Evaluating Macro-Modelling Systems: An Application to the Area Wide Model

Peter McAdam *

Ricardo Mestre

Research Department, European Central Bank,

Kaiserstraße 29, D-60311 Frankfurt am Main, Germany.

Preliminary

Abstract

This paper considers approaches to considering the validity of the overall structure of macro-econometric models – specifically, the Area Wide Model of the European Central Bank. By *structure*, we refer to the dynamic (business-cycle) and steady-state features that the model purports to capture. This leads to two types of tests. The first, drawing on the DSGE literature, is concerned with whether the model matches business-cycle (high-frequency) data characteristics. This is implemented by a Cholesky bootstrap whereby the steady state of the model is stochastically simulated using historically consistent co-variances. The generated data is analysed for stylised-facts fitting and, similarly, using the model's implied spectral characteristics, for congruence with the data in terms of persistence, periodicity and spectral fit. Moments matching, however, is only one aspect of overall model evaluation. Consequently, we move to tests that combine high-frequency aspects (short-horizon forecasts) with long-run features (such as the existence and identification of steady states and trends). Recursive forecasting tests form the second part. The forecasts attempt to measure the accuracy of model-based forecasts both simulated out-of-sample and in an in-sample exercise. The out-of-sample exercise analyses the 1- to 8-step-ahead forecasting ability of the model. For this, the model is re-estimated each time on a subset of the original sample, and the re-estimated model is used to generate a forecast over the remaining sample. The in-sample exercise omits the re-estimation step but performs a more thorough exercise, covering 1- to 12-steps-ahead forecasts over a larger part of the original sample. For this latter exercise, a thorough analysis of sources of forecast error is made. Both exercises incorporate alternative residual-projection methods, in order to assess the importance of unaccounted-for breaks in forecast accuracy. Conclusions reached are that, on the one hand, in-sample exercises should be preferred with systems of this size and typical samples; and, on the other, that mechanical residual adjustment or model re-estimation should be avoided except under strong evidence of mis-specification. The paper considers the testing procedure to be one applicable to the class of large, macro models and therefore of general interest.

Key words: Macro-model, Stylised Facts, Spectral analysis, Forecast Projections, Model Evaluation.
JEL: C52, E30, E32, E37.

* Corresponding author: peter.mcadam@ecb.int; Tel: +49 69 1344 6434; Fax: +49 69 1344 6575. The opinions expressed are not necessarily those of the ECB. We thank Gabriel Fagan for comments.

1. Introduction

In this paper, we document some recent testing framework that has been applied to the Area Wide model (AWM) of the European Central Bank (Fagan *et al,* 2001). The AWM can be considered as a mix of 'old' and 'new'. On the one hand, it is clearly embodies some Keynesian features: dynamics are mostly backward-looking and with the specific aim of 'fitting the data'. On the other hand, it can be considered modern-style in that it takes a system-wide view on long-run properties, which are based on clear, theoretical underpinnings. In common with models in this class, a large part of the steady state implicit in the model hinges on calibrated parameters. Therefore, the model embodies both short- and long-run mechanisms: the short run steers the model according to data-fitting criteria, and the long run according to theory.

Whatever its specific features, though, the AWM is a quantitative tool and, as such, should be subject to a testing apparatus. This apparatus should be designed in order to capture both aspects of data-fitting and theoretical coherence, and, in so doing, comment on the validity of the model's overall "structure". Structure is a word used in economic analysis to refer to many different concepts. The notion of structure, for instance, appears in the context of whether one can recover all the information from a structural VAR using the standard one by imposing enough restrictions on the latter. Thus, the term essentially refers to the imposition of economically meaningful restrictions (e.g., theoretical parameter constraints, causal structure, endo/exog distinction etc) that allows unique identification of a simultaneous system of equations. Hence, structure is a system property. Macro-models are not strictly speaking VARs but they, nevertheless, embody what we might consider structural features. Moreover, it certainly makes sense to think of proper tests on them being system-wide.

To proceed, we draw on a number of given methodologies. First, in line with the Dynamic Stochastic General Equilibrium (DSGE) literature, we are concerned with whether the model matches cyclical characteristics of the data. This is implemented by a Cholesky bootstrap whereby the steady state of the model is stochastically simulated using historically-consistent co-variances. The resulting outputs are flittered through both time and frequency-domain metrics. Moments matching, however, can only take us so far in system evaluation; "well-matching" models need not necessarily have any sensible structure or steady state. Thus, we move to tests that combined high-frequency aspects (short-horizon forecasts) with long-run features (such as the existence and identification of steady-states and trends). Recursive forecasting tests form the second part of our testing methodology. However, both types of analyses try to capture the same thing – the validity of the model's structure. To reiterate, the business-cycle tests capture short-run phenomena but are carried out on the model's steady state whilst the forecasting tests are conducted on historical samples but try to capture the validity of the steady state implicit in the model.

In analysing the model, it is thus necessary to exert particular care in evaluating the extent to which the model matches the data. The standard approach in the DSGE literature is to propose a fully articulated theoretical model of the economy predicated on the behaviour of an optimising representative agent in an economy under an evolving equilibrium. The first-order conditions of the problem, together with assumptions regarding the shocks hitting the economy, define the model's steady state and its dynamic responses. At first sight, drawing from such an approach might seem unattractive. The testing apparatus of DSGE models reflect specifications that are small, highly theoretical and calibrated to match "deep" parameters. That a typical macro-model can and should be evaluated in that vein can be motivated by some simple algebra. Assume a linear model summarised in (1), in the form of a (log-) linear approximation to the cyclical behaviour of the model, i.e. of the deviations from steady state.

$$A \cdot \mathrm{E}_t\left[x_{t+1}\right] = B \cdot x_t + C \cdot z_t \tag{1}$$

Model (1) postulates a dynamic, theory-based relationship between time t and t+1 among variables in vector x_t, some of whose elements are pre-determined but others have endogenous expectations, and variables in z_t assumed pre-determined.[1] Matrices in (1), as is standard, contain elements constrained by the theory underlying the model, including steady-state conditions. The number of parameters tends to be small and largely calibrated. In this case, any statistic generated by the model can be compared with its corresponding data-based statistic. Model (1) can be solved following a number of methodologies, of which linear techniques are widespread (e.g., Blanchard and Kahn, 1980, Klein, 1999, Juillard, 1999). Such techniques transform the system into the equivalent backward-looking system in (2), where y_t contains the non-pre-determined elements in x_t, and k_t the pre-determined ones.

[1] Expectations of pre-determined variables may differ from actual values, but only by an exogenous un-forecastable process.

Expression (2) can be taken to the data and doing so is becoming increasingly popular (e.g., Smets and Wouters, 2002).

$$\begin{aligned} y_t &= M \cdot k_t + N \cdot z_t \\ k_{t+1} &= P \cdot k_t + Q \cdot z_t \end{aligned} \tag{2}$$

In terms of (1) and (2), the AWM in its current version could be interpreted as a model in which all variables except one (the exchange rate) are pre-determined, a most unlikely assumption. In this respect, standard DSGE techniques might not seem to be of any particular interest in analysing the model. There is, however, an alternative view by which the AWM does not necessarily embody the unpalatable assumption mentioned before. A line of defence of the model, probably preferred by old-style econometricians, is that the model embodies endogenous expectations, but that these are embedded in its structure.[2] Heuristically, this can be understood if the first equation in system (2) leaded one period is used to solve for expectations of y_t in (1), leading to (3). System (3) is now a system in which endogenous expectations have disappeared, which means that it can be solved without reference to the future. (Leads of pre-determined variables pose no problem in this respect.)

$$A \cdot \mathrm{E}_t[x_{t+1}] = A \cdot \mathrm{E}_t \begin{bmatrix} y_{t+1} \\ k_{t+1} \end{bmatrix} = A \cdot \mathrm{E}_t \begin{bmatrix} M \cdot k_{t+1} + N \cdot z_{t+1} \\ k_{t+1} \end{bmatrix} = B \cdot \begin{bmatrix} y_t \\ k_1 \end{bmatrix} + C \cdot z_t \tag{3}$$

Within this framework, the AWM can be understood as a model in which endogenous expectations have been solved out (in terms of pre-determined variables) and embedded in the system. The problem with this approach, though, is that the embedding process implicitly transforms all the matrices in the system, whose true structural form cannot afterwards be recovered because there is no one-to-one mapping between the original structural form of the system and the mentioned semi-reduced form. In other words, the structural form of the model is a linear combination of the true structural matrices in which expectations and frictions in the economy can no longer be disentangled. The AWM in its backward-looking form, thus, becomes the representative of a *family* of forward-looking theoretical models: those whose structure after eliminating expectations would be compatible with the AWM. Nevertheless, the important point is that within this framework, it is now possible to

[2] In that sense, the model, though essentially a large-scale VAR or VECM, can be thought of as an empirical approximation to optimised decision rules and loss functions typical of DSGE models.

understand why applying standard DSGE approaches to the analysis of the AWM is not necessarily uninteresting.

The first set of model analysis exercises takes precisely this approach. As is done customarily in DSGE studies, the business cycle properties of the AWM model are analysed by means of statistics extracted from stochastic simulations, to investigate whether business-cycle characteristics present in the data can be replicated. This we examine in two ways: first, the standard moments and cross-correlations exercises, second, an examination of the periodicity and long-run variance properties implied by the model and data.

The second set of tests attempts to measure the model's predictive accuracy. Again, this exercise must be understood in the context of the family of theoretical models the AWM represents. Assuming that model (1) is the 'true' data generating process, in the sense that it is able to capture a sufficient set of relevant empirical features in the data, it is intuitive to expect that it would be able to provide forecasts of above-average quality, at least provided that these features remain in the future. Otherwise, it could be argued whether empirical features captured by the model are truly relevant. Obviously, structural breaks and important policy changes may render the model obsolete and severely affect its usefulness, and this has to be taken into account to the extent possible. However, claiming that a model is a correct representation of the data whose forecasts can systematically be improved upon with alternative models is a contradiction. To the extent that the AWM is a representative of the 'correct' model of the economy, therefore, it should inherit 'good' forecasting properties. The second exercise thus tries to exploit this intuition by exploring the model's forecasting performance, focusing in particular on possible gains made in forecasting by the presence of explicit steady-state conditions.

The organisation of the paper is as follows: Section 2 briefly discusses the Area Wide Model. Section 3 provides background to a discussion of the stylised facts in business-cycle analysis. Section 4 discusses the method of stochastic simulation used to derive and bootstrap the various outcomes against which we can judge the matching of historical stylised facts, with Section 5 reporting results of the exercise. Section 6 supplements the analysis by analysing spectral density measures. Section 7 describes the forecasting exercise. Section 8 concludes.

2. The AWM: Data and Specification

The AWM treats the euro area as a single economy. Its structure is standard, having a long-run classical equilibrium with a vertical Phillips curve but with some short-run frictions in price/wage setting, factor demands etc. Consequently, activity is determined in the short run by demand – given incomplete nominal adjustment – but supply determined in the longer run with employment having converged to a level consistent with the exogenously given level of equilibrium unemployment. In addition, stock-flow adjustments are accounted for, e.g. by the inclusion of a wealth / cumulated saving term in consumption. At present, the treatment of expectations in the model is limited; with the exception of the exchange rate (modelled by uncovered interest parity) the model embodies backward looking expectations. The AWM is estimated using the ECB's area-wide database. This data has been widely used in other publications: inter alia, Stock *et al*. (2002), Coenen and Wieland (2000), Galí *et al*. (2001), Camba-Méndez and Rodriguez-Palenzuela (2002). Applications of the model include examining the performance of policy rules and the monetary transmission mechanism (Dieppe and Henry, 2002; McAdam and Morgan, 2002), real exchange rate determination (Detken *et al*., 2002), Phillips curve estimation (McAdam and Willman, 2002).

3. Stylised Facts of Business-Cycle Analysis.

Documentation of "stylised facts" has a long tradition in empirical business-cycle research. Such facts summarise the main regularities in the macroeconomic variables and describe *volatility* (measured by standard deviations), *persistence* (auto-correlation) and *co-movements* (cross-correlations) between macro variables. Delineating these facts help develop and test theories of the business cycle (Bachus and Kehoe, 1992), build leading indicators (Stock and Watson, 1993) and, importantly in our context, calibrate and evaluate macro-econometric models (Watson, 1993, Canova, 1994, Brayton *et al*., 1997, Camba-Méndez and Pearlman, 1998, Roeger, 1999, Amano *et al*., 2000).

Standard macro stylised facts might include the following: [3]

[3] For other attempts to catalogue stylised facts, inter alia, Danthine and Donaldson (1993) and Fisher and Whitley (1998).

- Consumption is less volatile than output. This accords with the literature on consumption smoothing (i.e., Permanent Income and Life-Cycle Hypotheses) and is widely reported e.g. Kyland and Prescott (1990), Bachus and Kehoe (1992). It also appears that non-durable consumption is substantially less volatile than total consumption. Government consumption is considered more variable than output (Bachus and Kehoe, 1992) and acyclical (Cooley, 1997).
- Investment is around 2 to 4 times more volatile than and strongly correlated with output, (Mendoza, 1991). It is also appears to be pro-cyclical and co-incident. This accords, for example, with multiplier-accelerator models.
- The relationship between real and nominal magnitudes (i.e. output and prices) has been controversial. If demand shocks to output dominate those of supply then the price – output relationship would tend to be pro-cyclical, otherwise counter-cyclical. Indeed this type of partitioning has proved popular in VAR identifying work, Blanchard and Quah (1989). Judging by economic theory, under Keynesian reasoning such a relationship should be pro-cyclical given the Phillips Curve. In the Real Business Cycle literature, however, this correlation is assumed negative lending support to the supply-shocks interpretation (Cooley and Ohanian, 1991, Backus and Kehoe, 1992). Chada and Prasad (1994) however partly reconciled these views (and the demand v. supply debate) by suggesting that de-trended prices were strongly counter-cyclical but that there was a positive correlation between inflation and the trend deviation of output. In a similar vein, den Haan (2000) suggested that the correlation was positive in the short-run (i.e. demand dominated) but negative in the long-run (i.e. supply-shocks dominate). Overall, the counter-cyclicality of prices is now a widely accepted stylised fact. Cooley and Ohanian (1991), furthermore, find this negative correlation to be robust to different de-trending methods and Boone and Hall (1999) find prices counter-cyclical particularly in the post-war period using their stochastic trend method.[4] In terms of dynamic relations, Prices are generally less volatile than output, with the exception of high inflation countries, Christodoulakis *et al*. (1995) and Boone and Hall (1999). Cooley and Hansen (1994) furthermore suggested that prices lead output, and that inflation lags GDP.
- The (GDP deflator) mark-up over marginal costs is conventionally regarded as counter-cyclical (Rotemberg and Woodford, 1991, Oliveira Martins *et al*., 1996a) though with clear differences across sectors (Oliveira Martins *et al*., 1996b).
- There appears a weak money-output correlation over time, Garrat and Scott (1996). Furthermore, nominal short-term interest rates lead real output, Fuhrer and

[4] See also Ravn and Sola (1995).

Moore (1995). This might be consistent with borrowing rates effecting marginal investment decisions or authorities pre-emptively using monetary policy to control inflation and output. Cooley and Quadrini (1999) and Cooley and Hansen (1998) further suggest a strong negative monetary growth and nominal interest rates correlation (the so-called Liquidity Effect). Notably, for the euro area, Nicoletti-Altimari (2001) finds that money leads inflation. Finally, Diebold and Li (2002) suggest that short rates are more volatile and less persistent than long rates.

- On the labour-market side, unemployment (vacancies) is counter-cyclical (pro-cyclical) and lags (leads) real output, Fisher and Whitley (1998). This leads to the familiar negative relationship between unemployment and vacancies over the cycle: the Beveridge curve. Other notable stylised facts of labour markets have been the suggestion that productivity is more variable than real wages (Andolfatto, 1996), that private- is more volatile than public employment as well as the aspect of labour hoarding on employment variability over the cycle (Burnside *et al*. 1993). Notably, real wages also tend to be weakly correlated with output. This is in line with the criticism of early DSGE models: that the required intertemporal substitution of labour supply exceeded that found in data (Ball, 1990).

Of course, the existence of stylised facts at the euro-area level, presupposes the existence of a euro-area business cycle.[5] Whether there has been an improved correspondence of business cycles across European Union countries (i.e. effectively a euro-area business cycle) in recent times is still a matter of debate – compare the sceptical Dickerson *et al*. (1998) with Christodoulakis *et al*. (1995), Artis and Zhang (1997) and Peersman and Smets (2001). However, with a common monetary policy now operating among the EU11 geared towards euro-area aggregates, the question of whether we can identify aggregate business-cycle regularities that might inform policy and be reflected in models is, therefore, of special importance. Although there has been a considerable literature analysing stylised facts for other countries and areas (e.g., Basu and Taylor, 1999, Christodoulakis *et al*., 1995), there has been necessarily less work at the euro-area level and none in conjunction with euro-area modelling.

[5] See also Agresti and Mojon (2001).

4. Model Evaluation Using Stochastic Simulation

4.1 Methodology for Stochastic Simulations

To evaluate the model's ability to replicate these stylised facts we use stochastic simulation.[6] The technique for generating the stochastic shocks involves reproducing the characteristics of the observed random disturbances over the historical sample period – but adapted for use with non-linear rational expectations models. In particular, we used the Cholesky identification scheme.[7]

Given a generalised non-linear model, $\Gamma_i(y_t, x_t, \theta_i) = v_{it}$, *i=1,…,n* and *t=1,…T.* Where y_t is an n-vector of endogenous variables, x_t is a vector of predetermined variables (exogenous and lagged endogenous), θ_i is a vector of coefficients, v_{it} is an error term and the function Γ_i may be non-linear in the variables and coefficients . Assume that the first m equations (*m<n*) are stochastic, whilst the rest are identities or quasi-identities.[8] Given the vector of historical residuals in selected behavioural equations, we derive the variance-covariance matrix, *Ω* as $\frac{1}{T}\hat{v}\hat{v}'$ where $\hat{v}$ is the *m*T* matrix of values of v_{it} $\forall t = 1,..T$. A Cholesky decomposition is then performed on Ω, generating the lower-triangle matrix, *L: LL′ = Ω*. From a random number generator we make drawings from a standard Normal distribution, *k*, *k ~ N (0,1)*. The shocks applied to the model (v^*) are drawn from the pre-multiplication of the decomposition of the historical variances and the random number generated: $v^*=Lk$ – these have the same properties of the original historical co-variance matrix: E (v*v*') = Ω. Having drawn up the appropriate matrix of historical innovations, we can then set up a replication procedure. For each vector of shocks at period t, the model is simulated from the start date to the date of the simulation period, and agents form expectations of the future based on their information set at t. We then make the transition to period t+1 and update the information set and repeat over the whole replication sample. So for each replication, a set of shocks are drawn, the model is simulated, time is

[6] Adelman and Adelman (1959) provide an early reference for the stochastic business-cycle analysis of macro-models

[7] This method has been widely applied to the stochastic simulation of macro-models – inter alia, Frenkel *et al.* (1989), Masson and Symansky (1992), Bryant *et al.* (1993), Hughes-Hallett and McAdam (2001).

[8] Thus, in line with Fair (1998) and others, we do not shock identities or quasi-identities such as policy rules, the term structure or the uncovered interest parity condition etc. Such equations are typically not estimated but imposed or calibrated. Consequently, their residuals do not have the usual interpretation of recoverable innovations.

advanced and then another set of shocks drawn and so on.[9] The agents' information set (at t) contains all data up to and including t (that is, including any expectations formed at t). To deal with rational expectations, we use the *Stacked Time* algorithm, Juillard et al. (1998). In practice the *STACK* for the stochastic simulation is set well in advance of the actual simulation horizon to isolate the dynamic trajectory of the model from arbitrary terminal conditions. (We run 400 replications over 100 years). Note, that there was no truncation of the distribution of results due to failed replications (from non-linearity, numerical instability, illegal arithmetic etc) – see the discussions in Fisher and Salmon (1986), McAdam and Hughes-Hallett (1999).

For each replication, *j*, we thus obtain a solution for endogenous variable *i*: y_{it}^{j} with expected value over all replications, $\frac{1}{J}\sum_{j=1}^{J} y_{it}^{j}$. In constructing our results, we ran each simulation over a 100 year time period then discarded the first and last 35 yearly observations from each replication to remove any biases that might arise from starting or finishing each simulation from particular values. This is standard practice (e.g., Bryant *et al.*, 1993) and leaves a sample dimensioned to the historical one. Finally, and as is equally standard, the stochastic simulations are performed on the model's steady state horizon to remove the effect of arbitrary historical conditions.

Thus, shocks distributed according to the historical distribution of residuals are applied to selected behavioural equations. These comprise: real Consumption and real Investment; the GDP, Consumption, Investment, Export and Import deflators, real Import and Export demand, nominal Wages and Employment. Notably, we cannot expect the stochastically-simulated model to replicate all the variability in sample data since (a) only *m* of the *n* equations are shocked (b) the model is essentially a closed-economy one precluding some important sources of variability and, similarly, (c) certain variables (such as government expenditure, commodity prices etc) are exogenous.

[9] Thus, (real and nominal) exogenous variables remain exogenous and are extrapolated into the steady state using conventional balanced growth closures.

5. Historical and Simulated Stylised Facts

5.1 Historical Stylised Facts

In Table 1, we can see many of the afore-mentioned stylised facts for key variables (y) when transformed into covariance stationary processes. We use deviations from the well-known HP filter: Y= *y - HP(y)* [10]. Pro-cyclicality (counter-cyclicality) between Y_t and X_t implies *corr* $(Y_t, X_t)>0$ (<0). Furthermore, if a series is a "lagging indicator", its highest correlation with Real GDP occurs *after* the contemporaneous realisation of output – i.e., $\max_{k>0} \left| corr(Y_t, X_{t+k}) \right|$.

[Table 1]

Most series appear pro-cyclical with the exception of the nominal exchange rate, the price level, unit labour costs, the mark up, nominal wages, unemployment and all interest-rate series. Real Public consumption and nominal money balances, however, appear acyclical. Furthermore, most series appear to be contemporaneous. Series that appear to be leading (lagging) indicators of the output cycle are the price level, unit labour costs, nominal wages and all interest rate series (Inflation, the exchange rate series, real wages, employment and unemployment). On volatility, most series are less volatile than output with the exception of investment and the trade, exchange and interest rate variables. All series are highly persistent as judged by their AR (1) values.

5.2 Model-Generated Stylised Facts

Table 2 displays the moments generated by the stochastic simulation (described in section 4).[11] Broadly speaking, the model performs well. Those series, which were pro- or counter-cyclical in the data, remain so as do those that were leading or lagging

---

[10] The HP filter minimises the following loss function for a given smoothing parameter λ: $\min_{\gamma} \sum_{t}^{T} \left[(y_t - \gamma_t)^2 + \lambda((1-L)^2 \gamma_t)^2 \right]$. Where γ_t represents the trend of y_t series and L is the lag operator. Essentially this function trades off goodness-of-fit, $(y_t - \gamma_t)$ against trend smoothness, $(1-L)^2 \gamma_t$. The allowance for the latter is thus increasing in λ. At extremes, we have linear-trend (λ→∞) and perfect-interpolation (λ→0) cases. For a common comparison, we apply this filter to both our historical and stochastically-simulated data.

[11] The simulated results in Table 2 exclude money (real and nominal) and real government consumption since these series are either exogenous in the model or omitted.

the output cycle. Similarly, the ratio of second moments and AR properties are largely in line with the historical data.

[Table 2]

Comparing cross-correlations across data and models in a more exact manner, however, is a typically difficult exercise. For instance, any such fixed statistical metric is bound to suffer relatively low power.[12] Consequently, we make use of our various stochastic replications and its resulting bootstrap: that is to say, having accumulated each point estimate from the individual stochastic replications – i.e., $\sigma_{ij}, AR_{ij}(1), corr(Y_{ij}, X_{ij})$ – we then sort them and draw up the resulting distribution. We can then judge supporting hypotheses – e.g., $H_1 : corr(\cdot)^{\text{data}} \geq corr(\cdot)^{\text{model}}$ – and their associated probability-values etc. Thus, we construct the usual bootstrap test by finding τ_T^L, τ_T^U such that $\Pr(\rho \leq \tau_T^U) = 1 - \frac{\alpha}{2}, \Pr(\rho \geq \tau_T^L) = 1 - \frac{\alpha}{2}$, where $1-\alpha$ is the chosen confidence level for the Lower (L) and Upper (U) regions and T denotes sample size. Consequently, historical moments which appear in extreme values for the percentile (i.e., $>1-\alpha$ or $>1-\frac{\alpha}{2}$) indicate statistically significant differences.

It can be seen that in many cases the differences are insignificant (i.e., with probability values > 0.05). Table 2 reports a matrix of 209 *p*-values for which 135 are insignificantly different from the historical data – a "success" rate of 65%. This would appear to be at least competitive with other models in this class (e.g., Brayton *et al*., 1997, Roeger, 1999). There are, however, some important points. For instance, persistence values are generally lower (and statistically different) in the stochastically-generated data than in history; we might, however, expect there to be more persistence mechanisms (and deviations from trend) in-sample than could be simulated in a (largely) linear model which also incorporates stabilising mechanism (error-correction terms, stabilising policy feedback rules etc). Similarly, the variability of the simulated

[12] For instance, Walpole and Myers (1978) suggest testing the individual significance of the correlation coefficients at different lags and leads with the transform $\frac{\sqrt{T-3}}{2}\left\{\ln\left(\frac{(1+r^s)(1-r^m)}{(1-r^s)(1+r^m)}\right)\right\}$. Where T is sample size, and r^s and r^m are the sample and model-generated correlation coefficient respectively. Under the null of $r^s = r^m$, the statistic follows a standard normal distribution. One problem with this statistic, however, is that the r^m value comes from a distribution (i.e., the various Stochastic simulations) and thus has sampling distribution and size aspects. Failure to control for these leaves the statistic and its inference potentially un-robust and uncertain. For this reason, we use the bootstrapped comparison.

data – accounting for around 70% of actual output variability, $(0.8602/1.027)^2$ – falls below[13] that of history and thus would be statistically different. This lack of variability could be a casualty of the non-modelling of key variables (such as the foreign sector, commodity prices etc) as well as the fewer (compared to history) propagation mechanisms found in models. However, the cross-correlations are, in the large majority of cases, insignificantly different from the data and this is also true of the AR(1) parameters; interestingly, the interest series all fail this criteria.

Furthermore, , whilst in the historical data, both exchange rate series (nominal wages) lag (lead) the cycle, the same series from the stochastic simulation are contemporaneous. Although the interest rate series are robustly counter-cyclical (as judged by the stream of high and negative output cross-correlations), the actual contemporaneous value is relatively weak and – in the case of the short rate – of the wrong sign. Allied to this is the fact that the auto-regressions of the interest rate series are all significantly different (and lower) than history.[14] Plausible reasons for this include the fact that in the model's (Taylor) monetary policy rule the smoothing parameter on interest rates is set to zero; we know, however, from empirical studies – e.g., Clarida *et al*. (1998) – that this parameter is typically estimated at around 0.6 - 0.9. The fact that we do not represent this *gradualism* may imply too little persistence in the short rate.

6. Frequency Domain Analysis.

6.1 Overview

Hitherto we have worked in the time domain, but frequency-domain (or spectral analysis) has also been suggested as a means of model evaluation (e.g., Watson, 1993, Diebold *et al*., 1998).[15] Though any time series can be (Fourier) transformed into its spectral representation, working directly in the frequency domain gives further insight into the model's dynamic and business cycle features and, in particular, how

[13] Notably, King and Rebelo (1999) find that there model accounts for a similar degree of output variability.
[14] Given the strong link between interest rate movements and exchange rates – and hence trade volumes and relative prices – the trade and exchange rate variables are also significantly different in volatility.
[15] Of course, any covariance stationary process has both a time and frequency domain representation. However, explicitly dealing in the frequency domain is useful for expositional purposes and allows us to encompass some related literature.

important cycles of different frequencies are in accounting for its behaviour, periodicity and long-run variance.

At the outset, however, let us motivate the conditions under which their application to macro models is valid. Assume that linearising $\Gamma_i(y_t, x_t, \theta_i) = v_{it}$, yields, $y_t = \theta_1 y_{t-1} + \theta_1 x_{t-1} + v_t$;substituting back to an arbitrary date, we see more clearly the sources of cyclical behaviour involved: $y_t = \theta_1^t y_0 + \sum_{j=0}^{t-1} \theta_1^j \theta_2 x_{t-j} + \sum_{j=0}^{t-1} \theta_1^j v_{t-j}$. Thus, we have essentially three sources of cyclical behaviour: initial conditions, $\theta_1^t y_0$, and that associated with the cyclical path and characteristics of the truly exogenous variables and the underlying shocks: $\sum_{j=0}^{t-1} \theta_1^j \theta_2 x_{t-j}$ and $\sum_{j=0}^{t-1} \theta_1^j v_{t-j}$ respectively. Since the way in which the exogenous variables are set/extrapolated may or may not contain complex elements – depending on how the user defines them – our focus rests inevitably on the errors and initial conditions. On initial conditions, although the effect of initial impacts decay, $\lim_{t\to\infty} \theta_1^t \to 0$, there is an underlying cyclical dynamic – since $\theta_1^t = \psi \Lambda^t \psi^{-1}$, where $\Lambda^t = diag(\lambda_j)$, which implies that each endogenous variable is a linear combination of the powers of all the eigenvalues in the system whose complex part will induce cyclical behaviour. For the residuals, we have $E(v_t) = 0$ and $E(v_{t-i,} v_{t-j}^{'}) = \begin{cases} 0 & i \neq j \\ \Pi & i = j \end{cases}$. Thus, defining $h = \sum_{j=0}^{T} \theta_1^j v_{t-j}$, we have $E(h_t, h_t') = \Pi_0 = \Pi + \sum_{k=1}^{T-1} \theta_1^k \Pi \theta_1^{k}{}'$ and $E(h_{t,} h_{t-1}^{'}) = \theta_1 \Pi_0 - \theta_1^T \Pi \theta_1^{T-1}{}'$. Given $\lim_{T\to\infty} \theta^T \vert \theta < 1 \to 0$, we have $E(h_{t,} h_{t-m}^{'}) \approx \theta_1^m \Pi_0$. Consequently, if λ_j are complex so too will $E(h_{t,} h_{t-m}^{'})$ be.

The spectral representation of the variable y_t is $\int_{-\pi}^{\pi} e^{iwt} dy(w)$, where, as standard, w is frequency and $i = \sqrt{-1}$. And its model-based spectral representation would be $y_t = (1 - \theta_1 L)^{-1} \left[\theta_2 \int_{-\pi}^{\pi} e^{iwt} dx(w) + \int_{-\pi}^{\pi} e^{iwt} dv(w) \right]$. Thus, knowledge of the parameter vectors, underlying shocks and the paths for the exogenous variables allows us to take into account the cyclical aspects of the model and, in turn, build its spectral representation. (See also Chow, 1975).

The remaining issue is stationarity; spectral tools are stationary tools and should invariably be applied to stationary, stable models.[16] To continue, let us return to the more general state-space formulation of equation (1), $A \cdot \mathrm{E}_t[x_{t+1}] = B \cdot x_t + C \cdot z_t$. This generates $x_{t+1} = A_1 x_t + A_2 z_t$ where $A_1 = A^{-1} \cdot B, A_2 = A^{-1} \cdot C$ [17]. Here *x*, the state vector, is assumed to contain the N endogenous variable types partitioned into N-L predetermined variables and L lead variables. We know (Blanchard and Kahn, 1980) that the solution will be stable and unique if A_1 has $|\lambda_i| > 1, i = 1,...,L$ and $|\lambda_j| \leq 1, j = 1,..., N - L$; additionally A_2 is assumed $|\lambda_k| < 1, k = 1,..., K$. This ensures a stable, model solution but not necessarily a stationary one (since there may be growth, trends and non-stationary variables such as the capital stock, population growth etc). The alternatives thus are either to represent the model in equilibrium deviation form or to filter the output of the (stochastically-simulated) model to render it stationary. We prefer the latter since it ensures consistency between the data used to draw the correlation matrix in section 5.2 and that used to derive its spectral characteristics.

Typical applications of spectral methods to models include: model parameterisation based on minimising some distance function (of model and data spectra), supplementing model variability until the full stochastic moments and spectral properties of the data are met. For a given covariance stationary univariate series, $\{ x_{t,}\ t \in \mathbb{R} \}$ with mean $E(x_t) = \mu$ and auto-covariance function $E(x_t - \mu)(x_{t-k} - \mu) = \gamma_k = -\gamma_k$ (given $\sum_{k=0}^{\infty} \gamma_k < \infty$), the spectral density can be written as:

$$f(\omega_j) = \frac{1}{\pi} \left\{ \hat{\gamma}_0 + 2 \sum_{k=1}^{m} \lambda_k \hat{\gamma}_k \cos(k\omega_j) \right\} \tag{4}$$

Where frequency $\omega_j = \frac{\pi j}{m}, j \in [0, m]$, $\omega_j \in [0, \pi]$,[18] *m* is window size, and λ_k are lag window (Bartlett) weights. In defining *m*, we consider two solutions: bandwidth selection based on the Andrews (1991) procedure and a common default $m = 2\sqrt{T}$.[19]

[16] In the previous case, stability simply boiled down to whether the condition $|\theta_1| < 1$ was met.

[17] Rewriting in this way of course pre-supposes that the matrix *A* is invertible (e.g., see Klein, 2000).

[18] Given symmetry *w* can be $\in [-\pi, \pi]$ or $\in [0, \pi]$, where small ω relate to long-run relations and short-run near π.

Using the spectrum, we can judge how important cycles of different frequencies are in accounting for the behaviour of a series, its periodicity and long-run variance. The period (i.e., the length of a cycle) associated with each frequency is calculated as $\frac{2\pi}{\omega_j}$ and overall periodicity by $\frac{2\pi}{\omega_j}\Big| \max f(\omega_j)$. Furthermore, for long-run variance, we know that the standardised spectrum [20] evaluated at zero frequency, $f(0)$, provides a consistent estimate of Cochrane's (1988) measure of persistence (i.e., a measure of a series' deviation from long-run trend). [21]

6.2 Graphical Results

Table 3 and Figures 1 implement these techniques for historical and simulated data – using Output, Consumption, Investment (all real) and Employment[22]. Looking at historical data, we can see that all four series appear highly persistent and essentially I(1). The point estimates of long-run variance are relatively high (although persistence measures derived from such techniques tend to be in this range, e.g., Campbell and Mankiw, 1987)[23]. Their periodicity's range from between 4½-10 years (with Investment having the longest-lived cycle) which seems reasonable for macro data generally and in the euro-area case in particular (Stock and Watson, 1999, Orlandi and Pichelmann, 2000)

[Table 3]

[Figures 1]

Looking at the historical spectral densities and their standard errors (Figures 1 – i.e., $f(\omega_j) \pm 2\sigma_{f(\omega_j)}$), we see that all series display the typical Granger (1966) shape with

---

[19] These might be considered opposite ends of the range – from variable-specific to sample-specific bandwidth selection. See the discussion in Andrews and Monahan (1992).

[20] Where standardisation is by π / γ_0.

[21] Similarly, Campbell and Mankiw (1987) derive their persistence measure parametrically from the spectral density evaluated at $\omega = 0$ using the polynomial estimates from an ARMA representation.

[22] The same variables as in Watson (1993). As before (Tables 1 to 2), and as is necessary for spectral analysis, we transform the variables into covariance stationary processes by taking the difference of the log of the series and its HP filter: *y – HP(y)* .

[23] See also Romer (2000, ch 4).

high power at low frequencies and declining power thereafter.[24] The same can largely be said for the artificially generated data. We have the same familiar shape and, at typical business cycle (4½-10 years) frequencies, the stochastically generated results are largely contained within the standard errors of the historical data. Similarly, the persistent estimates of the model are typically in the range of +2 quarters of the data and the estimates of long-run variance encompass one another well over both bandwidth selection methods. Notably, the rate of decay of the model spectra appears slower than that of the data.[25] The notable exception is output: the model assigns a larger periodicity than that found in the model. Since the major demand components behave well in that respect, the culprit would appear to be either net trade modelling or (exogenous) public consumption. On a broad perspective, therefore, the model would appear to share similar spectral characteristics to the data; more similarities than differences at least.

7. Measures of Predictive Accuracy.

The goal of this section is to check to which extent the model captures relevant empirical regularities in the data, as was done in the last section, but following an alternative approach. An assessment of its performance will be obtained by simulating forecasts. The purpose is again broad enough to involve the model as a whole, i.e. not to analyse specific parts of the model, but rather the system in its entirety. The approach hinges on the idea that the economy is forecastable at least to some extent, which seems reasonable, and that only models able to capture relevant regularities can deliver above-average forecasting performance. Although backward-looking in nature, the model purports to describe structural features of the euro area economy: links between empirical regularities and structural priors are given by the data-based estimation of the parameters in the model. Thus, the AWM can be loosely described as a system whose links with the data are heavily constrained by theory. Whether this system can approach the forecasting performance of systems tailored to the data, i.e. avoiding these heavy prior constraints, is precisely the question at stake.

[24] Asymptotic standard errors are given by: $$\sigma_{f(\omega_j)} = \begin{cases} \sqrt{\frac{2}{v}} f(\omega_j) & \forall j \in [1, m-1] \\ \sqrt{\frac{4}{v}} f(\omega_j) & \forall j = 0, m. \end{cases} ; v = \frac{2T}{\sum_{k=-m}^{m} \lambda_k^2}.$$

[25] This was also a finding of Diebold *et al.*'s (1998) cattle-cycle model.

The approach taken is in spirit that of Stock and Watson (1999) and related literature, although with some important qualifications in order to accommodate the more complex structure of our forecasting model. In Stock and Watson's approach, recursive forecasts are performed using versions of models re-estimated ignoring, to the extent possible, information contained in the forecast period. Typical models in this literature are random walks, auto-regressive (AR) processes, VARs and indicator- or factor-based single equations etc. Although the AWM belongs to a class of relatively small macro-models, it certainly does not belong into the class of models mentioned previously. For instance, it is obviously impossible to drop a significant number of observations without seriously affecting the quality of the estimates, which means in turn that the number of simulated out-of-sample forecasts must be kept relatively small. Conclusions from this exercise, thus, are inevitably subject to important uncertainty. This factor has been countered in our approach by repeating the same forecast exercises using all available information in order to assess the impact of parameter re-estimation, i.e. performing in-sample dynamic simulations with the model in settings similar to the previous ones. The in-sample exercises have been run over the same period as the out-of-sample ones but also over a longer one, beyond what is possible (or advisable) in an out-of-sample setting. Links between the two exercises will be exploited in order to properly base our claim that the more extensive in-sample exercises contain useful information related to the model's forecasting performance, and not to possible data mining when the model was originally estimated.

In accordance with the mentioned literature, forecasting ability will be measured in terms of root mean square errors (RMSE) of the generated pseudo-forecasts. Another important qualification needed in the case of a complex structural model is a deeper analysis of how good a measure of forecasting ability are RMSE. This step can be normally omitted in the case of simpler models with no explicit structural content. When using a structural model, however, it turns out to be essential to gain an understanding of what factors are affecting most the forecasting performance. A simple decomposition of generated RMSE will be proposed with this in mind. The decomposition will be of a non-structural nature, in order to establish a comparison between the AWM and simpler benchmark models, but will be shown to have implications for a deeper understanding of the structural properties of the model.

Last but not least, in building forecasts with the AWM, residual adjustment will become an important element. In the exercises reported below, explicit use of add-factor approaches will be made, both in order to assess their importance in the model forecasts and also in terms of looking for possible reasons why add-factors could improve the outcome. Obviously, only simple residual-projection strategies will be adopted. It should be stressed that the mechanical rules for "add-factoring" considered here are stylised and bear little resemblance to the actual practices used in forecasting. For example, the setting of add-factors in practice, is usually based on information from off-model leading indicators and expert judgement. In addition, different approaches for setting add factors are used in different equations depending on the assessment of the nature of shocks hitting the economy.

7.1 Analysis of the RMSE of the forecasts

The framework

As said, simple residual-projection strategies will be adopted: from projections in which residuals will be set to their in-sample average (normally around zero) to more sophisticated rules. All these rules will be followed according to their expected impact on the forecasts, and will be based on a prior view on what adverse factors can be overcome by add-factoring. These views originate in a formal view of the forecasts that needs being expounded. Let us thus assume that variables to be forecast are stationary: a Wold decomposition for them can then be found, (5). In the expression, variable x_t is decomposed into a deterministic time-varying component μ_t and a stationary moving-average of possibly infinite dimension of the process ε_t, itself stationary and independently distributed. Expression (5) represents the true, data generation process of the variable to forecast, and is thus unknowable. Forecasts will be based on models that (hopefully!) approach the process (5) but will in general differ in one aspect or another from (5). Expression (8) expresses the forecasting models, again in the form of a Wold decomposition of the models.

$$x_t = \theta(L) \cdot \varepsilon_t + \mu_t, \quad \text{where} \sum_i^{\infty} \theta_i^2 < \infty, \ \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty \tag{5}$$

$$x_t = \varphi(L) \cdot \eta_t + v_t, \quad \text{where} \sum_i^{\infty} \varphi_i^2 < \infty, \ \mathrm{var}(\eta_t) = \sigma_\eta^2 < \infty \tag{6}$$

Forecasts with the models are as in (7), in which f_t^h is a stochastic forecast obtained from (8) h steps ahead. The error process in the forecast period is assumed to be given by a deterministic rule (such as $\eta_t^h = 0$). Note that v_t is made dependent on h to allow for model revisions between forecasts, but this is not *per se* a crucial element in the analysis. It is important to stress that forecasts are stochastic because (8) is a stochastic process, not because stochastic simulations will be run with the models: the fact that residual projection is deterministic means that the forecasts will sooner or later converge to a deterministic process with probability one. (We expand upon this later.)

$$f_t^h \equiv \mathrm{E}\left[x_t \mid \{x_{t-h}, x_{t-h-1}, \ldots\}, \varphi(L), \sigma_\eta\right] \equiv \mathrm{E}_{t-h}[x_t] = \varphi(L) \cdot \eta_t^h + v_t^h \tag{7}$$

One important aspect of the comparison between (5) and (8) is that they can be used to formalise some potentially important sources of forecast failure. Namely, there could be a false separation between deterministic and stochastic components of the variable on the one hand ($\mu_t \neq v_t$), and a false description of the law of motion of the variable on the other ($\theta(L) \neq \varphi(L)$). While the latter could lead to bad forecast performance in short-horizon forecasts, the former could have a longer-lasting influence. In what follows, strategies to project residuals will be explicitly linked to the sources of forecast failure tackled in each case.

The first set of sources of forecast failure, i.e. those stemming from wrongly modelling the deterministic component of the variable, are closely related to what in the forecasting literature has been identified as *deterministic shifts*, see Hendry and Clements (1998, 1999). As stated in Hendry and Clements (op.cit.), forecast failure can arise because of many factors, some potentially important being:

i) In-sample model mis-specification or data-based model selection;

ii) Poor estimation strategies or incorrect calculation of confidence intervals;

iii) Parameters depending on policy-regime changes;

iv) Structural un-forecastable breaks.

Not much will be said on points i) and ii), as no specific alternatives to the AWM's structure or used estimation techniques will be analysed. Point iii) will be tackled by assessing the degree of parameter instability in the model by the simple expedient of re-estimating the model with an artificially expanding sample. Point iv) will be assessed by suitable alternative residual adjustment and projection methods, with the goal in mind of replicating specific changes in μ_t by implicitly shifting v_t. Our understanding of point iv) is that it encompasses any change in the data-generating process (5) leading to a new distributional mean for x_t. In terms of the AWM, this means that the joint unconditional mean of the endogenous variables has to be identified. Considering that the AWM is mainly composed of accounting identities and stochastic equations expressed as equilibrium corrections mechanisms (EqCM), a log-linear approximation of the model could be expressed by (8), in which y_t is a set of I(1) endogenous variables, z_t a set of I(1) exogenous variables, d_t a set of deterministic variables and u_t a stochastic residual.

$$A(L)\cdot \Delta y_t = d_t + B(L)\cdot \Delta z_t - \alpha \cdot (\beta' x_{t-1} - \gamma' z_{t-1}) + u_t \tag{8}$$

In the framework set out in (5) and (8), forecasts would be made on a component of Δy_t, stationary by definition (i.e., x_t in (5) and (8) corresponds to Δy_t in (8)). Our definition of the deterministic part of x_t in (8) will then be given by the corresponding element of the unconditional expectation of Δy_t in (8), and the stochastic part by the difference between expectation and actual, as in (9). In this expression, the unconditional expectation $E(\Delta y_t)$ clearly results from solving the model as a deterministic set of equations, i.e. Δy_t-$E(\Delta y_t)$ is different from u_t in (8).[26] (Compare this with the expression for f_t^h in (7), an expectation made conditional on information from the recent past.) In other words, the expectation in (9) is given by the model's equilibrium condition for the *levels* of the variables and is thus a representation of the long run of the model. In consequence, our understanding of what deterministic shifts mean in the context of the forecasting model is: shifts to the steady-state conditions of the AWM. It is now clear that residual projection is a natural way to implement changes to these conditions, which is precisely what is attempted below. Note also that models expressed in difference terms, i.e. models that do not determine the level of the variable, will in general be in the short term more robust to deterministic shifts than models with a steady state embedded in them.

$$\begin{gathered} \Delta y_t = E(\Delta y_t) + [\Delta y_t - E(\Delta y_t)] \\ \text{where} \\ E(\Delta y_t) = A(1)^{-1} \cdot [d_t - B(1)\cdot E(\Delta z_t) - \alpha \cdot E(\beta' x_{t-1} - \gamma' z_{t-1})] \end{gathered} \tag{9}$$

A RMSE Decomposition

[26] In other words, $E(\Delta y_t) \equiv E_{-\infty}(\Delta y_t)$ will in general be different from $E_{t-h}(\Delta y_t)$.

As said previously, the forecast variable x_t will be assumed stationary and thus decomposable along the Wold decomposition, as in (5). Furthermore, the models used to forecast the variable all embody a (stationary) implicit distribution for the variable summarised in (8), where again the Wold decomposition has been used. Assuming that a h-step-ahead forecast f_t^h is generated using (7), the measures of accuracy of forecast used in the preceding sections are given by (10), the MSE (Mean Square Error). Reported RMSEs would simply be the square root of MSEs in (10). MSE is not made dependent on t because of the distributional assumptions made, but it depends on the forecasting horizon h.

$$\mathrm{MSE}^{\mathrm{h}} = \mathrm{E}\left(x_t - f_t^h\right)^2 \tag{10}$$

The expression for MSE leads naturally to a decomposition of the statistic using the Wold decomposition of actual and forecast. Taking into account that $\mathrm{E}\left(x_t - f_t^h\right)^2 = \mathrm{E}\left[\left(x_t - \mu_t\right) - \left(f_t^h - \nu_t^h\right) + \left(\mu_t - \nu_t^h\right)\right]^2$, $\mathrm{E}\left(x_t - \mu_t\right) = 0$ and $\mathrm{E}\left(f_t^h - \nu_t^h\right) = 0$, it is easy to show that expression (10) is equivalent to (11), which itself can be rewritten as (12). This expression is interesting because it decomposes the MSE along components about some of which we can a priori say something. Namely, the distributional assumptions made means that the first element is given and outside the control of the analyst. The only hope to reduce forecast error is by ensuring that: i) no bias is introduced in the forecast and ii) covariance between actual values and forecasts is such that their variances will be cancelled out.

$$\mathrm{MSE}^{\mathrm{h}} = \mathrm{E}\left(x_t - \mu_t\right)^2 + \mathrm{E}\left(f_t^h - \nu_t^h\right)^2 - 2 \cdot \mathrm{E}\left[\left(x_t - \mu_t\right) \cdot \left(f_t^h - \nu_t^h\right)\right] + \left(\mu_t - \nu_t^h\right)^2 \tag{11}$$

$$\mathrm{MSE}^{\mathrm{h}} = \mathrm{var}\left(x_t\right) + \mathrm{var}\left(f_t^h\right) - 2 \cdot \mathrm{cov}\left(x_t, f_t^h\right) + bias^2 \tag{12}$$

Furthermore, if the forecast is deterministic (i.e., $\mathrm{var}\left(\eta_t^h\right) = 0 \; \forall \; t > T - h$, where T is the forecast origin), it is easy to prove that the variance of the forecast will shrink as the forecast horizon is increased, as shown below.

$$\begin{aligned}&\operatorname{var}\left(f_t^h\right)=\mathrm{E}\left(f_t^h-v_t^h\right)^2=\mathrm{E}\left(\varphi(L)\cdot\eta_t^h\right)^2=\\&\sum_{s=0}^{\infty}\varphi_s^2\cdot\operatorname{var}\left(\eta_t^h\right)=\sum_{s=0}^{\infty}\varphi_s^2\cdot\sigma_\eta^2=\sum_{s=0}^{h}\varphi_s^2\cdot\sigma_\eta^2+\sum_{s=h+1}^{\infty}\varphi_s^2\cdot\sigma_\eta^2=\sum_{s=h+1}^{\infty}\varphi_s^2\cdot\sigma_\eta^2\end{aligned}\tag{13}$$

Hence, $\lim_{h\to\infty}\operatorname{var}\left(f_t^h\right)=0$, which means that $\operatorname{plim} f_t^h=v_t^h$. Obviously, the same shrinkage phenomenon applies to the covariance between actual and forecast. This means that in (12) all the elements except the first and last ones converge to zero as the forecast horizon is increased. The obvious implication is that in long-horizon forecasts avoiding biases is the only thing that matters. For empirically relevant horizons, this conclusion does not necessarily hold, as in their case a large variance for η_t may be the most important factor in looking for enhanced forecast accuracy. It is interesting to find whether a sample counterpart to (12) can be found, in order to assess the relative importance of all these factors. Unfortunately, this can be achieved only at some cost.

The sample expression for the MSE is given by (14), where it is assumed that the first forecast is done at time 0 and that the variable x_t spans until period T and is assumed to be present for at least h periods before time 0. A natural extension to express the MSE decomposition as in (12) would be to replace variances and means in the expression by their sample counterparts, as in (15). This expression, though, relies on the assumption that both means and variances do not depend on t, contrary to the rather eclectic underlying assumptions made in (13). We will nevertheless use (15) to decompose shown MSEs, in the conviction that such decomposition can lead to interesting insights. It is important to keep in mind, though, that conclusions reached may be affected.

$$\mathrm{MSE}^{\mathrm{h}}=\frac{\sum_{t=0}^{T-h}\left(x_t-f_t^h\right)^2}{T-h}\tag{14}$$

$$\text{MSE}^h = \frac{\sum_{t=0}^{T-h}(x_t - \bar{x})^2}{T-h} + \frac{\sum_{t=0}^{T-h}\left(f_t^h - \bar{f}^k\right)^2}{T-h} - 2 \cdot \frac{\sum_{t=0}^{T-h}(x_t - \bar{x}) \cdot \left(f_t^h - \bar{f}^k\right)}{T-h} + \left(\bar{x} - \bar{f}^h\right)^2 \quad (15)$$

where

$$\bar{x} = \frac{\sum_{t=0}^{T-h} x_t}{T-h} \quad \text{and} \quad \bar{f}^h = \frac{\sum_{t=0}^{T-h} f_t^h}{T-h}$$

7.2 A Description of the Exercise

The forecasts attempt to measure the 1- to *h*-step-ahead forecasting ability of the model, where *h* is set to 8 in the simulated out-of-sample exercise and to 12 in the in-sample exercise performed below. For the first exercise, the model is re-estimated each time on a subset of the original sample, and the re-estimated model is used to generate a forecast over the remaining part of the sample. Forecast errors are then calculated based on observed data. Exogenous variables for the AWM must be determined either using outside equations or taken from actual realisations. Although the first approach seems worth exploring, the second one was taken in order to simplify the analysis. As an informal benchmark, similar forecasts are run with simple alternative models based on pure time-series analysis of the data. Forecast accuracy will be analysed in terms of the first difference of the log of analysed variables. It is important to stress that this is a very tough criterion, because a forecast *h* steps ahead will imply that the rate of growth of the variable for a specific quarter will be forecast *h* quarters in advance. Many studies present forecast accuracy in terms of the difference between today's value and the value *h* quarters ahead of the level of the variable, thus exploiting the averaging properties inherent in a stationary process. In our approach, the forecaster has to face an *h*-step-ahead forecast without relying on the average accuracy of his/her forecasts for the periods in between.

Forecasts with the AWM involve the model as a whole, which means that forecasts for a large number of series are generated. Only three variables will be shown, though: GDP growth, GDP deflator inflation and consumption deflator inflation.[27] These three variables are a useful summary of model-based results, are relevant for the analysis of the euro area as a whole and are known to be forecastable with simple benchmark models alternative to the AWM. Furthermore, the three offer specific

[27] HICP has not been included due to the short sample available for the series in the data used, 1995 onwards. Linking HICP with national CPIs for previous periods is not an option in practice, due to definitional changes and detected important shifts in seasonal behaviour between HICP and (aggregated) CPIs.

challenges in a forecasting exercise. As shown in Figure 2 (upper panel), GDP growth (in the form of first difference of log of real GDP) shows a low degree of persistence and a noticeable degree of variability. The lower panel of Figure 2 shows GDP deflator inflation and consumption deflator inflation, again in the form of first difference of log of each variable. Two outstanding facts in the figure regarding these two series is their evident higher persistence, on the one hand, and the likely presence of shifts in their underlying means, on the other. It could even be claimed that inflation should be considered as an I(1) process, i.e. prices are I(2). This is not the view taken here, in which inflation will be considered a stationary process. Furthermore, although both variables broadly co-move, short but conspicuous separations take place here and there, mostly at times when oil prices were suffering a shock. All these factors, i.e. whether persistence, shifts in the underlying processes or the impact of exogenous information matters for a forecast, will be at the core of the following discussion.

Model re-estimation meant that making forecasts covering large spans of time could lead to small-sample estimation biases in the initial forecasts – a cost to pay in any out-of-sample exercise in which increasing parts of the observed sample are ignored. Instead, it was deemed preferable to focus on a shorter period in which the likelihood of structural breaks was probably high: 1996Q1 to 1999Q4. The period does not include any important turning point, which was seen as suitable for the exercise at hand for reasons that will be made clearer later. Each forecasts comprised two steps: a re-estimation of the model using information until the period previous to the start of the forecast; and a forecast covering as many periods as possible until 1999Q4. As said, three variables were taken to assess the forecasting performance: GDP growth and GDP deflator and consumption deflator inflation.[28] The forecast accuracy was measured by a number of standard statistics: the mean error of the forecast, its RMSE and the corresponding Theil's U and a non-parametric measure of the standard deviation of the latter.[29] Simple AR(2) models with the variables expressed in first difference and a constant were used as alternative benchmarks.[30] The use of the RMSE measures in a period devoid of important turning points means that both the no-change assumptions and the AR models used as benchmarks are at their best, due to the well-known fact that these models are bad in detecting turning points. If anything, thus, the chosen forecasting period should bias results against the AWM.

[28] Calculations shown later are actually for the first difference of the natural logarithm of these variables.
[29] Theil's U is the ratio between the RMSE of the model's forecast and the RMSE corresponding to forecasts in the assumption that the forecasted variable kept its last-observed value forever.
[30] Exercises were performed with one to four lags in the auto-regression without affecting conclusions.

The steps taken for the out-of-sample exercise were the following. For each model, a forecast was run starting in 1996Q1 with models (i.e., the AWM or the benchmark models) estimated using data until 1995Q4. Forecasts were run for the following eight quarters ahead and the corresponding forecast errors were collected. Thus, this forecast in particular generated one instance of 1-step-ahead to 8-step-ahead forecast errors. The models were then re-estimated using sample until 1996Q1 and a new forecast starting in 1996Q2 was run for eight quarters, from which a new instance of 1-step-ahead to 8-step-ahead forecast errors were collected. The exercise was repeated for each quarter between 1996Q1 to 1999Q3, although not all forecasts can be made for eight quarters ahead: forecasts made after 1998Q2 are run for the maximum number of periods ahead that actual observations allow, i.e. until 1999Q4. After the exercises were run, errors were collected according to the number of periods ahead to which they correspond. The final number of 1-step-ahead errors was thus 16, while the total number of 8-step-ahead errors was 9. For each of the 8 generated series of forecast errors the average mean and the average RMSE were calculated and are reported in the tables described below.

The steps taken for the in-sample exercises were similar, with the exception that the re-estimation step was skipped: the AWM was used in its standard specification and the benchmark models were used as estimated with their full sample. Rolling forecasts were started in 1990Q1 and were performed, as before, until 1999Q4. Taking advantage of the higher number of forecasts done, the longer forecast horizon was increased to 12 steps ahead, i.e. projections were run for 1- to 12-steps-ahead forecasts, generating from 40 1-step-ahead forecasts to 29 12-step-ahead forecasts.

Treatment of additional exogenous information

Important exogenous variables that could affect the outcome comprise foreign variables (output, output prices and oil and commodity prices) and fiscal variables. As a first approximation, these were simply taken as actually observed. Another set of series that are nominally endogenous but whose equations were dropped were the exchange rate and the interest rate. The two equations describing the law of motion for these two variables were calibrated from well-known theoretical constructs, i.e. the UIP condition and a standard Taylor rule, and are invariably not seen as appropriate for a forecasting exercise. Put simply, there is no warranty that these rules are a reflection of relevant structural features in the data. These two variables were thus also kept at observed values.[31]

As already explained, important exogenous variables were kept at their observed values, i.e. for each forecast they belonged to yet unobserved periods and thus the procedure was in disagreement with the rules set for the exercise. In order to establish a fair comparison, the benchmark AR models used in the RMSE comparison were expanded to include external variables as exogenous factors, which were then also taken from actual observations.[32]

Treatment of residuals

Among exogenous information needed by the model, one received a considerable amount of attention: the residuals of the estimated equations. As already stated, one of the crucial aspects of the exercise is to disentangle potential parameter instability from the presence of unaccounted-for structural breaks in assessing forecast failure. As already said, Hendry and Clements (1999) claim that "deterministic shifts" account for most systematic forecast failures, which they describe as un-forecastable breaks in equilibrium means. These breaks lead to large and persistent errors in forecasting precisely because they last for a long period, possibly forever. Regarding this description, the fact that all the estimated equations in the AWM include a long-run component in the form of an equilibrium-correction mechanism has to be stressed. All the equations have implicitly or explicitly an equilibrium mean embedded in them, and thus an intercept correction of these equations that is not subsequently reversed can be understood as reflecting a deterministic shift of this kind. Thus, how residuals are treated in these experiments is clearly a crucial part of the exercise.

One way to assess the importance of this factor in the AWM is by testing different residual-projection approaches and assessing their impact on forecast performance. In practice, what is done is projecting the residual of each equation using information available at the time of the forecast and a number of simple projection strategies. (Note that only stochastic equations in the model are shocked, identities are

[31] Note that this applies only to this part of the exercise, in which steady-state relationships are more the focus of attention. In the previous exercise the model needed to be treated as a whole system, which meant that closure rules for the exchange rate and the interest rate were needed. Shifts in the terminal values of either variable could be later washed away with the business-cycle frequency filters used. These terminal conditions, on the other hand, heavily affected the forecasts in this section.

[32] Changes in fiscal ratios over the period were very likely too small to affect results, and were not included in the AR models.

unaffected by the procedure.) A typical stochastic equation in the model can be described as:

$$\Delta y_t = c + a(L)\cdot \Delta y_{t-1} + b(L)\cdot \Delta z_t - \beta(y_{t-1} - \gamma' z_{t-1}) + u_t ,$$

where y_t is the variable set by the equation and z_t the determinants of y_t (themselves maybe endogenous), c is a constant and u_t is the residual to be projected. For the initial in-sample exercise, which was of an experimental nature, four simple extrapolation strategies were adopted:

1. Flat Projection Method: for each forecast, residuals were projected according to their mean over a long period of time (1985Q1 to the period previous to the start of the forecast in the out-of-sample exercise, same starting date to 1999Q4 in the in-sample exercise):

$$u_{t+k} = \overline{u} = \frac{\sum_{\tau=1}^{t} u_\tau}{t}, k = 1,\ldots,T-t$$

This method was used in two different forms. In the out-of-sample exercise, the residual was projected as its mean over the periods preceding the forecast. Thus, it was projected as a constant that varied slightly for each forecast, according to its initial date. In practice, residuals were not very different across forecasts, as model re-estimation did not lead to large variations in parameter values. In the in-sample exercise, instead the value was kept fixed across replications. Although it would seem that taking averages for the period 1985 to 1999 would imply that future information was taken on board this is not actually the case, as changing the period over which the average is calculated did not lead to noticeably different residuals nor, accordingly, results. Actually, most of the residuals' average for the period were very close to zero.

2. Smoothed-Outlier Projection Method: for each forecast, residuals were projected at the average value observed in the last four quarters before the start of the forecast:

$$u_{t+k} = \frac{1}{4}\sum_{i=0}^{3} u_{t-i}, k = 1,\ldots,T-t.$$

The method tries to strike a balance between considering that all residuals are shifts in the deterministic part of the forecast and acknowledging that true outliers might be present. The way to do that is by taking a moving four-quarter average of past residuals and assuming that the resulting average is the new terminal level for the variable.

3. Outlier Projection Method: for each forecast, residuals were at their last-observed value, i.e. the period preceding the first forecast period:

$$u_{t+k} = u_t, k = 1,...,T-k$$

This method is similar to the previous one, except for the extreme assumption that last period's equation residuals are *always* signalling the onset of a deterministic shift.

4. Smoothed-Trend Projection Method: for each forecast, residuals were projected according to their last-observed trend, defined as the average change over the four periods preceding the forecast.

$$\Delta u_{t+k} = \frac{1}{4}\sum_{i=0}^{3}\Delta u_{t-i}$$

This method is even more extreme than previous ones since the trend in the deterministic component of each forecast is revised each time an exercise is run. In practice, this approach implies that the model is not trusted at all.

For the 3rd and 4th methods, which gave the highest variability of residuals, a small smoothing mechanism had to be put in place in order to avoid large deviations in the forecasts. In the large number of exercises performed – some not reported in the text – it was found that some exceptionally large residuals led to implausibly large forecast errors, or outright simulation debacles. In order to ease this, residuals going above 5 times their historical – at the time of the forecast – variance were reset to their past mean. This solved the previous problem without critically altering results.[33]

As already announced, more sophisticated projection methods can be used: those selected above are probably too naïve to be realistic. A couple of alternative strategies

[33] It was ascertained that, whenever the no-smoothing rule did not lead to large mis-forecasts, results with the rule or without it did not change conclusions. The rule only struck for investment in 1996Q2 and for wages in 1998Q1 in both methods.

were selected (loosely) based on the form of the equations in the AWM, most of which are dynamic equations with a co-integration vector in a so-called equilibrium-correction mechanism (EqCM), of the form defined above. Dropping the dynamic part of a typical EqCM equation results in,

$$\Delta y_t = -\beta\left(y_{t-1} - \gamma' z_{t-1}\right) - u_t ,$$

Following Siviero *et al.* (2001), in this expression last period's residual u_0 could be a true one-off outlier to be corrected afterwards or it could be signalling that a true permanent shift has taken place. In turn, any of these hypotheses could lead to a gradual or immediate impact on the variable. Some of these hypotheses need a specific residual-projection assumption, as expressed in the schematic representation below. These approaches were tested for the in-sample exercise, results for the first two ones being reported.[34]

Type of Outlier	**Residual Projection**	**Impact on y_t**	**Impact on Δy_t**
Corrected Level Shift	$-\beta u_0$, $-\beta u_0$, $-\beta u_0$,…	u_0, u_0, u_0, …	u_0, 0, 0, …
Corrected Outlier	$-(1-\beta)u_0$, 0, 0, …	u_0, 0, 0, …	u_0, $-u_0$, 0, 0, …
Compensation of Outlier	$-\beta u_0$, $(1-\beta)\beta u_0$, 0,…	u_0, $-u_0$, 0, 0, …	u_0, $-2u_0$, u_0, 0, …

Re-estimation procedure

Re-estimating the model was a crucial step since in doing so it was important to use as little information as possible belonging to the period to be forecast. This was achieved by, among other things, restricting the number of parameters originally calibrated, as the calibration could be seen as reflecting prior information based on the knowledge of the full sample. Among parameters that were nevertheless taken at their original calibrated value, the capital-share parameter β was the most important. This parameter is calibrated based on the mean value of the wage-income share in GDP, which has varied somewhat over the last 20-30 years. Other calibrated parameters taken unchanged were parameters in the wage and price equations on price expectations and

[34] Note that what Siviero (2001) call *Error in Equilibrium* and *Gradual Vanishing* correspond to our smooth-outlier and the flat projection methods described above. Results for two strategies are only reported due to lack of space.

the ECM term in the wage equation. In addition, the NAIRU embodied in the model was left unchanged at its full-sample values. It is unlikely that this may have significantly affected the outcome of the exercise, due to the small scope for change in these parameters or variables in the relatively short period over which the forecasts were run. Finally, the VAR was also estimated and decomposed in a recursive way.

The equations were simplified by dropping dummy variables effective in or after 1995Q4, the first period estimated.[35] Another simplification was the harmonisation of the end-of-sample for all the equations. Although original equations were estimated with varying ending points, depending on data availability at the time of original estimation, in the present recursive exercises all equations were estimated until a common ending point, the last possible period before starting each forecast. Notwithstanding all these points, the equations were remarkably similar to those reported in Fagan *et al.* (2001), and relatively stable over the 4-year period used. Results for the estimations are reported in Annex 1.

For the in-sample exercises, the model was used as taken.

Alternative benchmark models

The AWM was tested against simpler alternative models: an AR(2) model of each variable, and an AR(2) augmented with exogenous information (foreign GDP and prices, exchange rate and interest rate) taken at their observed values. Both benchmark models were for the first difference of the log of the variable to forecast, and included a constant and two lags of the endogenous variable. The second also included the short-term interest rate (in levels) and the first difference of the exchange rate, foreign GDP, GDP deflator and oil and commodity prices. These variables entered contemporaneously and lagged once and twice respectively.

7.3 Results

A preliminary look: GDP deflator inflation

Table 4 shows the basic forecast statistics for the out-of-sample exercise GDP deflator quarter-on-quarter inflation. Although also GDP growth and consumption deflator

[35] An alternative is to alter the specification of the equations to re-introduce dummy variables after the period in which they become effective. The decision to keep equations unchanged over recursive re-estimations was taken to simplify the exercise. This means that outliers in the period are more disruptive than otherwise, thus possibly again biasing results against the AWM.

inflation will be analysed, it is better if some preliminary discussion of results is done based on a subset of the information. This is convenient since we will take advantage of the ensuing discussion to advocate, precisely, that not all generated information is relevant. What Table 4 shows is the mean error, RMSE and Theil's U statistic (together with its standard deviation) for exercises covering from 1- to 8-step ahead forecasts, run for the period 1996Q1 to 1999Q4. Models for which results are reported are the AR(2) model, the AR(2)-cum- exogenous-information model and the AWM.

[Table 4]

Results for the AWM are shown for the announced residual-projection strategies: no residual projection, residuals as smooth level shifts, residuals as non-smoothed level shifts and residuals as smooth trend shifts. Please note that Theil's U statistic is simply the ratio between the corresponding RMSE and the RMSE resulting from a naïve forecast in which the GDP deflator inflation is projected unchanged since it was last observed. The standard deviation attached to each U statistic is the non-parametric estimate of the ratio of the two RMSE involved in its calculation. Thus, an U statistic of less than one indicates that forecasts from the model under test is to be preferred to the naïve forecast.

A number of facts can be noticed:

- All the models behave similarly for short horizons (*h* less than 3 or 4), after which all models worsen substantially, in particular the AR(2).
- In terms of Theil's U statistic, the smooth-residual projection is clearly to be preferred to the no-residual strategy, and these two are much better than the other two residual-projection strategies.
- In terms of mean error (bias), the smooth level-shift projection strategy is by far the better. Actually, the AWM beats on this account any other model in the table.

It is worth pointing out to what extent these results apply to the other two variables analysed, GDP growth and consumption deflator inflation. GDP growth is clearly less persistent than inflation and thus less easily forecast, which leads to relatively worse behaviour across models than is the case for GDP deflator inflation. Consumption deflator inflation is forecast with accuracy similar to that of GDP deflator inflation, but the ranking of the models differs slightly. For shorter-term forecasts, i.e. those under 4 quarters ahead, the AR(2) and the AWM with no residual projection are better, but at longer horizons the AR(2)-cum-exogenous-information is best. Nevertheless, some of the resulting Theil's U's showed a great variability, underlining the uncertainty of these results due to the relatively small number of forecasts run. The finding that shows higher robustness, on the other hand, is the good performance of the AWM with smooth level shifts regarding forecast biases: this feature is present everywhere.

In summary, one could argue that the AWM, although no worse than standard forecasting benchmark models, needs some help from exogenous level shifts. In this respect, results from the exercise do not invalidate – far from it – results reported in Clements and Hendry (1999). The implications of this are that the model is not properly capturing stable long-run relationships, in line with the previous author's argument, but that dynamics in the data are (broadly) correctly portrayed. This line of reasoning is based on a casual look at mean forecast errors and RMSE of forecasts. As will be shown, though, a deeper analysis of the RMSE leads to partial reassessment of this conclusion. We will proceed to decompose the reported RMSE in order to show that although level shifts are important to counter model biases, their overall impact on forecast accuracy could be negative. Although deterministic shifts help reducing first-moment discrepancies between model and the data-generation process (and this seems to be a robust finding), this happens at a cost in terms of second-order moments than at times more than balances their potential benefit.

Table 5 shows the RMSE decomposition of the forecasts for the four models analysed in Table 4. In the table, the decomposition in (12) is shown relative to the variance of the variable, in order to enhance the readability of the table.[36]

[Table 5]

One striking feature in the table is the huge discrepancy between relative variance of the forecast and relative bias squared between the AWM with smooth level shifts and the rest of the models. The former has a very low contribution to MSE from bias,

[36] In other words, the raw numbers – which were very small – have been re-scaled by the variance of the observations used to calculate the corresponding RMSE. Please note that the column with the relative covariance, label 'cov(x,f)/var(x)', differs by construction from the correlation between actual and forecast.

clearly inferior to the rest of models shown. (The AR(2)-cum-exogenous-information model is clearly worse on this account, though, which explains its very bad behaviour at medium to long horizons.) On the other hand, the variance of the forecast is much higher for the AWM with smooth level shifts. Finally, the relatively good performance of the AR(2) at short horizons (less than 2 quarters ahead) is coming mostly from its low variance of the forecast. This low forecast variance means that AR(2)-based forecasts converge to a point faster than the other models, a feature which is not desirable *per se*.[37]

In a word, level shifts are adding to the AWM's forecast variance but this is balanced by a reduction in the forecast bias. Thus, level shifts seem to be positive to the extent that they get rid of bias faster than they add to forecast variability, increasing the probability of forecast debacles. Unfortunately, no firm conclusions can be drawn from the exercise due to the small number of forecasts run, as is made clear by the large standard deviations of the U statistics. Furthermore, forecast variance depends to an unknown extent on parameter re-estimation, with more complex models bound to lose more on this account due to higher number of parameters. In a word, results in Table 5 may bias the picture against the AWM (or indeed any structural macro-model) simply because of its higher complexity. It is thus necessary to reassess results in the light of constant-parameter models.

Table 6 presents results for a repetition of the exercise described in Table 4, but skipping the re-estimation step for all the models and with an extended period over which the exercise is run, i.e. 1990Q1 to 1999Q4. Table 7 presents the corresponding RMSE decomposition.

[Table 6]

[Table 7]

Results spanning 1996Q1 to 1999Q4, as in tables 5 and 6, were also assessed and lent weight to the view that model re-estimation adds significantly to the variance of the forecast without affecting much the bias. Hence, model re-estimation has in general a small beneficial impact on forecast bias, but adds to forecast variability to an extent

[37] A model may converge to a given point because actual data behaves in this manner, or simply because the dynamics of the model are mis-specified.

that often more than balances this factor. Testing macro-models using full re-estimation does not seem, accordingly, to be good practice.

Results in Table 6 have been drawn from simulated in-sample forecasts (i.e., dynamic simulations) spanning the period 1990Q1 to 1999Q4, as mentioned, which means that as much as 40 forecasts were available. This has the merit of significantly reduce the variability seen in the numbers in tables 5 and 6 and considerably clarify the picture. Taking advantage of that, forecast horizon has been increased to allow for forecasts up to 12 periods ahead.

One of the main conclusions from tables 7 and 8 is that the AWM with no residual projection is now clearly better than projecting with the smooth level shift projection. As can be seen in Table 7, this stems basically from a lower variance of the forecast in the latter case. Thus, although it is still the case that shifting the levels of the residuals leads to lower forecast bias, this is more than offset by forecast variability induced by shifting the levels. (In both cases the correlation between actual and forecasts is similar.) Comparing to the benchmark models, Table 6 shows that the AR(2)-cum-exogenous-information model now performs adequately, but with the caveat that this seems to come mostly from much lower bias. The AR(2) model performs relatively worse than previously, at least in comparison with the no-residual projection, in particular due to very high bias at horizons longer than 7/8 quarters.

Results for all the variables

Finally, tables 9 and 10 present results for the other two variables of interest: GDP growth and consumption deflator inflation. As said previously, GDP growth shows a relatively small degree of persistence, which means that the forecasting performance of all models is worse than for the inflation measures, see Table 8.

[Table 8]

The tables clearly indicate that the AWM with no residual-projection strategy fares better than when residuals are projected with smooth level shifts. Compared to benchmark models, the AWM (with no residual projection) fares very similarly to the AR(2) model, but somewhat worse than the AR(2) with exogenous variables. It turns out that the AR(2) cum exogenous variables produces forecasts that show a slightly higher correlation to actual GDP growth than the AWM. This relatively higher correlation seems to rest on foreign GDP growth and foreign GDP deflator inflation, oil prices and the exchange rate adding little to it.[38]

Finally, Table 9 shows results for consumption deflator inflation.

[Table 9]

For this variable, it is again the case that the AWM without residual projection fares better than the smooth-level-shift approach. In addition, this version of the AWM convincingly beats the AR(2) model but fares slightly worse than the AR(2) with exogenous information for horizons beyond 7/8 quarters ahead. This is explained basically by the lower variance of the forecast generated with the former at longer horizons, due probably to the faster rate of convergence of the AR(2) model to its implicit steady state compared to the AWM.

Last but not least, none of the other residual-projection strategies tested in the context of this exercise (i.e., the corrected-level-shift, the corrected-outlier and the compensation-of-outlier approaches – the last one not shown to save space) led to better results than the simpler approach.

7.4 *Dos* and *Don'ts* in Model Forecasting

Let us briefly summarise this exercise. First and foremost, add-factor adjustment in macro-models may lead to less biased forecasts, but at a cost in terms of increased variance of the forecasts (i.e., higher likelihood of forecasting debacles). They should thus be used only with care. Residual adjustment is obviously a useful device when dealing with information missing in the model, but adjustment not based on rigorous economic thinking should be avoided. Other source of increased forecast inaccuracy is parameter changes, which points to the possibility that model re-estimation should only be attempted when evidence of mis-specification is sufficiently strong. In particular, macro models should be left unchanged if re-estimating them would lead to barely changed model parameters. Another interesting conclusion is that basing an

[38] Dropping oil and commodity prices from the AR(2) model with exogenous information led to basically the same correlation, while dropping any of the other two variables – world GDP growth and world GDP deflator inflation – led to a much lower correlation.

analysis of macro models in simulated out-of-sample exercises if far from being wise. On the contrary, the loss of efficiency resulting from ignoring true sample information is of such a scale to distort conclusions. Very seemingly, out-of-sample analysis should be attempted only with care for macro models and samples of standard size.

8. Conclusions.

In this paper, we presented a testing framework on a macro-model of the euro area. It was not our primary purpose to draw specific conclusions about the AWM's failings or successes; indeed any such model is in an almost continual testing phase – reflecting the different uses and questions to which it is addressed as well as the environments and audiences to which it is exposed. While our work may provide important feedback in that respect, our interests were more general. Our analysis was launched in two directions. The first in terms of replicating high-frequency, business-cycle characteristics; the second at examining out-of-sample recursive forecast failure.

First, the business-cycle tests. Matching models to business-cycle properties is most readily associated with the DSGE programme. Our interpretation, however, differs in a number of respects. First, our model is larger than standard DSGE ones. Thus, instead of seeking to match a subset of empirical macro regularities, we take the model seriously in matching them all. Second, the model's estimation is given. Thus, we do not parameterise based on moments' matching. Third, rather than relying on highly-stylised replication procedures – such as AR (1) technology, monetary shocks etc – we employ a system-wide Cholesky covariance decomposition. This means we can be agnostic about the nature of historical shocks but still capture their moments. Once done, bootstrapping the repeated stochastic simulations allows a probabilistic formalisation of the match.

Overall, the model's moments-matching might be described as reasonably "good" – defined by the high number of insignificantly different cross-correlations. The model matches volatilities less well. We speculatively ascribed this *missing variability* to such things as the non-modelling of key foreign variables, lack of smoothing in the monetary policy rule, incorporating shocks from purely behavioural equations etc. Future work will seek to confirm that. Spectral properties are similarly encouraging; all (data and model) series considered share similar qualitative characteristics and the model data appeared to be encompassed by the confidence intervals of the historical data.

Our second set of tests reflected the concern that stylised fact tests could only take us so far in system evaluation. After all, a model estimated from the data *should* match many empirical regularities. Indeed, one can think of many classes of models that might replicate business-cycle features. However, such models may have no sensible structure, no steady state. Thus, we moved to tests that combined low-frequency aspects (short-horizon forecasts) with long-run features (e.g., the existence and identification of steady-states). Such concerns led us to consider forecast performance. Aspects of forecast failure identified in the academic literature (but typically employed with "small" systems) were applied. A number of tests were carried out trying to assess the relative predictive performance of the model, taking as benchmarks both simple time-series models and versions of the AWM where its residuals were projected according to well-specified, simple laws of motion. The first set of models are widely found in the literature and are useful in setting a well-known benchmark. The second set of benchmark models were close alternatives to the AWM with a focus on analysing the effects of the steady state in the model in terms of forecasting performance. The outcome of the exercise has been interesting in that: i) forecast accuracy in the model for inflation is visibly higher than for growth, up to the point of approaching the accuracy of widely used time-series models; and ii) in the tests, simple residual projection approaches are to be preferred to more complex ones. One result that would need a deeper analysis and further testing would be a more definite ranking of residual projection strategies, in particular whether taking into account in one or another way past forecast failures improves the accuracy of the forecasts or are better ignored. Moreover, given the stylised nature of the exercise, it is difficult to draw definitive conclusions. This would require an assessment of the contribution of actual judgmental adjustments made by forecasters to forecast performance, along the lines of McNess (1990).

REFERENCES

Adelman, I. And Adelman, F. L. (1959) "The Dynamic Properties Of The Klein-Goldberger Model", *Econometrica*, 27, 596-625.

Agresti, A. M. and Mojon, B. (2001) "Some stylised facts on the euro area business cycle", ECB Working Paper 95.

Amano, R., McPhail, K, Pioro, H and Rennison, A. (2000) "Evaluating QPM: A Preliminary Investigation", *Advances in Macroeconomic Model Building Workshop*, Bank of Canada 17-18 August.

Andolfatto, D. (1996) "Business Cycles and Labor-Market Search" *American Economic Review*, 86(1), 112-32.

Andrews, D. W. K and Monahan, J. C. (1992) "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica*, 60, 4, 953-966.

Andrews, D. W. K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59(3), 817-58.

Artis, M and Zhang, W (1997) "International Business Cycles and the ERM: is there a European Business Cycle?" *International Journal of Finance and Economics*, 2(1), 1-16

Bachus, D. K. and Kehoe, P. J. (1992) "International Evidence on the historical properties of Business cycles", *American Economic Review*, 82, 864-888.

Ball, L. (1990) "Intertemporal substitution and constraints on labor supply: evidence from panel data", *Economic Inquiry,* 28, 706-724.

Basu, S. and Taylor, A. M (1999) "Business Cycles in International Historical Perspective", *Journal of Economic Perspectives*, 13(2), 45-68.

Blanchard, O. J, and Quah, D. (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 79(4), 655-73.

Boone, L. and Hall, S. G. (1999) "Stylized Facts of the Business Cycle Revisited: A Structural Modelling Approach", *International Journal of Finance and Economics*, 4(3), 253-68.

Brayton, F., Levin, A., Tryon, R. W. and Williams, J. C. (1997) "The Evolution of Macro Models at the Federal Reserve Board", *Carnegie Rochester Conference Series on Public Policy,* 42, 115-167.

Bryant, R., Hooper, P. and Mann, C. (1993) *Evaluating Policy Regimes*, The Brookings Institute, Washington DC.

Burnside, C., Eichenbaum, M and Rebelo, S. (1993) "Labor Hoarding and the Business Cycle", *Journal of Political Economy*, 101(2), 245-73.

Camba-Méndez, G and Pearlman, J. (1998) "Can Real Equilibrium Models Account For The Fluctuations Of The UK Business Cycle?", National Institute of Economic and Social Research Discussion Paper 128.

Camba-Méndez, G and Rodriguez-Palenzuela, D. (2002) "Assessment criteria for output gap estimates", *Economic Modelling* forthcoming.

Campbell, J Y., Mankiw, N. Gregory (1987) "Are Output Fluctuations Transitory?", *Quarterly Journal Of Economics*, 102(4), 857-80.

Canova, F. (1994) "Statistical Inference in Calibrated Models." *Journal of Applied Econometrics,* 9 (supplement): S123-S144.

Chadha, B. and Prasad, E. (1994) "Are Prices Counter cyclical? Evidence from the G-7", *Journal of Monetary Economics*, 34(2), 239-57.

Chow, G. C. (1975) *Analysis and Control of Dynamic Economic Systems,* Wiley.

Christodoulakis, N., Dimelis, S. P. and Kollintzas, T. (1995) "Comparisons of Business Cycles in the EC: Idiosyncracies and Regularities", *Economica*, 62 (245), 1-27.
Clarida, R., Galí, J. and Gertler, M. (1998) "Monetary policy rules in practice. Some international evidence", *European Economic Review,* 42, 6, 1033-1067.
Clements, M. P. and Hendry, D. F. (1998) "Forecasting Economic Processes", *International Journal of Forecasting*, 14(1), 111-31.
Clements, M. P. and Hendry, D. F. (1999) *Forecasting Non-Stationary Economic Time Series,* Cambridge, Mass.: MIT Press.
Cochrane, J. H. (1988) "How Big Is The Random Walk In GNP?", *Journal of Political Economy*, 96(5), 893-920.
Coenen, G. and Vega, J. L. (2001) "The demand for M3 in the Euro area", *Journal of Applied Econometrics,* 16, 727–748.
Coenen, G. and Wieland, V. (2000) "A Small Estimated Euro Area Model With Rational Expectations and Nominal Rigidities", European Central Bank, *European Economic Review* Forthcoming.
Cogley, T. and Nason, J. (1995) "Effects of the Hodrick-Prescott filter on trend and difference stationary time series", *Journal of Economic Dynamics and Control,* 19, 253-278.
Cooley, T. F. and Hansen, G. D. (1994) "Money and the business cycle", in Cooley, T. F. (Ed.), *Frontiers of Business Cycle Research,* Princeton University Press.
Cooley, T. F. and Hansen, G. D. (1998) "The Role of Monetary Shocks in Equilibrium Business Cycle Theory: Three Examples", *European Economic Review*, 42(3-5), 605-17.
Cooley, T. F. and Ohanian, L. E. (1991) "The Cyclical Behavior of Prices", *Journal of Monetary Economics*, 28(1), 25-60.
Cooley, T. F. and Quadrini, V. (1999) "A Neo-classical Model of the Phillips Curve Relation", *Journal of Monetary Economics*, 44(2), 165-93.
Danthine, J. P., and Donaldson, J. B. (1993) "Methodological and Empirical Issues in Real Business Cycle Theory", *European Economic Review*, 37(1), 1-35.
den Haan, W.J. (2000) "The co-movement between output and prices", *Journal of Monetary Economics* 46, 3-30.
Detken, C., Dieppe, A., Henry J., Marin C., and F. Smets (2002) "Model uncertainty and the real effective exchange rate of the euro" *Australian Economic Papers,* forthcoming.
Dickerson, A, Gibson, H and Tsakalotos, E (1998) "Business cycle correspondence in the European Union", *Empirica*, 25, 51-77.
Diebold, F. X., Ohanian, L. E. and Berkowitz, J. (1998) "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data", *Review of Economic Studies*, 65(3), 433-51.
Diebold, F.X. and Li, C. (2002) "Forecasting the Term Structure of Government Bond Yields", Manuscript, University of Pennsylvania.
Dieppe, A and J. Henry (2002) "The euro area viewed as a single economy: how does it respond to shocks?", Paper presented at the CEPR conference on *Empirical models of the euro economy*, 7-8 June 2002, Bonn.
Fagan, G., Henry, J. and Mestre, R. (2001) "An area-wide model (AWM) for the euro area", ECB Working Paper 42.
Fair, R. C. (1998) "Estimated Stabilisation Costs of EMU", *National Institute Economic Review*, 164, 90-99.
Fisher, P. and Salmon, M. (1986) "On Evaluating the Importance of Non linearity in Large Macro econometric Models", *International Economic Review*, 27, 3, 625-

46.
Fisher, P. G and Whitley, J. D (1998) "Macroeconomic Models at the Bank of England", mimeo, Bank of England.
Frenkel, J. A., Goldstein, M and Masson, P. R (1989) "Simulating The Effects Of Some Simple Coordinated Versus Uncoordinated Policy Rules" in (Eds. Bryant, R, Currie D A, Frenkel, J A, Masson P R and R Portes) *Macroeconomic Policies In An Interdependent World,* Brookings Institute.
Galí, J., Gertler, M. and López-Salido, J. D (2001) "European Inflation Dynamics", *European Economic Review,* 45, 7, 1237-1270.
Garrat, A and Scott, A (1996) "Money and Output: a reliable predictor or an intermittent signal", Bank of England, mimeo.
Granger, C. W. J. (1966) "The Typical Spectral Shape of an Economic Variable", *Econometrica*, 34, 150-161.
Hansen, L. P. and Heckman, J. J. (1996) "The empirical foundations of calibration", *Journal of Economic Perspectives,* 10, 87-104.
Hendry, D. F (1996) *Dynamic Econometrics,* Oxford University Press.
Hughes Hallett, A. J. and McAdam, P. (2001) "The Stability Pact And The Interdependence Of Monetary And Fiscal Policy Rules" in *The Co-ordination Of Monetary Policies In Europe* (Eds. P. Mooslechner and M. Schurz), Kluwer Academic Press, Boston.
Juillard, M. (1999) "The dynamical analysis of forward-looking models", in *Analyses In Macroeconomic Modelling* (Eds. Hughes-Hallett, A. J. and P. McAdam), Kluwer Academic Press.
Juillard, M., Laxton, D., McAdam, P. and Pioro, H. (1998) "An Algorithm Competition: First-Order Iterations versus Newton-based Techniques", *Journal of Economic Dynamics and Control*, 22, 1291-1318.
King, R. and Rebelo, S. (1999) "Resuscitating Real Business Cycles" in J. B. Taylor and M. Woodford (Eds.) *Handbook of Macroeconomics*, Vol. 1B, North Holland).
Klein, P (2000) "Using the generalized Schur form to solve a multivariate linear rational expectations model", *Journal Of Economic Dynamics and Control*, 24, 10, 1405-1423.
Kyland, F. E. and Prescott, E. C. (1990) "Business cycles: real facts and a monetary myth", *Federal Reserve Bank of Minneapolis Quarterly Review*, 2, 3-18.
Masson, P. R. and Symansky, S. (1992) "Evaluating the EMS and EMU Using Stochastic Simulations", in *Macroeconomic Policy Co-Ordination In Europe: The ERM And Monetary Union* (Eds. R. Barrell and J. Whitley), London; Newbury Park.
McAdam, P. and Hughes Hallett, A. J. (1999) "Nonlinearity, Computational Complexity and Macroeconomic Modelling", *Journal of Economic Surveys*, 13, 577-618. <Reprinted in *Surveys in Economic Dynamics,* D. A. R. George, L. Oxley and S. Potter. (Eds.), Oxford: Blackwell Publishers, 2000>
McAdam, P. and Morgan J. (2002) "Analysing Monetary Policy Transmission at the Euro-Area Level using Structural Macroeconomic Models." in *Monetary Policy Transmission in the Euro Area* (Eds. Angeloni, I, Kashyap, A. and Mojon, B.), Cambridge University Press.
McAdam, P. and Willman, A. (2002) "New Keynesian Phillips Curves: a reassessment based on euro-area data", ECB mimeo.
McNees, S. K. (1990) "Man vs. Model? The Role of Judgement in Forecasting", *New England Economic Review*, Jul/Aug, 41-52.

Mendoza, E. G. (1991) “Real Business Cycles in a Small Open Economy”, *American Economic Review*, 81(4), 797-818.
Nicoletti-Altimari S. (2001) “Does money lead inflation in the euro area?”, ECB Working Paper 63.
Oliveira Martins, J., Scarpetta, S. and Pilat, D. (1996a) “Mark-Up Pricing, Market Structure and the Business Cycle”, *OECD Economic Studies*, 0, 27, 71-105.
Oliveira Martins, J., Scarpetta, S. and Pilat, D. (1996b) “Mark-Up ratios in manufacturing industries”, *OECD* WP 162.
Orlandi, F. and K. Pichelmann (2000) “Disentangling Trend and Cycle In The EUR-11 Unemployment Series”, *Economic Papers* 140, European Commission.
Ravn, M. O. and Sola, M. (1995) “Stylized facts and regime changes: are prices procyclical?”, *Journal of Monetary Economics*, 36, 497-526.
Röger, W. (1999) “Output, Prices and Interest Rates over the Business Cycle: How Well Does a Keynesian-Neo-classical Synthesis Model Perform?”, *Zeitschrift fur Wirtschafts und Sozialwissenschaften,* 119, 1, 57-97.
Romer, D. (2000) *Advanced Macroeconomics,* McGraw-Hill Higher Education.
Rotemberg, J. J. and Woodford, M. (1991) “Markups and the Business Cycle”, in Blanchard, O. J. and S. Fischer (Eds.) *NBER Macroeconomics Annual*. MIT Press, 63-129.
Siviero, S. (2001) “When and how to use add-factors: A practicioner's viewpoint”, mimeo.
Smets, F. and Wouters, R. (2002) “An estimated stochastic dynamic general equilibrium model of the euro area”, ECB Working Paper 171.
Stock J. H. and Watson, M. W. (1999) “Business cycle fluctuations in US macroeconomic time series” in J. B. Taylor and M. Woodford (Eds.) *Handbook of Macroeconomics*, Vol. 1A, North Holland).
Stock, J. H., M. Marcellino and M. W. Watson (2002) “Macroeconomic Forecasting In The Euro Area”, *European Economic Review* forthcoming.
Walpole, R. E. and Myers, R. H (1978) *Probability and Statistics for Engineers and Scientists*, 2nd Edition, Macmillan Publishing Co., New York.
Watson, M (1993) “Measures of Fit for Calibrated Models”, *Journal of Political Economy*, 101, 6, 1011-1041.

Table 1: Euro-Area Stylised Facts: Historical Data

	ST.DEV	ST.DEV Ratio	AR(1)	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
Real GDP	1.027	1.000	0.834	0.137	0.365	0.623	0.839	1.000	0.839	0.623	0.365	0.137
Consumption	0.897	0.873	0.751	0.238	0.388	0.521	0.646	0.718	0.518	0.376	0.217	0.088
Investment	2.721	2.649	0.883	0.234	0.405	0.583	0.741	0.866	0.795	0.675	0.507	0.331
Exports	2.509	2.442	0.756	0.167	0.318	0.503	0.64	0.743	0.605	0.401	0.202	-0.003
Imports	2.989	2.910	0.827	0.172	0.387	0.606	0.797	0.916	0.787	0.569	0.326	0.081
Nominal Exchange Rate	4.027	3.920	0.824	0.013	-0.032	-0.086	-0.172	-0.257	-0.327	-0.371	-0.332	-0.248
Real Exchange Rate	4.073	3.965	0.818	-0.09	-0.037	0.035	0.142	0.248	0.328	0.37	0.318	0.215
Inflation:Consumption Deflator	1.11	1.081	0.908	-0.55	-0.439	-0.247	-0.002	0.257	0.477	0.601	0.637	0.585
Price Level	1.196	1.164	0.955	-0.669	-0.713	-0.680	-0.591	-0.453	-0.279	-0.119	0.011	0.109
Unit Labour Costs	0.014	0.014	0.900	-0.589	-0.676	-0.725	-0.696	-0.605	-0.365	-0.121	0.117	0.288
Mark-up	0.862	0.839	0.794	-0.285	-0.389	-0.505	-0.554	-0.548	-0.319	-0.078	0.184	0.38
Nominal Wages	1.074	1.045	0.866	-0.635	-0.644	-0.586	-0.462	-0.265	-0.138	-0.009	0.12	0.205
Real (Producer) Wages	0.664	0.646	0.678	-0.132	-0.100	-0.039	0.046	0.195	0.165	0.159	0.198	0.223
Real Consumer Wages	0.834	0.811	0.737	0.167	0.190	0.207	0.245	0.321	0.235	0.171	0.151	0.123
Labour Productivity	0.719	0.700	0.705	0.228	0.379	0.563	0.69	0.799	0.512	0.226	-0.051	-0.26
Employment	0.621	0.604	0.928	-0.018	0.173	0.375	0.577	0.72	0.787	0.765	0.659	0.503
Unemployment	0.344	0.335	0.913	-0.094	-0.234	-0.398	-0.585	-0.734	-0.799	-0.76	-0.645	-0.471
Nominal Short Interest Rate	2.932	2.854	0.965	-0.567	-0.527	-0.434	-0.305	-0.175	-0.061	0.02	0.066	0.084
Nominal Long Interest Rate	2.394	2.330	0.983	-0.432	-0.424	-0.382	-0.316	-0.244	-0.181	-0.13	-0.098	-0.073
Real Short-Term Interest Rates	2.952	2.873	0.965	-0.564	-0.522	-0.425	-0.292	-0.156	-0.037	0.048	0.099	0.119
Nominal Money	0.393	0.383	0.827	-0.019	0.039	0.038	0.075	0.066	0.117	0.152	0.151	0.087
Real Money	0.817	0.796	0.909	0.531	0.533	0.487	0.403	0.302	0.23	0.136	0.047	-0.057
Real Gov. Consumption	0.59	0.574	0.637	-0.136	-0.155	-0.110	-0.05	0.019	0.037	0.066	0.194	0.282

Notes:
Sample for all series is 1970Q1 to 1999Q4 except for Nominal and Real Money Balances: 1980Q1 to 1999Q4. For Tables 1 to 2, we first log each series, then use the difference between the logged series and its HP filter to derive the associated moments and cross-correlations,

$$y - HP(y)$$

The exceptions to this are the interest rate series and unemployment for which we use levels. Finally, all interest rate series are shown as correlated with unfiltered GDP growth.

Table 2: Euro-Area Stylised Facts: Simulated Data

	ST.DEV	ST. DEV Ratio	AR(1)	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
Real GDP	0.8602 *[0.07]*	1	0.7139 *[0.01]*	0.1474 *[0.45]*	0.2798 *[0.27]*	0.4811 *[0.06]*	0.7141 *[0.00]*	1 *[1.00]*	0.7141 *[0.00]*	0.4811 *[0.06]*	0.2798 *[0.27]*	0.1474 *[0.45]*
Consumption	0.8471 *[0.31]*	0.9859	0.7525 *[0.56]*	0.1279 *[0.23]*	0.2552 *[0.15]*	0.4428 *[0.28]*	0.6655 *[0.63]*	0.938 *[0.01]*	0.7287 *[0.00]*	0.5351 *[0.12]*	0.3455 *[0.2]*	0.2136 *[0.07]*
Investment	2.3595 *[0.09]*	2.7608	0.7458 *[0.00]*	0.2567 *[0.52]*	0.3346 *[0.48]*	0.4514 *[0.5]*	0.5719 *[0.05]*	0.7132 *[0.45]*	0.4815 *[0.47]*	0.2814 *[0.48]*	0.1206 *[0.52]*	0.0092 *[0.52]*
Exports	1.7464 *[0.00]*	2.0485	0.638 *[0.03]*	0.1515 *[0.057]*	0.2293 *[0.28]*	0.3504 *[0.13]*	0.4844 *[0.05]*	0.6509 *[0.15]*	0.38 *[0.02]*	0.2083 *[0.09]*	0.0595 *[0.23]*	0.0012 *[0.42]*
Imports	1.9113 *[0.00]*	2.2451	0.7091 *[0.00]*	0.2261 *[0.4]*	0.2914 *[0.27]*	0.3844 *[0.07]*	0.4801 *[0.00]*	0.5956 *[0.00]*	0.3778 *[0.00]*	0.2135 *[0.00]*	0.0638 *[0.04]*	-0.0124 *[0.29]*
Nominal Exchange Rate	1.3735 *[0.00]*	1.6023	0.7879 *[0.68]*	-0.2933 *[0.02]*	-0.3812 *[0.00]*	-0.5129 *[0.00]*	-0.6554 *[0.00]*	-0.8205 *[0.00]*	-0.5586 *[0.00]*	-0.3248 *[0.41]*	-0.1377 *[0.07]*	-0.0028 *[0.04]*
Real Exchange Rate	1.294 *[0.00]*	1.5102	0.7262 *[0.05]*	0.2328 *[0.00]*	0.3274 *[0.00]*	0.4758 *[0.00]*	0.6451 *[0.00]*	0.8662 *[0.00]*	0.6059 *[0.00]*	0.3688 *[0.51]*	0.1619 *[0.08]*	0.0128 *[0.05]*
Inflation:Consumption Deflator	0.4571 *[0.00]*	0.5374	0.841 *[0.00]*	-0.1813 *[0.00]*	-0.1653 *[0.04]*	-0.1435 *[0.26]*	-0.1114 *[0.28]*	-0.021 *[0.04]*	0.1012 *[0.02]*	0.2117 *[0.00]*	0.2833 *[0.00]*	0.282 *[0.00]*
Price Level	0.4825 *[0.00]*	0.5665	0.9134 *[0.06]*	-0.296 *[0.02]*	-0.3342 *[0.00]*	-0.363 *[0.03]*	-0.3758 *[0.125]*	-0.3248 *[0.25]*	-0.2382 *[0.44]*	-0.153 *[0.4]*	-0.0916 *[0.31]*	-0.0377 *[0.25]*
Unit Labour Costs	0.0078 *[0.00]*	0.009	0.8131 *[0.06]*	-0.4104 *[0.13]*	-0.4585 *[0.29]*	-0.525 *[0.2]*	-0.5763 *[0.14]*	-0.6112 *[0.46]*	-0.2842 *[0.32]*	-0.0775 *[0.41]*	0.0897 *[0.45]*	0.1473 *[0.19]*
Mark-up	0.6405 *[0.00]*	0.7465	0.6555 *[0.2]*	-0.3342 *[0.34]*	-0.3954 *[0.48]*	-0.4928 *[0.55]*	-0.5928 *[0.3]*	-0.7306 *[0.02]*	-0.3748 *[0.7]*	-0.1479 *[0.3]*	0.0715 *[0.22]*	0.1556 *[0.00]*
Nominal Wages	0.6045 *[0.00]*	0.7102	0.8105 *[0.19]*	-0.1554 *[0.00]*	-0.1277 *[0.00]*	-0.0565 *[0.00]*	0.0331 *[0.00]*	0.1668 *[0.01]*	0.0119 *[0.27]*	-0.0681 *[0.43]*	-0.1 *[0.125]*	-0.1432 *[0.03]*
Real (Producer) Wages	0.4544 *[0.00]*	0.5334	0.6418 *[0.37]*	0.02 *[0.23]*	0.0477 *[0.22]*	0.1101 *[0.24]*	0.1666 *[0.24]*	0.1994 *[0.46]*	-0.0393 *[0.13]*	-0.1662 *[0.01]*	-0.1779 *[0.01]*	-0.2154 *[0.00]*
Labour Productivity	0.7052 *[0.39]*	0.8219	0.6189 *[0.13]*	0.3201 *[0.3]*	0.3941 *[0.46]*	0.5219 *[0.41]*	0.6468 *[0.23]*	0.7911 *[0.47]*	0.3173 *[0.00]*	0.0312 *[0.00]*	-0.1765 *[0.05]*	-0.279 *[0.46]*
Employment	0.5298 *[0.47]*	0.6125	0.924 *[0.49]*	-0.178 *[0.09]*	-0.0672 *[0.00]*	0.0855 *[0.00]*	0.2953 *[0.00]*	0.5653 *[0.00]*	0.746 *[0.16]*	0.7558 *[0.48]*	0.7051 *[0.7]*	0.6257 *[0.86]*
Unemployment	0.4809 *[0.05]*	0.556	0.9239 *[0.22]*	0.178 *[0.00]*	0.0674 *[0.00]*	-0.0853 *[0.00]*	-0.2953 *[0.00]*	-0.5653 *[0.00]*	-0.7461 *[0.125]*	-0.7558 *[0.53]*	-0.705 *[0.21]*	-0.6255 *[0.057]*
Nominal Short Interest Rate	2.1591 *[0.05]*	2.5364	0.7744 *[0.00]*	-0.4102 *[0.096]*	-0.3534 *[0.067]*	-0.2549 *[0.06]*	-0.1292 *[0.058]*	0.021 *[0.08]*	0.0682 *[0.18]*	0.0929 *[0.33]*	0.1011 *[0.38]*	0.1265 *[0.37]*
Nominal Long Interest Rate	2.0567 *[0.24]*	2.414	0.944 *[0.02]*	-0.4232 *[0.48]*	-0.407 *[0.52]*	-0.3483 *[0.44]*	-0.2505 *[0.43]*	-0.1169 *[0.17]*	-0.0166 *[0.12]*	0.0564 *[0.077]*	0.1033 *[0.00]*	0.1429 *[0.07]*
Real Short-Term Interest Rates	2.1468 *[0.05]*	2.5217	0.7708 *[0.00]*	-0.4122 *[0.96]*	-0.3541 *[0.67]*	-0.2552 *[0.58]*	-0.128 *[0.67]*	0.024 *[0.96]*	0.0689 *[0.24]*	0.0923 *[0.38]*	0.0999 *[0.5]*	0.1241 *[0.49]*

Notes:
Probability-values in *[]*s.
See also note to Table 1.

Table 3: Spectral Comparison

	Output	Consumption	Investment	Employment
	Historical Data (1970:1-1999q4)			
Bandwidth Method	Persistence			
Andrews	2.2026 (0.98504)	3.0053 (1.1422)	3.2732 (1.5811)	1.1683 (0.69663)
$2\sqrt{T}$	1.5954 (0.78876)	2.1652 (1.0705)	3.0097 (1.4881)	2.7219 (1.3457)
	Periodicity			
	18 quarters [ω=0.34907]	26 quarters [ω=0.24166]	42 quarters [ω=0.14960]	32 quarters [ω=0.19635]
	Stochastic Simulation (Model)			
	Persistence			
Andrews	2.8189 (1.1836)	2.7789 (1.0915)	2.3450 (1.1281)	1.0032 (0.61407)
$2\sqrt{T}$	1.9817 (0 .97574)	1.9931 (0 .98131)	2.2023 (1.0843)	4.1322 (2.0345)
	Periodicity			
	32 quarters [ω=0.19635]	28 Quarters [ω=0.22440]	42 Quarters [ω=0.14960]	34 quarters [ω=0.18480]

Notes:

- Standard errors in ()s.
- Periodicity: $\frac{2\pi}{\omega_j}\Big|\max f(\omega_j)$.

Table 4. Simulated Out-Of-Sample Projections

GDP Deflator inflation

AR(2)						AR(2) with exogenous variables						AWM forecasts, no residual projection					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000802	0.001878	1.04	0.12	16	1	-0.001366	0.002248	1.25	0.26	16	1	-0.001153	0.001967	1.09	0.17	16
2	-0.001317	0.002342	1.11	0.13	15	2	-0.002094	0.002902	1.37	0.24	15	2	-0.001874	0.002737	1.29	0.18	15
3	-0.002044	0.002640	1.32	0.18	14	3	-0.003261	0.003743	1.87	0.32	14	3	-0.002926	0.003740	1.87	0.19	14
4	-0.002479	0.002821	1.86	0.09	13	4	-0.004147	0.004394	2.90	0.42	13	4	-0.003120	0.003682	2.43	0.29	13
5	-0.002988	0.003507	1.80	0.12	12	5	-0.005132	0.005496	2.82	0.63	12	5	-0.003665	0.004078	2.09	0.32	12
6	-0.003342	0.003922	1.69	0.20	11	6	-0.006013	0.006435	2.77	0.68	11	6	-0.003745	0.004213	1.82	0.31	11
7	-0.003766	0.004096	1.96	0.32	10	7	-0.006876	0.007193	3.44	0.93	10	7	-0.004063	0.004754	2.28	0.37	10
8	-0.004238	0.004396	2.47	0.28	9	8	-0.007807	0.008052	4.52	0.78	9	8	-0.004041	0.004658	2.61	0.19	9

AWM forecasts, residual projection by smooth level shift						AWM forecasts, residual projection by level shift						AWM forecasts, residual projection by smooth trend shift					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000278	0.001878	1.04	0.16	16	1	-0.000263	0.002473	1.38	0.10	16	1	-0.000194	0.002933	1.63	0.12	16
2	-0.000468	0.002352	1.11	0.20	15	2	-0.000401	0.002061	0.97	0.11	15	2	-0.000265	0.003082	1.45	0.17	15
3	-0.000688	0.003054	1.53	0.19	14	3	-0.000926	0.003761	1.88	0.19	14	3	-0.000810	0.005658	2.83	0.29	14
4	-0.000626	0.002632	1.74	0.20	13	4	-0.000890	0.003259	2.15	0.53	13	4	-0.000193	0.005885	3.88	0.82	13
5	-0.000767	0.002934	1.51	0.22	12	5	-0.001171	0.004629	2.38	0.57	12	5	0.000057	0.009132	4.69	0.94	12
6	-0.000671	0.003015	1.30	0.34	11	6	-0.000392	0.003908	1.68	0.32	11	6	0.002323	0.009639	4.15	0.80	11
7	-0.000975	0.003246	1.55	0.47	10	7	-0.000398	0.004936	2.36	0.33	10	7	0.002851	0.012033	5.76	1.13	10
8	-0.001155	0.002964	1.66	0.36	9	8	0.000567	0.005213	2.93	0.37	9	8	0.007080	0.011380	6.39	1.14	9

Statistics based on one- to twelve-step-ahead forecasts of 1st difference of log of variable, from 1996Q1 to 199Q4.

Theil's U is ratio of shown RMSE to RMSE of no-change assumption.

Standard Deviation of Theil's U is calculated by a delta-method expansion based on the HAC-corrected variance-covariance of the residuals squared, using a window with as many lags as steps ahead reported.

AR(2): AR(2) in first difference with a constant.

AR(2) with exogenous variables: AR(2) in first difference with a constant and fixed exogenous regressors as used for the AWM.

AWM forecasts, no residual projection: AWM used with flat residual projection

AWM forecasts, residual projection by smooth level shift: AWM with residuals projected as average of 4 quarters previous to beginning of forecast.

AWM forecasts, residual projection by level shift: AWM with residuals projected as quarter previous to beginning of forecast.

AWM forecasts, residual projection by smooth trend shift: AWM with residuals projected as a trend from average trend of 4 quarters previous to beginning of forecast.

Table 5. Simulated Out-Of-Sample Projections, RMSE Decomposition

GDP Deflator, Relative RMSE Decomposition

AWM forecasts, no residual projection						AWM forecasts, residual projection by smooth level shift					
Step Ahead	MSE/var(x)	var(x)/var(x)	var(f)/var(x)	cov(x,f)/var(x)	$bias^2$/var(x)	MSE/var(x)	var(x)/var(x)	var(f)/var(x)	cov(x,f)/var(x)	$bias^2$/var(x)	No. Obs.
1	1.966	1.000	1.288	0.499	0.675	1.792	1.000	1.488	0.368	0.039	16
2	3.882	1.000	1.411	0.174	1.819	2.865	1.000	2.048	0.148	0.113	15
3	7.184	1.000	1.457	-0.165	4.397	4.789	1.000	3.232	-0.157	0.243	14
4	6.669	1.000	1.131	0.125	4.789	3.408	1.000	2.937	0.361	0.193	13
5	8.023	1.000	1.066	0.261	6.478	4.154	1.000	3.656	0.393	0.284	12
6	8.046	1.000	1.035	0.174	6.360	4.121	1.000	3.756	0.419	0.204	11
7	9.613	1.000	1.255	-0.168	7.022	4.481	1.000	4.027	0.476	0.405	10
8	8.812	1.000	0.971	-0.105	6.632	3.568	1.000	3.206	0.590	0.542	9

AR(2)-cum-exogenous forecasts						AR(2) forecasts					
Step Ahead	MSE/var(x)	var(x)/var(x)	var(f)/var(x)	cov(x,f)/var(x)	$bias^2$/var(x)	MSE/var(x)	var(x)/var(x)	var(f)/var(x)	cov(x,f)/var(x)	$bias^2$/var(x)	No. Obs.
1	2.568	1.000	1.073	0.227	0.948	1.792	1.000	0.949	0.242	0.327	16
2	4.363	1.000	1.257	0.083	2.271	2.840	1.000	0.983	0.021	0.899	15
3	7.196	1.000	1.495	0.379	5.459	3.578	1.000	0.618	0.092	2.144	14
4	9.496	1.000	1.793	0.879	8.460	3.916	1.000	0.652	0.380	3.024	13
5	14.571	1.000	1.981	0.559	12.707	5.934	1.000	0.686	0.029	4.306	12
6	18.774	1.000	1.989	0.303	16.391	6.974	1.000	0.671	-0.119	5.065	11
7	22.006	1.000	1.819	0.460	20.108	7.136	1.000	0.542	0.219	6.033	10
8	26.333	1.000	1.908	0.666	24.757	7.847	1.000	0.573	0.510	7.294	9

Step Ahead: Number of steps ahead in the forecasts.

MSE: Mean-square errors of the forecasts.

var(x): (Sample) variance of actual observations.

var(f): (Sample) variance of forecasts.

cov(x,f): Covariance between actual and forecasts.

bias: Bias in the forecasts, i.e. average forecast error.

Table 6. Simulated In-Sample Projections

GDP Deflator inflation

AR(2)						AR(2) with exogenous variables						AWM forecasts, no residual projection					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000537	0.002065	0.97	0.05	40	1	-0.000502	0.001786	0.84	0.11	40	1	-0.000686	0.002437	1.15	0.15	40
2	-0.000929	0.002464	0.96	0.04	39	2	-0.000772	0.002106	0.82	0.09	39	2	-0.000970	0.002705	1.05	0.09	39
3	-0.001377	0.002669	1.02	0.05	38	3	-0.001091	0.002240	0.86	0.09	38	3	-0.001348	0.002976	1.14	0.13	38
4	-0.001713	0.002919	1.02	0.07	37	4	-0.001355	0.002395	0.84	0.10	37	4	-0.001489	0.003024	1.06	0.14	37
5	-0.001990	0.003018	1.10	0.08	36	5	-0.001545	0.002530	0.92	0.14	36	5	-0.001623	0.003110	1.14	0.17	36
6	-0.002454	0.003326	1.15	0.07	35	6	-0.001856	0.002803	0.97	0.15	35	6	-0.001938	0.003245	1.12	0.14	35
7	-0.002948	0.003647	1.12	0.08	34	7	-0.002226	0.003031	0.93	0.18	34	7	-0.002337	0.003361	1.03	0.12	34
8	-0.003341	0.003797	1.21	0.10	33	8	-0.002518	0.003160	1.01	0.24	33	8	-0.002567	0.003489	1.11	0.12	33
9	-0.003733	0.004130	1.22	0.15	32	9	-0.002733	0.003362	0.99	0.28	32	9	-0.002789	0.003630	1.07	0.12	32
10	-0.004070	0.004487	1.18	0.15	31	10	-0.002880	0.003515	0.93	0.29	31	10	-0.002939	0.003769	1.00	0.09	31
11	-0.004415	0.004767	1.23	0.15	30	11	-0.003037	0.003701	0.96	0.30	30	11	-0.003075	0.003917	1.01	0.10	30
12	-0.004815	0.005174	1.20	0.14	29	12	-0.003204	0.003920	0.91	0.27	29	12	-0.003368	0.004117	0.96	0.10	29

AWM forecasts, residual projection by smooth level shift						AWM forecasts, residual projection by corrected level shift						AWM forecasts, residual projection by corrected outlier					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000194	0.002783	1.31	0.17	40	1	-0.001008	0.002577	1.21	0.18	40	1	-0.001930	0.003996	1.88	0.34	40
2	-0.000386	0.003059	1.19	0.12	39	2	-0.001477	0.002954	1.15	0.15	39	2	-0.001955	0.003257	1.27	0.16	39
3	-0.000608	0.003141	1.21	0.14	38	3	-0.002114	0.003425	1.31	0.14	38	3	-0.002876	0.004019	1.54	0.21	38
4	-0.000702	0.003299	1.15	0.13	37	4	-0.002315	0.003569	1.25	0.13	37	4	-0.002753	0.003820	1.33	0.16	37
5	-0.000735	0.003303	1.21	0.16	36	5	-0.002624	0.003677	1.34	0.15	36	5	-0.003310	0.004240	1.55	0.18	36
6	-0.000984	0.003759	1.30	0.15	35	6	-0.003032	0.004192	1.45	0.15	35	6	-0.003739	0.004642	1.61	0.17	35
7	-0.001363	0.004315	1.33	0.12	34	7	-0.003626	0.004684	1.44	0.13	34	7	-0.004430	0.005116	1.57	0.13	34
8	-0.001593	0.004535	1.44	0.14	33	8	-0.003936	0.005102	1.63	0.17	33	8	-0.004630	0.005364	1.71	0.17	33
9	-0.001872	0.004883	1.44	0.09	32	9	-0.004278	0.005460	1.61	0.11	32	9	-0.005048	0.005769	1.70	0.11	32
10	-0.002091	0.005247	1.39	0.07	31	10	-0.004554	0.005842	1.54	0.11	31	10	-0.005311	0.006084	1.61	0.08	31
11	-0.002305	0.005478	1.42	0.10	30	11	-0.004826	0.006218	1.61	0.19	30	11	-0.005641	0.006423	1.66	0.13	30
12	-0.002692	0.005906	1.37	0.16	29	12	-0.005301	0.006867	1.60	0.20	29	12	-0.006032	0.006816	1.59	0.16	29

Statistics based on one- to twelve-step-ahead forecasts of 1^{st} difference of log of variable, from 1996Q1 to 199Q4.

Theil's U is ratio of shown RMSE to RMSE of no-change assumption.

Standard Deviation of Theil's U is calculated by a delta-method expansion based on the HAC-corrected variance-covariance of the residuals squared, using a window with as many lags as steps ahead reported.

AR(2): AR(2) in first difference with a constant.

AR(2) with exogenous variables: AR(2) in first difference with a constant and fixed exogenous regressors as used for the AWM.

AWM forecasts, no residual projection: AWM used with flat residual projection

AWM forecasts, residual projection by smooth level shift: AWM with residuals projected as average of 4 quarters previous to beginning of forecast.

AWM forecasts, residual projection by corrected level shift: AWM with residuals projected as quarter previous to beginning of forecast, corrected in subsequent periods by EqCM of corresponding equation.

AWM forecasts, residual projection by corrected outlier: AWM with residuals projected as quarter previous to beginning of forecast corrected by the EqCM of the equation.

Table 7. Simulated In-Sample Projections, RMSE Decomposition

GDP Deflator, Relative RMSE Decomposition

AWM forecasts, no residual projection						AWM forecasts, residual projection by smooth level shift					
Step Ahead	**MSE/var(x)**	**var(x)/var(x)**	**var(f)/var(x)**	**cov(x,f)/var(x)**	**bias^2/var(x)**	**MSE/var(x)**	**var(x)/var(x)**	**var(f)/var(x)**	**cov(x,f)/var(x)**	**bias^2/var(x)**	**No. Obs.**
1	0.440	1.000	0.931	0.763	0.035	0.574	1.000	1.176	0.803	0.003	40
2	0.564	1.000	0.861	0.685	0.072	0.720	1.000	1.300	0.796	0.011	39
3	0.693	1.000	0.657	0.553	0.142	0.771	1.000	1.286	0.772	0.029	38
4	0.723	1.000	0.690	0.571	0.175	0.860	1.000	1.420	0.799	0.039	37
5	0.747	1.000	0.609	0.533	0.204	0.843	1.000	1.550	0.874	0.042	36
6	0.927	1.000	0.705	0.554	0.331	1.245	1.000	2.036	0.938	0.085	35
7	1.187	1.000	0.810	0.599	0.574	1.958	1.000	2.797	1.017	0.195	34
8	1.428	1.000	0.934	0.640	0.773	2.412	1.000	3.409	1.147	0.298	33
9	1.819	1.000	1.051	0.653	1.074	3.293	1.000	4.124	1.157	0.484	32
10	2.062	1.000	1.118	0.655	1.253	3.995	1.000	4.523	1.081	0.634	31
11	2.280	1.000	1.129	0.627	1.405	4.460	1.000	4.755	1.042	0.789	30
12	2.842	1.000	1.342	0.701	1.901	5.848	1.000	5.617	0.992	1.215	29
AR(2)-cum-exogenous forecasts						**AR(2) forecasts**					
Step Ahead	**MSE/var(x)**	**var(x)/var(x)**	**var(f)/var(x)**	**cov(x,f)/var(x)**	**bias^2/var(x)**	**MSE/var(x)**	**var(x)/var(x)**	**var(f)/var(x)**	**cov(x,f)/var(x)**	**bias^2/var(x)**	**No. Obs.**
1	0.236	1.000	0.834	0.808	0.019	0.316	1.000	0.857	0.782	0.021	40
2	0.342	1.000	0.811	0.758	0.046	0.468	1.000	0.836	0.717	0.066	39
3	0.392	1.000	0.746	0.724	0.093	0.557	1.000	0.728	0.659	0.148	38
4	0.453	1.000	0.720	0.706	0.145	0.673	1.000	0.689	0.623	0.232	37
5	0.495	1.000	0.677	0.683	0.185	0.704	1.000	0.621	0.612	0.306	36
6	0.692	1.000	0.665	0.638	0.304	0.974	1.000	0.649	0.602	0.530	35
7	0.966	1.000	0.646	0.600	0.521	1.398	1.000	0.691	0.603	0.913	34
8	1.171	1.000	0.640	0.606	0.744	1.691	1.000	0.714	0.666	1.309	33
9	1.561	1.000	0.586	0.528	1.031	2.355	1.000	0.790	0.679	1.924	32
10	1.793	1.000	0.474	0.442	1.204	2.921	1.000	0.755	0.619	2.403	31
11	2.035	1.000	0.425	0.380	1.370	3.377	1.000	0.660	0.590	2.896	30
12	2.577	1.000	0.384	0.264	1.721	4.488	1.000	0.637	0.517	3.886	29

Step Ahead: Number of steps ahead in the forecasts.

MSE: Mean-square errors of the forecasts.

var(x): (Sample) variance of actual observations.

var(f): (Sample) variance of forecasts.

cov(x,f): Covariance between actual and forecasts.

bias: Bias in the forecasts, i.e. average forecast error.

Table 8. Simulated In-Sample Projections

GDP growth

AR(2)

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000713	0.004162	0.85	0.14	40
2	-0.001055	0.004307	0.77	0.15	39
3	-0.001200	0.004528	0.78	0.17	38
4	-0.001343	0.004692	0.65	0.11	37
5	-0.001366	0.004695	0.67	0.16	36
6	-0.001413	0.004756	0.67	0.17	35
7	-0.001503	0.004817	0.64	0.14	34
8	-0.001485	0.004863	0.69	0.16	33
9	-0.001441	0.004908	0.73	0.13	32
10	-0.001639	0.004920	0.67	0.09	31
11	-0.001387	0.004708	0.68	0.09	30
12	-0.001250	0.004683	0.63	0.06	29

AR(2) with exogenous variables

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000621	0.003882	0.79	0.12	40
2	-0.000763	0.003917	0.70	0.10	39
3	-0.000798	0.004050	0.69	0.11	38
4	-0.000922	0.004065	0.57	0.08	37
5	-0.001019	0.004100	0.59	0.11	36
6	-0.001272	0.003941	0.55	0.11	35
7	-0.001441	0.003924	0.52	0.09	34
8	-0.001643	0.003878	0.55	0.10	33
9	-0.001692	0.003938	0.58	0.07	32
10	-0.001929	0.003870	0.53	0.06	31
11	-0.001771	0.003740	0.54	0.07	30
12	-0.001739	0.003772	0.51	0.08	29

AWM forecasts, no residual projection

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000705	0.005441	1.11	0.16	40
2	-0.000936	0.005647	1.01	0.16	39
3	-0.001068	0.005263	0.90	0.13	38
4	-0.001143	0.005229	0.73	0.09	37
5	-0.001045	0.005324	0.76	0.12	36
6	-0.001114	0.005375	0.76	0.13	35
7	-0.001242	0.005368	0.72	0.10	34
8	-0.001057	0.005175	0.73	0.11	33
9	-0.000938	0.005129	0.76	0.08	32
10	-0.001170	0.005201	0.71	0.07	31
11	-0.000983	0.005169	0.74	0.07	30
12	-0.000978	0.005245	0.71	0.07	29

AWM forecasts, residual projection by smooth level shift

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000135	0.005423	1.11	0.15	40
2	-0.000292	0.005546	1.00	0.14	39
3	-0.000323	0.005313	0.91	0.12	38
4	-0.000545	0.006119	0.85	0.10	37
5	-0.000446	0.006611	0.94	0.10	36
6	-0.000538	0.006803	0.96	0.07	35
7	-0.000728	0.006788	0.91	0.06	34
8	-0.000559	0.006527	0.92	0.06	33
9	-0.000466	0.006422	0.95	0.08	32
10	-0.000717	0.006544	0.89	0.08	31
11	-0.000608	0.006532	0.94	0.07	30
12	-0.000617	0.006374	0.86	0.08	29

AWM forecasts, residual projection by corrected level shift

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	0.000085	0.005856	1.19	0.16	40
2	-0.000282	0.006044	1.09	0.17	39
3	-0.000781	0.006251	1.07	0.17	38
4	-0.000945	0.006029	0.84	0.12	37
5	-0.000978	0.006686	0.95	0.14	36
6	-0.001023	0.007187	1.01	0.15	35
7	-0.001465	0.006725	0.90	0.09	34
8	-0.001075	0.006911	0.97	0.13	33
9	-0.001094	0.006281	0.93	0.14	32
10	-0.001374	0.006225	0.85	0.11	31
11	-0.001275	0.006388	0.92	0.09	30
12	-0.001493	0.006889	0.93	0.12	29

AWM forecasts, residual projection by corrected outlier

Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000189	0.008076	1.65	0.29	40
2	-0.001187	0.005761	1.03	0.16	39
3	-0.002039	0.005915	1.01	0.16	38
4	-0.001738	0.005834	0.81	0.09	37
5	-0.001934	0.005851	0.83	0.15	36
6	-0.002044	0.005949	0.84	0.16	35
7	-0.002426	0.006121	0.82	0.13	34
8	-0.001723	0.005935	0.84	0.15	33
9	-0.002073	0.005845	0.87	0.12	32
10	-0.002372	0.006037	0.82	0.09	31
11	-0.002271	0.005880	0.85	0.08	30
12	-0.002324	0.005896	0.79	0.07	29

Statistics based on one- to twelve-step-ahead forecasts of 1st difference of log of variable, from 1996Q1 to 199Q4.

Theil's U is ratio of shown RMSE to RMSE of no-change assumption.

Standard Deviation of Theil's U is calculated by a delta-method expansion based on the HAC-corrected variance-covariance of the residuals squared, using a window with as many lags as steps ahead reported.

AR(2): AR(2) in first difference with a constant.

AR(2) with exogenous variables: AR(2) in first difference with a constant and fixed exogenous regressors as used for the AWM.

AWM forecasts, no residual projection: AWM used with flat residual projection

AWM forecasts, residual projection by smooth level shift: AWM with residuals projected as average of 4 quarters previous to beginning of forecast.

AWM forecasts, residual projection by corrected level shift: AWM with residuals projected as quarter previous to beginning of forecast, corrected in subsequent periods by EqCM of corresponding equation.

AWM forecasts, residual projection by corrected outlier: AWM with residuals projected as quarter previous to beginning of forecast corrected by the EqCM of the equation.

Table 9. Simulated In-Sample Projections

Consumption Deflator inflation

AR(2)						AR(2) with exogenous variables						AWM forecasts, no residual projection					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000308	0.001909	0.92	0.04	40	1	-0.000276	0.001801	0.86	0.08	40	1	-0.000188	0.001447	0.69	0.14	40
2	-0.000590	0.002090	0.94	0.04	39	2	-0.000448	0.002104	0.95	0.11	39	2	-0.000435	0.001944	0.87	0.15	39
3	-0.000831	0.002212	0.99	0.06	38	3	-0.000596	0.002390	1.07	0.12	38	3	-0.000700	0.002329	1.04	0.16	38
4	-0.001196	0.002508	1.03	0.04	37	4	-0.000706	0.002528	1.04	0.09	37	4	-0.000917	0.002464	1.01	0.16	37
5	-0.001568	0.002841	1.01	0.06	36	5	-0.000807	0.002667	0.95	0.08	36	5	-0.001091	0.002602	0.93	0.13	36
6	-0.001949	0.002920	1.10	0.08	35	6	-0.000980	0.002571	0.97	0.09	35	6	-0.001384	0.002695	1.01	0.14	35
7	-0.002326	0.003262	1.11	0.08	34	7	-0.001132	0.002550	0.87	0.08	34	7	-0.001690	0.002898	0.99	0.16	34
8	-0.002793	0.003472	1.14	0.09	33	8	-0.001333	0.002398	0.79	0.08	33	8	-0.001999	0.003035	1.00	0.14	33
9	-0.003147	0.003691	1.16	0.10	32	9	-0.001504	0.002417	0.76	0.11	32	9	-0.002220	0.003188	1.01	0.15	32
10	-0.003448	0.003947	1.20	0.08	31	10	-0.001651	0.002533	0.77	0.12	31	10	-0.002444	0.003323	1.01	0.16	31
11	-0.003814	0.004309	1.17	0.09	30	11	-0.001785	0.002708	0.73	0.11	30	11	-0.002658	0.003465	0.94	0.13	30
12	-0.004049	0.004525	1.19	0.10	29	12	-0.001801	0.002774	0.73	0.12	29	12	-0.002849	0.003646	0.96	0.16	29

AWM forecasts, residual projection by smooth level shift						AWM forecasts, residual projection by corrected level shift						AWM forecasts, residual projection by corrected outlier					
Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs	Step	Mean Error	RMS Error	Theil U	Std. Dev.	N. Obs
1	-0.000220	0.001555	0.75	0.15	40	1	-0.000181	0.001468	0.71	0.14	40	1	-0.000182	0.002099	1.01	0.20	40
2	-0.000321	0.002282	1.03	0.18	39	2	-0.000663	0.002048	0.92	0.16	39	2	-0.001224	0.002780	1.25	0.20	39
3	-0.000462	0.002843	1.27	0.21	38	3	-0.001130	0.002513	1.12	0.17	38	3	-0.001692	0.002929	1.31	0.19	38
4	-0.000557	0.003024	1.24	0.17	37	4	-0.001469	0.002806	1.15	0.15	37	4	-0.001953	0.003047	1.25	0.15	37
5	-0.000602	0.003172	1.13	0.18	36	5	-0.001756	0.002979	1.06	0.13	36	5	-0.002286	0.003261	1.16	0.13	36
6	-0.000748	0.003415	1.28	0.17	35	6	-0.002195	0.003345	1.26	0.16	35	6	-0.002911	0.003756	1.41	0.17	35
7	-0.000986	0.003724	1.27	0.13	34	7	-0.002688	0.003761	1.28	0.15	34	7	-0.003431	0.004219	1.44	0.17	34
8	-0.001271	0.004014	1.32	0.09	33	8	-0.003123	0.004222	1.39	0.14	33	8	-0.003833	0.004535	1.49	0.15	33
9	-0.001531	0.004163	1.31	0.12	32	9	-0.003454	0.004464	1.41	0.19	32	9	-0.004186	0.004874	1.54	0.19	32
10	-0.001809	0.004360	1.32	0.18	31	10	-0.003832	0.005108	1.55	0.30	31	10	-0.004593	0.005278	1.60	0.27	31
11	-0.002100	0.004673	1.27	0.15	30	11	-0.004182	0.005384	1.46	0.24	30	11	-0.004991	0.005660	1.54	0.22	30
12	-0.002368	0.005124	1.35	0.23	29	12	-0.004555	0.005845	1.54	0.29	29	12	-0.005305	0.006003	1.58	0.27	29

Statistics based on one- to twelve-step-ahead forecasts of 1^{st} difference of log of variable, from 1996Q1 to 199Q4.

Theil's U is ratio of shown RMSE to RMSE of no-change assumption.

Standard Deviation of Theil's U is calculated by a delta-method expansion based on the HAC-corrected variance-covariance of the residuals squared, using a window with as many lags as steps ahead reported.

AR(2): AR(2) in first difference with a constant.

AR(2) with exogenous variables: AR(2) in first difference with a constant and fixed exogenous regressors as used for the AWM.

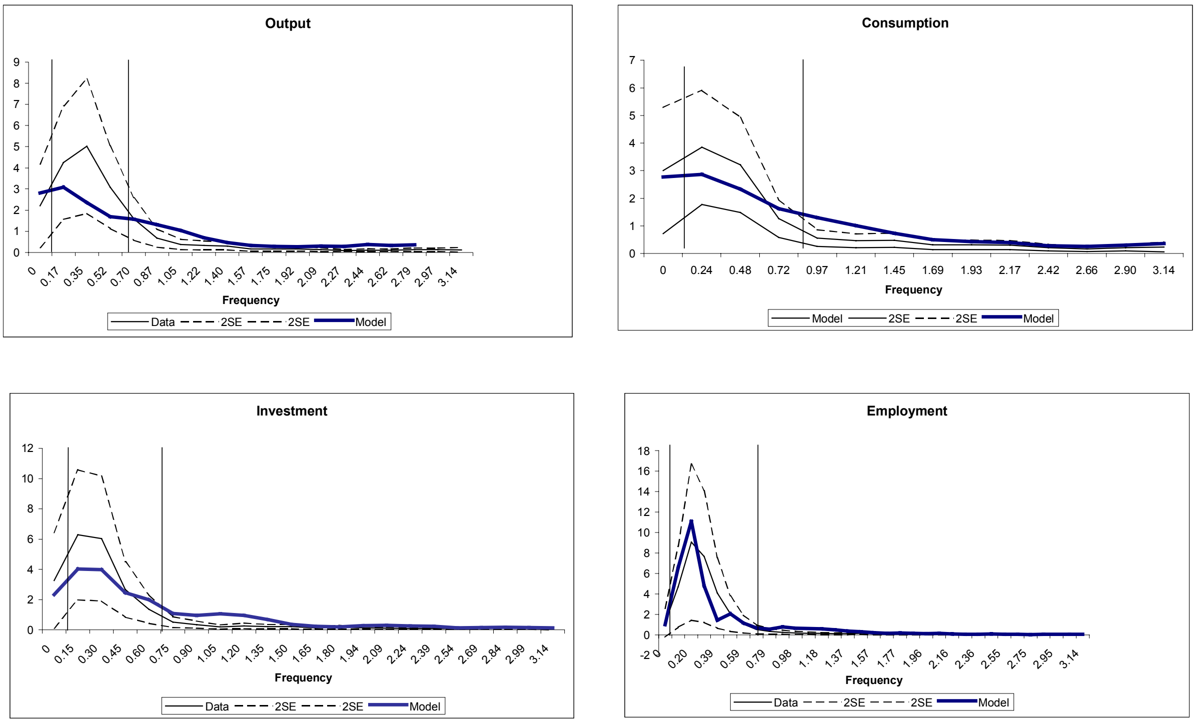
AWM forecasts, no residual projection: AWM used with flat residual projection

AWM forecasts, residual projection by smooth level shift: AWM with residuals projected as average of 4 quarters previous to beginning of forecast.

AWM forecasts, residual projection by corrected level shift: AWM with residuals projected as quarter previous to beginning of forecast, corrected in subsequent periods by EqCM of corresponding equation.

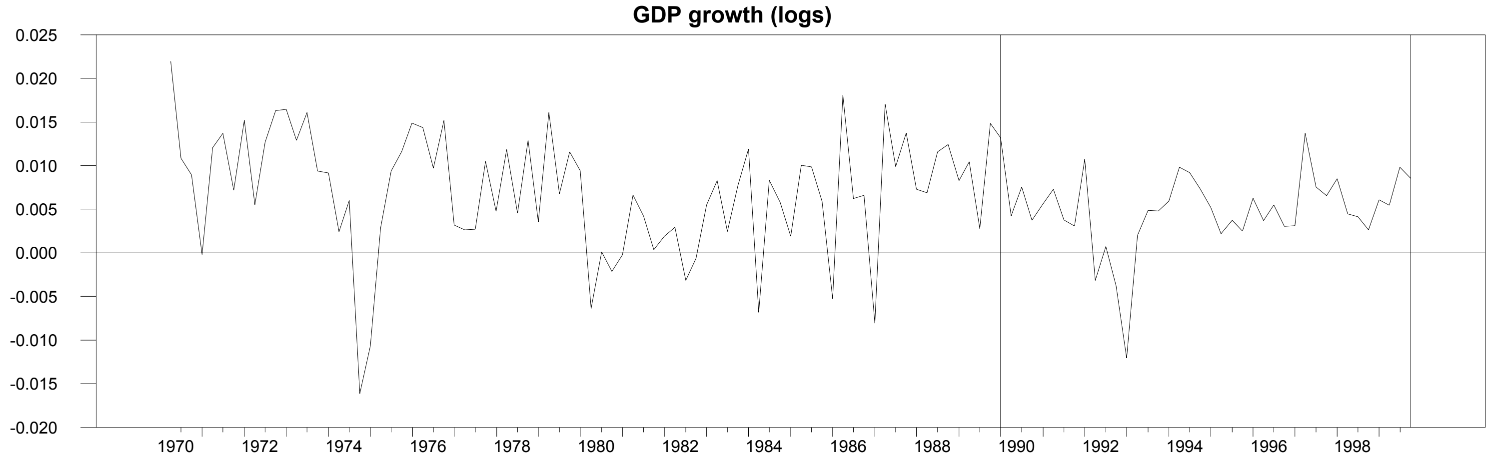
AWM forecasts, residual projection by corrected outlier: AWM with residuals projected as quarter previous to beginning of forecast corrected by the EqCM of the equation.

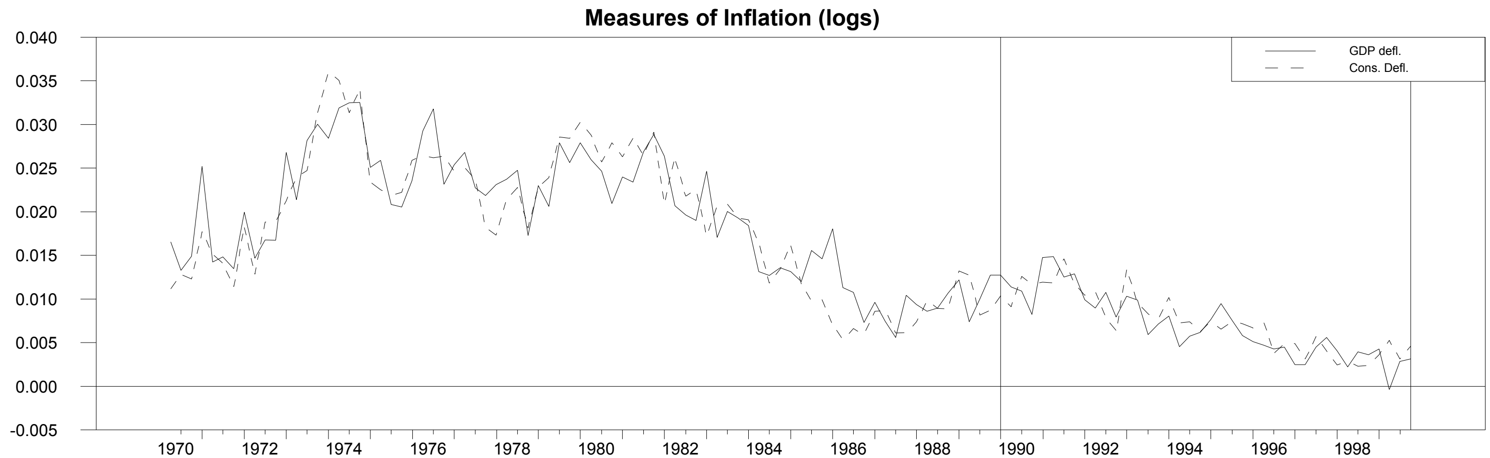
Figures 1: Spectral Densities (Historical and Model-Generated Data)



Note: Here we present results from the Andrews Bandwidth procedure, though the conclusions are the same across both methods.

Figure 2. Forecast Series





ANNEX: RECURSIVELY ESTIMATED PARAMETERS

Simulated out-of-sample forecasts are conventionally made using no information from periods within the forecast horizon, except maybe for the use of final, fully-revised datasets maybe not available at the time. The exercise described in the main text actually did use information coming from the forecast period, i.e. the exchange and interest rates and external and fiscal variables, but re-estimating the model was nevertheless seen as a crucial part of the exercise. This was so because of the desire to separate as far as possible forecast failure due to unstable parameters and due to deterministic shifts in the data. In order to achieve this, the model was re-estimated for the periods 1995Q4 to 1999Q3 using either the standard equations in the model or simplified versions in which dummy variables for the period were dropped. The model was found to be broadly stable for the period, with parameters in general changing not much. As a consequence, forecast performance compared to benchmark models probably depends considerably more on residual projection than on parameter instability.

Some insight into stability properties of the parameters of the model can be gained with the table and graph included in this annex. The table records the value of the recursively re-estimated parameters for the period, showing only 4th quarter results to condense information. Results reported under the heading 'original' recall original values for the parameters. Results are reported naming parameters following the convention in Fagan *et al*. (op. cit.), with the exception of consumption and stock equations, which have been re-specified. Parameters shown under 'original' may differ slightly from those reported in the previous reference due to slight changes in the database.[39]

The graph shows, for the period under analysis, recursive estimation residuals superimposed on a same graph. Equations are labelled according to the name of the variable they define, again following conventions set in Fagan *et al*. (op. cit.). The evident overlap of most residuals is another indication that equations were relatively stable throughout the period.

[39] The database used was a version of the original database distributed with Fagan *et al.* (2001) and downloadable from www.ecb.int. The database used here was extended until 1999Q4 and was used in Detken *et al.* (2002). The extension involved slight historical revisions to some variables, which explains for the mentioned differences.

Table A.1. Selection of Recursive Estimated Parameters

Equation	Parameter	Original	1995q4	1996q4	1997q4	1998q4	1999q4
Employment	LNN.DLLNT	0.692276	0.683772	0.682438	0.691779	0.713435	0.691929
	LNN.DLYERADJ	0.180177	0.183676	0.183327	0.180310	0.167283	0.170650
	LNN.DLWRRADJ	-0.120938	-0.129171	-0.127726	-0.121410	-0.120343	-0.119579
	LNN.DLWRRADJ1	-0.125018	-0.127474	-0.127635	-0.125020	-0.121545	-0.128524
	LNN.D872	0.004483	0.004585	0.004609	0.004490	0.004496	0.004343
	LNN.D841	-0.004429	-0.004487	-0.004503	-0.004431	-0.004187	-0.004536
	LNN.ECM	-0.081495	-0.082260	-0.082408	-0.081525	-0.082152	-0.068280
Investment	ITR.D894	0.020823	0.020511	0.020930	0.020983	0.021269	0.021455
	ITR.DITY1	0.168826	0.247538	0.174441	0.167245	0.142753	0.141033
	ITR.ECM	0.539096	0.504877	0.558784	0.554734	0.591342	0.606284
	ITR.ADJ	0.012310	0.012285	0.012396	0.012489	0.012645	0.012848
Wages	WRN.DLWRCQ4	0.273892	0.263043	0.255321	0.250201	0.252917	0.282311
	WRN.DDLPCD	-0.919736	-0.797606	-0.810661	-0.814063	-0.781790	-0.764833
	WRN.DDLPCD1	-0.571055	-0.390019	-0.391060	-0.401781	-0.397747	-0.371002
	WRN.DDLPCD2	-0.470158	-0.370688	-0.374635	-0.376440	-0.422863	-0.426678
	WRN.DDLPCD3	-0.334405	-0.268985	-0.280349	-0.294451	-0.289109	-0.301979
	WRN.DDLPROD	-0.562628	-0.549744	-0.545439	-0.537485	-0.547781	-0.554636
	WRN.DDLPROD1	-0.456273	-0.437333	-0.436445	-0.430985	-0.447532	-0.460392
	WRN.DDLPROD2	-0.399817	-0.392926	-0.391432	-0.393959	-0.410047	-0.429397
	WRN.DDLPROD3	-0.261245	-0.272927	-0.272842	-0.272394	-0.297906	-0.315672
	WRN.LURX_GAP1	-0.014744	-0.021272	-0.021843	-0.022551	-0.022536	-0.021635
	WRN.I81Q1	-0.004171	-0.010014	-0.010113	-0.010121	-0.010444	-0.010449
	WRN.I84Q2	-0.012449	-0.016239	-0.016169	-0.015979	-0.016072	-0.015970
GDP deflator	YFD.CST	0.003920	0.003995	0.003993	0.003920	0.003963	0.004016
	YFD.DLYFD1	0.229684	0.226243	0.225394	0.229684	0.227293	0.219794
	YFD.DLULT	0.246086	0.251419	0.249390	0.246086	0.245890	0.230702
	YFD.DLULT1	0.083541	0.078281	0.082691	0.083541	0.090131	0.102890
	YFD.DLULT2	0.155765	0.157398	0.155377	0.155765	0.151721	0.159442
	YFD.DLMTD1	0.031486	0.032161	0.031956	0.031486	0.031346	0.032186
	YFD.ECM	-0.045006	-0.042939	-0.043548	-0.045006	-0.043234	-0.041669
Consumption deflator	PCD.ECM.LYED	0.940751	0.992626	0.991629	0.990656	0.989898	0.989107
	PCD.ECM.LMTD	0.059249	0.007374	0.008371	0.009344	0.010102	0.010893
	PCD.CST	0.001263	0.000652	0.000717	0.000709	0.000740	0.000763
	PCD.DLPCD4	0.188246	0.208861	0.208447	0.207154	0.205507	0.203585
	PCD.DLYED	0.445458	0.450697	0.448685	0.450976	0.448738	0.441644
	PCD.DLYED1	0.226345	0.243509	0.242692	0.242073	0.243997	0.252726
	PCD.DLMTD	0.071909	0.069102	0.069417	0.070809	0.070087	0.069731
	PCD.DLMTD1	0.025380	0.026242	0.026219	0.024868	0.025318	0.024661
	PCD.DLCOMPREEN	0.004465	0.004483	0.004479	0.004373	0.004510	0.004672
	PCD.ECM	-0.060559	-0.042746	-0.043246	-0.043550	-0.043238	-0.043549
	PCD.DI82Q1	-0.003529	-0.003579	-0.003572	-0.003584	-0.003577	-0.003566
	PCD.DI92Q4	-0.002979	-0.003081	-0.003081	-0.003084	-0.003086	-0.003108
	PCD.I77Q4I78Q1	-0.003981	-0.004023	-0.004018	-0.004017	-0.004007	-0.004006
Investment deflator	ITD.CST	0.004648	0.005854	0.004583	0.004642	0.004919	0.004811
	ITD.DDLYFD	-0.368482	-0.367276	-0.363803	-0.367348	-0.355295	-0.365059
	ITD.DDLMTD	0.111071	0.109045	0.109335	0.110476	0.111505	0.113977
	ITD.DLITDYFD1	0.313957	0.309011	0.309662	0.310311	0.306023	0.296158
	ITD.DLITDYFD4	0.207904	0.206759	0.211355	0.210999	0.219258	0.238621
	ITD.DLITDYFD7	0.169020	0.168403	0.165539	0.168244	0.165021	0.166495
	ITD.DLITDMTD1	-0.116393	-0.118348	-0.116749	-0.116844	-0.116678	-0.118108
	ITD.DLITDMTD5	0.047888	0.046081	0.048674	0.048114	0.048097	0.049652
	ITD.DLITDMTD6	0.037784	0.035204	0.036635	0.037602	0.037548	0.040126
	ITD.ECM1	-0.052662	-0.062607	-0.051976	-0.052571	-0.055305	-0.054783
	ITD.ECM2	-0.012261	-0.012873	-0.012128	-0.012233	-0.012725	-0.012951
Consumption	PCR.CST	-0.179412	-0.214655	-0.219753	-0.223782	-0.219428	-0.204340
	PCR.DLPYR	0.650734	0.643854	0.645894	0.650466	0.646259	0.638657
	PCR.DLSTR	-0.342388	-0.355225	-0.365362	-0.374164	-0.370326	-0.367107
	PCR.DURX	-0.629524	-0.788679	-0.782355	-0.788942	-0.822881	-0.810950
	PCR.ECM.LPYR	-0.101508	-0.113101	-0.112870	-0.114631	-0.110789	-0.102438
	PCR.ECM.LWLR	-0.051959	-0.062463	-0.064086	-0.065255	-0.064121	-0.059840
	PCR.I93Q1	-0.008909	-0.008790	-0.008816	-0.008692	-0.008790	-0.009047
Variation of Stocks	LSR.CST	-0.000243	-0.005395	-0.005560	-0.005646	-0.005734	-0.005653
	LSR.DLSR1	0.737271	0.754211	0.755744	0.750514	0.752620	0.755519
	LSR.DLSTR	-0.146929	-0.143087	-0.144643	-0.125457	-0.130754	-0.136393
	LSR.DYER	0.314350	0.306871	0.308657	0.306798	0.302035	0.303639
	LSR.ECM	-0.034750	-0.025073	-0.025249	-0.025958	-0.026715	-0.026384
	LSR.LSRYER	-0.098435	-0.097457	-0.097609	-0.098435	-0.099109	-0.098928
Exports	XTR.CST	0.222518	0.280269	0.247217	0.202111	0.216763	0.219860
	XTR.DLXTDYWDX1	-0.377152	-0.379655	-0.369044	-0.368404	-0.358968	-0.363958
	XTR.DLXTRYWRX7	0.156410	0.172929	0.197856	0.148231	0.143344	0.135790
	XTR.ECM	-0.121453	-0.148866	-0.133110	-0.111407	-0.118791	-0.120393
	XTR.LXTDYWDX1	-0.097625	-0.074571	-0.084366	-0.104097	-0.100995	-0.101605
	XTR.TIME	0.000989	0.001090	0.001029	0.000937	0.000979	0.000990
Imports	MTR.ECM.LMTDYED	-0.285416	-0.266180	-0.271016	-0.285416	-0.303068	-0.321489
	MTR.ECM.TIME	0.003407	0.003288	0.003311	0.003407	0.003469	0.003528
	MTR.CST	-0.158258	-0.169305	-0.171293	-0.158258	-0.138373	-0.133373
	MTR.DLFDD	2.021422	1.986599	1.995031	2.021422	2.034089	2.038034
	MTR.ECM	-0.086131	-0.092455	-0.093544	-0.086131	-0.075083	-0.072275
	MTR.D743	0.016717	0.016817	0.016916	0.016717	0.016185	0.016313
Export deflator	XTD.CST	-0.004465	-0.003598	-0.004147	-0.004465	-0.004735	-0.004804
	XTD.DLXTD1	0.236281	0.246397	0.242103	0.236281	0.249249	0.260540
	XTD.DLYED	0.720610	0.671128	0.701749	0.720601	0.725697	0.725177
	XTD.DLEEN	0.119249	0.122203	0.122809	0.119249	0.120750	0.119498
	XTD.DLMTD	0.219660	0.221619	0.219710	0.219660	0.213592	0.205837
	XTD.ECM	-0.034589	-0.037129	-0.036290	-0.034589	-0.032859	-0.031714

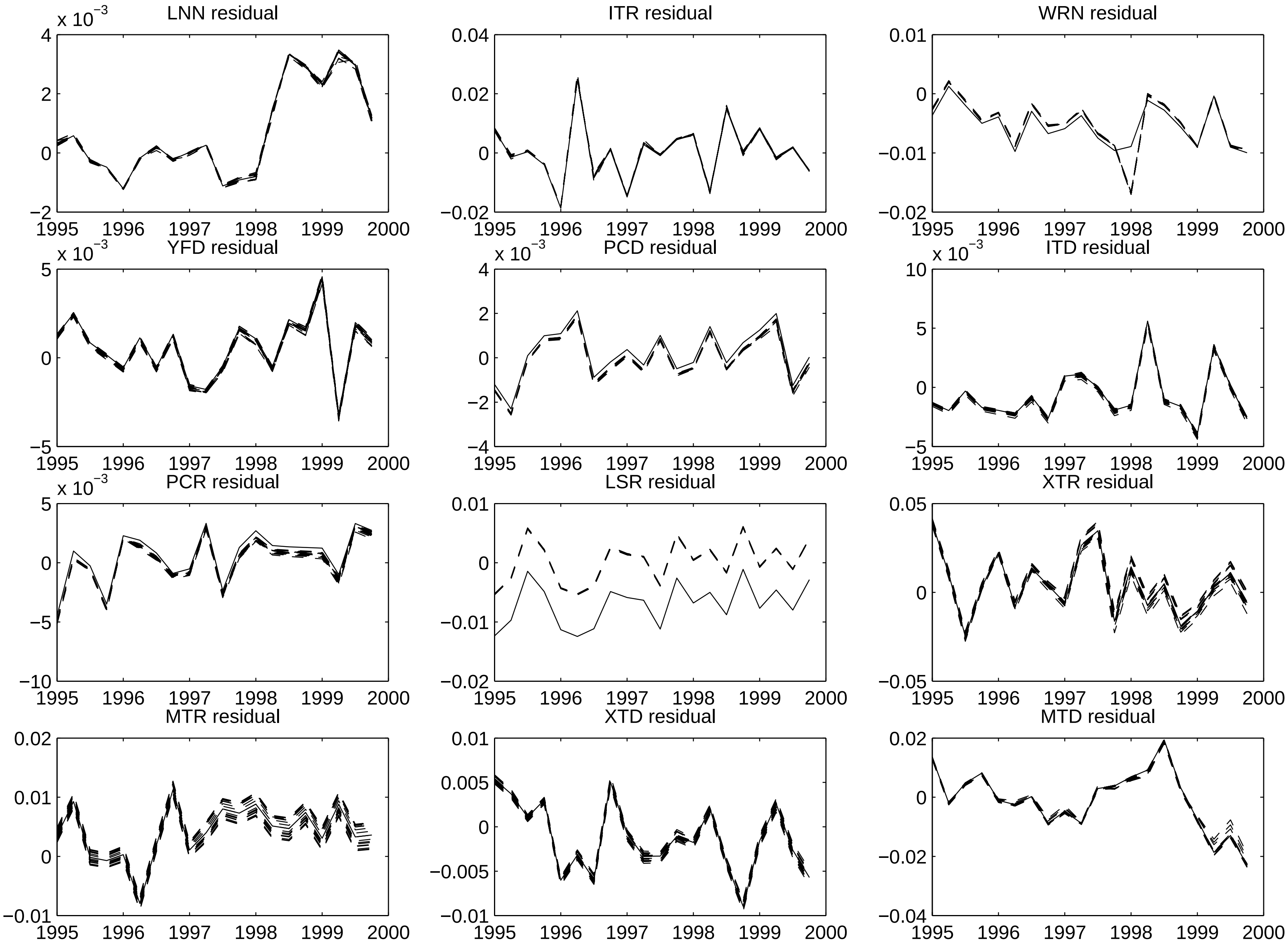


Figure A.1. Recursive Residuals