RESERVE BANK
OF NEW ZEALAND

DP2005/02

Mind your ps and qs! Improving ARMA forecasts with RBC priors

Kirdan Lees and Troy Matheson

October 2005

JEL classification: C11, C22, E37

Discussion Paper Series

DP2005/02

Mind your Ps and Qs! Improving ARMA forecasts with RBC priors*

Kirdan Lees† and Troy Matheson

Abstract

We utilise prior information from a simple RBC model to improve ARMA forecasts of post-war US GDP. We develop three alternative ARMA forecasting processes that use varying degrees of information from the Campbell (1994) flexible labour model. Directly calibrating the model produces poor forecasting performance whereas a model that uses a Bayesian framework to take the model to the data, yields forecasting performance comparable to a purely statistical ARMA process. A final model that uses theory only to restrict the order of the ARMA process (the ps and qs), but that estimates the ARMA parameters using maximum likelihood, yields improved forecasting performance.

*The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Reserve Bank of New Zealand. We thank Stephen Burnell, Aaron Drew, Thomas Lubik, Christie Smith, Shaun Vahey and seminar participants at the Reserve Bank of New Zealand's DSGE Modelling workshop, August 2005, for comments. We are also grateful for Matlab code by Juan Rubio-Ramirez, which we adapted for a section of the paper. Any remaining errors are the responsibility of the authors.

†Kirdan Lees, Economics Department, RBNZ, P.O.Box 2498, Wellington, New Zealand, Tel: +64-4-471-3666; Fax +64-4-473-1209; *email address: kirdan.lees@rbnz.govt.nz*

1 Introduction

DSGE models imply imposing restrictions on underlying reduced form statistical models. In-sample, such restrictions cannot improve the empirical fit of the model relative to unrestricted models. However, in small samples, restricted models may produce superior out-of-sample forecasts for reasons of parsimony. This implies it is an empirical question whether information from a specific theoretical model provides restrictions that improve forecasts over an unrestricted statistical model.

We forecast post-war US output with information from the Campbell (1994) flexible labour model. This parsimonious model reduces to an ARMA representation for output, which we compare with a standard ARMA forecasting method.

Our specific application utilises the Campbell (1994) model to define three specific theoretical ARMA forecasting models. First, we form priors over key parameters in the Campbell (1994) model and calibrate the underlying ARMA parameters from these priors. This model forecasts poorly.

Second, we adopt Bayesian methods that deliver a posterior that uses the data to slant our prior beliefs according to the likelihood principle. This produces forecast performance comparable with the statistical ARMA forecasting model.

Finally, we allow the theoretical model to select only the ARMA order and estimate the reduced form ARMA parameters by maximum-likelihood. The weight of evidence suggests this model performs well: restricting the order of the ARMA model according to RBC theory, appears to help forecasting.

Section 2 details ARMA forecasting, how RBC theory might help forecasting, and outlines the Campbell (1994) model. Section 3 presents the data and section 4 details our four forecasting models. Section 5 details results of the forecasting experiments, and concluding remarks are offered in section 6.

2 Forecasting with ARMA processes

2.1 Statistical ARMA estimation

Wold (1938) shows a covariance stationary processes y_t, can be expressed as:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{1}$$

where ε_t is the error from forecasting y_t solely with lagged values of y_t. A Wold representation requires an infinite number of parameters. However, in practice $\psi(L)$ can be approximated by the ratio:

$$\sum_{j=0}^{\infty} \psi_t = \frac{\Omega(L)}{\Xi(L)} \equiv \frac{1 + \omega_1 L + \omega_2 L^2 + \cdots + \omega_q L^q}{1 + \xi_1 L + \xi_2 L^2 + \cdots + \xi_p L^p}. \tag{2}$$

ARMA modelling seeks a finite parameter approximation to the Wold representation. Box and Jenkins (1976) emphasise parsimony in the choice of the AR order p and the MA q order and note that purely statistical univariate ARMA models often prove useful forecasting models.[1] The general ARMA model is:

$$y_t = \sum_{i=1}^{p} \xi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \omega_j \varepsilon_{t-j} \tag{3}$$

where ε_t is a disturbance term and y_t is the stationary forecast variable. The typical strategy implies:

1. Select an appropriate maximum p, q order for the ARMA (p, q) model.

[1] Frequently large-scale macroeconometric models fail to outperform univariate ARMA models with a small number of arguments (Hamilton, 1994, p. 109).

2. Estimate $\xi(L)$ and $\omega(L)$ associated with each choice of p and q.
3. Select the best model (choice of p and q) based on model diagnostics.

Researchers examine the behaviour of the ACF and PACF of the data to guide appropriate maximum (positive integer) values of p and q and use information criteria (such as the Akaike Information Criterion (AIC) and the Schwarz-Bayesian Criterion (BIC)) to trade-off explanatory power with parsimony to help selection. In practice, different information criteria frequently lead to different ARMA orders.[2]

2.2 Why theory might help

Several papers show that RBC theory can be useful for forecasting. DeJong, Ingram, and Whiteman (2000) emphasize the Bayesian framework for combining theoretical and statistical models and show improved forecasting performance from using the Greenwood, Hercowitz, and Huffman (1988) RBC model. Ireland (2004) finds superior forecasting performance from adopting prior information from the Hansen (1985) RBC model.[3]

While an infinite order VAR can approximate a VARMA model arbitrarily well, Chari, Kehoe, and McGrattan (2005) argue VARs of order length typically used in applied work can fail to match the structure implied by theoretical models. Christiano, Eichenbaum, and Vigfusson (2005) confirm this finding but argue these models are rejected by the data.

[2] Hamilton (1994), p. 112-113, suggests log changes of quarterly US real GNP data 1947 to 1988 appear consistent with both AR(1) and MA(2) processes.

[3] Out-of-sample forecast improvements are not restricted to real variables. Del Negro and Schorfheide (2004) find information from a tri-variate New-Keynesian DSGE model improves VAR forecasts for nominal variables, namely inflation and the federal funds rate, in addition to US output growth.

Fernández-Villaverde, Rubio-Ramírez, and Sargent (2005) provide a simple test for conditions where a VAR can recover the theoretical model. But if we are agnostic about the true DGP (Data Generating Process), we cannot test whether the DGP can be adequately represented by a VAR. To put the forecasting comparison on an even footing, we argue it is appropriate to include moving average terms within the statistical model.

These applications work in small samples because the restrictions these particular RBC models impose slant the forecasting models towards the DGP, improving out-of-sample forecasts. Just as the Minnesota prior usefully slants a forecasting model towards persistent dynamics, Ingram and Whiteman (1994) show that RBC theory can improve forecasting.

Both the statistical model and theoretical model misspecify the true DGP process. Given that we do not know the true underlying DGP, only the data can determine whether information from a particular theoretical model will help forecasting.

Prior theoretical information can be incorporated in both the choice of the order of the ARMA process and the estimation of the ARMA process. Theory may prove helpful with:

1. Selecting an appropriate order for the autoregressive and moving average components (the *ps* and *qs*); and
2. Informing $\xi(L)$ and $\omega(L)$, the autoregressive and moving average terms associated with the selected model.

To test whether RBC theory is helpful, we compare forecasts from a purely statistical baseline ARMA model with three ARMA processes that incrementally use less information from the RBC model:

1. A *Calibrated model*, where the RBC model picks the ARMA order; point mass priors for the deep parameters are used to calibrate ξs and ωs.

2. A *Bayesian model*, where the RBC model picks the ARMA order; and the prior about the deep parameters of the RBC model is combined with the likelihood using the Metropolis-Hastings simulator, to generate ξs and ωs;
3. A *ML model*, where the RBC model picks the ARMA order (*ps* and *qs*) and the ξs and ωs are estimated using maximum likelihood;

2.3 The Campbell model

Campbell's (1994) flexible labour model is a useful RBC model to test the ability of economic theory to improve forecasting, because it is parsimonious and well known within the literature.[4]

In this model the representative agent has log utility for consumption and power utility for leisure:

$$E_t \sum_{i=0}^{\infty} \beta^i [\log(C_{t+i}) + \theta \frac{(1-N_{t+i})^{1-\gamma}}{1-\gamma}] \tag{4}$$

where γ is the coefficient of relative risk aversion, and the intertemporal elasticity of substitution of leisure is defined to be $\sigma \equiv 1/\gamma$.

Firms produce with a constant returns-to-scale Cobb-Douglas production function:

$$Y_t = (A_t N_t)^{\alpha} K_t^{1-\alpha} \tag{5}$$

where N_t is the number of hours of labour, K_t the stock of capital and A_t technology. Capital accumulates according to:

$$K_{t+1} = (1-\delta)K_t + Y_t - C_t \tag{6}$$

[4]For example, Campbell (1994) uses the model to show that output may increase initially in the face of highly persistent negative technology shocks; Ludvigson (1996) shows that the persistence of the government debt process may obviate any crowding out effects from deficit-financed cuts in distortionary income taxation; Lettau (2003) uses a model variant to show why the equity premium is small, and Campbell and Ludvigson (2001) reconcile the small intertemporal elasticity of substitution implied by the model by introducing a household sector.

where δ is the depreciation rate.

Campbell (1994) derives a log-linear representation of the model with an autoregressive process for the natural logarithm of technology shocks, a_t:

$$a_t = \phi a_{t-1} + \epsilon_t \tag{7}$$

Campbell shows that an analytical solution exists for the model that reduces the natural logarithm of deviations of output from the steady-state y_t, to an ARMA(2,1) process:

$$y_t = (\phi + \eta_{kk})y_{t-1} - \eta_{kk}\phi y_{t-2} + \eta_{ya}\epsilon_t - (\eta_{yk}\eta_{ka} - \eta_{ya}\eta_{kk})\epsilon_{t-1} \tag{8}$$

where ϵ_t is the idiosyncratic shock to (log) technology and η_{ya}, η_{yk}, η_{ka} and η_{kk} are elasticities that are functions of the deep parameters of the model. The appendix gives a summary of the derivation of the ARMA representation.

2.4 The steady-state

Campbell (1994) derives the dynamic properties of the model using log-linearisations and first order Taylor approximations around steady state values. Typically the data is expressed in terms of deviations from the steady state — but what are the appropriate steady state values?

Quantitative properties of the data can be dependent on the detrending method. Canova (1998) suggests that at a minimum, the data generated by the theory should be passed through a variety of different detrending filters to check the implications of the theoretical model over a range of frequencies. We detrend our data using two methods — we extract a linear trend (LT), and we take log first differences (FD).

3 Data

The calibration literature (for example, see Cooley and Prescott 1995 and King and Rebelo 2000) emphasises matching the concept variable in the model to the corresponding concept of the variable in the data.[5]

Our theoretical economy is closed, it has no government sector, and consumption goods only last for one period. This suggests removing net exports, removing the government sector and defining consumption narrowly to exclude durable goods since durables can exist for more than one period. Thus we consider a narrow definition of output, that we call total private demand, that sums private consumption and private investment. We also work with a broad definition of output that is total real GDP.[6]

4 Four forecasting models

4.1 The baseline statistical ARMA model

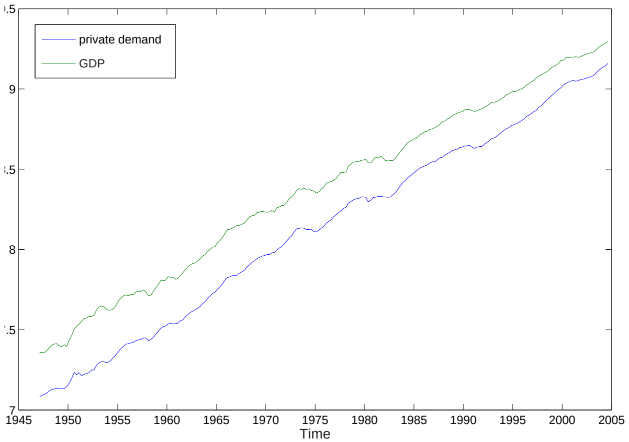
Stock and Watson (2001) suggest that only "state-of-the-art" VARs outperform univariate forecasting models. This appears true with our data. For example, we forecast with a random walk model and a three variable VAR model (with lag length selected according to the BIC) over the sample 1947Q1 – 2002Q3. We look at the linearly detrended private demand output measure.

We use the standard procedure to estimate our baseline ARMA model. This forms the statistical model that forms the baseline for comparison

[5]This point is not restricted to RBC models but applies to DSGE models in general (see Lubik and Schorfheide (2005), for an example of a model model with nominal rigidities.

[6]We use quarterly data obtained from the Bureau of Economic Analysis (in chain-linked 2000 dollars) from 1947q1 to 2004q3. Private consumption is private non-durables consumption (PCNDGC96) plus private services consumption (PCESVC96). Private investment is defined to be total private fixed investment (FPIC1) plus private durables consumption (PCDGCC96). The broad output measure is total real GDP (GDPC1).

Figure 1
US Output Definitions: 1947Q1 – 2004Q3



for the three theory-based forecasting models that follow. We estimate p, q combinations (where $p = 0, 1, 2, 3$ and $q = 0, 1, 2$), and select the best model using the BIC.[7] Table 1 shows our ARMA model yields lower root mean squared errors (RMSEs) than either model up to eight quarters ahead.

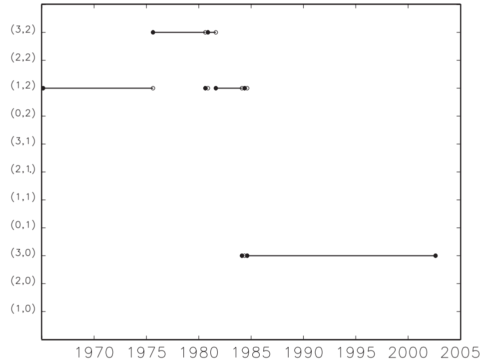
Table 1
MSEs: Private demand, % deviation from trend, 1947Q1–2002Q3

h	1	2	3	4	5	6	7	8
ARMA	0.71	1.18	1.62	2.06	2.43	2.75	3.04	3.27
RW	0.76	1.30	1.79	2.23	2.62	2.96	3.26	3.51
VAR	1.38	1.70	2.10	2.52	2.88	3.17	3.44	3.68

[7] Throughout the paper we adopt the BIC simply to conserve space. Results based on the AIC are quantitatively different but yield identical qualitative conclusions. These results are available on request.

If the DGP is time invariant, sampling error in estimates of autocorrelation coefficients suggests misspecification in the ARMA process. Figure 2 shows that the order of our purely statistical ARMA varies over time. The Bayesian Information Criterion initially selects an ARMA(1,2). By the end of the sample period, the order selected has switched to an ARMA (3,0). This indicates some small sample misspecification and possibly, a structural break in the early 1980s. This misspecification may be associated with a second oil price shock in the late 1970s, the Volcker disinflation, or changes in fiscal policy.[8]

Figure 2
The ARMA order changes over time



NB: The y-axis gives the ARMA(p,q) order. Initially the BIC chooses a (1,2) order but this switches to a (3,0) specification over the latter part of the sample.

[8] An extensive array of papers tests for structural breaks in the growth rate of US output. Perron (1989) rejects a random walk model in favour of a one-time break in the change in US output in 1973. Several papers (Banerjee, Lumsdaine, and Stock 1992, Christiano 1992, Zivot and Andrews 1992) show that critical values are much larger when the breakpoint is determined by the data and this leads to fewer rejections of the null of no break. Bai, Lumsdaine, and Stock (1998) sharpen inference with multivariate tests and find evidence of a break around 1967 (predating the oil shock in 1973). Ireland (2004) estimates an RBC model for two subsamples 1960Q1–1979Q4 and 1980Q1–1997Q3 and finds only small differences in estimates of RBC parameters, but smaller and less persistent shocks, when a structural break is allowed.

4.2 The calibrated model

Our first theoretical model uses a particular calibration of the flexible labour model.[9] The mean of our priors for particular structural parameters are used to form the parameters of the reduced form ARMA(2,1) representation for output; namely ξ_1, ξ_2 and ω_1, which produces forecasts for output for $h = 1$ to 8 quarters ahead.

Our goal is not to explore the theoretical implications of alternative calibrations of the flexible labour model. Instead, we seek to make quantitative inference statements about the model's forecasting ability. *A priori* we regard certain parameter values as more likely than other values and calibrate the model with point priors of $\phi = 0.9$ and $\sigma = 1$.[10]

4.3 The Bayesian model

Our second model allows the data to inform the parameters of the flexible labour model and hence the ARMA(2,1) representation used for forecasting.

We estimate the posterior distribution of the key deep parameters of the flexible labour model by combining our prior distribution (used for forecasting for the first theoretical model) with the likelihood function estimated using the Kalman filter. We fix some deep parameters. Specifically, we set the underlying growth parameters: α, labour's share of income, is two-thirds; δ the rate of depreciation is 2.5 percent; g the quarterly per capita growth rate in log technology is 1.5 percent; and r, the quarterly log real return on capital is 1.5 percent.

[9] Campbell (1994) takes an agnostic view regarding the model parameterization, tabulating results across a range of combinations of ϕ=0.00 ,0.50 ,0.95 ,1.00 and σ=0.0, 0.2, 1, 5, ∞. Rather than make a statement regarding the plausibility of particular values, Campbell's aim is to explore the implications of different parameter values.

[10] That particular parameter values are more plausible than others is almost implicit in Campbell's (1994) exploration of the model's implication for highly autocorrelated ($\phi = 0.95$) technology shocks.

We focus estimation on ϕ and σ for forecasting purposes rather than matching the moments of observed macroeconomic times series — the objective of traditional calibration (see Campbell 1994, *inter alia*). We generate 50,000 draws from the joint density of the parameters using the Metropolis-Hastings posterior simulator.[11] The first 25,000 draws from the posterior density are discarded, and we use the posterior means of the remaining draws as our estimates of the model's deep parameters.

Table 2
Prior and posterior distributions for Campbell (1994) model

	Prior				Posterior			
	Mean	Density	Range	SD	Mean	Median	CI(low)	CI(high)
ϕ	0.9	Beta	[0,1)	0.1	0.956	0.979	0.942	0.994
σ	1	Gamma	[0,∞)	0.2	0.345	0.619	0.150	0.956
σ_a	0.01	Inv Gam	[0,∞)	4	0.012	0.012	0.010	0.013

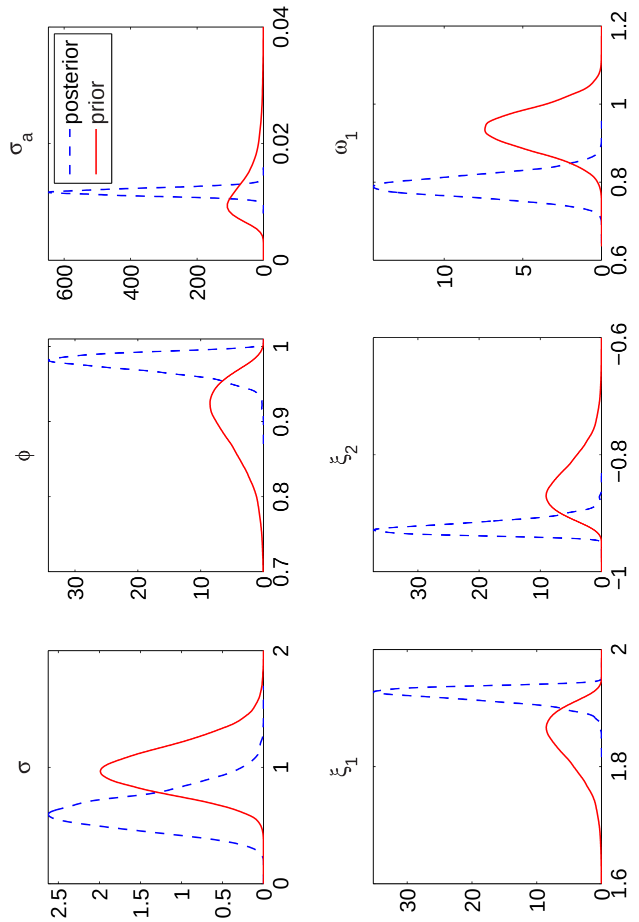
NB: ϕ is autocorrelation in technology shocks; σ is intertemporal elasticity of substitution for leisure and σ_a is the variance of technology shocks. The Inverse Gamma prior is of the form $p(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-s/2\sigma^2}$ and we report the parameters s and ν.

Our pseudo real-time forecasting exercise requires us to recursively estimate the model each quarter—for every combination of data definition and detrending method.[12] Table 2 presents priors and posteriors from the results of the Bayesian estimation exercise when private demand is detrended using a linear trend over the longest sample.

Figure 3 shows the relationship between our priors and the posteriors for this particular case. The distributions of the parameters of the RBC model are given in the first row of the table. These parameters map into the reduced form ARMA parameters (that is, ξ_1, ξ_2 and ω_1) in the second row of the figure.

[11] See (Schorfheide 2000) for a description of the Bayesian estimation methodology. We begin our chain by evaluating the Hessian at the posterior mode using Chris Sims's algorithm (csminwel),

Figure 3
Priors and posteriors



The data shift the prior for ϕ towards a relatively tight posterior, centred on highly correlated technology shocks. The data shift the mass of the prior for σ towards much lower values. These deep parameters generate different reduced-form ARMA parameters compared with the purely calibrated model. The AR(1) parameter ξ_1 is higher; the AR(2) parameter ξ_2 is lower and the MA(1) parameter ω_1 is much higher when we take our priors to the data.

4.4 Maximum likelihood estimation

Our final model uses theory to restrict the ARMA order to a (2,1) representation, but estimates the ARMA parameters by ML. Thus the RBC model is used to inform the choice of ARMA order only.

Note that the ARMA representation implied by the flexible labour model, equation (7), prevents recovery of the deep parameters from the reduced form ARMA representation. However, solving the autoregressive parameters for either ϕ or η_{kk} reduces to a quadratic equation with two unknowns. This implies we can reduce possible solutions for ϕ or η_{kk} to two values. For the linearly detrended private demand measure of output estimated over the full sample, either ϕ=0.735 and $\eta_{kk} = 0.896$; or $\phi = 0.896$ and $\eta_{kk} = 0.735$. Both alternatives appear plausible *a priori*, although these parameters fall outside the range of the posterior distribution of the Bayesian model for the particular case illustrated in the preceding section.

5 Our forecasting experiment

We estimate the ML and Bayesian models recursively and construct out-of-sample forecasts for each quarter from 1965Q1 to 2001Q3. We

available from his website.

[12]We do not consider data revisions and hence use the most recent vintage of data.

compare the baseline ARMA model to our three theoretical models. We collect the sets of pseudo real-time forecasts and compare them to the ex-post series detrended over the entire sample.[13]

For each of the forecast horizons (h), we define the 'loss', ($\varepsilon_{i,t}$), from forecasting model i at time t, to be the model's squared forecast error:

$$\varepsilon_{i,t} = (\hat{y}^h_{i,t+h} - y^h_{t+h})^2. \quad (9)$$

We follow Diebold and Mariano (1995) and compute loss differentials (d_t) between model i and model j as:

$$d_t = \ln(\varepsilon_{i,t}) - \ln(\varepsilon_{j,t}) \quad (10)$$

where d_t can be thought of as the percentage difference in squared error between the two models.

This yields a series of loss differentials $\{d_t\}^T_{t=1}$, where T-8 is the last date at which the out-of-sample forecast comparisons are made. If the mean of this series is negative, model i produces more accurate forecasts than model j. Ideally, it is desirable to report p-values testing the statistical significance of the mean loss differential. If the two models i and j are non-nested, this can readily be accomplished using standard methods (see West 1996, for example).

However, because we use a recursive lag length selection procedure under our null hypothesis, at some dates the two models are nested, implying a non-standard distribution for d_t (Clark and McCracken 2001). Thus far, the distribution theory related to testing mean squared forecast error differences in nested models is limited to special cases only

[13]The forecasts for the variables that are detrended using (log) differences are cumulated over the forecast period: a one-quarter-ahead forecast is a quarterly growth forecast, a two-quarter-ahead forecast is a forecast for growth over two quarters, and so on. These forecasts are then compared to the cumulative growth that occurred in the ex-post data.

(Clark and McCracken 2001 and Clark and West 2005). For this reason, we report mean loss differentials and make no assertions regarding statistical significance.

5.1 Forecast comparisons

Table 3 details the US output forecast comparisons between the baseline statistical ARMA model and the three theoretical models (Calibrated, Bayesian and Maximum Likelihood). The table presents results for each data definition and each detrending procedure.

Focusing initially on the results in the first half of the table, the calibrated model performs particularly poorly and results in marked reductions in forecasting ability relative to the statistical ARMA. The cells in row "*sum*" of the top half of the table show that the forecasting performance of the calibrated model is much worse than the statistical model across all horizons. Of course, this is not surprising: Campbell (1994) aims at exploring the implication of alternative parameter values for consumption and investment behaviour and not at identifying plausible forecasting models.

However, in contrast, the performance of the Bayesian model is surprising. The percentage differences in loss in the table show forecasting performance broadly comparable with the statistical ARMA model. Under the linear trend, the Bayesian model performs relatively poorly at shorter forecast horizons but yields forecasting improvements at longer horizons. This improvement is probably due to increased persistence in the forecasting model driven by the highly autocorrelated technology shocks.[14]

In contrast, when the data are first differenced, the initial five quarters yield improvements in forecasting with the Bayesian model over the statistical ARMA model. At the third and fourth quarter in particular,

[14]Ingram and Whiteman (1994) find persistent technology shocks are important for forecasting within a VAR model augmented with a prior from the King, Plosser, and Rebelo (1988) model.

Table 3
Loss Differential, US Output Definitions: 1965Q1 – 2004Q3

	Narrow Output Measure: Total Private Demand					
	Linear Trend			First Differences		
h	Calibrated	Bayesian	ML	Calibrated	Bayesian	ML
1	30.77	11.58	-5.11	20.32	-6.09	-11.81
2	58.41	50.51	-6.00	27.53	-6.46	-13.11
3	39.04	9.79	6.11	-8.26	-32.39	-4.63
4	36.71	11.02	-13.39	-18.36	-37.38	-11.83
5	19.49	-4.95	-0.05	17.27	-4.37	-9.21
6	10.69	-9.11	-15.72	47.56	27.62	4.50
7	5.75	-18.74	-5.24	60.38	36.59	-1.18
8	12.88	-18.13	-5.93	29.61	44.76	-9.72
sum	17.92	1.61	-2.09	30.57	15.65	-0.49
	Broad Output Measure: Total Real GDP					
	Linear Trend			First Differences		
h	Calibrated	Bayesian	ML	Calibrated	Bayesian	ML
1	28.24	0.70	12.34	34.41	23.82	-2.80
2	14.70	17.12	-0.36	62.82	26.24	14.81
3	3.79	8.23	-5.57	35.58	19.17	-2.20
4	-2.07	-15.48	-5.04	29.77	17.88	8.09
5	8.63	-8.74	-14.88	68.62	39.58	17.92
6	-6.81	-14.68	-10.40	76.15	40.99	27.46
7	-1.40	-13.90	5.59	93.77	24.84	20.34
8	8.08	11.10	-0.24	106.75	49.25	40.80
sum	6.09	-3.81	-0.60	64.63	22.56	10.90

NB. All three models are compared to the baseline statistical ARMA model with ARMA order selected according to the BIC. The row 'sum' gives the mean loss differential cumulating squared errors over all horizons.

mean improvements of 32% and 37% appear non-trivial. However, the performance of the model deteriorates at longer horizons. Still, compared with the statistical ARMA model — explicitly designed to forecast well — the Bayesian model appears competitive under both detrending methods.

Retaining focus on the results for total private demand (in the top half of the table), the ML model, which uses the Campbell's flexible labour model to inform the choice of ARMA order only, performs particularly well. Detrending the data using an ex-post linear trend, the ML model outperforms the statistical model for all but the three quarter horizon. Detrending the data using first differences, only the ML model outperforms the statistical model for all but the six quarter horizon. The weight of evidence suggests restricting the ARMA order to the (2,1) order implied by the flexible labour model helps forecast output.

The results in the bottom half of the table shows some deterioration in relative performance for the ARMA models that use information from the flexible labour model. Using linearly detrended data, both the Calibrated and Bayesian models show some forecasting improvements at longer horizons but these results are mixed. Both models perform badly under the broad data definition using first differenced data with substantial percentage differences in MSEs across all forecast horizons: no quarter yields forecasts superior to the statistical ARMA model.

The ML model again performs relatively well when the broad output measure is linearly detrended. Only the one- and seven-quarter ahead forecasts are worse when the ARMA order is restricted according to the theoretical model. However, when the data is detrended using first differences this result is almost the opposite: only the one- and three-quarter ahead forecasts are superior to the statistical ARMA model.

The deterioration in forecasting performance using a broad output measure is to be expected. Campbell's (1994) flexible labour model is a model of consumption and investment, not the aggregate of output our "broad" measure represents. Quantitatively, thinking about the relevant dataset for theoretical model seems to matter. Richer models that explicitly treat each element of output may improve output forecasts from ARMA models.

6 Conclusion

Our results show that information from a parsimonious RBC model can improve on forecasts from a baseline statistical ARMA model. This confirms RBC theory can improve forecasting when both the theoretical and statistical model contain moving average terms and are not restricted to VAR representations. However, the amount of theory-based information that is used for forecasting has a critical robustness impact.

We show that a naive ARMA model derived from a prior over deep parameters of the flexible labour model forecasts poorly.

Our other two theory-based models perform better. We use the likelihood principle to combine this prior with the likelihood and find that forecasting with reduced-form ARMA parameters implied by the mean of the posterior estimates produces forecasts comparable with that of a standard statistical ARMA model. This improvement in performance appears to be due to increased persistence in technology shocks.

A forecasting model that uses theory to restrict the ARMA order of the forecasting model, but uses maximum likelihood to estimate the reduced form ARMA parameters yields forecasts where the weight of observations are strongly suggestive of improvements over the statistical model.

These results are sensitive to the definition of output the model attempts to forecast. The model is only capable of forecasting a narrow definition of output based on private demand. A richer model that seeks to explain additional components of output may yield similar improvements. Future work could explore the implications of the RBC

model for forecasting consumption and the other variables contained in the Campbell (1994) model.

References

Bai, J, R L Lumsdaine, and J H Stock (1998), "Testing for and dating common breaks in multivariate time series," *The Review of Economic Studies*, 65(3), 395–432.

Banerjee, A, R L Lumsdaine, and J H Stock (1992), "Recursive and sequential tests of the unit-root and trend-break hypothesis: Theory and international evidence," *Review of Economic Studies*, 10(3), 271–87.

Box, G E P and G M Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Campbell, J Y and S Ludvigson (2001), "Elasticities of substitution in real business cycle models with home production," *Journal of Money, Credit and Banking*, 33(4), 847–875.

Canova, F (1998), "Detrending and business cycle facts," *Journal of Monetary Economics*, 31(3), 475–512.

Chari, V V, P J Kehoe, and E R McGrattan (2005), "A critique of structural VARs using real business cycle theory," *mimeo*, (631).

Christiano, L J (1992), "Searching for a break in GNP," *Journal of Business and Economic Statistics*, 10(3).

Christiano, L J, M Eichenbaum, and R Vigfusson (2005), "Assessing structural VARs," *manuscript*.

Clark, T E and M W McCracken (2001), "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, 105(1), 85–110.

Clark, T E and K D West (2005), "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis," *Journal of Econometrics*, forthcoming.

Cooley, T F and E C Prescott (1995), "Economic growth and business cycles," in *Frontiers of Business Cycle Research*, chap 1.

DeJong, D N, B F Ingram, and C H Whiteman (2000), "A Bayesian approach to dynamic macroeconomics," *Journal of Econometrics*, 98(2), 203–23.

Del Negro, M and F Schorfheide (2004), "Priors from general equilibrium models for VARs," *International Economic Review*, 45(2), 643–73.

Diebold, F X and R S Mariano (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3).

Fernández-Villaverde, J, J Rubio-Ramírez, and T Sargent (2005), "A,B,C's (and D)'s for understanding VARs," *mimeo*.

Greenwood, J, Z Hercowitz, and G Huffman (1988), "Investment, capacity utilization, and the real business cycle," *American Economic Review*, 78(3), 402–417.

Hansen, G D (1985), "Indivisible labor and the business cycle," *Journal of Monetary Economics*, 16, 309–327.

Ingram, B F and C H Whiteman (1994), "Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors," *Journal of Monetary Economics*, 34(3), 497–510.

Ireland, P (2004), "A method for taking models to the data," *Journal of Economic Dynamics and Control*, 28(6), 1205–26.

King, R G, C I Plosser, and S T Rebelo (1988), "Production, growth and business cycles: I. The basic neoclassical model," *Journal of Monetary Economics*, 21(2/3), 195–232.

King, R G and S T Rebelo (2000), "Resuscitating real business cycles," *NBER Working Paper 7534*, 7534.

Lettau, M (2003), "Inspecting the mechanism: Closed-form solutions for asset price in real business cycle models," *Economic Journal*, 15, 550–75.

Lubik, T and F Schorfheide (2005), "A Bayesian look at new open economy macroeconomics," *2005 NBER Macroeconomics Annual*, 20.

Ludvigson, S (1996), "The macroeconomic effects of government debt in a stochastic growth model," *Journal of Monetary Economics*, 38, 25–45.

Perron, P (1989), "The great crash, the oil-price shock, and the unit root hypothesis," *Econometrica*, 57(6), 1361–401.

Schorfheide, F (2000), "Loss function-based evaluation of DSGE models," *Journal of Applied Econometrics*, 15(6), 645–70.

Stock, J H and M W Watson (2001), "Vector autoregressions," *Journal of Economic Perspectives*, 15(4), 101–15.

West, K D (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64(5), 1067–8.

Wold, H (1938), *A Study in the Analysis of Stationary Time Series*, Almqvist and Wiksell, Uppsala, Sweden.

Zivot, E and D W K Andrews (1992), "Further evidence on the great crash, the oil-price shock and the unit-root hypothesis," *Journal of Business and Economic Statistics*, 10, 251–70.

Appendix

A Campbell Model

The gross rate of return on a one-period investment in capital R_{t+1} is the marginal product of capital plus the capital stock without depreciation:

$$R_{t+1} \equiv (1-\alpha)\left(\frac{A_{t+1}N_{t+1}}{K_{t+1}}\right)^{\alpha} + (1-\delta) \quad (11)$$

The first order conditions that result from maximising the objective function (4) subject to the constraints (5) and (6) are:

$$C_t^{-1} = \beta E_t\{C_{t+1}^{-1}R_{t+1}\} \quad (12)$$

$$\theta(1-N_t)^{-\gamma} = \frac{W_t}{C_t} = \alpha\frac{A_t^{\alpha}}{C_t}\left(\frac{K_t}{N_t}\right)^{1-\alpha} \quad (13)$$

where the marginal utility of leisure is set equal to the wage W_t times the marginal utility of consumption. Given competitive labour markets, the wage also equals the marginal product of labour from the production function.

In steady-state, technology, capital, output, and consumption all grow at a common constant rate $G \equiv A_{t+1}/A_t$. In steady state, the gross rate of return on capital R_{t+1} is a constant R, so that the first-order condition (12) is:

$$G = \beta R \quad (14)$$

or in logs (denoted by lower-case letters),

$$g = \log(\beta) + r \quad (15)$$

The definition of the return to capital (11) and the first-order condition (12) imply that the technology-capital ratio is a constant:

$$\frac{A}{K} = [G/\beta - (1-\delta)1-\alpha]^{1/\alpha} \approx \left[\frac{r+\delta}{1-\alpha}\right]^{1/\alpha} \quad (16)$$

using $G/\beta = R \approx 1 + r$. The production function and the technology-capital ratio (16) also imply a constant steady-state output-capital ratio:

$$\frac{Y}{K} = \left(\frac{A}{K}\right)^{\alpha} \approx \frac{r+\delta}{1-\alpha} \quad (17)$$

The steady state consumption-output ratio is then a constant given by:

$$\frac{C}{Y} = \frac{C/K}{Y/K} \approx 1 - \frac{(1-\alpha)(g+\delta)}{r+\delta} \quad (18)$$

When the model is outside its steady state, time variation in the consumption-output ratio and incomplete capital depreciation implies the model is a system of non-linear equations, in logs of technology, capital, output, and consumption. Campbell (1994) obtains an approximate analytical solution for the model by transforming the model into a system of approximate log-linear difference equations and taking first-order Taylor series approximations around the steady state:

$$y_t = \alpha(a_t + n_t) + (1-\alpha)k_t \quad (19)$$

$$k_{t+1} \approx \lambda_1 k_t + \lambda_2(a_t + n_t) + (1-\lambda_1-\lambda_2)c_t \quad (20)$$

$$r_{t+1} = \lambda_3(a_{t+1} + n_{t+1} - k_{t+1}) \quad (21)$$

$$E_t\triangle c_{t+1} = \lambda_3 E_t[a_{t+1} + n_{t+1} - k_{t+1}] \quad (22)$$

$$n_t = \nu[(1-\alpha)k_t + \alpha a_t - c_t] \quad (23)$$

To close the model, technology is assumed to follow an AR(1) process:

$$a_t = \phi a_{t-1} + \epsilon_t \tag{24}$$

Campbell solves the model analytically using the method of undermined coefficients, and shows that the dynamic behavior of the economy can be characterised by the following equations.

$$c_t \equiv \eta_{ck} k_t + \eta_{ca} a_t \tag{25}$$

$$n_t \equiv \eta_{nk} k_t + \eta_{na} a_t \tag{26}$$

$$y_t \equiv \eta_{yk} k_t + \eta_{ya} a_t \tag{27}$$

$$k_{t+1} \equiv \eta_{kk} k_t + \eta_{ka} a_t \tag{28}$$

where:

$$\eta_{ck} \equiv \frac{-Q_1 - \sqrt{Q_1^2 - 4Q_0Q_2}}{2Q_2}$$

$$Q_0 \equiv -\lambda_3((1-\alpha)\nu - 1)(\lambda_1 + \lambda_2(1-\alpha)\nu)$$

$$Q_1 \equiv (1+\lambda_3\nu)(\lambda_1 + \lambda_2(1-\alpha)\nu) - \lambda_3((1-\alpha)\nu - 1)(1 - \lambda_1 - \lambda_2(1+\nu)) - 1$$

$$Q_2 \equiv (1+\lambda_3\nu)(1 - \lambda_1 - \lambda_2(1+\nu))$$

and:

$$\eta_{ca} \equiv \frac{(1+\alpha\nu)(\lambda_3\phi - \lambda_2(\eta_{ck}(1+\lambda_3\nu) - \lambda_3((1-\alpha)\nu - 1)))}{(\eta_{ck}(1+\lambda_3\nu) - \lambda_3((1-\alpha)\nu - 1))(1 - \lambda_1 - \lambda_2(1+\nu)) - (1 - \phi(1+\lambda_3\nu))}$$

$$\eta_{nk} \equiv \nu(1 - \alpha - \eta_{ck})$$

$$\eta_{na} \equiv \nu(\alpha - \eta_{ca})$$

$$\eta_{kk} \equiv \lambda_1 + \lambda_2(1-\alpha)\nu + \eta_{ck}[1 - \lambda_1 - \lambda_2(1+\nu)]$$

$$\eta_{ka} \equiv \lambda_2(1+\alpha\nu) + \eta_{ca}[1 - \lambda_1 - \lambda_2(1+\nu)]$$

$$\eta_{yk} \equiv (1-\alpha) + \alpha\nu(1 - \alpha - \eta_{ck})$$

$$\eta_{ya} \equiv \alpha + \alpha\nu(\alpha - \eta_{ca})$$

$$\lambda_1 \equiv \frac{1+r}{1+g}, \quad \lambda_2 \equiv \frac{\alpha(r+\delta)}{(1-\alpha)(1+g)}, \quad \lambda_3 = \frac{\alpha(r+\delta)}{r+1}$$

$$\nu \equiv \frac{(1-N)\sigma}{N + (1-\alpha)(1-N)\sigma}$$