better than non-periodic models in-sample, non-periodic models that ‘over-difference’ the data may have the edge for out-of-sample forecasting.

We use the testing procedure suggested by Franses and McAleer (1995) to assess the evidence for mean shifts in the consumption series. Their procedure is designed to discriminate between a PIAR(p) and a NPIAR(p) (non-periodically integrated PAR) with deterministic shifts. The motivation for testing between these alternatives is that a PIAR with the α_s all close to 1 may be mistakenly chosen when the non-periodic $(1-L)$ filter is more appropriate if there are mean shifts, and vice versa. Since our empirical results give support to non-periodic models, a simple extension to their testing procedure allows us to test whether once we allow for structural change the periodic variation in the slope parameters is redundant. We make use of nested tests because the nesting model is itself a possible candidate model for the series.

The PIAR(p) model is given by (7) (with $\gamma_s = 0$). The NPIAR with deterministic mean shifts (NPIAR $_{\tau}$) is (7) with $\alpha_s = 1$ and the addition of the terms in δ_s^* :

$$\Delta y_t = \delta_s + \delta_s^* \mathbf{I}_T^{\tau} + \sum_{j=1}^{p-1} \beta_{js} \Delta y_{t-j} + \zeta_t \quad (9)$$

where $\mathbf{I}_T^{\tau} = 1$ when $t \geq \tau$. Consequently, the nesting model is:

$$y_t - \alpha_s y_{t-1} = \kappa_s + \kappa_s^* \mathbf{I}_T^{\tau} + \sum_{j=1}^{p-1} \beta_{js} (y_{t-j} - \alpha_{s-j} y_{t-j-1}) + \xi_t \quad (10)$$

with $\alpha_1 \alpha_2 \alpha_3 \alpha_4 = 1$.

Under the null that the PIAR is appropriate, an F -test based of (7) against (10) is asymptotically distributed as $F_{4, T-(4p+7)}$, where the 4 restrictions result from setting $\kappa_s^* = 0$ in (10). This test is denoted F_{PIAR} .

The null that (9) is correct can be tested by an F -test that compares the residual sum of squares (RSS) from (9) against the unrestricted RSS from (10). This is asymptotically distributed as $F_{3, T-(4p+7)}$,

where the 3 restrictions result from setting $\alpha_s = 1$ for each s , where the maintained hypothesis is that $\alpha_1\alpha_2\alpha_3\alpha_4 = 1$. This test is denoted $F_{NPIAR\tau}$. Testing the null of no periodic variation in the slope parameters once we allow for mean shifts:

$$\Delta y_t = \delta_s + \delta_s^* \mathbf{I}_T^\tau + \sum_{j=1}^{p-1} \beta_j \Delta y_{t-j} + \zeta_t \quad (11)$$

is again accomplished by an F -test (which we denote $F_{AR\Delta_1,\tau}$) that compares the RSS from (11) and the RSS from (10), and is asymptotically distributed as $F_{3+3(p-1), T-(4p+7)}$, where the $3(p-1)$ restrictions result from setting $\beta_{js} = \beta_j, j = 1, \dots, p-1$.

To make these tests operational, τ has to be chosen. We follow Franses and McAleer (1995) by plugging in a prior estimate based on:

$$\arg \max_{\tau} LR(\tau) = T \log \left(\frac{RSS}{RSS_1 + RSS_2} \right)$$

where RSS results from the full-sample (1955-94) estimation of the PIAR(p) model in (7), and RSS_1 and RSS_2 are obtained by splitting the sample at τ and estimating models of the form of (7) on each sub-sample. The grid of values Γ , such that $\tau \in \Gamma$, is recorded in the notes to table 7, which reports the estimated values of τ . A test that a structural change has occurred can be conducted by comparing the largest LR statistic (the supLR) to the critical values in Andrews (1993), Table 1. These values are also recorded in table 7, along with the outcomes of the nested tests described above. It is apparent that the null of no structural break is clearly rejected for all the series. A comparison of the estimated break dates in table 7 with the time series plots of the separate quarters for each series (figures 1 and 2) is informative.

For food, the PIAR model is clearly rejected, while neither the $F_{NPIAR\tau}$ or $F_{AR\Delta_1,\tau}$ tests reject at the 5% level. Testing the restriction of no periodic variation in the slope parameters conditional on the $NPIAR_\tau$ model (i.e., $\beta_{js} = \beta_j$) yields a p -value of 0.15 (the last column of table 7). Hence (11) appears to be adequate for explaining food. The estimated break date of 1973 is supported by the visual evidence. Around that date quarter 1 food consumption falls markedly relative to annual average consumption, quarter 4 consumption is more closely matched by that in quarters 2 and 3.

The nested testing outcomes for services and total consumption are similar to food, although the rejection of the PIAR for services is marginal, suggesting the nesting model may be appropriate. The estimated break date for services (1978) is not so obvious visually, although at that time the relationship between consumption in quarters 1 and 4 appears to alter. After 1974 total consumption in quarters 2 and 3 appears to fall relative to that in the other two quarters.

Alcohol requires the nesting model. Here the break date appears to reflect lower quarter 2 and 3 consumption relative to quarters 1 and 4. Clothing is adequately characterised by the PIAR model, and visually there is less evidence of a structural break. Energy can be explained by the NPIAR_τ , but the rejection of the PIAR is marginal. The seasonal pattern appears to alter at the end of the 1960's, as indicated by the estimated break date. Other goods fail to reject either the PIAR or the NPIAR_τ , suggesting that either is adequate and the tests lack power. Since the PIAR is rejected at the 10% level, there is some evidence that the NPIAR_τ model is more appropriate. The break is not as clearly evident as it is for some of the other series.

The outcomes of the nested tests suggest little support for the periodic models (especially the PIAR model) as compared to non-periodic models with mean shifts, or to composite models. They accord reasonably well with the poor empirical forecast performance of the periodic models. An exception is clothing, where we fail to reject the PIAR model (and the structural change models are rejected) but the PIAR forecasts are inferior to the non-periodic model forecasts.

In the next sub-section we report the results of a Monte Carlo study that attests to the usefulness of 'over-differenced' non-periodic models when the process is generated by a model in first differences with mean shifts (as appears to be the case for food, services and total consumption). In the final sub-section some simple algebraic manipulations are presented to help interpret the results.

5.1 A Monte Carlo study

We conducted a Monte Carlo study to see if the empirical forecasting results we obtained are broadly compatible with the models of the variables suggested by the nested tests. We took the preferred models

for each variable as the DGP (food, services and total: $AR_{\Delta_1, \tau}$; alcohol: nesting model; energy, other goods: $NPIAR_{\tau}$; clothing: PIAR) and simulated data for a period corresponding to that of the empirical exercise. We then estimate $PIAR^3$ and AR_{Δ_4} models for the simulated data, and carry out a forecasting exercise just as we did for the observed data. That is, we estimate models on a number of observations corresponding to the original in-sample data period (1955 – 1978), and then calculate rolling sequences of 1-to 10-step ahead forecasts. The model specifications (e.g., lag lengths) are those of the empirical models. We then move the origin forward by 1 observation, as in the empirical exercise, re-estimate and forecast. In this way we build up a sample of multi-step forecast errors. The above is then repeated on each of the 100 replications of the Monte Carlo, yielding 55×100 simulated multi-step forecast errors for $h = 1, \dots, 10$. The models are compared on the basis of RMSFE and the Diebold and Mariano (1995) test of equal forecast accuracy is implemented on the MSFEs.

The results are summarised in table 8. Note the higher levels of significance are because there are one hundred times as many forecast errors as in the empirical work. It is apparent that when the DGP is an $AR_{\Delta_1, \tau}$ process, as for food, services and total consumption, the AR_{Δ_4} model outperforms the PIAR. For these three series we also recorded the number of replications for which the test of no periodic variation in the slope parameters rejected. On the full simulated sample (corresponding to 1959–94) the rejection frequencies (out of 100) were: food 86, services 73 and total consumption 89. Notwithstanding the small Monte Carlo sample these results attest to the likelihood of incorrectly selecting a periodic model when the process is non-periodic (given by the $AR_{\Delta_1, \tau}$ model).

Thus our empirical results for these three series can be explained if the $AR_{\Delta_1, \tau}$ model is a reasonable empirical characterisation, as suggested by the outcomes of the testing procedures.

Not surprisingly, when the DGP is a PIAR, as for clothing, the PIAR model fares best. We noted above that the finding in favour of the PIAR for clothing on the nested tests was surprising given the empirical forecast performance of the periodic models.

³In the nested testing procedures, the PIAR model was estimated with seasonal intercepts, while in the simulation study we estimate intercepts when this leads to an improved forecast performance.

We find that for the nesting model DGP for alcohol the $AR_{\Delta_1, \tau}$ forecasts better than the PIAR model.

Both energy and other goods are characterised as $NPIAR_{\tau}$ processes: for the former the PIAR model yields more accurate forecasts, and for the latter, the $AR_{\Delta_1, \tau}$ model. We suspect this is because the periodic variation for other goods is less marked than for energy: conditional on the $NPIAR_{\tau}$ model we are unable to reject the $AR_{\Delta_1, \tau}$ model for other goods (see the last column of table 7), although $F_{AR_{\Delta_1, \tau}}$ does reject. Moreover, the sup LR statistic for energy, although significant, is the smallest for all the series, suggesting the break may be less marked than for other goods.

For some of the series characterised as $AR_{\Delta_1, \tau}$ processes we experimented with an additional break, of a similar magnitude to the historical (estimated) break, assumed to occur in the forecast period. As expected, this further favoured the AR_{Δ_4} model. In the following sub-section we illustrate the effects of structural breaks in the forecast period.

5.2 Seasonal mean shifts

Firstly, consider a seasonal mean shift at the beginning of the forecast period, so the effect on the estimated model parameters can be ignored. Assume the DGP is the $AR_{\Delta_1, \tau}$:

$$\varphi(L) \Delta y_t = \sum_{s=1}^4 \delta_s Q_{s,t} + \sum_{s=1}^4 \delta_s^* I_T^{\tau} + \epsilon_t \quad (12)$$

where we introduce explicit seasonal dummy variables, $Q_{s,t} = 1$ if $t = s + 4j$, $j = 0, 1, 2, \dots$. Multiplying (12) by $S(L) = 1 + L + L^2 + L^3$ results in a model for the fourth differences:

$$\varphi(L) \Delta_4 y_t = \sum_{s=1}^4 S(L) \delta_s Q_{s,t} + \sum_{s=1}^4 S(L) \delta_s^* I_T^{\tau} Q_{s,t} + S(L) \epsilon_t.$$

Let $S(L) \delta_s = \delta$, and $S(L) \delta_s^* = \delta^*$. If we assume that $\delta^* = 0$, so that the expected annual growth rate ($E[\Delta_4 y_t]$) is unchanged by the shift in seasonal means at τ , then:

$$\varphi(L) \Delta_4 y_t = \delta + \sum_{s=1}^4 \delta_s^* I_{\tau+2}^{\tau} Q_{s,t} + S(L) \epsilon_t.$$

The $S(L)$ operator converts the level changes to blips. A model of this process that ignores the structural change would be:

$$\varphi^+(L) \Delta_4 y_t = \delta^+ + \epsilon_t^+ \quad (13)$$

where $\delta^+ \simeq \delta$.

Consider forecasting Δy_t using (12) with the δ_i^* terms absent (the shifts are not modelled), and setting $\varphi(L) = 1$. Forecasts of $\Delta y_{\tau+j}$, $j \geq 0$ made pre-break at time $\tau - 1$ rapidly go off course (we are forecasting δ_s instead of $(\delta_s + \delta_s^*)$ for season s). The outcome is moderated a little if the forecast origin is moved forward to $\tau + 1, \tau + 2, \dots$ and the model estimates updated on the extended sample, but the impact of the post-break observations on the estimated model parameters will initially be slight. By way of contrast, consider forecasting using (13) with $\varphi^+(L) = 1$. At time $\tau - 1$, the expected value of the forecast is:

$$\begin{aligned} E[\Delta y_\tau] &= E[(1 - S(L)) \Delta y_\tau] + \delta^+ \\ &= - \sum_{i=1}^3 \delta_{s-i} + \delta^+ = -(\delta - \delta_s) + \delta^+ \simeq \delta_s \end{aligned}$$

when τ falls in season s . The forecasts are biased, and similarly for origins up to $\tau + 1$, but for forecasts made at $\tau + j$, $j \geq 2$:

$$E[\Delta y_{\tau+j}] = - \sum_{i=1}^3 (\delta_{s-i} + \delta_{s-i}^*) + \delta^+ = -(\delta - \delta_s) - (-\delta_s^*) + \delta^+ \simeq \delta_s + \delta_s^*$$

assuming $\tau + j$ falls in season s . This demonstrates that forecasts made with the AR_{Δ_4} model after the mean shift has occurred are approximately unbiased.

Secondly, we can demonstrate the effect of a seasonal mean shift during the estimation period. The (13) parameters will be little affected provided $\delta^* = 0$. But suppose we estimate a second-order periodic model for the $AR_{\Delta_1, \tau}$ DGP:

$$y_t - \alpha_s y_{t-1} = \mu_s + \beta_s (y_{t-1} - \alpha_{s-1} y_{t-2}) + v_t \quad (14)$$

and we impose $\alpha_s = 1 \forall s$, then:

$$\Delta y_t = \mu_s + \beta_s \Delta y_{t-1} + v_t \quad (15)$$

where $0 < \beta_s < 1$ and as $\delta_s^* - \delta_s \rightarrow \infty$, $\beta_s \rightarrow 1$. The forecast function is (ignoring parameter estimation uncertainty):

$$\widehat{\Delta y_{T+h}} = \mu_s \sum_{i=0}^{h-1} (\beta_s)^i + (\beta_s)^h \Delta y_T \rightarrow \frac{\mu_s}{1 - \beta_s} \text{ as } h \rightarrow \infty \text{ and for } |\beta_s| < 1.$$

Let $\mu_s^+ = \frac{\mu_s}{1 - \beta_s}$. Then μ_s^+ is the long-run mean of the process in season s from the historical data, which incorporates the two regimes. The long-run mean of the process after $t = \tau$ is μ_s , and $\mu_s^+ \neq \mu_s$, so the forecasts from the NPIAR model are biased for large h .

6 Conclusions

The in-sample support for periodic models (the null of no periodic variation in the slope parameters is clearly rejected for all series) does not translate in to clear forecast gains, even at short horizons, in contrast to the findings of Osborn and Smith (1989). We find that for a number of series the periodic model can be rejected in favour of a model with a shift in the seasonal means, and we provided simulation and analytical evidence that such processes could explain some of the results of our empirical forecast accuracy comparisons.

References

- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821–856.
- Boswijk, H. P., and Franses, P. H. (1996). Unit roots in periodic autoregressions. *Journal of Time Series Analysis*, **17**, 221–245.
- Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.
- Clements, M. P., and Hendry, D. F. (1997a). An empirical study of seasonal unit roots in forecasting. *International Journal of Forecasting*. Forthcoming.
- Clements, M. P., and Hendry, D. F. (1997b). Forecasting economic processes. *International Journal of Forecasting*. Forthcoming.
- Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.
- Fildes, R., and Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, **63**, 289–308.
- Franses, P. H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*. Oxford: Oxford University Press.
- Franses, P. H., and Koehler, A. B. (1994). Model selection strategies for time series with increasing sea-

sonal variation. revised version of Econometric Institute Report 9308, Erasmus University Rotterdam.

Franses, P. H., and McAleer, M. (1995). Testing nested and non-nested periodically integrated autoregressive models. Center for Economic Research Discussion Paper No. 9510, Tilburg University.

Franses, P. H., and Paap, R. (1994). Model selection in periodic autoregressions. *Oxford Bulletin of Economics and Statistics*, **56**, 421–439.

Franses, P. H., and Vogelsang, T. J. (1995). Testing for seasonal unit roots in the presence of changing seasonal means. Econometric Institute Report 9532, Erasmus University Rotterdam.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

Hylleberg, S., Engle, R. F., Granger, C. W. J., and Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, **44**, 215–238.

Hylleberg, S., Jørgensen, C., and Sørensen, N. K. (1993). Seasonality in macroeconomic time series. *Empirical Economics*, **18**, 321–325.

Jarque, C. M., and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, **6**, 255–259.

Osborn, D. R. (1988). Seasonality and habit persistence in a life cycle model of consumption. *Journal of Applied Econometrics*, **3**, 255–266.

Osborn, D. R. (1990). A survey of seasonality in UK macroeconomic variables. *International Journal of Forecasting*, **6**, 327–336.

Osborn, D. R., and Smith, J. P. (1989). The performance of periodic autoregressive models in forecasting seasonal UK consumption. *Journal of Business and Economic Statistics*, **7**, 117–127.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.

- Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics*, **8**, 153–162.
- Proietti, T. (1996). Spurious periodic autoregressions. Mimeo, Dipartimento di Scienze Statistiche, Università di Perugia.
- 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.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 462–464.
- Smith, J., and Otero, J. (1997). Structural breaks and seasonal integration. *Economic Letters*. Forthcoming.
- Taylor, A. M. R. (1997). On the practical problems of computing seasonal unit root tests. *International Journal of Forecasting*. Forthcoming.
- Tiao, G. C., and Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, **67**, 365–373.

Table 1 Shares for Consumption Components in Total and Across Year.

Variables	Mean	Q1	Q2	Q3	Q4
Food	0.210	0.241	0.251	0.250	0.258
Alcohol	0.151	0.215	0.245	0.255	0.285
Clothing	0.069	0.207	0.240	0.240	0.313
Energy	0.090	0.290	0.232	0.209	0.269
Other goods	0.125	0.225	0.238	0.247	0.290
Services	0.355	0.232	0.251	0.274	0.243

Table 2 HEGY tests for unit roots.

Variable	p	$t(\pi_1)$	$t(\pi_2)$	$F(\pi_3, \pi_4)$	AR(5)	ARCH(4)	NORM	RESET
Food	6	-2.620	-2.104	3.612	0.552	0.034	0.585	0.532
Alcohol	1	-2.323	-2.505	4.993	0.793	0.236	0.591	0.140
Clothing	2	-2.944	-2.460	7.221*	0.427	0.018	0.168	0.893
Energy	0	-3.108	-4.642**	2.984	0.375	0.421	0.697	0.979
Other goods	4	-2.659	-1.739	3.649	0.299	0.851	0.628	0.506
Services	5	-1.801	-2.649	7.112*	0.094	0.463	0.373	0.394
Total	5	-3.605*	-2.563	3.016	0.270	0.190	0.669	0.631

* denotes significance at the 5% level, and ** at the 1% level using the appropriate (non-standard) critical values (see Hylleberg *et al.*, 1990).

AR(5) is a test for 5th-order residual serial correlation: see Godfrey, 1978.

ARCH(4) is a 4th-order residual autoregressive conditional heteroscedasticity test: see Engle, 1982.

NORM is a χ^2 test for normality: see Jarque and Bera, 1980.

The RESET test is a test of appropriate functional form: see Ramsey, 1969.

The elements in the last 4 columns are p -values.

Table 3 SARMA Model. $(1 - \phi_1 L)(1 - \phi_4 L^4)\Delta_1\Delta_4\varphi(L)y_t = (1 + \theta_1 L)(1 + \theta_4 L^4)\varepsilon_t$.

Variable	μ	ϕ_1	ϕ_4	θ_1	θ_4	AR(5)	ARCH(4)	NORM
Food	$8.93e^{-5}$ (0.002)			-0.567 (0.082)	-0.654 (0.060)	0.435	0.156	0.285
Alcohol	$1.44e^{-4}$ (0.000)		0.368 (0.093)	-0.451 (0.090)	-0.849 (0.061)	0.011	0.001	0.234
Clothing	$1.52e^{-4}$			-0.330 (0.079)	-0.589 (0.092)	0.053	0.028	0.319
Energy	$-1.80e^{-4}$ (0.001)	0.093 (0.051)	-0.789 (0.083)	-0.886 (0.042)	0.746 (0.116)	0.190	0.056	0.307
Other goods	$5.56e^{-4}$ (0.001)	-0.353 (0.154)		0.104 (0.160)	-0.286 (0.137)	0.0120	0.480	0.460
Services	$-2.68e^{-4}$ (0.001)				-0.365 (0.095)	0.317	0.329	0.221
Total	$4.98e^{-4}$ (0.000)				-0.426 (0.091)	0.742	0.488	0.997

Table 4 Periodic Models.

Variable	p	SC	$F(\phi_{p,s})$	AR(4)	ARCH(4)	NORM	F_{PAR}	$F(t)_{PAR}$	RPAR lags
Food	4	-1179.22	0.000	0.006	0.158	0.421	2.414**	1.926*	1-4
Alcohol	4	-1027.41	0.000	0.159	0.303	0.976	5.199**	3.266**	1,4
Clothing	4	-1041.68	0.011	0.171	0.256	0.035	5.918**	2.070*	1-4
Energy	4	-910.90	0.000	0.461	0.786	0.512	2.196*	3.255**	1-4
Other goods	5	-1087.03	0.000	0.205	0.959	0.087	2.984**	1.538	1,4,5
Services	5	-1165.96	0.000	0.170	0.182	0.388	2.431**	2.872**	1,4,5
Total	5	-1300.46	0.000				3.222**	2.072*	1,5

SC is the Schwarz criteria: see Schwartz (1978).

$F(\phi_{p,s})$ is an F -test of the null hypothesis $H_0 : \phi_{p,s} = 0$, i.e., a test of whether the order of the model can be further reduced.

F_{PAR} tests the null of no periodic variation (in the slopes) and is $F_{3p,n-4p-4}$.

$F(t)_{PAR}$ tests the null of no periodic variation allowing for seasonal trends.

Table 5 Periodic Integration Tests.

Variable	p	LR	$F_{(1-B)}$	$LR(t)$	$F(t)_{(1-B)}$
Food	4	2.553	2.630*	0.000	5.326**
Alcohol	4	0.207	4.233**	1.152	1.983
Clothing	4	2.967	8.004**	1.258	2.445*
Energy	4	0.001	3.637**	1.221	7.359**
Other goods	5	0.009	1.801	7.692	0.491
Services	5	0.015	2.606*	11.337*	1.207
Total	5	4.947	3.753**	10.247	3.880**

The LR and $LR(t)$ statistics test the null of periodic integration.

The $F_{(1-B)}$ and $F(t)_{(1-B)}$ statistics test for 'non-periodic' integration.

Critical values for LR are 9.24 and 7.52 and for $LR(t)$ 12.96 and 10.50 at the 5% and 10% significance levels, the corresponding critical values for F -statistic are around 2.69 and 2.13.

Table 6 RMSFEs for seasonal UK consumption.

h	Food	Alcohol	Clothing	Energy	Other goods	Services	Total
BIC_{μ_s}							
1	2.13	3.33*	3.79**	4.85	2.35**	2.38**	1.34**
2	2.40*	4.06*	4.37**	5.37	3.13**	3.34**	2.01**
5	3.13*	5.53	6.20	6.46*	6.11	6.91**	4.12
10	4.25*	7.77	10.29	8.17	10.69	13.76**	7.58
PAR_{μ_s}							
1	2.26**	3.01	3.51**	4.15	2.28**	2.51**	1.24
2	2.43*	3.34	4.02**	3.75*	2.90**	3.43**	1.76
5	3.12*	4.62	5.86	5.22**	5.35	7.09**	3.53
10	3.94	6.33	10.16	6.02	9.65	14.19*	6.98**
$RPAR_{\mu_s}$							
1	2.26**	2.98	3.51**	4.15	2.12	2.30**	1.30
2	2.43*	3.40	4.02**	3.75*	2.74	3.26**	1.97*
5	3.12*	4.68	5.86	5.22**	5.13	6.47**	3.81
10	3.94	6.45	10.16	6.02	9.40	13.22**	7.33
$PIAR_{\mu=0}$							
1	2.12**	2.77	3.53**	3.87	2.25**	2.24**	1.30
2	2.44**	3.32	3.88*	3.53*	2.71	3.14**	1.89
5	2.74	4.55	5.43	4.92**	5.04	5.89	3.54
10	3.37**	6.57	8.25	5.72	8.98	10.44	6.54**
$NPIAR_{\mu}$							
1	2.28**	2.77	3.44**	3.92	2.44**	2.22**	1.34
2	2.41**	3.37	3.80*	3.55*	2.99*	3.11**	1.89
5	2.89	4.82	5.79	5.07**	5.62	6.10*	3.49
10	3.51	7.55	8.72	5.75	10.10	10.84	6.63**
AR_{Δ_4}							
1	1.92	3.03*	3.12*	3.75	1.82	2.06*	1.21
2	2.18	3.59*	3.76	3.62*	2.28	2.91	1.70
5	2.58	5.97*	5.35	5.36**	4.32	5.53	3.36
1	3.33*	9.60**	8.03	6.84	7.47*	10.72	6.65*
$HEGY_{\mu}$							
1	1.95	3.48	3.03*	3.76	1.81	2.01	1.15
2	2.19	4.15*	3.51	3.53*	2.29	2.82	1.71
5	2.65	6.94**	5.22	4.93**	4.28	5.52	3.30
10	3.52	10.92**	7.96**	5.70	7.47*	10.65	6.98
$SARMA_{\mu=0}$							
1	1.92	2.76	2.79	4.21	1.92	1.90	1.18
2	2.15	3.30	3.45	4.98	2.40	2.68	1.68
5	2.73	5.18	5.32	15.85	4.88	5.43	3.58
10	3.79	8.02	9.34	65.45	9.32	11.03	7.34

** (*) denotes the forecast model is more accurate (on MSFE) than SARMA at the 10% (20%) level. ** and * imply less accurate. The test is implemented as in Diebold and Mariano (1995) (i.e., using a uniform window).

Table 7 Nested tests.

	p	$supLR(\tau)$	τ	F_{PIAR}	$F_{NPIAR\tau}$	$F_{AR_{\Delta_1,\tau}}$	$F_{NPIAR\tau,AR_{\Delta_1,\tau}}$
Food	4	77.29	1973	6.67 [0.000]	2.33 [0.077]	1.74 [0.065]	1.50 [0.154]
Alcohol	4	75.07	1980	7.15 [0.000]	3.57 [0.016]	3.65 [0.000]	3.48 [0.001]
Clothing	4	57.41	1979	1.58 [0.183]	6.75 [0.000]	3.57 [0.000]	2.23 [0.024]
Energy	4	44.68	1969	2.50 [0.046]	1.13 [0.339]	2.43 [0.007]	2.85 [0.004]
Other goods	5	55.08	1974	2.05 [0.091]	1.52 [0.213]	2.09 [0.022]	1.69 [0.076]
Services	5	78.29	1978	2.46 [0.049]	1.31 [0.274]	1.04 [0.417]	0.71 [0.740]
Total	5	73.59	1974	6.04 [0.000]	0.47 [0.704]	1.59 [0.102]	1.49 [0.136]

p denotes the order of the PIAR model within which the test for a structural break is carried out. The candidate breakpoints are 1968, 1969, through to 1983. Thus, for the 1968 breakpoint the PIAR model is estimated on the two sub-samples 1956:1 – 1967:4 and 1968:1 – 1994:4 (for $p = 4$), and the sum of the residual sums of squares is compared to the overall (1956:1 – 1994:4) residual sum of squares. The reported value of τ maximises the value of the LR test for a structural change – $supLR(\tau)$ is the value of the test statistic. From Andrews (1993), Table 1 an approximate 1% critical value for the PIAR model with $p = 4$ is 44.76 (since there are 19 parameters to be estimated, and we are approximately considering breaks in the range $[0.3, 0.7]$ of the sample). Thus the test outcomes are significant at the 1% level. Critical values are not tabulated for $p = 5$ (23 parameters) but crude extrapolation suggests that the test outcomes would again be significant at the 1% level. The final column reports the test of the NPIAR model with seasonal mean shifts versus a model in first differences with mean shifts but no seasonal variation ($AR_{\Delta_1,\tau}$). The values in squared brackets are p -values.

Table 8 RMSFEs for simulated data.

h	Food	Alcohol	Clothing	Energy	Other goods	Services	Total
PIAR							
	No μ_s	μ_s	No μ_s	No μ_s	μ_s	No μ_s	No μ_s
1	1.77**	4.16**	4.73**	4.58**	3.01**	2.04**	1.20**
2	1.84**	4.87**	5.26**	4.89**	3.70**	2.70**	1.65**
5	2.31*	6.22	7.03**	6.16**	5.78**	4.58**	2.76**
AR $_{\Delta_4}$							
1	1.72	3.99	4.93	4.81	2.52	1.87	1.13
2	1.80	4.50	5.57	5.00	3.21	2.49	1.56
5	2.28	6.19	7.33	6.47	5.48	4.31	2.67

** (*) denotes the PIAR model is more accurate (on MSFE) than the AR $_{\Delta_4}$ model at the 1% (5%) level.

** and * imply less accurate. The test is that of Diebold and Mariano (1995).

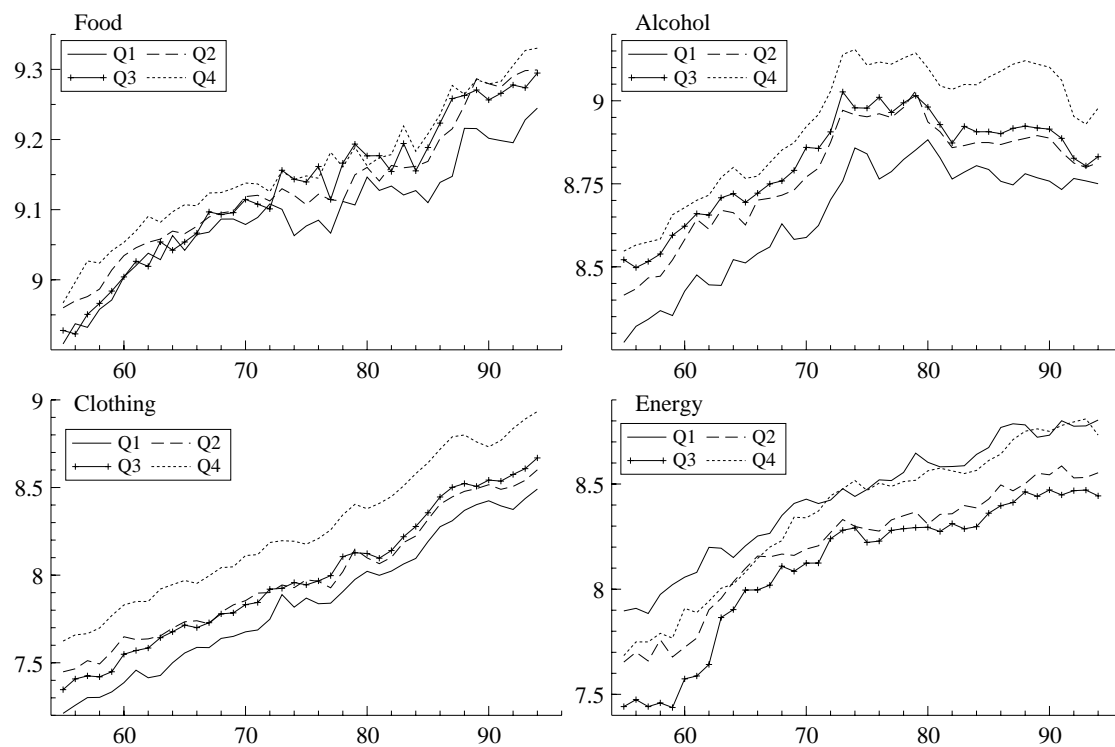


Figure 1 Time series of quarters: Food, alcohol, clothing, energy.

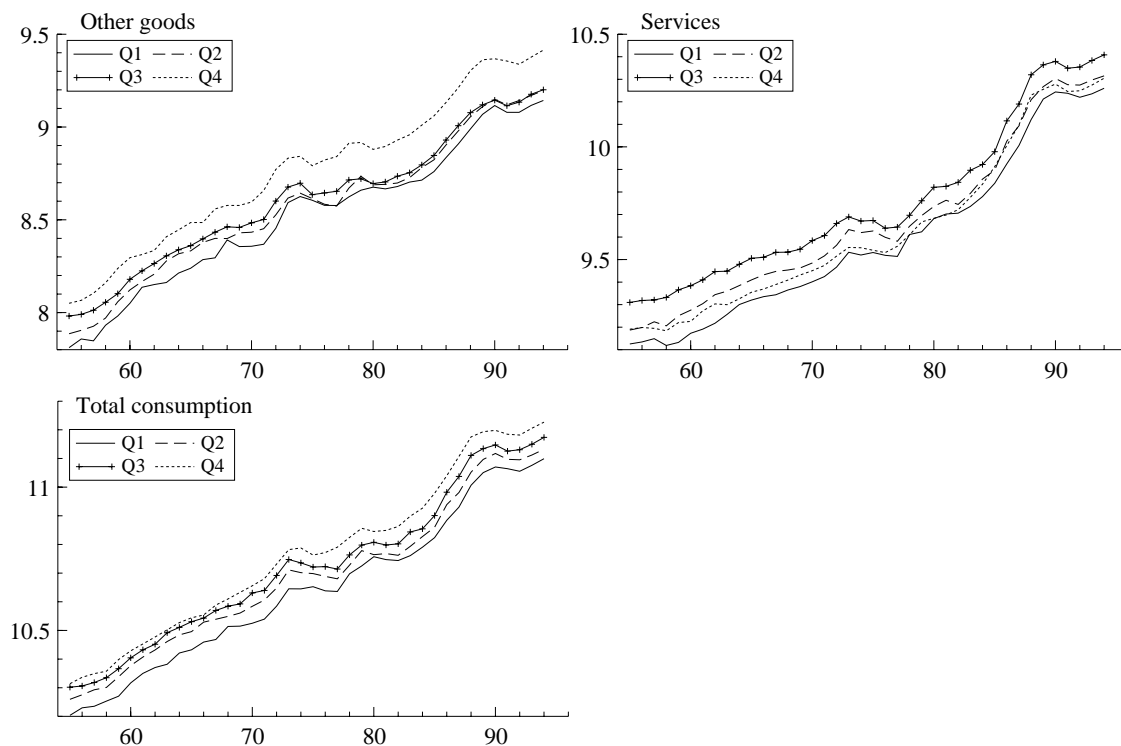


Figure 2 Time series of quarters: Other goods, services, total consumption.