CARDIFF BUSINESS SCHOOL
WORKING PAPER SERIES

Cardiff Economics
Working Papers

Vo Phuong Mai Le, David Meenagh, Patrick Minford and Michael Wickens

How much nominal rigidity is there in the US economy? Testing a New Keynesian DSGE model using indirect inference

E2008/32

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
www.cardiff.ac.uk/carbs

ISSN 1749-6101
December 2008/Updated March 2011

How much nominal rigidity is there in the US economy? Testing a New Keynesian DSGE Model using indirect inference*

Vo Phuong Mai Le (University of Sheffield)† David Meenagh (Cardiff University)‡

Patrick Minford (Cardiff University and CEPR)§

Michael Wickens (Cardiff University, University of York and CEPR)¶

March 3, 2011

Abstract

We evaluate the Smets-Wouters model of the US using indirect inference and the bootstrap with a VAR representation of the main US data series. We find that the New Keynesian SW model is strongly rejected by the data's dynamic properties and in particular cannot match the variability of the data. An alternative (New Classical) version of the model with flexible wages and prices and a one-period information lag fares no better. A 'weighted' model (mostly New Classical but part New Keynesian) is better able to match the data variability, though it too is rejected overall. Allowing for structural breaks in the monetary regime we find a model from 1984 onwards fits fairly well dynamically; this has a high New Keynesian weight, suggesting much greater nominal stickiness during the 'great moderation'. Our results are robust to a variety of concerns about the bootstrap, parameter uncertainty, and numerical procedures.

JEL Classification: C12, C32, C52, E1,

Keywords: Bootstrap, US model, DSGE, VAR, New Keynesian, New Classical, indirect inference, Wald statistic, regime change, structural break, great moderation

---

*An earlier version of this paper was presented in Tokyo, January 2010, at the JEDC conference on Frontiers in Structural Macroeconomic Modelling. We thank Lawrence Christiano, Laurence Copeland, Huw Dixon, Martin Eichenbaum, Christopher Sims and especially Masao Ogaki, our discussant at the conference, for their helpful comments. We also thank the editors of this volume and three anonymous referees for extremely useful feedback. This work was supported by the UK's Economic and Social Research Council under grants RES-165-25-0020 and PTA-026-27-1623. A full set of results can be found in the Annexes to this paper at www.cf.ac.uk/carbs/faculty/minfordp/E2008_32Annex.pdf

†V.Le@sheffield.ac.uk; University of Sheffield, Department of Economics, 9 Mappin Street, Sheffield, UK

‡Meenaghd@cf.ac.uk; Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

§Patrick.minford@btinternet.com; Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

¶Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

1 Introduction

In this paper we propose a new way to test dynamic stochastic general equilibrium (DSGE) models that is based on indirect inference and apply this to one of the leading DSGE models of the US economy, the model of Smets and Wouters (2007, SW). A key feature of this model is that it has sticky prices and wages, i.e. it is a New Keynesian (NK) model. The extent of nominal rigidity is a major area of disagreement between economists. We therefore compare the SW model with a New Classical (NC) version of their model which has flexible prices and wages but lagged information in the form of a one-quarter delay for households in receiving macro information. We also consider the possibility that the economy consists of a mixture of the two in which some parts of the economy display nominal rigidities and other parts do not.

We find that for post-war data a hybrid model, in which most of the economy enjoys price and wage flexibility, but a non-negligible part of the economy is subject to nominal contracts, comes closest to matching the data, whereas the NK and NC models are seriously at odds with the data. If, however, we use only data for the last part of the sample, from the mid-1980s to the mid-2000s, then a model with a high degree of nominal rigidity is able to match key aspects of the data. Our results suggest that the state-dependency of pricing could dominate its time-dependency for the bulk of the post-war period but during the later period of the 'great moderation', when the economy was more stable, time-dependency could have dominated. Though we do not consider evidence from micro data here, we note that both time dependence (Bils and Klenow, 2004) and state dependence (Gertler and Leahy, 2008) have been found in such data.

How to test a calibrated, or even a partially Bayesian estimated DSGE model, such as the SW model, is a long-standing problem. Early work compared particular features of data simulated from the calibrated or estimated model with the actual data. Our method, based on indirect inference, formalises this approach. It exploits the fact that the solution to a log-linearised DSGE model can be represented as a restricted vector autoregressive-moving-average (VARMA) model either in levels or in first differences (if there are permanent shocks), and this can be closely represented by a VAR. When identified, the *a priori* structural restrictions of the DSGE model impose restrictions on the VAR. The DSGE model can be tested by comparing unrestricted VAR estimates (or some function of these estimates such as the value of the log-likelihood function or the impulse response functions) derived using data simulated from the DSGE model with unrestricted VAR estimates obtained from actual data. In practice, we use a Wald test based on the VAR estimates. If the DSGE model is correct then the simulated data, and the VAR estimates based on these data, will be close to the actual data.

One advantage of this procedure over a classical likelihood ratio test is that we do not have to specify a different DSGE model as the alternative hypothesis. An unrestricted VAR model based on the actual data automatically generates an alternative hypothesis suitable for testing of the specification of the model. The procedure requires that the DSGE model generates an identified VAR. We argue that the SW model does this.

A further issue is the probability distribution of the test statistic — in our case a Wald statistic. Instead of using the asymptotic distribution of the Wald statistic, we use an empirical estimate of its small sample distribution obtained by bootstrap methods.

This paper joins a large and rapidly expanding literature on the evaluation of DSGE models — see Minford et al. (2009) and Theodoridis (2006) for recent accounts. Two related issues stand out in this literature. First, how to measure the closeness of DSGE models to the data — see, for example, Watson (1993), Canova (1994, 1995, 2005), Del Negro and Schorfheide (2004, 2006), Corradi and Swanson (2007) and Del Negro et al. (2007a). Second, how well simulations of the model compare with various descriptions of the data such as moments, cross-moments and impulse response coefficients. Elsewhere (Le et al, 2010) we have referred to this as the 'puzzles methodology', as a poor match is often treated as a puzzle to be resolved by further model development; a recent example for a two-country world macro-model is Chari et al. (2002).

This paper contributes to this literature in two main ways. With respect to the puzzles methodology, it provides a formal statistical basis for comparing simulations of a calibrated or previously estimated DSGE model with key features of actual data. It also provides more detailed information on which features of the data the model is able and unable to capture, thereby supplementing available closeness measures.

The paper is organised as follows. In section 2 we review the key features of the SW model of concern to us. We explain our test procedure in detail in section 3. In section 4 we report our test findings and compare the performance of alternative flexible price versions of the SW model, including a hybrid model that combines flexible and sticky price versions of the model. In section 5 we examine whether changes in monetary regimes are a possible source of mis-specification. In section 6 we consider the robustness of our test and various other related issues. We summarise our conclusions in section 7.

2 The Smets-Wouters model of the US economy

One of the main issues that emerged from the first type of calibrated DSGE model, the real business cycle (RBC) model, was its failure to capture the stylised features of the labour market observed in actual data. Employment was found to be not nearly volatile enough in the RBC model compared with observed data, and the correlation between real wages and output was found to be much too high (see, for example, King, Plosser and Rebelo, 1988). The clear implication is that in the RBC model real wages are too flexible. The Smets-Wouters model (2007) marks a major development in macroeconometric modelling based on DSGE models. Its main aim is to construct and estimate a DSGE model for the United States in which prices and wages, and hence real wages, are sticky due to nominal and real frictions arising from Calvo pricing in both the goods and labour markets, and to examine the consequent effects of monetary policy which is set through a Taylor rule. It may be said, therefore, to be a New Keynesian model. They combine both calibration and Bayesian estimation methods and use data for the period 1966Q1–2004Q4.

Unusually, the SW model contains a full range of structural shocks. In the EU version — Smets and Wouters (2003) — on which the US version is based, there are ten structural shocks. These are reduced to seven in the US version: for total factor productivity, the risk premium, investment-specific technology, the wage mark-up, the price mark-up, exogenous spending and monetary policy. These shocks are generally assumed to have an autoregressive structure. The model finds that aggregate demand has hump-shaped responses to nominal and real shocks. A second difference from the EU version is that in the US version the Dixit-Stiglitz aggregator in the goods and labour markets is replaced by the aggregator developed by Kimball (1995) where the demand elasticity of differentiated goods and labour depends on their relative price. A third difference is that, in order to use the original data without having to detrend them, the US model features a deterministic growth rate driven by labour-augmenting technological progress.

Smets and Wouters report that their model is able to compete well with standard VAR and BVAR models in forecasting the main US macro variables at business cycle frequencies. They verify this by comparing the marginal likelihood of out-of-sample predictions of the model with Bayesian VAR models. They find that price and wage rigidities are important in explaining the data. They also find that demand shocks, such as those to the risk premium and to exogenous spending, and investment specific technology shocks explain a significant fraction of the short-run forecast variance in output, but wage mark-up and productivity shocks contribute little to explaining output variation in the medium to long run. They also confirm that productivity shocks have a significant short-run negative impact on hours worked. Inflation developments are mostly driven by the price mark-up shocks in the short run and wage mark-up shocks in the long run. The model can capture the cross correlation between output and inflation at business cycle frequencies. As an ultimate check of the model's performance, they estimate the model for two subsamples: the 'Great Inflation' period from 1966Q2 to 1979Q2 and the 'Great Moderation' period from 1984Q1 to 2004Q4, and find that most of the structural parameters are stable over the two periods. The exceptions are a fall in the standard deviation of the productivity, monetary policy and price mark-up shocks, which reflect the decrease in output growth and inflation volatility, and a fall in the monetary policy response to output in the second subsample.

Del Negro, Schorfheide, Smets and Wouters (2007a, DSSW) have conducted further tests on the specification of the SW model. Estimating the SW model using Bayesian methods with detrended data, they approximate this by a VAR and compare this with an unrestricted VAR fitted to actual data that ignores cross-equation restrictions. They introduce a hyper-parameter λ to measure the relative weights of the two VARs. λ is chosen to maximise the marginal likelihood ratio of the combined models. DSSW find that this estimate of λ is a reasonable distance away from $\lambda = 0$, its value when the restrictions are ignored, and is far away from $\lambda = \infty$, its value when the SW restrictions are correct. In addition, DSSW show that the forecasting capacity of the weighted model at horizons up to 2 years ahead is better than that of the unrestricted VAR, while that of the DSGE-restricted VAR is generally worse. DSSW suggest that their λ

measure can act as a form of test. Strictly, this would require knowing the distribution of the estimate of λ. In his comment on DSSW, Christiano (2007) provides preliminary evidence on this using Montecarlo methods.

It should be noted that none of these exercises in evaluating the SW model, useful as they may be in various respects, are a test of specification in the sense we propose ours to be. Thus poor forecast performance out of sample is different from in-sample mis-specification. Stability within sample also does not imply the model is well-specified in the first place. The λ measure is not, at least as it stands, a test with a known distribution; it does, however, come closest to one and we will refer back to it below when discussing our own tests.

In addition to testing for possible misspecification of the SW model, we examine an alternative version in which prices and wages are fully flexible but there is a simple one-period information delay for labour suppliers. This may be expected to make the labour supply less elastic and make output more responsive to supply shocks. We refer to this as a New Classical (NC) model.

We also propose a hybrid model that merges the NK and NC models by assuming that wage and price setters find themselves supplying labour and intermediate output partly in a competitive market with price/wage flexibility, and partly in a market with imperfect competition. We assume that the size of each sector depends on the facts of competition and does not vary in our sample but we allow the degree of imperfect competition to differ between labour and product markets. The basic idea is that economies consist of product sectors where rigidity prevails and others where prices are flexible; essentially this reflects the degree of competition in these sectors. Similarly with labour markets; some are much more competitive than others. An economy may be more or less dominated by competition and therefore more or less flexible in its wage/price-setting. The price and wage setting equations in the hybrid model are assumed to be a weighted average of the corresponding NK and NC equations.

3 Model evaluation by indirect inference

Indirect inference provides a classical statistical inferential framework for judging a calibrated or already, but maybe partially, estimated model whilst maintaining the basic idea employed in the evaluation of the early RBC models of comparing the moments generated by data simulated from the model with actual data. Using moments for the comparison is a distribution free approach. Instead, we posit a general but simple formal model (an auxiliary model) — in effect the conditional mean of the distribution of the data — and base the comparison on features of this model estimated from simulated and actual data.

Indirect inference on structural models may be distinguished from indirect estimation of structural models. Indirect estimation has been widely used for some time, see Smith (1993), Gregory and Smith (1991,1993), Gourieroux et al. (1993), Gourieroux and Monfort (1995) and Canova (2005). In estimation the parameters of the structural model are chosen so that when this model is simulated it generates estimates of the auxiliary model similar to those obtained from actual data. The optimal choice of parameters for the structural model are those that minimise the distance between a given function of the two sets of estimated coefficients of the auxiliary model. In the use of indirect inference for model evaluation the parameters of the structural model are taken as given. The aim is to compare the performance of the auxiliary model estimated on simulated data derived from the given estimates of a structural model — which is taken as the true model of the economy, the null hypothesis — with the performance of the auxiliary model when estimated from actual data. If the structural model is correct then its predictions about the impulse responses, moments and time series properties of the data should match those based on actual data. The comparison is based on the distributions of the two sets of parameter estimates of the auxiliary model, or of functions of these estimates.

Our choice of auxiliary model exploits the fact that the solution to a log-linearised DSGE model can be represented as a restricted VARMA model and this can be closely represented by a VAR. For further discussion on the use of a VAR to represent a DSGE model, see for example Canova (2005), Dave and DeJong (2007), Del Negro and Schorfheide (2004, 2006) and Del Negro et al. (2007a,b) together with the comments by Christiano (2007), Gallant (2007), Sims (2007), Faust (2007) and Kilian (2007). A levels VAR can be used if the shocks are stationary, but a VAR in differences may be needed if the shocks are permanent, such as productivity (real) or money supply shocks (nominal). The *a priori* structural restrictions of the DSGE model impose restrictions on the VAR; see Canova and Sala (2009) for an example of lack of identification based on a simple three equation model consisting of a new Keynesian IS function, a Phillips Curve and a Taylor Rule. Provided this VAR is over-identified, the DSGE model can be tested by comparing unrestricted VAR estimates (or some function of these estimates such as the value of the log-likelihood function or the impulse response functions) derived using data simulated from the DSGE model with unrestricted VAR estimates obtained from actual data. If the solved VAR is not identified then it is not restricted; the null and alternative hypotheses are then indistinguishable in the VAR. The SW model is clearly over-identified; changes in its parameter values imply quite different simulation properties, see Minford and Peel (2002, pp.436–7).

The model evaluation criterion we use is the Wald test of the difference between the vector of relevant VAR coefficients from simulated and actual data. If the DSGE model is correct (the null hypothesis) then the simulated data, and the VAR estimates based on these data, will not be significantly different from those derived from the actual data. One advantage of this procedure over a classical likelihood ratio test is that we do not have to specify a different DSGE model as the alternative hypothesis. An unrestricted VAR model based on the actual data automatically generates an alternative hypothesis suitable for testing the specification of the model. In contrast, Hall et al. (2009) propose a criterion based on a comparison of impulse response functions. As previously noted, Del Negro et al. (2007a) propose a hybrid model that combines the VAR generated under the null and alternative hypotheses and use a scalar index λ to combine and test the models.

A formal statement of the inferential problem is as follows. Using the notation of Canova (2005) which was designed for indirect estimation, we define y_t an $m \times 1$ vector of observed data ($t = 1, ..., T$), $x_t(\theta)$ an $m \times 1$ vector of simulated time series of S observations generated from the structural macroeconomic model, θ a $k \times 1$ vector of the parameters of the macroeconomic model. $x_t(\theta)$ and y_t are assumed to be stationary and ergodic. We set $S = T$ since we require that the actual data sample be regarded as a potential replication from the population of bootstrapped samples. The auxiliary model is $f[y_t, \alpha]$; an example is the $VAR(p)$ $y_t = \Sigma_{i=1}^{p} A_i y_{t-i} + \eta_t$ where α is a vector comprising elements of the A_i and of the covariance matrix of y_t. Under the null hypothesis H_0: $\theta = \theta_0$, the stated values of θ whether obtained by calibration or estimation; the auxiliary model is then $f[x_t(\theta_0), \alpha(\theta_0)] = f[y_t, \alpha]$. We wish to test the null hypothesis through the $q \times 1$ vector of continuous functions $g(\alpha)$. Such a formulation includes impulse response functions. Under H_0 $g(\alpha) = g[\alpha(\theta_0)]$.

Let a_T denote the estimator of α using actual data and $a_S(\theta_0)$ the estimator of α based on simulated data for θ_0. We may therefore obtain $g(a_T)$ and $g[a_S(\theta_0)]$. Using N independent sets of simulated data obtained using the bootstrap we can also define the bootstrap mean of the $g[a_S(\theta)]$, $\overline{g[a_S(\theta_0)]} = \frac{1}{N}\Sigma_{k=1}^{N} g_k[a_S(\theta_0)]$. The Wald test statistic is based on the distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$ where we assume that $g(a_T) - \overline{g[a_S(\theta_0)]} \xrightarrow{p} \mathbf{0}$. The resulting Wald statistic (WS) may be written as

$$WS = (g(a_T) - \overline{g[a_S(\theta_0)]})' W(\theta_0) (g(a_T) - \overline{g[a_S(\theta_0)]})$$

where $W(\theta_0)$ is the inverse of the variance-covariance matrix of the distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$.

$W(\theta_0)^{-1}$ can be obtained from the asymptotic distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$ and the asymptotic distribution of the Wald statistic would then be chi-squared. Instead, we obtain the empirical distribution of the Wald statistic by bootstrap methods based on defining $g(\alpha)$ as a vector consisting of the VAR coefficients and the variances of the data.[1]

The following steps summarise our implementation of the Wald test by bootstrapping:

Step 1: Estimate the errors of the economic model conditional on the observed data and θ_0.

Estimate the structural errors ε_t of the DSGE macroeconomic model, $x_t(\theta_0)$, given the stated values θ_0 and the observed data. The number of independent structural errors is taken to be less than or equal to the number of endogenous variables. The errors are not assumed to be normally distributed. Where the equations contain no expectations the errors can simply be backed out of the equation and the data. Where there are expectations estimation is required for the expectations; here we carry this out using the robust instrumental variables methods of McCallum (1976) and Wickens (1982), with the lagged endogenous data as instruments — thus effectively we use the auxiliary model VAR.

[1] We use these in preference to the variances of the VAR errors as these are directly affected by the a, which are tested separately; if inaccurate they would induce inaccuracy in the error size. The data variance distribution can be obtained from the bootstraps independently of the distribution of a.

Step 2: Derive the simulated data

On the null hypothesis the $\{\varepsilon_t\}_{t=1}^T$ are the structural errors. The simulated disturbances are drawn from these errors. In some DSGE models, including the SW model, many of the structural errors are assumed to be generated by autoregressive processes rather than being serially independent. If they are, then under our method we need to estimate them. We derive the simulated data by drawing the bootstrapped disturbances by time vector to preserve any simultaneity between them, and solving the resulting model using Dynare (Juillard, 2001). To obtain the N bootstrapped simulations we repeat this, drawing each sample independently. We set $N = 1000$.

Step 3: Compute the Wald statistic

We estimate the auxiliary model — a VAR(1) — using both the actual data and the N samples of simulated data to obtain estimates a_T and $a_S(\theta_0)$ of the vector α. The distribution of $a_T - \overline{a_S(\theta_0)}$ and its covariance matrix $W(\theta_0)^{-1}$ are estimated by bootstrapping $a_S(\theta_0)$. The bootstrapping proceeds by drawing N bootstrap samples of the structural model, and estimating the auxiliary VAR on each, thus obtaining N values of $a_S(\theta_0)$; we obtain the covariance of the simulated variables directly from the bootstrap samples. The resulting set of a_k vectors ($k = 1, ..., N$) represents the sampling variation implied by the structural model from which estimates of its mean, covariance matrix and confidence bounds may be calculated directly. Thus, the estimate of $W(\theta_0)^{-1}$ is

$$\frac{1}{N}\Sigma_{k=1}^{N}(a_k - \overline{a_k})'(a_k - \overline{a_k})$$

where $\overline{a_k} = \frac{1}{N}\Sigma_{k=1}^{N} a_k$. We then calculate the Wald statistic for the data sample; we estimate the bootstrap distribution of the Wald from the N bootstrap samples.

We note that the auxiliary model used is a VAR(1) and is for a limited number of key variables: the major macro quantities which include GDP, consumption, investment, inflation and interest rates. By raising the lag order of the VAR and increasing the number of variables, the stringency of the overall test of the model is increased. If we find that the structural model is already rejected by a VAR(1), we do not proceed to a more stringent test based on a higher order VAR[2].

Rather than focus our tests on just the parameters of the auxiliary model or the impulse response functions, we also attach importance to the ability to match data variances, hence their inclusion in α. As highlighted in the debates over the 'Great Moderation' and the recent banking crisis, there is a major concern over the scale of real and nominal volatility. In this way our test procedure is within the traditions of RBC analysis.

Figure 1 illustrates the joint distribution for just two parameters of the auxiliary equation for two cases: assuming that the covariance matrix of the parameters is diagonal and that it is not. One can think of estimation via indirect inference as pushing the observed data point as far into the centre of the distribution as possible. The Wald test, however, takes the structural parameters as given and merely notes the position of the observed data point in the distribution.[3]

[2] This increasing stringency is illustrated by the worsening performance of the model tested in Table 7 below for higher order VARs, as noted in footnote 6.

In fact the general representation of a stationary loglinearised DSGE model is a VARMA, which would imply that the true VAR should be of infinite order, at least if any DSGE model is the true model. However, for the same reason that we have not raised the VAR order above one, we have also not added any MA element. As DSGE models do better in meeting the challenge this could be considered.

[3] To understand why DSGE models will typically produce high covariances and so distributions like those in the bottom panel of Figure 1, we can give a simple example in the case where the two descriptors are the persistence of inflation and interest rates. If we recall the Fisher equation, we will see that the persistence of inflation and interest rates will be highly correlated. Thus in samples created by the DSGE model from its shocks where inflation is persistent, so will interest rates be; and similarly when the former is non-persistent so will the latter tend to be. Thus the two estimates of persistence under the null have a joint distribution that reflects this high correlation.

In Figure 1, we suppose that the model distribution is centred around 0.5 for each VAR coefficient; and the data-based VAR produced values for their partial autocorrelations of 0.1 and 0.9 respectively for inflation and interest rates — the two VAR coefficients. We suppose too that the 95% range for each was $0 - 1.0$ (a standard deviation of 0.25) and thus each is accepted individually. If the parameters are uncorrelated across samples, then the situation is as illustrated in the top panel. They will also be jointly accepted.

Now consider the case where there is a high positive covariance between the parameter estimates across samples, as implied by the DSGE model (with its Fisher equation). The lower panel illustrates the case for a 0.9 cross-correlation between the two parameters. The effect of the high covariance is to create a ridge in the density mountain; and the joint parameter combination of $0.1, 0.9$ will be rejected even though individually the two parameters are accepted.

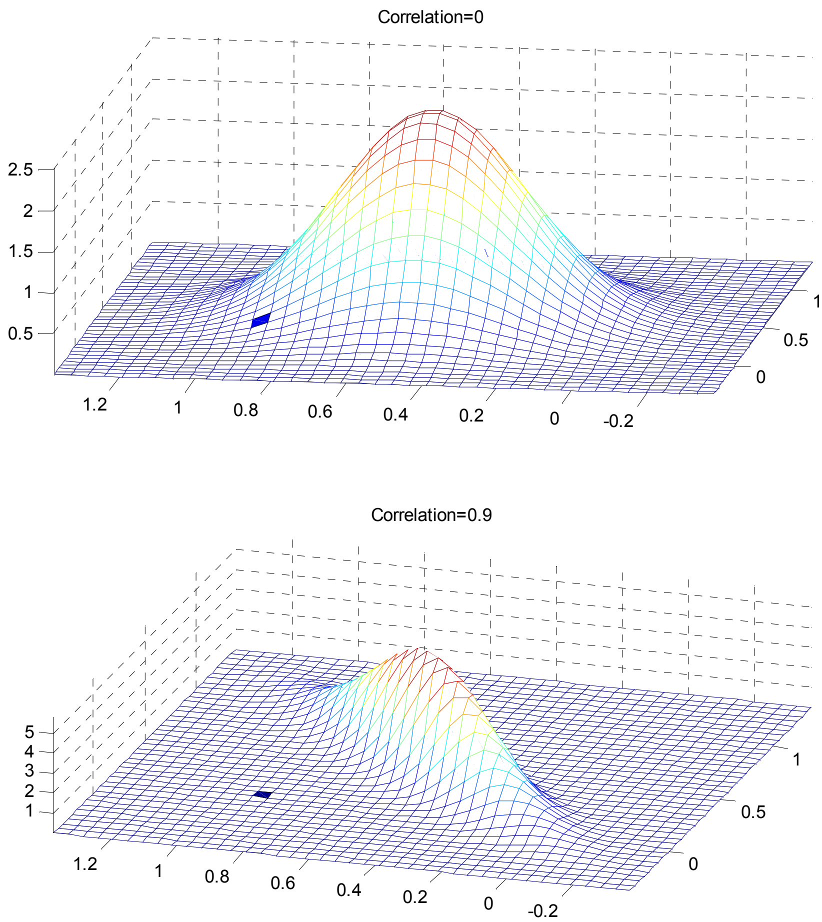

Figure 1: Bivariate Normal Distributions $(0.1, 0.9$ shaded) with correlation of 0 and 0.9.

We refer to the Wald statistic based on the full set of variables as the Full Wald test; it checks whether the a vector lies within the DSGE model's implied joint distribution and is a test of the DSGE model's specification in a wide sense. We show where in the Wald bootstrap distribution the Wald based on the data lies (the Wald percentile). We also show the Mahalanobis Distance based on the same joint distribution, normalised as a t-statistic, and also the equivalent Wald p-value, as an overall measure of closeness between the model and the data.[4].

We also consider a second Wald test, which we refer to as a 'Directed Wald statistic'. This focuses on more limited features of the structural model. Here we seek to know how well a particular variable or limited set of variables is modelled and we use the corresponding auxiliary equations for these variables in the VAR as the basis of our test. For example, we may wish to know how well the model can reproduce the behaviour of US output and inflation by creating a Wald statistic based on the VAR equation for these two variables alone.

A Directed Wald test can also be used to determine how well the structural model captures the effects of a particular set of shocks. This requires creating the joint distribution of the IRFs for these shocks alone. For

[4]The Mahalanobis Distance is the square root of the Wald value. As the square root of a chi-squared distribution, it can be converted into a t-statistic by adjusting the mean and the size. We normalise this here by ensuring that the resulting t-statistic is 1.645 at the 95% point of the distribution.

example, to determine how well the model deals with supply shocks, we construct the joint distribution of the IRFs for the supply shocks and calculate a Wald statistic for this. Even if the full model is misspecified, a Directed Wald test provides information about whether the model is well-specified enough to deal with specific aspects of economic behaviour.

We are implicitly assuming, therefore, that the auxiliary model can distinguish between different structural models. This has been challenged recently by Canova and Sala (2009). They argue that the identification of different DSGE models is 'weak' and may give rise to the same VAR. As already noted, they explain the point analytically using a three-equation New Keynesian model. They show that several parameters of the DSGE model cannot be identified in the solution, a VAR. As they note, their example is rigged in particular ways. For example, the shocks in the three equations are i.i.d. and there are no lagged endogenous variables (either from adjustment costs or indexation). In DSGE models like that of SW shocks are generally autocorrelated and lagged endogenous variables enter widely. As a result, DSGE models like SW's are generally found to be over-identified through the rational expectations mechanism (see Minford and Peel, 2002, pp.436–7). Changes in the parameters usually imply quite different simulation properties. This is illustrated in our results as the Wald statistic appears to distinguish (apparently rather starkly) between DSGE models. Canova and Sala argue that this is not the case with the λ-measure of Del Negro et al. (2007a) which we discussed above. Our results show that the position of different models in the tail of the Wald distribution (as measured by the normalised Mahalanobis Distance) varies with modest changes in model specification and so reflects the model's complete specification in a sensitive way.

In this paper we focus on testing a particular specification of a DSGE model and not on how to respecify the model should the test reject it. Rejection could, of course, be due to sampling variation in the original estimates and not because the model is otherwise incorrect. This is an issue we discuss further below under the heading of 'parameter uncertainty' in the final section. For further discussion of estimation issues see Smith (1993), Gregory and Smith (1991,1993), Gourieroux et al. (1993), Gourieroux and Monfort (1995), Canova (2005), Dridi et al. (2007), Hall et al. (2010), and Fukac and Pagan (2010).

4 Testing the SW Model using the method of indirect inference

4.1 Filtering the data

We apply the proposed testing procedure to the SW model for the whole post-war period of 1947Q1–2004Q4; later we look at sub-periods. First we need to filter the data to make them stationary. We consider several ways of doing this, the results from which are reported in the Annexes: SW's own filter (log differencing of all variables except inflation, log of hours worked and interest rates which are left in levels — Annex A); differencing all variables (log differencing as SW but differencing the remaining three — Annex B); a Hodrick-Prescott (HP) filter (Annex C); and linear detrending as followed by SW for their EU work (SW, 2003) — Annex D. SW's filter has the property of transforming the model dynamics, which may well affect the tests we carry out here.

Filter/Walds	Variances	Coefficients	Variances and Coefficients
Original	97	100	100
Differenced	100	100	100
HP	100	100	100
Linear Detrend	100	100	100

Table 1: The Wald percentiles for the SW model for each filter.

In order to check the effect of using different filters, we compared the rejection probabilities for the Wald test of the SW model when data variances and VAR coefficients are included in the test statistic. We found that there was very little difference in the results as the model was rejected 100% of the time for each filter. Even if we consider just the model's ability to replicate the data variances, the model was still rejected 100% of the time except when using the original SW filter when it was rejected 97% of the time — Table 1. As the choice of filter seems not to be a significant issue, in our more detailed investigations of the model we have used log-linear detrending as this extracts the least information from the raw data. As shown in Table 2, this was found to be sufficient to make the data stationary.

Variables	t-statistic	p-values
C	-2.5274	0.0114
I	-3.9830	0.0001
Y	-3.0486	0.0024
L	-3.2679	0.0012
π	-3.5307	0.0005
W	-2.6802	0.0074
R	-2.3397	0.0019
K	-3.9830	0.0001
Q	-3.5336	0.0005
RK	-4.3397	0.0000

$C =$ consumption; $I =$ investment; $Y =$ output; $L =$labour
$\pi =$ inflation; $W =$ real wage; $R =$ interest rate;
$K =$capital; $Q =$Tobin's q; $RK =$return to capital

Table 2: ADF test for stationarity of linear detrended data

As previously noted, the auxiliary VAR model has five main observable variables: output, investment, consumption, the quarterly interest rate and the quarterly inflation rate; and the VAR is of order one. As we will see below this VAR(1) is sufficient for discrimination.

4.2 Evaluating the SW model using SW's own assumed error properties

First we test the original SW model using their Bayesian estimate posterior means for the error variances and their autoregressive coefficients. The model is rejected as the Full Wald test percentile is 100; its normalised Mahalanobis Distance is 3.4, indicating that the data's dynamic properties are not close to that implied by the model. This can be explained by the large number (9 out of 25) of the unrestricted VAR parameters that lie outside the 95% bounds for the VAR based on simulating the model — see Table 3 which contains the parameter and variance estimates and their 95% bounds. The t-stats for some of these coefficients lie a long way outside the confidence intervals; in particular for the partial autocorrelations of consumption and inflation, the model's bounds lie above the unrestricted VAR estimates. One could interpret this as indicating excessive inflation and consumption persistence in the model.

Further, the data variances (the bottom 5 entries in Table 3) for the nominal variables are too low compared with the data; for interest rates the data variance lies outside the model bounds while for inflation it lies just above the 95% bound. Overall, therefore, even with SW's own assumed error properties, their model is considerably out of line with the data. If we compare our rejection with the Del Negro et al. (2007a)'s estimate of the λ weight for the model when combined with an unrestricted VAR, we can see that their estimate is quite consistent with our rejection of the model's overall specification.

VAR coeffs*	Actual Estimate	Lower Bound	Upper Bound	T-stats
A_Y^Y	0.99908	0.71104	0.96272	2.02349
A_Y^R	0.01503	-0.00557	0.04018	0.07322
A_Y^π	-0.00417	-0.0068	0.0673	-1.41540
A_Y^C	0.10174	-0.07815	0.02091	5.03459
A_Y^I	0.22591	-0.27355	0.18519	2.34051
A_R^Y	-0.64529	-1.28857	-0.40445	0.84625
A_R^R	0.85001	0.66138	0.86763	1.60632
A_R^π	0.15154	-0.11021	0.18262	1.56321
A_R^C	-0.5553	-0.83083	-0.2264	-0.16999
A_R^I	-1.7064	-2.33113	0.39231	-1.14775
A_π^Y	0.11612	-0.44029	0.32551	0.93503
A_π^R	0.02374	0.07066	0.26195	-2.89958
A_π^π	0.59496	0.59853	0.85809	-2.05169
A_π^C	-0.38833	-0.56528	-0.04657	-0.63310
A_π^I	-0.25917	-1.90858	0.40689	0.83059
A_C^Y	-0.08009	-0.12788	0.08505	-0.74255
A_C^R	-0.02553	-0.03697	0.00303	-0.98493
A_C^π	0.0121	-0.04785	0.01597	1.56170
A_C^C	0.78488	0.85948	0.95736	-5.18686
A_C^I	-0.4296	-0.36543	0.08677	-2.66676
A_I^Y	0.02034	0.01692	0.08499	-1.53642
A_I^R	0.01022	-0.00484	0.00905	2.21138
A_I^π	0.01159	-0.01534	0.00714	2.63929
A_I^C	0.01957	0.01241	0.04757	-1.12305
A_I^I	1.02924	0.94769	1.08301	0.38801
σ_Y^2	18.32858	8.71183	47.42615	-0.32176
σ_R^2	0.65276	0.19035	0.56812	3.22089
σ_π^2	0.44451	0.18584	0.46733	1.96505
σ_C^2	10.3888	6.19987	45.31804	-0.74834
σ_I^2	71.79914	65.12685	269.8422	-1.23001
Wald percentile 100		M-distance (Normalised) 3.4 $\left(p = 3.4e^{-4}\right)$		

(*) A_X^Z- coefficient of a variable Z_{t-1} on a variable X_t; σ_X^2 = variance of X
Y = output; R = interest rate; π = inflation; C = consumption; I = investment

Table 3: VAR Parameters, Data Variances and Model Bootstrap Bounds of the SW Model with SW's error properties

4.3 Evaluating the SW model using actual errors

Our simulations of the SW model have so far been based on the errors derived from Bayesian estimates of the model which are conditioned on the prior distribution. We now consider errors derived still from the model parameters but based on the actual data, which we will call the actual or structural errors. For equations with no expectations these can be backed out as the residuals. For equations with expectations we estimate the residuals using the instrumental variable procedures suggested by McCallum (1976) and Wickens (1982). The instruments are the lagged values of the endogenous variables. Thus, in effect, the generated expectations used in deriving the residuals are the predictions of the VAR based on the actual data.

Seven behavioural residuals are estimated by this means: consumption, investment, productivity, monetary policy, wages, prices, and one exogenous process, government spending, which enters the goods market clearing condition. These residuals are shown in Figure 2.

Figure 2: Single Equation Errors from SWNK model

We proceed as though five of these residuals follow an AR(1) and the price and wage residuals follow ARMA(1,1) processes, as assumed by SW. In all cases the standard deviations of the estimated error innovations are larger than those assumed by SW; for investment and the price mark-up they are nearly twice as large (see Table 4). Further, the actual preference, investment and monetary shocks exhibit markedly less persistence than SW assumed. Hence, though the properties of the residuals estimated from the data are recognisably similar to those assumed by SW, there are differences whose effects we go on to investigate in our subsequent bootstrap exercise. We use a vector bootstrap to preserve any dependence between the structural innovations.

	Govt Spend	Cons Prefce	Investmt	Monetary	Prodty	Price Mark-up	Wage Mark-up
SW stdev	0.53	0.23	0.45	0.24	0.45	0.14	0.24
Data stdev	0.673	0.371	0.704	0.344	0.553	0.239	0.311
SW $AR(1)$	0.97	0.22	0.71	0.15	0.95	0.89	0.96
SW $MA(1)$	0.52					-0.69	-0.84
Estimated $AR(1)$	0.944	-0.064	0.530	-0.062	0.971	0.925	0.915
Estimated $MA(1)$	0.553					-0.709	-0.848

Table 4: Standard deviations of innovations and coefficients of shocks (actual vs. assumed)

Table 4 reports the assumed and then the actual standard deviations of the innovations followed by the assumed and the actual estimates of the AR and MA parameters. Table 5 reports the estimates of the VAR parameters based on the estimated structural errors. The Full Wald Statistic, for all of the VAR coefficients and data variances, strongly rejects the model at the 5% level. Seven of the VAR coefficients lie outside their 95% bounds, together with the interest rate data variance. Again, as with SW's own assumed errors, the model fails to capture the scale of the nominal data variances; for the interest rate the data variance is now roughly double the model's upper bound while for inflation it remains around the model upper bound. The model's Mahalanobis Distance is 3.6 which is not much different from the SW model's 3.4 using their assumed error properties.

We find, therefore, that there is little change in the results when using structural errors obtained from the original data using our Limited Information estimation method, instead of errors derived from SW's Bayesian estimates. This finding is reassuring as it suggests that our test procedure is robust to different ways of extracting the residuals. It also echoes the finding of Le et al. (2010) in their investigation of the results of Chari et al. (2002). Nonetheless, how best to extract the residuals implied by an estimated structural model may still merit further work.

4.4 Evaluating the New Classical model using actual errors

Next we consider the New Classical version of the SW model proposed above. Once more the results are poor, see Table 6. The main problem is the model's massive overprediction of the inflation variance (3rd last entry, Table 6). This occurred regardless of variations in the Taylor Rule. (We adopted the NK rule except for setting potential output, y_t^p, to a constant.) For example, a larger reaction to inflation causes the interest rate variance to blow up but without bringing the inflation variance down sufficiently. Thus the NC model fails on the basic preliminary test of data variance matching.

The NC model's Full Wald percentile is again 100. In addition to the model overpredicting the inflation variance, 13 of the 25 VAR coefficients lie outside their 95% bounds. The model wrongly predicts all the partial autocorrelation coefficients, except for that of investment. Of the 13 coefficients that do not fit, five are related to the inflation rate. Further, the cross effects from the main macroeconomic variables to the interest rate, the inflation rate and consumption are badly predicted. The cross-effect from inflation to interest rates in the model is negative; theoretically the interest rate should react to offset a rise in the inflation rate. The Mahalanobis Distance is 4.7 which is considerably worse than for SW's New Keynesian model.

The model's IRFs also perform poorly (see Annex). The dominant shocks on the real variables are productivity and labour supply shocks, and on the nominal variables are preference, monetary, productivity and labour supply shocks. The responses of all of the variables to these shocks lie outside the model 95% bounds. Further, the model fails to replicate the cross-correlations of many of the main macroeconomic variables; it underpredicts the autocorrelations of interest and inflation rates, and their cross-correlations with output; it overpredicts the effect of investment on future output; due to excessive inflation variation, it also fails to replicate the correlation between inflation and output.

VAR coeffs *	Actual Estimate	Lower Bound	Upper Bound	T-stats
A_Y^Y	0.99908	0.75267	1.00004	1.52155
A_Y^R	0.01503	0.00444	0.05065	−0.85584
A_Y^π	−0.00417	−0.00459	0.06545	−1.69669
A_Y^C	0.10174	−0.02606	0.07359	2.97157
A_Y^I	0.22591	−0.19814	0.26058	1.59418
A_R^Y	−0.64529	−1.06427	−0.27522	0.12295
A_R^R	0.85001	0.53595	0.73313	3.90309
A_R^π	0.15154	−0.13061	0.15756	1.94297
A_R^C	−0.5553	−0.84942	−0.40004	0.60623
A_R^I	−1.7064	−1.80976	0.3887	−1.85185
A_π^Y	0.11612	−0.55029	0.07888	2.23055
A_π^R	0.02374	0.11413	0.27924	−4.05015
A_π^π	0.59496	0.47347	0.72517	−0.16778
A_π^C	−0.38833	−0.40915	−0.0401	−1.78108
A_π^I	−0.25917	−2.2204	−0.48423	2.53019
A_C^Y	−0.08009	−0.15347	0.11375	−0.78433
A_C^R	−0.02553	−0.06976	−0.01127	0.87955
A_C^π	0.0121	−0.06965	0.02018	1.70294
A_C^C	0.78488	0.8106	0.94474	−2.99327
A_C^I	−0.4296	−0.3507	0.1845	−2.51936
A_I^Y	0.02034	−0.00146	0.06908	−0.48864
A_I^R	0.01022	−0.00373	0.01252	1.44447
A_I^π	0.01159	−0.01151	0.01217	1.76852
A_I^C	0.01957	0.00317	0.03533	−0.02989
A_I^I	1.02924	0.90238	1.03372	1.86272
σ_Y^2	18.32858	9.67374	45.56006	−0.41926
σ_R^2	0.65276	0.16837	0.37665	7.52511
σ_π^2	0.44451	0.22269	0.47431	1.84365
σ_C^2	10.3888	4.62427	35.15967	−0.46506
σ_I^2	71.79914	63.05612	258.1966	−1.24068
Wald percentile 100		M-distance (Normalised) 3.6 $(p = 1.6e^{-4})$		

(*) A_X^Z- coefficient of a variable Z_{t-1} on a variable X_t; σ_X^2 = variance of X

Y = output; R = interest rate; π = inflation; C = consumption; I = investment

Table 5: VAR Parameters, data variances and Model Bootstrap Bounds of the SW Model with Estimated Coefficients

VAR coeffs *	Actual Estimate	Lower Bound	Upper Bound	T-stats
A_Y^Y	0.99908	0.75964	0.98742	1.81743
A_Y^R	0.01503	−0.0305	0.06499	−0.02425
A_Y^π	−0.00417	−0.03652	0.25225	−1.23410
A_Y^C	0.10174	−0.05527	0.07763	2.72888
A_Y^I	0.22591	−0.18755	0.24126	1.84048
A_R^Y	−0.64529	−0.77893	−0.00061	−1.29619
A_R^R	0.85001	0.16679	0.55328	4.77475
A_R^π	0.15154	−1.70697	−0.51682	4.01721
A_R^C	−0.5553	−0.49638	0.04422	−2.35582
A_R^I	−1.7064	−1.53308	0.39447	−2.34629
A_π^Y	0.11612	−0.14849	0.09029	2.45278
A_π^R	0.02374	0.01012	0.12289	−1.43555
A_π^π	0.59496	0.10313	0.43618	3.75306
A_π^C	−0.38833	−0.15269	0.00878	−7.46875
A_π^I	−0.25917	−0.43894	0.13493	−0.79497
A_C^Y	−0.08009	−0.11079	0.13547	−1.25499
A_C^R	−0.02553	−0.14256	−0.02468	1.75061
A_C^π	0.0121	−0.42415	−0.06702	2.51545
A_C^C	0.78488	0.84915	1.00249	−3.95797
A_C^I	−0.4296	−0.30533	0.16651	−2.94383
A_I^Y	0.02034	−0.00818	0.06222	−0.31562
A_I^R	0.01022	−0.00083	0.03162	−0.58207
A_I^π	0.01159	−0.02859	0.07379	−0.40644
A_I^C	0.01957	−0.00999	0.03741	0.44958
A_I^I	1.02924	0.90735	1.03984	1.47843
σ_Y^2	18.32858	9.43786	62.14178	−0.60189
σ_R^2	0.65276	0.36337	0.76928	1.30337
σ_π^2	0.44451	2.33699	3.60734	−7.64773
σ_C^2	10.3888	7.39139	63.64939	−1.00088
σ_I^2	71.79914	60.45211	284.8093	−1.12994
Wald percentile 100		M-distance (Normalised) 4.7 $(p = 1.3e^{-6})$		
(*) A_X^Z− coefficient of a variable Z_{t-1} on a variable X_t; σ_X^2 = variance of X				
Y = output; R = interest rate; π = inflation; C = consumption; I = investment				

Table 6: VAR Parameters, data variances and Model Bootstrap Bounds of the NC Model with Estimated Coefficients

Overall, therefore, the New Classical version of the original SW model also fails to match the data, and in quite serious ways.

4.5 Evaluating a hybrid model: a weighted combination of New Keynesian and New Classical models

We have analysed two rather different macroeconomic models with a view to understanding the mechanisms behind each of them, particularly the degree of price and wage flexibility. The NK model is highly rigid with Calvo price and wage setting, while the NC model is a flexible wage/price model with only a simple one-period information delay for labour suppliers.

In SW's NK model, because capacity utilisation is fairly flexible, output is strongly affected by shocks to demand and this in turn — via the Phillips Curve — moves inflation and then — via the Taylor Rule — interest rates. Supply shocks can affect demand directly (e.g. productivity shocks change the return on capital and so affect investment) and also play a role as 'cost-push' inflation shocks (e.g. price/wage mark-up shocks). Persistent shocks to demand raise 'Q' persistently and produce an 'investment boom' which, via demand effects, reinforces itself. Thus the model acts as a 'multiplier/accelerator' of shocks both on the demand and the supply side.

In the NC model an inelastic labour supply causes fluctuations in output to be dominated by supply shocks (productivity and labour supply), while investment and consumption respond to output in a standard RBC manner. These reactions, together with demand shocks, create market-clearing movements in real interest rates and — via the Taylor Rule — in inflation. Supply shocks are prime movers of all variables in the NC model, while demand shocks add to the variability of nominal variables. In order to mimic real variability and persistence, suitably sized and persistent supply shocks are needed. But to mimic the limited variability in inflation and interest rates only a limited variance in demand shocks is required; and to mimic their persistence, the supply shocks must be sufficiently autocorrelated.

Thus both the NK and NC versions of the SW model fail to match the data. Essentially, the NK model generates too little nominal variation while the NC model delivers too much. Given that each model fails in an opposite way, we propose a hybrid model that merges the NK and NC models by assuming that wage and price setters find themselves supplying labour and intermediate output partly in a competitive market with price/wage flexibility, and partly in a market with imperfect competition. We assume that the size of each sector depends on the facts of competition and does not vary in our sample but we allow the degree of imperfect competition to differ between labour and product markets. The basic idea is that economies consist of product sectors where rigidity prevails and others where prices are flexible; essentially this reflects the degree of competition in these sectors. Similarly with labour markets; some are much more competitive than others. An economy may be more or less dominated by competition and therefore more or less flexible in its wage/price-setting[5]. We also assume that the monetary authority pursues a Taylor Rule that reflects the properties of the hybrid model.

In the hybrid model the price and wage setting equations are assumed to be a weighted average of the corresponding NK and NC equations. This weighting process is an informal use of indirect inference, the idea being to find the combination of the weights and Taylor coefficients that make the combined model perform best when compared with the auxiliary model. To find the optimal set of weights we carry out a grid search over the two weights (one for the product market and one for the labour market) using the criterion of minimising the normalised Mahalanobis distance.

We find that the optimal weights are $v_w = 0.1$ (the NK share for wages) and $v_p = 0.2$ (the NK share for prices). That is, only 10% of labour markets and only 20% of product markets are imperfectly competitive. Therefore, the model requires only a small amount of nominal rigidity in order to match the data. The Taylor rule then becomes: $R_t = 0.6R_{t-1} + (1-0.6)\{2.3\pi_t + 0.08y_t\} + 0.22\,(y_t - y_{t-1}) + \varepsilon_t$. This is a somewhat more aggressive response to inflation than either the NK ($R_t = 0.81R_{t-1} + (1-0.81)\left\{2.04\pi_t + 0.08(y_t - y_t^P\right\} + 0.22\left[\left(y_t - y_t^P\right) - \left(y_{t-1} - y_{t-1}^P\right)\right] + \varepsilon_t^r$) or NC rules (the NC is the same as NK except that it sets ‘potential

[5] Formally, we model this as follows. We assume that firms producing intermediate goods have a production function that combines in a fixed proportion labour in imperfect competition (‘unionised’) with labour from competitive markets — thus the labour used by intermediate firms becomes $n_t = n_{1t} + n_{2t} =$ $\left\{\left[\int_0^1 (n_{1it})^{\frac{1}{1+\lambda_{w,t}}} di\right]^{1+\lambda_{w,t}} + \left[\int_0^1 (n_{2it}) di\right]\right\}$ where n_{1it} is the unionised, n_{2it} the competitive labour provided by the ith household at t; we can think of n_t as representing the activities of an intermediary ‘labour bundler’. Note that $n_{1t} = v_w n_t$, where v_w is the share of unionised labour in the total, thus $n_{2t} = (1 - v_w)n_t$ so that $W_t = v_w W_{1t} + (1 - v_w)W_{2t}$. Each household’s utility includes the two sorts of labour in the same way, that is $U_{it} = ... - \frac{n_{1it}^{1+\sigma_n} \epsilon_{1nt}}{1+\sigma_n} - \frac{n_{2it}^{1+\sigma_n} \epsilon_{2nt}}{1+\sigma_n} ...$ W_{1t} is now set according to the Calvo wage-setting equation, while W_{2t} is set equal to current expected marginal monetary disutility of work; in the latter case a 1-quarter information lag is assumed for current inflation but for convenience this is ignored in the usual way as unimportant in the Calvo setting over the whole future horizon.

These wages are then passed to the labour bundler who offers a labour unit as above at this weighted average wage. Firms then buy these labour units off the manager for use in the firm.

Similarly, retail output is now made up in a fixed proportion of intermediate goods in an imperfectly competitive market and intermediate goods sold competitively. Retail output is therefore $y_t = y_{1t} + y_{2t} =$ $\left\{\left[\int_0^1 y_{j1t}^{\frac{1}{1+\lambda_{p,t}}} dj\right]^{1+\lambda_{p,t}} + \left[\int_0^1 y_{j2t} dj\right]\right\}$. The intermediary firm prices y_{1t} according to the Calvo mark-up equation on marginal costs, and y_{2t} at marginal costs.

Note that $y_{1t} = v_p y_t$, where analogously v_p is the share of imperfectly competitive goods market; thus $y_{2t} = (1 - v_p)y_t$ so that $P_t = v_p P_{1t} + (1 - v_p)P_{2t}$. The retailer combines these goods as above in a bundle which it sells at this weighted average price.

Notice that apart from these equations the first-order conditions of households and firms will be unaffected by what markets they are operating in.

output' to a constant). Notice that if one substitutes for the interest rate from a simple money demand function with an exogenous money supply growth process, then one obtains a 'Taylor Rule' that has the form $\Delta R_t = \frac{1}{\beta}\{\pi_t + \gamma \Delta y_t) + v_t$ where β is the semi-log interest rate elasticity of money demand, γ is the corresponding income elasticity and v_t is a combination of the money supply growth process and the change in the money demand error. This is fairly close to the rules adopted in these models when the lagged term in interest rates is large and the term in the output gap is small compared with the term in the rate of change of output.

The VAR results for the hybrid model are reported in Table 7. The main difference between the hybrid and the NK and NC models is the hybrid model's ability to reproduce the variances of the data. Using the structural errors from the model and the observed data, we find that all of the data variances lie within the model's 95% bounds (Table 7, last 5 entries). Furthermore, only nine of the 25 VAR coefficients lie outside their 95% confidence intervals. While the Full Wald percentile of 100 rejects this hybrid version of the model, as it does the others, the Mahalanobis Distance of 2.8 implies that the hybrid model is substantially closer to the data[6].

Since the optimal combination indicates that the majority of the market participants behave in a competitive manner, it is not a surprise that the variance decomposition (Table 8) shows that the supply shocks — productivity and labour supply shocks — explain most of the movements of the real variables. They also explain a large part of the nominal variables. While the demand shocks also contribute quite a lot to movements in the interest rate, they do so less for movements in inflation. So why are these results different from those of the NK and NC models?

The hybrid model mostly acts like the NC model, where the supply shocks explain most of the variation and the demand shocks play a small in part in the variability of real variables due to the one period information lag but add also to the variability of nominal variables. Since, however, some economic agents behave in the New Keynesian manner, aggregate supply and labour supply are more elastic, and demand shocks have a greater impact on real variables. Most importantly, inflation variability is dampened down to better reflect actual data variability. It is remarkable how large the reduction in the lower bound is by the introduction of only small Calvo shares (10% in wages, 20% in prices — or 30% rigidity overall); the model's lower bound on inflation's standard deviation falls no less than 57%. The reason appears to be that the variability of inflation also reacts to the variability of expected inflation. Thus, as the Calvo element rises, expected inflation varies less which, in turn, reduces the variability of actual inflation and, again in turn, reduces the variability of expected inflation, and so on in a sort of 'multiplier' process. This is an effect anticipated by Dixon (1992, 1994).

Next we investigate the VAR impulse response functions to three main shocks: investment, labour supply, and productivity shocks, where we identify the VAR shocks by using the structural model. The main differences from the data are in the long-run responses of interest and inflation rates to the shocks; also the response of consumption is much more aggressive in the data than in the model. Nonetheless, these responses lie only just outside the 95% bounds. We can therefore say that the performance of the hybrid model, based on the IRFs, is relatively good when compared to that of the NK and NC models (to be found in the Annex).

[6] This result is for a VAR(1) on all five macro variables. As we noted earlier, raising the lag order of the VAR worsens the fit to the data because of the greater complexity in the behaviour being captured. The Mahalanobis Distance rises on the VAR(2) to 4.55 and on the VAR(3) to 5.14. In both cases the Wald percentiles are 100.

VAR coeffs *	Actual Estimate	Lower Bound	Upper Bound	T-stats
A_Y^Y	0.99908	0.76761	0.99753	1.58317
A_Y^R	0.01503	-0.04394	0.01945	1.55148
A_Y^π	-0.00417	-0.02898	0.06556	-0.76799
A_Y^C	0.10174	-0.03905	0.10284	2.05348
A_Y^I	0.22591	-0.2093	0.28019	1.59406
A_R^Y	-0.64529	-0.97523	-0.14419	-0.40543
A_R^R	0.85001	0.49302	0.75838	3.23533
A_R^π	0.15154	-0.27545	0.1136	2.20900
A_R^C	-0.5553	-0.69948	-0.10079	-1.14167
A_R^I	-1.7064	-1.69603	0.44879	-1.94286
A_π^Y	0.11612	-0.29996	0.30247	0.72094
A_π^R	0.02374	0.07486	0.27241	-2.85460
A_π^π	0.59496	0.51488	0.78029	-0.76721
A_π^C	-0.38833	-0.27412	0.18435	-2.98726
A_π^I	-0.25917	-1.31646	0.30961	0.54641
A_C^Y	-0.08009	-0.142	0.09477	-0.82072
A_C^R	-0.02553	-0.06465	0.00228	0.19880
A_C^π	0.0121	-0.09455	0.00559	2.01856
A_C^C	0.78488	0.81382	0.96991	-2.92486
A_C^I	-0.4296	-0.36452	0.1334	-2.63914
A_I^Y	0.02034	-0.00329	0.07082	-0.56989
A_I^R	0.01022	0.00293	0.02417	-0.45314
A_I^π	0.01159	-0.01199	0.02095	0.83386
A_I^C	0.01957	-0.00749	0.03827	0.27887
A_I^I	1.02924	0.89898	1.04153	1.49582
σ_Y^2	18.32858	9.69749	61.85333	-0.61346
σ_R^2	0.65276	0.29191	0.76451	1.58414
σ_π^2	0.44451	0.43895	0.89102	-1.65685
σ_C^2	10.3888	7.30487	72.01693	-0.99793
σ_I^2	71.79914	61.41478	301.772	-1.17817
Wald percentile 100		M-distance (Normalised) 2.8 ($p = 0.0026$)		

(*) $A_X^Z -$ coefficient of a variable Z_{t-1} on a variable X_t; $\sigma_X^2 =$ variance of X
$Y =$ output; $R =$ interest rate; $\pi =$ inflation; $C =$ consumption; $I =$ investment

Table 7: VAR Parameters, data variances and Model Bootstrap Bounds of the Weighted Model with Estimated Coefficients

Shocks	Govt. Spend	Cons Prefce	Investmt	Monetary	Prodty	Price mark-up	Wage mark-up	Labour supply	Total
Output	2.6796	0.9823	1.9547	0.6995	48.2598	0.5086	0.00003	44.9154	100
Interest rate	11.8312	16.2245	17.4343	2.2156	15.3872	3.5352	0.000695	33.3713	100
Inflation	2.0282	7.0541	3.7657	33.3303	17.6394	4.9596	0.000769	31.2218	100
Consumption	5.0587	1.0009	1.7749	0.6429	34.3637	0.34915	0.00004	56.8097	100
Investment	11.298	0.0508	28.0270	0.1305	32.6701	0.2853	0.00001	27.5383	100

Table 8: Variance Decompositions of the weighted Model

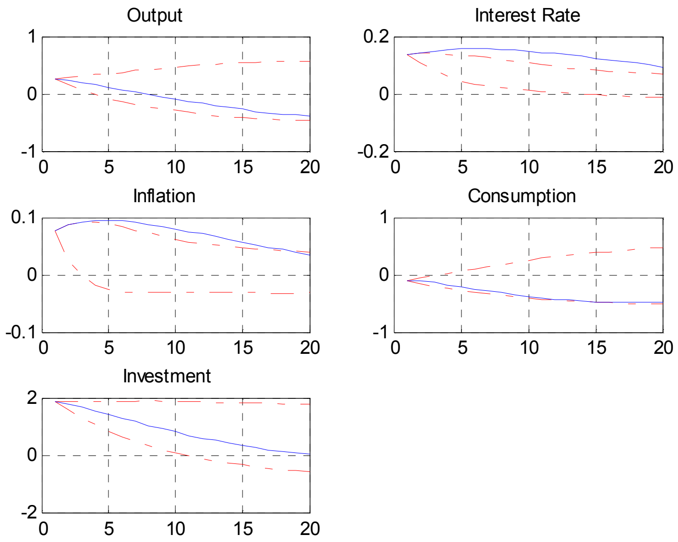

Figure 3: Investment Shock

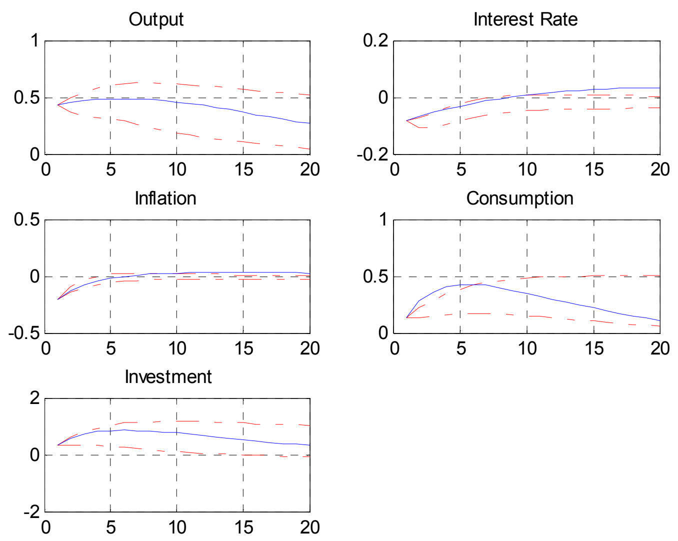

Figure 4: Productivity Shock

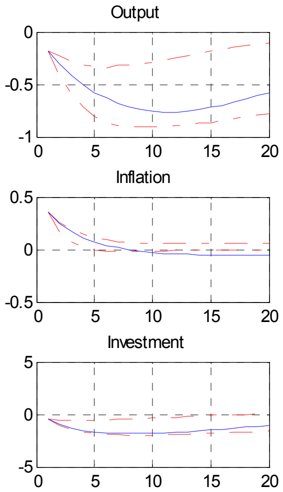

Figure 5: Labour Supply Shock

The cross-correlations are accepted in a number of cases (Figure 6). The actual autocorrelations and cross-correlations of the variables lie within the model's bounds, though the correlation of investment with future output lies outside the bound. The performance of the cross-correlations among the nominal variables is, however, poor. The autocorrelations of interest and inflation rates are underpredicted by the model, even though the differences are much smaller than those for the NK and NC models. These failures are consistent with the overall rejection of the hybrid model.

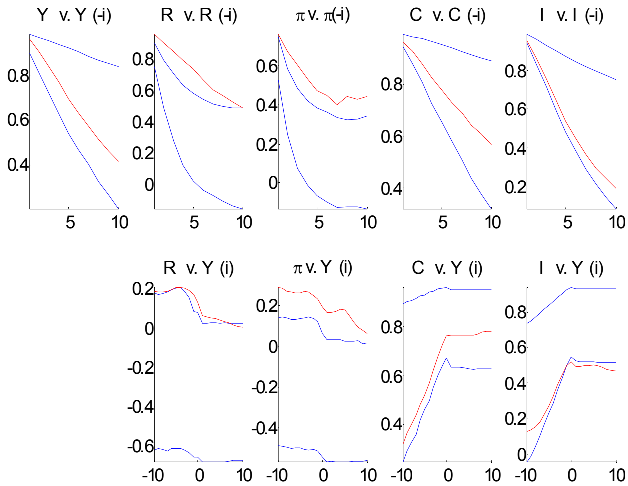

Figure 6: Cross-Correlations for Weighted Model; $Y = \text{output}$; $R = \text{interest rate}$; $\pi = \text{inflation}$; $C = \text{consumption}$; $I = \text{investment}$

We now consider the model's performance for particular aspects of the data, using the Directed Wald test. Our method is to focus first on individual variables and then in groups by estimating the best $ARMA(i, j)$ in the case of a single variable and a VAR(1) for a group of variables. We then apply the Directed Wald test. To assess the individual shocks we take the IRFs (we use the IRF average) of the shock for the variables where they have a major impact and generate the model-implied joint distribution of these IRFs, computing the Wald statistic for the joint values in the data. We also look at the joint distribution of the variances to confirm our earlier judgement from the individual variances. Tables 9 and 10 report these Wald statistics.[7]

Table 9 shows that first, the model does fit the data variances jointly but only at the 99% level. Second, for individual variables, the responses of all are accepted at the 99% level; inflation is accepted at the 95% level. As observed earlier, many of the VAR coefficients involving interest rates are rejected individually. It therefore seems clear that this is the area in which we can improve the specification of the model. Third, the real variables fit the data taken as a group, though again only at the 99% level, as do nominal variables taken as a group. When, however, nominal and real variables are combined the dynamic fit deteriorates sharply and the model is rejected at the 99% level; only if we restrict ourselves to output and inflation does the model pass this Wald test at the 99% level. This is mirrored in the individual shocks (as shown in Table 10); the responses to both productivity and labour supply, the two key shocks in this model, are close to the 99% rejection borderline. We can also look at shock/variable combinations where the ones we have selected emerge from the variance decomposition analysis as those the model finds are non-trivial (thus for example the government spending shock only has non-trivial effects on interest rates and investment). One can see how certain relationships match the data very easily — such as the effects of productivity and labour supply shocks on output, or the same on consumption — while others fail badly — such as the effects of the five shocks that non-trivially affect it on the interest rate.

Variable combinations	Direct Wald
$\sigma_Y^2, \sigma_R^2, \sigma_\pi^2, \sigma_C^2, \sigma_I^2$	97
$Y\,(AR\,(3))$	96.2
$R\,(ARMA\,(1,1))$	98.4
$\pi\,(AR\,(3))$	90.3
$C\,(AR\,(3))$	98.8
$I\,(AR\,(2))$	95.2
Y, C, I	98.3
Y, C, I, R	99.0
Y, C, I, π	100
Y, R, π	99.6
Y, π	97.6
R, π	96.2

Table 9: Directed Wald statistics by variable combinations- all shocks

Shocks	Variables	Directed Wald
$Prod$	Y, R, π, C, I	98.2
$LabSup$	Y, R, π, C, I	99.1
$GovSpend$	R, I	84.1
$Invt$	R, I	100.0
$Prod, LabSup$	Y	37.3
$Prod, LabSup, GovSpend, Invt, Cons$	R	100.0
$Prod, LabSup, Mon$	π	95.5
$Prod, LabSup$	C	34.0
$Prod, LabSup, GovSpend, Invt$	I	98.8

Table 10: Directed Wald statistics- by shock/variable combinations

[7] The notation follows that of the previous Tables.

5 Regime change as a possible source of mis-specification

In view of the apparently crucial role of interest rates in the rejection of the hybrid model, the implication is that the problem could lie in the specification of monetary policy, and in particular the use of one monetary regime for the whole sample from 1950s to the 2000s. We therefore tested for structural change during this period following the procedure of Perron and Qu (2007) designed to test for multiple breaks in VAR parameters; we found evidence of parameter breaks in two places: 1965 and 1984 (Table 11) .

The estimated breaks are:	1965.02	1984.02
The 95% C.I. for the 1st break is	(1964.04–1965.04)	
The 95% C.I. for the 2nd break is	(1983.02–1985.02)	

Table 11: Perron-Qu Multivariate Structural Break Test

These are natural places to find such breaks due to changes that occurred in the monetary regime. The earlier break is associated with the emergence of serious inflation for the first time; the later break is associated with the shift towards interest rate setting that followed from the adoption of (implicit) inflation targeting[8].

In moving to three sub-periods we tripled the size of our testing problem. Furthermore linear detrending no longer proved sufficient to make the data stationary; we therefore used an H-P filter. So far we have been unable to locate acceptable versions of the model for the first two sub-periods. However for the third and latest sub-period (1984.03–2004.02), we found good results in our grid search over the weights in the hybrid model when we moved them greatly towards the New Keynesian end of the spectrum $(0.8, 0.8)$. It may well be that in the 'Great Moderation' price-setting was far less disturbed by shocks to the state and was dominated instead by time dependence.

The model is still rejected on the full VAR with a Wald percentile of 100, and its Mahalanobis Distance is 3.97 (similar to the full sample weighted model of 3.9 on H-P-filtered data). While it is also rejected for interest rates alone for its best $AR(2)$ representation, for the combined variable set of output, inflation and interest rates it is now jointly accepted at the 99% level and nearly accepted at the 95% level (Table 12). For just output and inflation it is easily accepted at the 95% level (83.8)(Table 13). Significantly, this is the first time that any model we have examined over the full data period has passed the test embracing real GDP and both nominal variables.

6 Robustness and other issues

We consider a number of issues concerning our procedures. First, we discuss the use of the bootstrap; second we consider the issue of parameter uncertainty; and finally we then turn to some issues of numerical robustness.

6.1 Potential problems with the bootstrap

6.1.1 Reliability of the bootstrap

Several authors (e.g. Basawa et al, 1991, Hansen (1999) and Horowitz, 2001a,b) have noted that asymptotic distribution theory is unlikely to provide a good guide to the bootstrap distribution of the AR coefficient if

[8] In a recent paper Bianchi (2010) uses a Markov-switching model of monetary regime and volatility to estimate when breaks occurred; he finds a big shift to 'Hawkish' monetary policy (high response of interest rates to inflation) in 1965, with consistently High Volatility between 1970 and 1985; and a big shift to Low Volatility in 1985, after which monetary policy is also generally Hawkish, with a high weight on inflation similar to that in our Taylor Rule. Allowing for the quite different modelling approach and the fact that his model has not been subjected to the same overall tests as those here, our breaks and his findings appear broadly consistent.

As we report below, we are unable to find a good model structure for the two early sub-periods; it may be that part of the reason lies in variations of monetary regime and volatility as Bianchi claims. For the last period, Bianchi finds not so much variation in policy regime and consistently Low Volatility. It is for this period that we find a well-performing model with a structure fairly close to New Keynesian and a Taylor Rule with a high weight on inflation like Bianchi's.

VAR coeffs *	Actual	Lower	Upper	State
A_Y^Y	0.92801	0.60815	0.94046	$TRUE$
A_Y^R	0.06743	-0.04591	0.04223	$FALSE$
A_Y^π	0.05812	-0.03086	0.06354	$TRUE$
A_R^Y	-0.47031	-0.90001	0.54311	$TRUE$
A_R^R	0.75002	0.53789	0.88332	$TRUE$
A_R^π	0.0901	-0.04968	0.38778	$TRUE$
A_π^Y	0.13438	-1.08851	0.69176	$TRUE$
A_π^R	0.1202	-0.21217	0.25296	$TRUE$
A_π^π	0.00798	-0.14963	0.31854	$TRUE$
σ_Y^2	0.9136	0.53617	2.03963	$TRUE$
σ_R^2	0.06939	0.03261	0.11177	$TRUE$
σ_π^2	0.03058	0.02693	0.05061	$TRUE$
Wald percentile 96.1		M-distance (Normalised) 1.785 ($p = 0.0371$)		

(*) $A_X^Z -$ coefficient of a variable Z_{t-1} on a variable X_t; $\sigma_X^2 =$ variance of X
$Y =$ output; $R =$ interest rate; $\pi =$ inflation

Table 12: VAR coefficients and variances

Variable Combinations	Direct Walds
$Y\,(AR\,(3))$	97.4
$R(AR(1))$	97.1
$R\,(AR\,(2))$	100
$\pi\,(AR\,(1))$	54.8
Y, π	83.8
Y, R, π	96.1
$Var(Y), Var(R), Var(\pi)$	47.33

Table 13: Direct Walds for different combinations of output, inflation and interest rate

the leading root of the process is a unit root or is close to a unit root. This is also likely to apply to the coefficients of a VAR when the leading root is close to unity and may therefore affect indirect inference where a VAR is used as the auxiliary model, as here. Our results suggest, however, that this is not a problem as the bootstrap samples are all stationary and do not generate roots on or outside the unit circle and the original disturbances of the SW model were stationary. We also found that not one of the VARs estimated from the 1000 bootstraps generated such roots: the VAR coefficients all lie within ranges for the roots inside the unit circle. Nevertheless it is possible that the roots of the VAR coefficients might be sufficiently close to unity to cause problems. We therefore check the properties of the bootstrap in the context of indirect inference which, to our knowledge, has not been done previously.

To see whether the bootstrap distributions were generating the correct size of test we carried out a Montecarlo experiment. In the experiment, we set up the model we chose as best for the post-1984 sample, with the errors implied by that sample and we gave these a normal distribution with the same variance as estimated — note that the model includes the ARMA parameters estimated from these errors. We then drew random samples from the innovations in these error processes, creating 1000 artificial samples of the same length as the original data — 76 observations, a small sample. We then bootstrapped each of these samples 1000 times, exactly following the procedures used above. For each sample we computed the Wald statistic generated by the bootstraps to check whether the model is accepted or rejected at the 95% confidence level. This gave us 1000 acceptances/rejections at the 95% level — we repeated these a number of times, giving several thousand.

Following Horowitz (2001a), we also implemented a more elaborate bootstrap procedure as follows: 1) generate 1000 bootstrap samples, calculate the covariance matrix from this, and generate 1000 Wald values 2) repeat this step 1000 times, thus generating $1,000,000$ WS values; from these calculate the 'bootstrap distribution' of WS and note its 95% value. 3) use this 95% critical value to accept/reject the data-based WS. We checked this also by Montecarlo methods. This gave a very slight improvement on the original procedure used above (which simply used step 1) and calculated the percentile where the data-based WS fell in that distribution). Table 14 gives the true rejection probabilities (RPs) found in these Montecarlo experiments against the nominal RP corresponding to the confidence levels. We also report the true RPs based on the critical values from the asymptotic distribution which is $\chi^2_{(30)}$ as discussed below. These give somewhat greater accuracy at 10% and 5% but less at 1%. The asymptotic distribution rejects more frequently than the small-sample bootstrap distribution (see Table 14 below) which may compensate for the bootstrap's tendency to under-reject. For all these distributions the inaccuracy is, however, quite small.

Nominal RP(%):	10	5	1
Original procedure			
True RP (%)	5.1	2.6	0.4
Revised procedure			
True RP (%)	5.4	2.8	0.7
Asymptotic RP (%)	6.9	5.1	2.3

Table 14: True versus nominal Rejection Probabilities- n=76; Montecarlo experiment (post-1984 model and sample)

When there is a difference between nominal and true RPs across a variety of different possible Montecarlo experiments then it is possible to obtain more accuracy in the bootstrap procedure by substituting for the bootstrap distribution of the Wald statistic, used here, an estimate of it provided by a Montecarlo experiment. This estimate generates the distribution of WS by taking repeated samples of the population (calculating WS for each). Such refinement is appropriate when the estimate is robust to the choice of population distribution. One can check across a variety of such population distributions to determine whether the resulting distribution yields true RPs close to the nominal RPs across them.

We may thus use the information in the Montecarlo experiments in the manner described above, by substituting the WS distribution found by repeated samples from a typical population. Our typical population has the estimated third and fourth moments of the model shock-innovations. We use this to generate the estimated distribution of WS. This can then be used to assess the probability of WS for the data sample.

To test the accuracy of using this estimated distribution based on the typical population we repeat our Montecarlo experiments (again with 1000 replications) under alternative assumed populations with third and fourth moments ranging between their upper and lower 95% bounds. This gives a range of True RPs which are tabulated in Table 15. It can be seen that the procedure achieves a high degree of accuracy — we discuss the robustness of our results to this refined procedure below (6.1.3).

Nominal RP(%):	10	5	1
Range across alternative 3rd and 4th moments			
True RP (%)	9.34 – 11.06	5.08 – 5.80	0.92 – 1.18

Table 15: True versus nominal Rejection Probabilities - n=76; Montecarlo experiments (1000 replications) with different skewness/kurtosis assumptions within the 95are calculated on the estimated skewness/ kurtosis values for this sample

6.1.2 The consistency of the bootstrap for the Wald statistic

It is also possible to check the consistency of our bootstrapped Wald statistic, i.e. whether, as the sample goes to infinity, the bootstrap distribution converges on the true distribution in some measurable way — in particular, whether the size of the test is correct. Presumably, if the bootstrap fails to produce a distribution close enough to the true distribution in the limit as the sample increases then it is unlikely to do so for a small, or abnormal, sample. We examine this by a Montecarlo experiment. If in the Montecarlo experiment the bootstrap does badly then asymptotic analysis may be able to suggest why.

It is not easy to state necessary and sufficient conditions for test consistency that readily translate to practical situations. Horowitz (2001a) gives two theorems on consistency by Beran and Ducharme (1991) and by Mammen (1992); and then illustrates their applicability in several examples. Essentially these theorems set conditions on the statistic's distribution which relate to smoothness, continuity and convergence asymptotically to some defined distribution such as the normal. Horowitz comments that 'the conditions that cause inconsistency are unusual in econometric practice'. Unfortunately, because our Wald statistic is a nonlinear functional of the population draws, none of these theorems, or others in Mammen, appear to be directly applicable to the indirect inference here; for such a functional the conditions for consistency are complex (Mammen, pp11–13), and their applicability to our case is unclear.

It appears from the discussion in Horowitz (2001a**,** pp. 3164–5) that a key condition (which in most situations is sufficient) is that the bootstrapped test statistic, B, be an 'asymptotically pivotal' statistic, meaning its asymptotic distribution does not depend on the distribution of the population from which the sample is drawn. The t statistic and the chi-squared statistic are examples: their distributions in large samples are fixed. So if B has one of these two distributions asymptotically it satisfies this criterion. However, a bootstrapped statistic is not generally 'pivotal', i.e. such that its bootstrapped distribution in small samples as well as large is independent of the population distribution from which the sample is drawn. Self-evidently, in general, this will not be the case since, for example, fat tails will influence the bootstrap distribution in small samples.

We establish the consistency of the bootstrapped Wald statistic proposed in this paper as follows[9]. The asymptotic distribution of the Wald statistic is a chi-squared (k) where k is the size of the vector α. It is asymptotically pivotal, since the chi-squared distribution converges on a fixed distribution regardless of the underlying population (where it is the sum of k squared standard normal variables) as the sample size, T, goes to infinity, due to the central limit theorem. Thus the bootstrap distribution will converge on this chi-squared distribution as T goes to infinity.

We now elaborate the steps in this brief statement:

1) The Wald statistic has a chi-squared (k) distribution: The Wald statistic is calculated based on the bootstrap distribution (implied by the DSGE model coefficients θ_0) of a (the VAR coefficients and data variances) around their bootstrap means, as $WS = [a_T - \overline{a_S(\theta_0)}]'W(\theta_0)[a_T - \overline{a_S(\theta_0)}]$, where $W(\theta_0)$ is the inverse of the var-covar matrix of the a. a_T are the values from the data. Thus the bootstrap distribution is found by replacing a_T in this expression by a_S and calculating the value of WS for every bootstrap sample a_S. WS can be interpreted as a sum of squared t-statistics; for example if the covariance matrix inverted in $W(\theta_0)$ is diagonal (so that the covariances are all zero) then $WS = t_1^2 + t_2^2 + ... + t_{30}^2]$ where t_i is the ith VAR coefficient (or data variance) normalised as a deviation from its model-implied mean divided by its model-implied standard error. In the more general case where the covariances are non-zero, the expression is adjusted for these covariances; for example in the case of just 2 coefficients it is $WS = \frac{t_1^2+t_2^2-2t_1t_2\rho_{12}}{1-\rho_{12}^2}$ where ρ_{12} is the correlation between coefficients 1 and 2.

2) The chi-squared distribution is asymptotically pivotal: i.e. the distribution tends to a fixed asymptotic distribution as the sample size, T, grows. In particular, the distribution is invariant to the characteristics of the population from which the samples are drawn. The reason for this, of course, is that the t-values in WS have the dimension of means; and means of random variables tend to the normal distribution as n increases. Hence their summed squares tend to their value under the normal distribution — which is thus now the asymptotic distribution of the chi-squared.

3) The bootstrap distribution must also converge to this same distribution: the bootstrap values, α_S, are estimated from random draws from the underlying population. Thus they too are t-values when normalised; and, by the central limit theorem, they too tend to normality as n tends to infinity. Thus their sum of squares in WS also tends to the same fixed asymptotic chi-squared distribution.

Numerical properties of the bootstrap Wald distribution as sample size increases To illustrate how the bootstrapped Wald distribution behaves as the sample size increases, we conduct a Montecarlo

[9] Our Montecarlo results for small samples appear to show we can assume that the near-unit root problems do not arise; clearly if they did the following would not apply.

experiment on the US model since 1984 ($T = 76$). We report the bootstrap distribution of the Wald statistic (based on repeated resampling of the data sample) for $T = 76$ and 500 together with the asymptotic distribution and various confidence levels of the distribution. As expected from our consistency results, the numerical distribution clearly converges on the asymptotic distribution. (Figure 7 and Table 16).

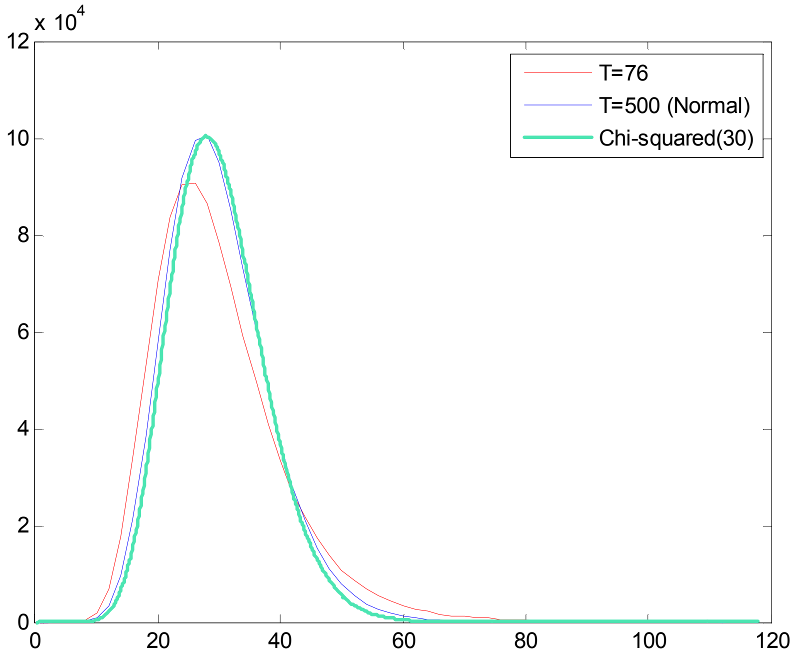

Figure 7: Histogram of distributions from Monte Carlo Bootstrap

	$T = 76$	$T = 500$	$T = \infty$
90%	43.293	41.134	40.26
95%	49.245	45.390	43.77
99%	63.586	54.766	50.89

Table 16: Convergence of bootstrap distribution critical values as n increases (Montecarlo experiment, model as above)

6.1.3 Using the refinement procedure on our results:

As noted above, it is possible to increase the accuracy of the bootstrap by using Montecarlo simulation in a manner we describe as a refinement procedure. Here we apply this procedure to our final set of results, just described. For all the cases where we obtained a Wald percentile of 100, clearly we need no refinement. Our full period results, where we rejected the best-fitting model would, on this refinement procedure, come out marginally more on the side of rejection and so our conclusions there are robust. For the post-1984 sample the result for the trio of output, inflation and interest rates now becomes rather less favourable — the Wald moves from 96.3 to 98.37, implying that it is (still) accepted only at the 99% confidence. Only the output and inflation combination, at 91.58 (moved from 84.4), remains unambiguously accepted at 95% — see Table 17 for refined Wald statistics for the post-1984 sample. Our conclusions are not changed materially.

	Wald (with variances)	
Variables	Original	Refined procedure
Y, C, INV	99.40	99.86
π, R	93.80	97.00
Y, π, R	96.30	98.37
INV, π, R	95.60	98.07
C, π, R	100.00	100.00
$Y(AR(3))$	98.40	99.34
π	51.20	67.53
$R(AR(1))$	97.00	98.67
$R(AR(2))$	100.00	100.00
$Var(Y, R, \pi)$	38.00	55.85
Y, π	84.40	91.58

Table 17: Directed Walds

6.1.4 The bootstrap variance-covariance matrix

In our Wald statistic we used the variance-covariance matrix implied by our model bootstrap samples. An alternative procedure that would avoid bootstrapping is simply to use the variance-covariance matrix implied by the VAR based on the data; under the null this would correspond to the true variance-covariance matrix implied by the model. If, however, the model is false then this variance-covariance matrix would not correspond to the true variance-covariance matrix. The bootstrap variance-covariance matrix corresponds to the true matrix, as restricted by the model, whether the model is true or false. When the model is false, therefore, we conjecture that the null is more likely to be rejected if we use the restricted variance-covariance matrix than if we use the data variance-covariance matrix. Effectively the Wald statistic would not be based on the behaviour of the model being tested except in respect of the bootstrap mean. Yet we have seen that the model restricts the distribution of the VAR coefficients in very definite ways, with a joint distribution 'footprint'. It seems much easier for a false model to get closer to the data if it is allowed to have a footprint from some true but less theory-based model. Indeed we found this to be the case: usually the off-diagonal elements in the data-based matrix are very small, so that the test becomes close to the Wald based only on the no-covariance case, which is, in general, a much easier test to pass. This mirrors our basic findings in this and related papers: that DSGE models are rejected because the data does not reflect the fine detail imposed by the model's restrictions. Thus, specifically, the data-based variance-covariance matrix mirrors in its small off-diagonal elements the lack of tight restrictions between VAR coefficients in the data, whereas the model-restricted variance-covariance matrix with its large off-diagonal elements reflects the tight restrictions imposed between the VAR coefficients by the DSGE model (for an illustration of this see the discussion above of Figure 1).

We have opted for the restricted variance-covariance matrix because we want a test of maximum power for the alternative hypothesis that the model is misspecified. In future work we plan to investigate this issue further using Montecarlo methods, as we need to know the power of various procedures such as this alternative method.

6.2 Parameter uncertainty

In our Direct Wald tests we have tried to pin down potential sources of misspecification by asking whether a particular fixed set of structural parameters θ_0 could have generated the data. It is possible, however, that another set of parameters might be needed to explain how the data are generated. If no set of parameters can be found under which the model passes, then the model itself is rejected. This is, in effect, the basis of indirect inference: its role is to search for the parameter set that gets closest to matching the data, and then to determine whether this set passes our test. Thus the role of Indirect Inference estimation in our procedure (which minimises the Wald value) is to maximise the chances of the model passing the test. While our procedure begins with testing a single parameter set, with the addition of a full Indirect Inference search over the whole parameter space, it provides a test of the full model. Consequently, our test procedure may be interpreted as a way of addressing the issue of parameter uncertainty.

The issue of parameter uncertainty can also addressed more directly by examining whether we could find parameter values for the SW model that reduced the Wald statistic. We calculated the minimum-value full Wald statistic for the Smets and Wouters' original model over the whole sample period using a powerful algorithm based on Simulated Annealing in which search takes place over a wide range around the initial values, with optimising search accompanied by random jumps around the space[10]. We found that while we could improve the Wald substantially it still remained well above the rejection threshold with a Wald percentile of 100.

In view of these results based on a full Wald test, we then peformed a similar exercise for Directed Wald tests (for output, inflation and interest rates) for the full sample and for the post-1984 sample. We found that this resulted in non-trivial changes in the parameters which reduced the value of the Wald statistic[11]. Nevertheless, our conclusions are not changed; the weighted model for the full sample is still rejected and the post-1984 model remains not rejected, only more decisively — see Table 18.

This failure to overturn our original results is not, perhaps, too surprising in view of the effort that went into SW's original Bayesian estimation process. An implication of these findings is that the SW model may be rejected based not only the original parameter estimates, but also on estimates derived from indirect estimation.

We have used the Direct Wald test to examine parameter uncertainty where the original parameters are treated as exact and found that we have been able to pinpoint the parameter range within which the model is not rejected rather precisely. The test also appears have good power against false models. To further examine its power properties we conduct the following Montecarlo experiment.

We generate data from a true model and compute both the Full and the Directed Wald values. We check the rejection rate (at 5%) of models where we arbitrarily raise all the even coefficients, including those of the error time-series, and lower all the odd ones, by $x\%$, where x rises steadily from 1 up to 7 (we keep the innovations that are bootstrapped in the mis-specified models the same, for each data set, as in the true model). Thus x steadily increases the degree of model numerical mis-specification. The results are shown in Figure 8. It can be seen that the method has considerable power against poor parameter values. We can extend this test to model mis-specification and ask how often a model that is arbitrarily mis-specified in structure will be rejected. The wrong model we choose is a weighted average of the NK and the NC models with a weight on the NK of 0.9 in both labour and goods markets and the SW estimates for both the NK and NC parameters. This model is only moderately mis-specified in the sense that it has only a 10% New Classical element; it is dominated by the True NK model. But, not surprisingly, on our test which is based on exact parameter values, it is rejected 100% of the time on the Full Wald— see Figure 9 where the Wald distribution for the False Model on True Model data hardly overlaps with the Wald distribution for the True Model on its own data. The False Model would have to be very much closer to the True Model to fail to be rejected 100% of the time; *a fortiori* more mis-specified models are similarly rejected all the time.

[10] We use a Simulated Annealing algorithm due to Ingber (1996). This mimics the behaviour of the steel cooling process in which steel is cooled, with a degree of reheating at randomly chosen moments in the cooling process — this ensuring that the defects are minimised globally. Similarly the algorithm searches in the chosen range and as points that improve the objective are found it also accepts points that do not improve the objective. This helps to stop the algorithm being caught in local minima. We find this algorithm improves substantially here on a standard optimisation algorithm. We set the range to be examined around the initial parameter values as + or - 10% of these.

[11] The weights on the NK model become substantially higher, in this case where the Directed Wald is minimised. However, they move far less when the Full Wald is minimised. This suggests that, using the fuller information from the sample, the model that gets closest to the data is closer to the one with our original weighted coefficients.

	Original Coefficients	Full Sample Wald Minimised Coefficients	Post 1984 Wald Minimised Coefficients
φ	5.74	5.1710	6.2170
σ_c	1.38	1.5172	1.4719
λ	0.71	0.7808	0.7123
ξ_w	0.70	0.6701	0.7150
σ_L	1.83	1.6544	1.9920
ξ_p	0.66	0.5962	0.7116
ι_w	0.58	0.6140	0.6280
ι_p	0.24	0.2635	0.2385
ψ	0.54	0.5497	0.5350
Φ	1.50	1.3535	1.3727
r_p	2.3/2.5*	2.0747	2.5932
ρ	0.60	0.5400	0.6591
r_y	0.08	0.0878	0.0817
$r_{\Delta y}$	0.22	0.2417	0.2106
$\bar{\pi}$	0.78	0.8562	0.7805
$100(\beta^{-1}-1)$	0.16	0.1445	0.1563
$\overline{L}$	0.53	0.4792	0.5185
$\bar{\gamma}$	0.43	0.4720	0.4627
α	0.19	0.2089	0.1715
υ_w	0.1/0.8*	0.8700	0.9966
υ_p	0.2/0.8*	0.8185	0.9928
Wald percentile	99.6/96.1*	98.7	83.8
M-distance (Normalised)	3.939/1.785*	2.685	0.767
p-value	$4.1e^{-5}$/0.0371*	0.0036	0.2215

* The left hand number applies to the full sample, the right hand to the post 1984.

Table 18: Indirect Inference parameter estimation for weighted model (full sample and post 1984 sample, Directed Wald)

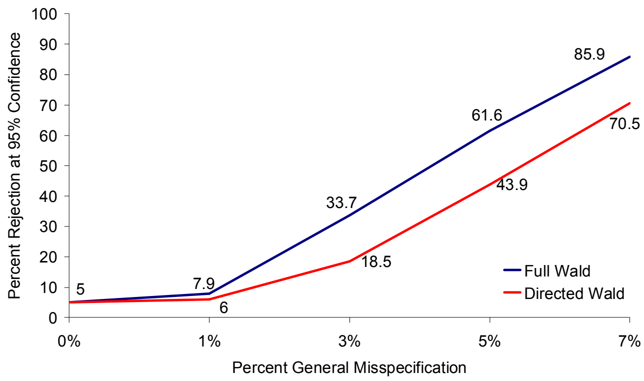

Figure 8: Power Function-Full and Directed Walds

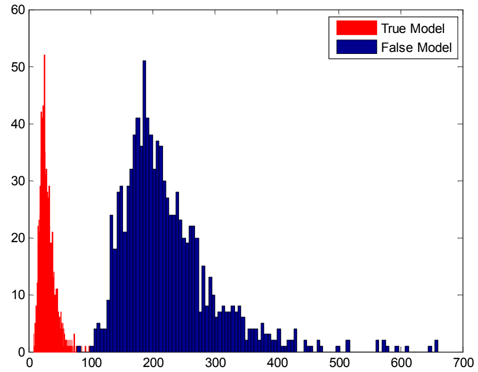

Figure 9: Histograms of True Model and False Model (Full Wald)

An alternative way of examining parameter uncertainty would be to embed the estimation process within the testing procedure. Thus we could estimate θ by Indirect Inference, Bayesian or some other estimation process, and then test the unknown true model by carrying out the test on the estimated parameters. The idea is that then we will be testing a model with unknown θ, solely using its estimated θ. We call this method the estimation-augmented test. To find the distribution of this test we need to do a Montecarlo with the following steps on a True Model with parameters θ

1. Generate a random sample from this model and estimate the auxiliary equation parameters from this as $\widehat{a}$.

2. Estimate $\widehat{\theta}$ from this data sample by the chosen estimation method.

3. Generate the bootstrap distribution of the $\widehat{a}$ from this $\widehat{\theta}$ and calculate the Wald value for this sample.

4. Repeat steps 1–3 for a large number of samples; calculate the resulting Wald distribution and its critical values.

Step 4 would thus give us the distribution of a test statistic that tests whether a model is true based on its estimated parameters.

We investigate this proposal here using the Bayesian estimator, since that is how the Smets-Wouters model was actually estimated[12]. We carried out this Montecarlo experiment for the full sample (we used the modal estimator which is fairly close to the FIML estimator). We found that the distribution of the Walds expanded very substantially — see Figure 10 which compares the Wald distribution for the estimation-

[12] We use the Bayesian estimation code in Dynare (Juilliard, 2001).

augmented test with our standard distribution. Essentially the Bayesian estimates of the model have a wide joint distribution; similarly the corresponding VARs, with many model-simulated VARs a long way from the corresponding data-estimated VARs. With such a wide distribution any VAR generated by a data sample is very likely to lie within its 95% bound, simply because it is not one of these very poorly-fitting cases[13]. Thus this estimation-based test, though elegant, has rather low power. This is illustrated here by the fact that every one of our models would pass this test easily on the full VAR auxiliary model. Since these models differ radically in specification, ranging from New Keynesian through weighted to New Classical, this test's low discrimination is evident.

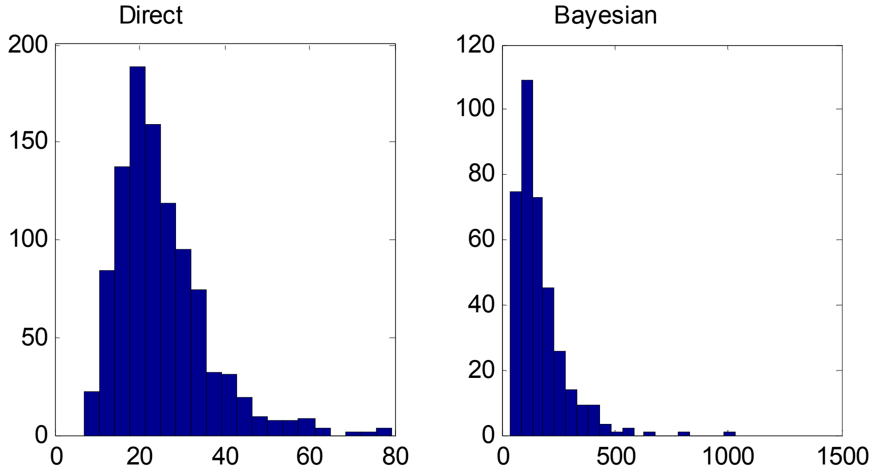

Figure 10: Direct and Bayesian Histograms

To investigate its power formally we generate the Wald distributions for a different and thus mis-specified model where the Wald values come from estimating this wrong model on the data generated by the true model; from this we can calculate the rejection rate of the wrong model at the 5% level. We choose the weighted model as the true model but with weights on NK of unity (thus it is as close as possible to the SW original NK model); we choose the same model as the wrong model but with a weight of 0.3 on NK in both labour and goods markets. These weights are treated as fixed in estimation in both models. Even this seriously mis-specified model is rejected only around 50% of the time. This compares with a 100% rejection rate on our Direct Method for a model with a weight of 0.9 and hence very much closer to the true model. It would appear therefore that the test has only limited power against a highly mis-specified model. More generally, while the idea of allowing for possible parameter variation occurring through estimation is attractive in that it permits one to test models merely on their estimated parameters, its power seems rather low.

Thus we conclude from this preliminary discussion that our Direct method is viable as a powerful test not just of exact parameters but also of models; and that an estimation-based alternative, while viable, seems to lack power, at least in the cases we examine here. This is, of course, an area where further work would be most desirable.

6.3 Some issues of numerical robustness:

6.3.1 The use of H-P filter for the post-1984 sample

The aim of H-P filtering is to detrend the variables thereby leaving the data stationary. The drawback with the widely-used H-P filter is that it is an ad-hoc and arbitrary solution to the problem which over-differences, and therefore induces cycles in, the data; and being a two-sided filter, it distorts the expectations structure

[13]We also carried out the same Montecarlo experiment with the Indirect Inference estimator, with only a limited number of replications, and found a similar distribution to the Bayesian.

of the model by incorporating future outcomes in the current filtered value. Moreover, if each series is filtered in this way then, by implication, the structural errors are also H-P filtered and hence have induced serial correlation. In other words, the filtered structural errors are not the original errors (or shocks) of interest. Nonetheless, if each series is filtered and the same H-P filter is used on each series, and if the distortion to the expectations structure is ignored, then the impulse response functions of the original variables to the original errors are the same as those of the filtered data to the filtered errors.[14]

Arguably, a better solution is to remove a linear trend as this does not induce cycles or distort the expectations structure. For our post-1984 sample, however, we could not use a linear detrending method because it failed to make the data stationary. Despite its deficiencies, we therefore used the H-P filter; our use of it was standard, with both data and bootstraps filtered in the same way. We checked the sensitivity of our results to using another widely-used filter, the Band-Pass filter. We found the Band-Pass filter gave smoother series than the H-P filter but otherwise the results were similar. The main exception was the variance of inflation which could no longer be matched. If, however, we raised the Taylor Rule coefficient on inflation in order to dampen inflation so that it is in line with the data, then we obtained the same results. In particular, we found the same degrees of nominal rigidity in both markets: rejection overall, but acceptance for the limited trio of output, inflation and interest rates. Hence our conclusion, that for the post-1984 data there exists a high-rigidity version of the SW model that fits the business cycle in broad terms, is robust to the choice of this type of filter.

6.3.2 VAR estimation bias

The use of OLS — as here — to estimate the VAR auxiliary equation produces biased coefficients in general owing to the use of lagged endogenous variables as regressors. To check whether this bias is serious enough to affect our results, we re-estimated our post-1984 model sample with a bootstrap estimation procedure that corrects for any bias (this involves estimating the VAR by OLS, then bootstrapping the estimated VAR, comparing the mean VAR coefficients of these bootstraps with those estimated from the data to give the bias, then correcting the estimated coefficients for this — for this method see Kilian, 1998). We found the biases to be small; when bias-correction was applied both to the data and the bootstraps, the Wald statistic was trivially affected.

6.3.3 Initial conditions for the bootstraps

The early part of the actual sample contains the effects of lagged shocks (via the one period lag in the data) while the early part of the bootstrap samples (in our method here) contains the same actual lagged values given by history for each bootstrap sample; this has the disadvantage that each sample is to some (small) degree tied down to artificial similarity. An alternative method would be to randomly perturb the initial values by drawing them randomly from the later parts of the bootstrap samples; the bootstrap samples are then recalculated with randomised initial values. This alternative has the advantage that it is consistent with the idea of each bootstrap corresponding to a different 'potential history'. To check robustness, we recalculated our main results using this method. We found the results to be virtually identical.

7 Conclusion

We have used the method of indirect inference to test a well-known DSGE model of the New Keynesian type based on its dynamic performance for US post-war data. We compared this model with a flexible wage/price version with a short information lag (New Classical) and found that if we use the structural errors jointly implied by each model and the data, then neither model can fit the data variances. The NK produces too little variation in interest rates and the NC model generates an excessive variation in inflation rates. But when the two models are combined in a weighted combination to give a hybrid model which is a mixture

[14] Plainly therefore were we to H-P filter our bootstrap data yet again, this would be spurious. Out of curiosity however we examined what would happen on our post-1984 sample if we did H-P filter the bootstraps produced from our already H-P filtered errors. Consistently with what we suggest above, the variances of the bootstrapped data fall, of course because some of the variation is taken out by filtering again. Hence the variation of the bootstrapped data now is less close to that of the data though generally still inside the 95% bounds. However, the dynamics fit as well as before; and overall the Wald statistic is the same.

of imperfectly competitive and flexible-price markets, then with high flexibility in both the product and the labour market the hybrid model comes much closer to matching the data, even though it too is rejected especially in respect of interest rate behaviour. One possible reason is monetary regime change, for which there is evidence in the mid-1960s and the mid-1980s. When we examined the period since 1984 we found that in respect of output and nominal variables the data did not reject a model with a high degree of nominal rigidity. This suggests that the situation with regard to price or wage rigidity in the US economy may have changed over time; during most of the period state dependence in pricing could have complemented time dependence due to the economy's large fluctuations, but, during the later period, the 'Great Moderation', time dependence seems to have heavily dominated.

We have undertaken an extensive examination of the robustness of our test procedure. We find that it is a consistent test, that it has good power and that our results are not due to uncertainty in the parameters in the original model. The significance of our test statistic appears not to be due to any detectable weakness in the test procedure, or to poor estimates of the original model, but to the specification of the model.

There is a widespread view that tests of DSGE models are not useful as the models are misspecified. One reason why likelihood-based tests of a DSGE model tend to reject might be that they involve all features of the model. Often theory has more to say about some aspects of a model rather than others yet the test takes into account all aspects. For example, we might claim to know more about the long-run properties of a model than the short-run lag structure. Even if a model is rejected we would also like to know which aspects of the model are causing rejection and which aspects are robust to the data. (See, for example, studies of policy robustness to models such as that of Cogley et al, 2010, at this conference). It may therefore make sense to consider particular aspects of a model, such as those that would affect policymaking objectives — see, for example, An and Schorfheide (2007). This view might be used to justify constraining core elements of DSGE models through prior distributions (subject only to variation within the posterior distribution). The model can then be tested through marginal improvements in parameter values using the likelihood criterion. We have proposed what we refer as a Directed Wald statistic to address this issue. Using this test we found that a weighted model with high rigidity does well for the three key macro variables, GDP, inflation and interest rates, after 1984. This makes it a good model for policymakers who are mainly interested in these three things.

Bayesian estimation methods are an effective practical tool for improving DSGE model estimates by incorporating prior information about the macroeconomy. This paper shows that indirect inference is a useful companion to this approach.

References

[1] An, S. and F. Schorfheide, 2007. 'Bayesian Analysis of DSGE Models', Econometric Reviews, 26(2), March 2007, 113–172

[2] Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991) 'Bootstrapping unstable first-order autoregressive processes', Annals of Statistics, 19, 1098–1101.

[3] Beran, R. and G.R. Ducharme (1991) Asymptotic Theory for Bootstrap Methods in Statistics, Publications CRM, Universite de Montreal.

[4] Bianchi, F. (2010) 'Regime switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics', Duke University, downloadable from http://mpra.ub.uni-muenchen.de/24251/ .

[5] Bils, M. and P. J. Klenow, 2004, 'Some Evidence on the Importance of Sticky Prices,' Journal of Political Economy, 112 (5), 947–985.

[6] Canova, F. (1994) 'Statistical inference in calibrated models', Journal of Applied Econometrics, 9, S123–144.

[7] Canova, F. (1995) 'Sensitivity analysis and model evaluation in Dynamic Stochastic General Equilibrium models', International Economic Review, 36, 477–501.

[8] Canova, F., 2005. Methods for Applied Macroeconomic Research, Princeton University Press, Princeton.

[9] Canova, F. and L. Sala (2009) 'Back to square one: Identification issues in DSGE models', Journal of Monetary Economics, 56, 431–449.

[10] Chari, V., P. J. Kehoe and E. McGrattan (2002) 'Can sticky price models generate volatile and persistent real exchange rates?' Staff Report 277 from Federal Reserve Bank of Minneapolis, published in Review of Economic Studies, Vol. 69, No. 3, July 2002, pp. 533–563.

[11] Christiano, L. (2007) Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,143–151.

[12] Christiano, L., M. Eichenbaum, and C. Evans, 2005. 'Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,' Journal of Political Economy, University of Chicago Press, vol. 113(1), pages 1–45, February.

[13] Cogley, T., B. De Paoli, C. Matthes, K. Nikolov and A. Yates (2010) "A Bayesian Approach to Optimal Monetary Policy with Parameter and Model Uncertainty", paper presented at this conference.

[14] Dave.C., and D.N. De Jong (2007) Structural Macroeconomics, Princeton University Press.

[15] Del Negro, M. and F. Schorfheide (2004) 'Priors from general equilibrium models for VARs', International Economic Review, 45, 643–673.

[16] Del Negro, M. and F. Schorfheide (2006) 'How good is what you have got? DSGE-VAR as a toolkit for evaluating DSGE models', Federal Reserve Bank of Atlanta Economic Review, 91, 21–37.

[17] Del Negro, M., F. Schorfheide, F. Smets, R. Wouters (2007a) 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,123–143.

[18] Del Negro, M., F. Schorfheide, F. Smets, R. Wouters (2007b) Rejoinder to Comments on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,159–162.

[19] Dixon, H. 1992. 'Nominal wage flexibility in a partly unionised economy,' The Manchester School of Economic and Social Studies, 1992, 60, 295–306.

[20] Dixon, H. 1994. 'Macroeconomic price and quantity responses with heterogeneous product markets,' Oxford Economic Papers, 1994, 46, 385–402.

[21] Dridi, R., A. Guay and E. Renault (2007) 'Indirect inference and calibration of dynamic stochastic general equilibrium models', Journal of Econometrics, 136, 397–430.

[22] Faust, J. (2007) Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,154–156.

[23] Fukac, M. and A. Pagan (2008) 'Limited Information Estimation and Evaluation of DSGE Models', DP2008/11, dowloadable from www.rbnz.govt.nz/research/discusspapers/.

[24] Gallant, A.R. (2007) Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,151–152.

[25] Gertler, M. and J. Leahy (2008) 'A Phillips curve with an Ss foundation", Journal of Political Economy, 118 (3), 533–572.

[26] Gourieroux, C. and A., Monfort, 1995. Simulation Based Econometric Methods, CORE Lectures Series, Louvain-la-Neuve.

[27] Gourieroux, C., A. Monfort, and E. Renault, 1993. 'Indirect inference", Journal of Applied Econometrics, 8, 85–118.

[28] Gregory, A. and G. Smith, 1991. ' Calibration as testing: Inference in simulated macro models,' Journal of Business and Economic Statistics, 9, 293–303.

[29] Gregory, A. and G. Smith, 1993. 'Calibration in macroeconomics", in G. Maddala, ed., ' Handbook of Statistics", vol. 11, Elsevier, St. Louis, Mo., 703–719.

[30] Hall, A., A. Inoue, J.M. Nason and B. Rossi (2009) 'Information Criteria for Impulse Response Function Matching Estimation of DSGE Models' Centre for Growth and Business Cycle Research, Economic Studies, University of Manchester, discussion paper no. 127; downloadable from http://www.socialsciences.manchester.ac.uk/cgbcr/discussionpapers/index.html

[31] Hansen, Bruce E, (1999) 'The Grid Bootstrap And The Autoregressive Model', The Review of Economics and Statistics, 1999, 81, (4), 594–607.

[32] Horowitz, Joel L. (2001a) 'The bootstrap', Handbook of Econometrics , eds. J.J.Heckman and E. Leamer, vol.5, ch. 52, 3159–3228, Elsevier.

[33] Horowitz, Joel L. (2001b) "The Bootstrap and Hypothesis Tests in Econometrics," Journal of Econometrics, 100, 37–40; "Response to Comments," Journal of Econometrics, 100, 97–98..

[34] Ingber, L. (1996) 'Adaptive simulated annealing (ASA): Lessons learned,' Special issue on Simulated Annealing Applied to Combinatorial Optimization, Control and Cybernetics, 25(1), 33–54.

[35] Juillard, M. (2001) 'Dynare: a program for the simulation of rational expectations models' Computing in economics and finance 213. Society for Computational Economics.

[36] Kilian, L. (1998) 'Small-sample confidence intervals for impulse response functions' Review of Economics and Statistics, 80, 218–230.

[37] Kilian, L. (2007) Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,156–159.

[38] Kimball, M., 1995. 'The quantitative analytics of the basic neomonetarist model", Journal of Money, Credit and Banking, 27(4), Part 2, 1241–1277.

[39] King, R., C. Plosser and S. Rebelo (1988) 'Production, growth and business cycles: I' Journal of Monetary Economics, 21, 195–232.

[40] Le, V.P.M., P. Minford and M. Wickens (2010) 'The 'Puzzles' methodology: en route to indirect inference?' Economic Modelling, in press..

[41] Mammen, E. (1982) 'When Does Bootstrap work? Asymptotic results and simulations', Lecture Noes in Statistics 77, Springer.

[42] McCallum, B. T. (1976) 'Rational expectations and the natural rate hypothesis: some consistent estimates", Econometrica, 44, 43–52.

[43] Minford, P. and D. Peel (2002) Advanced Macroeconomics — A Primer, Edward Elgar.

[44] Minford, P., K. Theodoridis, and D. Meenagh (2009), 'Testing a model of the UK by the method of indirect inference', Open Economies Review, vol. 20(2), pages 265–291, April.

[45] Perron, P. and Qu, Z. (2007) 'Estimating and Testing Structural Changes in Multivariate Regressions,' Econometrica, 75(2), 459–502.

[46] Sims, C. A. (2007) Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics, 25,152–154.

[47] Smets, F. and R. Wouters, 2003. 'An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,' Journal of the European Economic Association, MIT Press, vol. 1(5), pages 1123–1175, 09.

[48] Smets, F. and R. Wouters, 2007. 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,' American Economic Review, American Economic Association, vol. 97(3), pages 586–606, June

[49] Smith, A. , 1993. ' Estimating nonlinear time-series models using simulated vector autoregressions", Journal of Applied Econometrics, 8, S63–S84.

[50] Theodoridis, K. (2006) 'Methods of evaluating DSGE models: a literature review', chapter 2, PhD thesis, Cardiff University, available at http://www.cf.ac.uk/carbs/econ/minfordp/ under 'Estimation and Testing'.

[51] Watson, M. (1993) 'Measures of fit for calibrated models,' Journal of Political Economy, 101, 1011–1041.

[52] Wickens, M.R. (1982) 'The efficient estimation of econometric models with rational expectations", Review of Economic Studies, 49, 55–67.

Appendix A Listing of models — SWNK and SWNC

SWNK MODEL

Consumption Euler equation

$$c_t = \frac{\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}} c_{t-1} + \frac{1}{1+\frac{\lambda}{\gamma}} E_t c_{t+1} + \frac{(\sigma_c - 1)\left(W_*^h L_*/C_*\right)}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}(l_t - E_t l_{t+1}) - \left(\frac{1-\frac{\lambda}{\gamma}}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}\right)(r_t - E_t\pi_{t+1}) + eb_t \quad (1)$$

Investment Euler equation

$$inn_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}} inn_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}} E_t inn_{t+1} + \frac{1}{\left(1+\beta\gamma^{1-\sigma_c}\right)\left(\gamma^2\right)\varphi} qq_t + einn_t \quad (2)$$

Tobin Q equation

$$qq_t = \frac{1-\delta}{1-\delta+R_*^K} E_t qq_{t+1} + \frac{R_*^K}{1-\delta+R_*^K} E_t rk_{t+1} - (r_t - E_t\pi_{t+1}) + \frac{1}{\frac{1-\frac{\lambda}{\gamma}}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}} eb_t \quad (3)$$

Capital Accumulation equation

$$k_t = \left(1 - \frac{\delta}{\gamma}\right) k_{t-1} + \frac{\delta}{\gamma} inn_t + \left(1-\frac{\delta}{\gamma}\right)\left(1+\beta\gamma^{1-\sigma_C}\right)\gamma^2\varphi\,(enn_t) \quad (4)$$

Price Setting equation

$$\pi_t = \begin{bmatrix} \frac{\beta\gamma^{1-\sigma_C}}{1+\beta\gamma^{1-\sigma_C}\iota_p} E_t\pi_{t+1} + \frac{\iota_p}{1+\beta\gamma^{1-\sigma_C}\iota_p}\pi_{t-1} - \left(\frac{1}{1+\beta\gamma^{1-\sigma_C}\iota_p}\right) \\ \left(\frac{\left(1-\beta\gamma^{1-\sigma_C}\xi_p\right)\left(1-\xi_p\right)}{\xi_p((\Phi_p-1)\varepsilon_p+1)}\right)((\alpha)\, rk_t + (1-\alpha)\, w_t - ea_t) \end{bmatrix} + ep_t \quad (5)$$

Wage Setting equation

$$w_t = \begin{bmatrix} \frac{\beta\gamma^{1-\sigma_C}}{1+\beta\gamma^{1-\sigma_C}} E_t w_{t+1} + \frac{1}{1+\beta\gamma^{1-\sigma_C}} w_{t-1} + \frac{\beta\gamma^{1-\sigma_C}}{1+\beta\gamma^{1-\sigma_C}} E_t\pi_{t+1} - \frac{1+\beta\gamma^{1-\sigma_C}\iota_w}{1+\beta\gamma^{1-\sigma_C}}\pi_t \\ +\frac{\iota_w}{1+\beta\gamma^{1-\sigma_C}}\pi_{t-1} - \frac{1}{1+\beta\gamma^{1-\sigma_C}}\left(\frac{\left(1-\beta\gamma^{1-\sigma_C}\xi_w\right)(1-\xi_w)}{(1+\varepsilon_w(\Phi_w-1))\xi_w}\right) \\ \left(w_t - \sigma_L l_t - \left(\frac{1}{1-\frac{\lambda}{\gamma}}\right)\left(c_t - \frac{\lambda}{\gamma}c_{t-1}\right)\right) \end{bmatrix} + ew_t \quad (6)$$

Labour demand

$$l_t = -w_t + \left(1 + \frac{1-\psi}{\psi}\right) rk_t + k_{t-1} \quad (7)$$

Market Clearing condition in goods market

$$y_t = c_y c_t + i_y inn_t + R_*^k k_y \frac{1-\psi}{\psi} rk_t + eg_t \quad (8)$$

Aggregate Production equation

$$rk_t = \frac{1}{\phi_p \alpha \frac{1-\psi}{\psi}}\left(y_t - \phi_p \alpha k_{t-1} - \phi_p (1-\alpha)\, l_t - \phi_p ea_t\right) \quad (9)$$

Taylor Rule

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_p\pi_t + r_y\left(y_t - y_t^f\right)\right) + r_{\Delta y}\left(\left(y_t - y_t^f\right) - \left(y_{t-1} - y_{t-1}^f\right)\right) + er_t \quad (10)$$

SWNC MODEL

Consumption Euler equation

$$c_t = \frac{\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}}c_{t-1} + \frac{1}{1+\frac{\lambda}{\gamma}}E_t c_{t+1} + \frac{(\sigma_c - 1)\left(W_*^h L_*/C_*\right)}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}\left(l_t - E_t l_{t+1}\right) - \left(\frac{1-\frac{\lambda}{\gamma}}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}\right)(r_t) + eb_t \tag{11}$$

Investment Euler equation

$$inn_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}}inn_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}}E_t inn_{t+1} + \frac{1}{\left(1+\beta\gamma^{1-\sigma_c}\right)\left(\gamma^2\right)\varphi}qq_t + einn_t \tag{12}$$

Tobin Q equation

$$qq_t = \frac{1-\delta}{1-\delta+R_*^K}E_t qq_{t+1} + \frac{R_*^K}{1-\delta+R_*^K}E_t rk_{t+1} - \left(r_t - E_t\pi_{t+1}\right) + \frac{1}{\frac{1-\frac{\lambda}{\gamma}}{\left(1+\frac{\lambda}{\gamma}\right)\sigma_c}}eb_t \tag{13}$$

Capital accumulation equation

$$k_t = \left(1-\frac{\delta}{\gamma}\right)k_{t-1} + \frac{\delta}{\gamma}inn_t + \left(1-\frac{\delta}{\gamma}\right)\left(1+\beta\gamma^{1-\sigma C}\right)\gamma^2\varphi\left(enn_t\right) \tag{14}$$

Marginal Product of Labour

$$\alpha rk_t + (1-\alpha)\,w_t = ea_t \tag{15}$$

Labour supply

$$w_t = \ \sigma_L l_t + \left(\frac{1}{1-\frac{\lambda}{\gamma}}\right)\left(c_t - \frac{\lambda}{\gamma}c_{t-1}\right) - \left(\pi_t - E_{t-1}\pi_t\right) \tag{16}$$

Labour Demand

$$l_t = -w_t + \left(1+\frac{1-\psi}{\psi}\right)rk_t + k_{t-1} \tag{17}$$

Market clearing condition

$$y_t = c_y c_t + i_y inn_t + R_*^K k_y \frac{1-\psi}{\psi}rk_t + eg_t \tag{18}$$

Production function

$$rk_t = \frac{1}{\phi_p \alpha \frac{1-\psi}{\psi}}\left(y_t - \phi_p \alpha k_{t-1} - \phi_p (1-\alpha)\,l_t - \phi_p ea_t\right) \tag{19}$$

Taylor Rule

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_p\pi_t + r_y y_t\right) + r_{\Delta y}\left(y_t - y_{t-1}\right) + er_t \tag{20}$$