Comparing behavioural and rational expectations for the US post-war economy*

Chunping Liu Nottingham Trent University
Patrick Minford Cardiff University and CEPR

October 4, 2015

Abstract

The banking crisis has caused a resurgence of interest in behavioural models of expectations in macroeconomics. Here we evaluate behavioural and rational expectations econometrically in a New Keynesian framework, using US post-war data and the method of indirect inference. We find that after full reestimation the model with behavioural expectations is strongly rejected by the data, whereas the standard rational expectations version passes the tests by a substantial margin.

Key words: behavioural expectation, rational expectation, bank crisis, indirect inference

1 Introduction

Since the banking crisis of 2007 there has been a resurgence of interest in macroeconomic models embodying expectations-formation other than rational expectations. Evidence of biases in expectations, of herd behaviour and of chart-following has been found by a number of researchers in behavioural economics-for example, Kagel and Roth (1995), McCabe (2003), Camerer et al. (2005) and Della Vigna (2009). Kirman (2011) and De Grauwe (2010) have suggested that such behaviour can be found at the macroeconomic level also (they reject the 'rational learning' models of Sargent (1993) and Evans and Honkapohja (2001), in which for many cases learning converges on rational expectations).

There is also work on behavioral switching models fitting various time series data. The reinforcement learning mechanism with switching between forecasting strategies has proven to be successful in describing individual expectations using both survey data (see, e.g., Branch (2004)) and experimental data (see, e.g., Hommes (2011)). Moreover, recent empirical applications of reinforcement learning models fit data and reproduce stylized facts in the S&P500 market index

*We are grateful to Huw Dixon and Paul de Grauwe for helpful comments.

(see, e.g., Boswijk, et al. (2007)), in the DAX30 index options (see, e.g., Frijns et al. (2010)) and in the Asian equity market (see, e.g., De Jong et al. (2009)). Furthermore, recent empirical papers question the assumption of rational expectations. For example, Rudd and Whelan (2006) estimate a New Keynesian Phillips Curve and they find no evidence in post-war US data that inflation dynamics reflect the rational behavior hypothesized by the standard model. As another example, Carriero (2008) tests the assumption of rational expectations in the setting of a New Keynesian Phillips Curve and he finds no combinations of structural parameters consistent with both the restrictions imposed by the model under rational expectations and US data. However, as pointed it out by ap Gwilym (2010), it is hard to empirically distinguish a behavioural model of stock prices from a rational expectations one.

There is therefore a wide range of work that supports the presence of some type of behavioural expectations in the economy. However, there is no overall test of how far behavioural expectations can account for macroeconomic behaviour in general, as compared with the usual workhorse of DSGE models, rational expectations. Here we focus on this issue, which is clearly of great importance for policymakers.

In this paper, we test a particular model of bounded rationality (that of De Grauwe (2010)), characterized by one specific set of forecasting strategies, within the standard New Keynesian model; in parallel we test the same model with rational expectations. We examine how far these two models can account for US business cycle behaviour over the past few decades including the recent crisis period. Clearly there is a whole spectrum of behavioural expectations assumptions we could have tested instead of the De Grauwe model; whereas rational expectations are tightly defined, behavioural expectations are by definition ad hoc, the point being that people have essentially unexplained biases. We chose the De Grauwe model as our exemplar because De Grauwe has been a well-known, widely-cited and persuasive advocate of the behavioural position in macroeconomics over recent years; clearly, our tests cannot be the end of the story since there is a large if not infinite variety of alternative ways that behavioural expectations could be specified and so tested. It would be well beyond the scope of this one paper to investigate anything approaching this variety; our aim is simply to test a prominent variant to start a debate.

Our (indirect inference) procedure asks whether each model can match US business cycle behaviour, as described by the variances of the three main variables, output, inflation and interest rates, and a VAR embodying their interrelationships. The match is gauged by a Wald statistic that has a well-defined distribution, enabling us to assess the statistical significance of fit. To enable each model to achieve its best possible performance, we allow its model coefficients to be reestimated and only perform the final tests after this has been done.

In using indirect inference to test the two models we deviate from the popular use of Bayesian methods in evaluating models. However, what is not often explained is that such Bayesian evaluation (by marginal likelihood and odds ratio tests) does not test any model as a whole against the data; indeed Bayesians dismiss the idea of 'testing models'. What Bayesian evaluation does is to estimate the model assuming the truth of the prior distributions and the model structure; then one variant of the model may, on those assumptions, turn out to be more probable. But the model in question may still be rejected, assumptions and all, by the data. Furthermore, a model which is 'less probable' under these assumptions than another model, may not be rejected, or may be rejected at a lower confidence level, than the other by the data.

Thus Bayesian methods cannot be used to test models against the data-our aim here. As an alternative to indirect inference for testing models against the data one may use the direct inference likelihood, as in the Likelihood Ratio test. However, as we elaborate below, this alternative method has considerably less power in small samples than the indirect inference test we use here; by implication indirect inference will provide more powerful discrimination between the models.

Bayesians may still argue that it is wrong to do what we do here: that one should not test models as a whole against the data but rather only check improvements conditional on prior assumptions which should not be challenged. However, in macroeconomics it is hard to argue that any set of prior assumptions can be taken for granted as true and beyond challenge. This can be seen from the number of 'schools of thought' still in existence in macroeconomics; this situation of a wide divergence in beliefs has been exacerbated by the financial crisis of the late 2000s. Whether one likes it or not as a macroeconomist one must recognise that to establish a model scientifically to the satisfaction of other economists and policymakers, it needs to be shown that the model being proposed for policy use is consistent with the data in a manner that enables it to be used for that purpose. We show below that indirect inference fulfills that need.

The models we test are identical in form, conforming to a standard New Keynesian model, with a forward-looking IS curve, a Phillips Curve, and a Taylor Rule governing interest rates. The only difference lies in expectations-formation. Thus the comparison precisely tests the different specification of expectations, allowing each model the benefit of reestimation of the exact parameter values. In the standard model these are rational expectations whereas in the alternative ('behavioural') version they are determined by groups of speculators who follow 'fundamentalist' and 'extrapolative' expectations patterns, as set out by De Grauwe (2010). While initially we calibrate these models with typical parameters found in the New Keynesian literature and we report these results in passing, the results we attach importance to are after reestimation (by indirect estimation) to allow each model to get as close as possible to the data, within the bounds set by its theory.

It might well be thought, given the events of recent years, that the standard model would perform badly over the recent post-war period, while the behavioural version would do well. However, we find exactly the opposite: the behavioural version is strongly rejected by the data (including the crisis period), while the standard version is not rejected at the usual significance levels. This apparently surprising result is of some importance to the macroeconomics debate of the current time and so we feel it deserves to be properly exposed to a broad economist audience.

In the rest of this paper, we first explain the models (section 2); we then set out our testing and reestimation procedure (section 3); we turn next to our results, first on calibrated (section 4) and then on reestimated parameters (section 5); section 6 concludes.

2 The Two Models

The behavioural model is a stylized DSGE model similar to the model in De Grauwe (2010). It includes a standard aggregate demand equation, an aggregate supply function, and a policy rule equation, as follows:

$$\tilde{Y}_t = \tilde{E}_t \tilde{Y}_{t+1} - a_1(R_t - \tilde{E}_t \pi_{t+1}) + \varepsilon_{1t} \tag{1}$$

$$\pi_t = b_1 \tilde{Y}_t + \beta \tilde{E}_t \pi_{t+1} + k\varepsilon_{2t} \tag{2}$$

$$R_t = (1 - c_1)(c_2 \pi_t + c_3 \tilde{Y}_t) + c_1 R_{t-1} + u_t \tag{3}$$

where $\tilde{Y}_t$ is the output gap, π_t is the rate of inflation, R_t is the nominal interest rate, and ε_{1t}, ε_{2t}, and u_t are the demand, supply and policy errors respectively. These errors are assumed to be autoregressive processes. They are extracted from the model and the data; thus the model implies the errors, conditional on the data. Equation 1 is the aggregate demand equation with as the expectations operator in the behavioural model where the tilde above $\tilde{E}$ refers to expectations that are not formed rationally. The aggregate demand function is standard, with demand determined by the expectation of the output gap in the next period and by the real interest rate. Equation 2 is the aggregate supply function, derived from profit maximization by individual producers, is the familiar New Keynesian Phillips Curve, a function of the output gap and of expected inflation in the next period. Equation 3 includes a lagged interest rate in Taylor's (1993) original interest rate rule to achieve smoothing of interest rate reactions over time. As all our data is stationary and has the dimension of log deviations from trend, all constants, including the inflation and interest rate targets, are suppressed.

The difference between the behavioural and rational expectations model lies in expectations formation. The expectation term in the behavioural model, $\tilde{E}$ is the weighted average of two kinds of forecasting rule. One is the fundamental forecasting rule, by which agents forecast the output gap or inflation at their steady state values. The other one is the extrapolative rule, by which individuals extrapolate the most recent value into the future. Thus:

$$\tilde{E}_t^f \tilde{Y}_{t+1} = 0 \tag{4}$$

$$\tilde{E}_t^e \tilde{Y}_{t+1} = \tilde{Y}_{t-1} \tag{5}$$

$$\tilde{E}_t^{tar}\pi_{t+1} = \pi^* = 0 \tag{6}$$

$$\tilde{E}_t^{ext}\pi_{t+1} = \pi_{t-1} \tag{7}$$

Equation 4 and 5 are the forecasting rules for the output gap, while Equation 6 and 7 are the equivalents for inflation. The steady state output gap is zero, while the inflation target in the Taylor Rule is the steady state inflation rate, π^* $(= 0)$.

In De Grauwe (2010), it is assumed that the market forecast is the weighted average of the fundamentalist and extrapolative rules. Equation 8 is the market forecast for the output gap, while Equation 9 is the one for inflation.

$$\tilde{E}_t\tilde{Y}_{t+1} = \alpha_{f,t} * 0 + \alpha_{e,t}\tilde{Y}_{t-1} = \alpha_{e,t}\tilde{Y}_{t-1} \tag{8}$$

$$\tilde{E}_t\pi_{t+1} = \beta_{tar,t} * 0 + \beta_{ext,t}\pi_{t-1} = \beta_{ext,t}\pi_{t-1} \tag{9}$$

where $\alpha_{f,t}$ and $\alpha_{e,t}$ are the probabilities that agents will use a fundamentalist and extrapolative rule for forecasting the output gap, while $\beta_{tar,t}$ and $\beta_{ext,t}$ are the equivalents for inflation. These probabilities sum to one:

$$\alpha_{f,t} + \alpha_{e,t} = 1 \tag{10}$$

$$\beta_{tar,t} + \beta_{ext,t} = 1 \tag{11}$$

These probabilities are defined according to discrete choice theory (see Anderson, de Palma, and Thisse 1992 and Brock and Hommes 1997). Agents' utilities are given by the negative of the forecast performance (measured by the squared forecast error) of the different rules as follows:

$$U_{f,t} = -\sum_{k=1}^{\infty}\omega_k(Y_{t-k} - \tilde{E}_{t-k-1}^{f}\tilde{Y}_{t-k})^2 \tag{12}$$

$$U_{e,t} = -\sum_{k=1}^{\infty}\omega_k(Y_{t-k} - \tilde{E}_{t-k-1}^{e}\tilde{Y}_{t-k})^2 \tag{13}$$

$$U_{tar,t} = -\sum_{k=1}^{\infty}\omega_k(\pi_{t-k} - \tilde{E}_{t-k-1}^{tar}\pi_{t-k})^2 \tag{14}$$

$$U_{ext,t} = -\sum_{k=1}^{\infty}\omega_k(\pi_{t-k} - \tilde{E}_{t-k-1}^{ext}\pi_{t-k})^2 \tag{15}$$

where $U_{f,t}$ and $U_{e,t}$ are the utilities for the output gap of the fundamentalists and extrapolators, respectively; while $U_{tar,t}$ and $U_{ext,t}$ are the equivalents for inflation; ω_k are geometrically declining weights, defined as

$$\omega_k = (1-\rho)\rho^k \tag{16}$$

where ρ, between zero and one, is the memory coefficient.

The probabilities of the fundamentalist and extrapolator in forecasting output are given by the relative utility of their forecasts:

$$\alpha_{f,t} = \frac{\exp(\gamma U_{f,t})}{\exp(\gamma U_{f,t}) + \exp(\gamma U_{e,t})} \tag{17}$$

$$\alpha_{e,t} = \frac{\exp(\gamma U_{e,t})}{\exp(\gamma U_{f,t}) + \exp(\gamma U_{e,t})} \tag{18}$$

while the probabilities of the inflation targeting rule and extrapolative rule are

$$\beta_{tar,t} = \frac{\exp(\gamma U_{tar,t})}{\exp(\gamma U_{tar,t}) + \exp(\gamma U_{ext,t})} \tag{19}$$

$$\beta_{ext,t} = \frac{\exp(\gamma U_{ext,t})}{\exp(\gamma U_{tar,t}) + \exp(\gamma U_{ext,t})} \tag{20}$$

where γ is defined as the 'intensity of choice', assumed to be one in De Grauwe (2010); it measures the degree to which the deterministic component of utility determines actual choice.

Equation 17-18 show that the probability of fundamentalists increases as the forecast performance of the fundamental rule improves relative to the extrapolative rule; and similarly with inflation, as shown by Equation 19-20, where we can interpret the weight on the target in the inflation forecasting rule as a measure of the central bank's credibility in inflation targeting. These mechanisms driving the selection of the rules introduce a dynamic element to the model, rather like adaptive expectations in the old Neo-Keynesian Synthesis models.

Dealing with the infinite sum in Equation 12 to 15, we can transform them into recursive representations of the sum, so that the model can be solved. Then Equation 12-15 can be transformed by the following:

$$U_{f,t} = -(1-\rho)\rho(Y_{t-1})^2 - \rho U_{f,t-1} \tag{21}$$

$$U_{e,t} = -(1-\rho)\rho(Y_{t-1} - Y_{t-3})^2 - \rho U_{e,t-1} \tag{22}$$

$$U_{tar,t} = -(1-\rho)\rho {\pi_{t-1}}^2 - \rho U_{tar,t-1} \tag{23}$$

$$U_{ext,t} = -(1-\rho)\rho(\pi_{t-1} - \pi_{t-3})^2 - \rho U_{ext,t-1} \tag{24}$$

The solution to the behavioural model is obtained by substituting the expectation formation of Equation 8 and 9 into Equation 1 and 2, so that the model becomes

$$\tilde{Y}_t = \alpha_{e,t} Y_{t-1} - a_1(R_t - \beta_{ext,t}\pi_{t-1}) + \varepsilon_{1t} \quad (25)$$

$$\pi_t = b_1 \tilde{Y}_t + \beta(\beta_{ext,t}\pi_{t-1}) + k\varepsilon_{2t} \quad (26)$$

$$R_t = (1 - c_1)(c_2\pi_t + c_3\tilde{Y}_t) + c_1 R_{t-1} + u_t \quad (27)$$

with the definition for the probabilities in Equation 12-20. This model is a pure backward model, which can be solved in an overlapping sequence for each set of innovations.

The stylized DSGE model with rational expectation is defined as Equation 1-3 except that now the expectations are formed rationally. The only specification difference between the two models is in the nature of these expectations. Thus the comparison precisely tests the different specification of expectations, allowing each model the benefit of reestimation of the exact parameter values. This rational expectation version of the model can be solved in the standard way; we use Dynare (Juillard 2001) for this.

It should be noted that the errors in each model are determined by some wider model which will almost certainly generate autocorrelation. Thus by construction the errors will typically be autocorrelated. These errors are extracted from the model and the data in each case; thus the model implies the errors, conditional on the data. The overall model dynamics are a compound of the model's structure, including forward expectations and the persistence mechanism of the smoothed Taylor Rule and the error dynamics; this is true also of the behavioural expectations model for which the behavioural expectations produce an additional persistence mechanism.

3 The Testing Procedure

Indirect Inference provides a framework for judging whether a model with a particular set of parameters could have generated the behaviour found in a set of data. The procedure provides a statistical criterion for rejecting the model as the data generating mechanism.

Indirect inference has been well known in the estimation literature, since being introduced by Smith (1993); see also Gregory and Smith (1991, 1993), Gourieroux et al. (1993), Gourieroux and Montfort (1995) and Canova (2005). In indirect estimation the behaviour of the data is first described by some atheoretical time-series model such as a Vector Auto Regression, the 'auxiliary model'; then the parameters of the structural model are chosen so that this model when simulated generates estimates of the auxiliary model as close as possible to those obtained from actual data. In what follows we give a brief account of the method; a full account, together with Monte Carlo experiments checking its accuracy and power and comparing it with other methods in use for evaluating DSGE models, can be found in Le et al. (2011, 2012).

The test is based on the comparison of the auxiliary model as estimated on the actual data with that estimated on the data simulated from the structural model. We choose a VAR as our auxiliary model and base our tests on the VAR coefficients and also the variances (of the variables in the VAR). The reason for choosing a VAR as the auxiliary model is that a DSGE model like the ones here have as their solution a restricted vector autoregressive-moving-average (VARMA), which can be closely represented by a VAR. The VAR captures the dynamic inter-relationships found in the data between the variables of the model. The test statistic is based on the joint distribution of the chosen descriptors-here the VAR coefficients and the variances. The null hypothesis is that the macroeconomic model is the data generating mechanism.

The test statistic for this joint distribution is a Wald statistic. Following the notation of Canova (2005), y_t is defined as an $m \times 1$ vector of observed data $(t = 1, ..., T)$ and $x_t(\theta)$ is an $m \times 1$ vector of simulated data with S observations from the model, θ is a $k \times 1$ vector of structural parameters from the model. We set $S = T$, because we want to compare simulated data and actual data using the same size of sample. y_t and $x_t(\theta)$ are assumed to be stationary and ergodic. The auxiliary model is $f[y_t, \alpha]$, where α is the vector of descriptors. Under the null hypothesis $H_0 : \theta = \theta_0$, the auxiliary model is then $f[x_t(\theta_0), (\theta_0)] = f[y_t, \alpha]$. The null hypothesis is tested through the $q \times 1$ vector of continuous functions $g(\alpha)$. Under the null hypothesis, $g(\alpha) = g(\alpha(\theta_0))$. a_T is defined as the estimator of α using actual data and $\alpha_S(\theta_0)$ as the estimator of α based on simulated data for θ_0. Then we have $g(a_T)$ and $g(\alpha_S(\theta_0))$.The simulated data is obtained by bootstrapping N times of structural errors, so there are N sets of simulated data. We can calculate the bootstrapped mean by $\overline{g(\alpha_S(\theta_0))} = \frac{1}{N}\sum_{k=1}^{N} g_k(\alpha_S(\theta_0))$. The Wald statistic ($WS$) using the bootstrapped distribution of $g(a_S) - \overline{g(\alpha_S(\theta_0))}$ can be specified as

$$WS = (g(a_T) - \overline{g(\alpha_S(\theta_0))})'W^{-1}(\theta_0)(g(a_T) - \overline{g(\alpha_S(\theta_0))})$$

where $W(\theta_0)$ is the variance-covariance matrix of the bootstrapped distribution of $g(a_S) - \overline{g(\alpha_S(\theta_0))}$. Here we use a, the descriptors themselves, as $g(a)$.

The testing procedure involves three steps. The first step is to back out the structural errors from the observed data and parameters of the model. If the model equations have no future expectations, the structural errors can be simply calculated using the actual data and structural parameters. If there are expectations in the model equations, we calculate the rational expectation terms using the robust instrumental variables methods of McCallum (1976) and Wickens (1982); we use the lagged endogenous data as instruments and hence use the auxiliary VAR model as the instrumental variables regression. The errors are treated as autoregressive processes; their autoregressive coefficients and innovations are estimated by OLS. [1]

[1] The idea of using these backed-out errors is that they should be consistent with the model

Secondly, these innovations are then bootstrapped and the model is solved by Dynare. The innovations are repeatedly drawn by time vector to preserve any contemporaneous correlations between them. By this method we obtain N (usually set at 1000) sets of simulated data, or bootstrap samples. These represent the sampling variation of the data implied by the structural model.

Finally, we compute the Wald statistic. By estimating the VAR on each bootstrap sample, the distribution of the VAR coefficients and data variances is obtained, the α_S. Thus, the estimates of α from the data and the model estimates can be compared. We examine separately the model's ability to encompass the dynamics (the VAR coefficients) and the volatility (the variances) of the data. 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, as an overall measure of closeness between the model and the data.[2]

We use a VAR(1) as the auxiliary model. With a VAR(1), α contains 12 elements, the 9 VAR coefficients and the 3 data variances. This number of descriptors provides a strong requirement for the structural model to match. Raising the VAR order would increase the number of VAR coefficients (eg with a VAR(2) the number would double to 18, making 21 elements in α in total); the requirement of the test arguably becomes excessive, since we do not expect our structural models to replicate data dynamics at such a high level of refinement.

The steps above detail how a given model, with particular parameter values, is tested. These values would typically be obtained in the first place by calibration. However, the power of the test is high and the model will be rejected if the numerical values chosen for the parameters are inaccurate. Therefore, to test a model fully one needs to examine its performance for all (theoretically permissible) values of these parameters. This is where we introduce Indirect Estimation; in this we search for the numerical parameter values that minimise the Wald statistic and then test the model on these values. If it is rejected on these, then the model itself is rejected, as opposed merely to its calibrated parameter values. We discuss details of this further below.

---

and the data: otherwise the model being tested could be considered rejected by the data at the structural stage. As noted by Le et al (2012), an alternative way to estimate the errors in equations with rational expectations terms is to use the model (including the lagged errors) to generate the expectations and iterate to convergence but in Monte Carlo experiments the LIML method is slightly more accurate (if we knew the true model including the true ρs, then we could back out the exact errors by using the model to solve for the expectations; but of course we do not).

Once the errors and their autoregressive coefficients (ρ) are estimated, they become part of θ_0 and are fixed for the testing process therefore. In indirect estimation the search algorithm finds the structural parameters, the backed-out errors and the ρs that jointly get closest to the α found in the data. If they are also not rejected by these α, then we may treat this model as the data generating mechanism.

[2] 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.

3.1 Comparison of Indirect Inference with Other Testing Methods

It may be asked why we use Indirect Inference rather than the now widely-used Bayesian approach to estimating and testing DSGE models. We considered this frequently-asked question in the opening section above. As noted there, the Bayesian approach does not test a model as a whole against the data since it operates entirely under the assumptions that the prior distributions and the model structure are correct. It would be appropriate to use this approach if we knew that this was the case so that the assumptions themselves do not need to be tested; but because of well-known controversies in macroeconomics, policymakers and others using these models need to be assured that the model overall is consistent with the data and is not imposed on the data. Even a major model like the Smets-Wouters (2007) model of the U.S., that has been carefully estimated by Bayesian methods, is rejected by our indirect inference test, see Le et al. (2011).

It may then be asked why we use Indirect Inference rather than other available tests of overall specification, given as explained above that normal Bayesian methods are simply not available for the task here[3]. The alternatives are based on the likelihood of the estimated model, and a widely-used one is the Likelihood Ratio test (LR). This test is examined carefully in Le et al. (2012) who find that LR is much less powerful in small samples as a test of specification than a Wald test based on indirect inference. This is presumably related to the nature of the two tests: the LR test is based on a model's in-sample current forecasting ability whereas the Wald is based on the model's ability to replicate data behaviour, as found in the VAR coefficients and the data variances, reflecting the causal processes at work in the data. Models that are somewhat mis-specified may still be able to forecast well in sample as the error processes will pick up the effects of mis-specification but mis-specified models will imply a reduced form that differs materially from the true one, and so therefore a VAR approximation to this that similarly deviates from the VAR given by the true model.

The Table below reproduces the findings by Le et al. (2012) comparing the two tests (on the Smets-Wouters model, for a 3-variable VAR(1)) as the degree of mis-specification rises.

In sum, we could use LR instead of indirect inference as a test of our two models. But it would be a much weaker test and hence we would get much less discrimination between the models. As will be seen below, the indirect inference Wald test discriminates powerfully between the two models, strongly rejecting one model while not rejecting the other.

[3] If strong priors are not appropriate, then Bayesian methods could still be used but with flat priors. These amount to FIML estimation so that then ordinary likelihood tests such as the likelihood ratio, discussed next, can be applied.

Percent Mis-specified	Wald	LR
True	5.0	5.0
1	19.8	6.3
3	52.1	8.8
5	87.3	13.1
7	99.4	21.6
10	100.0	53.4
15	100.0	99.3
20	100.0	99.7

Table 1: Rejection Rates (for 3 Variable VAR(1). Source: Let et al. (2012))

4 Data, Calibration and Calibrated Results

4.1 Data

We apply the models to quarterly US data from 1981Q4 to 2013Q4 on the output gap ($\tilde{Y}_t$), the inflation rate (π_t), and the interest rate (R_t) [4], collected from Federal Reserve Bank of St. Louis. The data include the recent financial crisis up to the most recent observations. The output gap ($\tilde{Y}$) is defined by the percentage gap between real GDP and potential GDP, for which we use the HP filter so that this variable is stationary by construction. Inflation (π) is defined as the quarterly change in the log of the CPI. The interest rate is the federal funds rate, expressed as a fraction per quarter.

We firstly test the trend-stationarity of inflation and interest rates. We can see from the Table 3[5] that both have significant deterministic trends. The sample concerned in this paper includes a few turbulent periods, such as the early 1980s recession, the early 1990s recession, the dot-com bubble, the September 11th attacks and the recent global financial crisis. Therefore we test trend-stationarity using unit root tests, after taking into account possible structural breaks in the deterministic trend. The Chow test is used for testing known break points, while the Bai and Perron test is for unknown multiple break points. In the following we check for breaks due to the September 11th attacks (2001Q3 and 2001Q4) and the recent global financial crisis (2007Q1-2009Q1) using the Chow test based on the unit root testing equations. It can be seen in the table 2 that all the F-statistics are insignificant at 5% significance level (the critical value is 3.9), which implies that no structural break is detected in the two crises. However, it is difficult to identify the possible break points for the early crisis periods; instead we can test these unknown structural breaks by Bai and Perron tests. Table 2 shows that the F-statistics are lower than the 5% critical value of 8.58. This indicates that there is no structural break both in inflation and interest rate across the whole sample, which confirms the Chow test results.

[4] To calculate the lagged variables $U_{f,t}$, $U_{e,t}$, $U_{tar,t}$, $U_{ext,t}$, we go back to 1970Q2.

[5] In this paper, * represents 5% significance level. ** denotes 1% significance level.

Test	Event		F-statistics		Conclusion	
			π	R	π	R
Chow	Sep11th attack	2001Q3	0.051575	0.501155	no	no
		2001Q4	0.202721	0.434604	no	no
	financial crisis 2007-2008	2007Q1	0.067102	1.097126	no	no
		2007Q2	0.012518	1.622606	no	no
		2007Q3	0.152927	2.242481	no	no
		2007Q4	0.130049	2.679775	no	no
		2008Q1	0.561341	2.518808	no	no
		2008Q2	0.961695	1.382308	no	no
		2008Q3	1.883081	1.086589	no	no
		2008Q4	3.622815	1.812170	no	no
		2009Q1	0.000897	0.291587	no	no
Bai-Perron	-		5.703069	2.781012	no	no

Table 2: Structural Breaks Test

Since no structural breaks are detected in the linear trends, the unit root tests remain valid. Table 3 gives the stationarity property for each variable using the augmented Dickey–Fuller (ADF) and the Phillips-Perron (PP) tests, both of which strongly reject the unit root for both inflation and interest rates after allowing for a linear trend. Therefore, π_t and R_t are simply linearly detrended. Figure 1 displays the resulting data. It is notably volatile in the early 1980s, a turbulent period. With inflation in double digits, Paul Volcker was appointed as Fed chairman in 1979 to bring it under control. With the resulting policies, which included spells of both monetary base and credit controls, interest rate volatility reached a peak, not exceeded even in the recent bank crisis.

Variable	Coeff on the Trend	p-value of ADF	p-value of PP	Implication
π	-4.44E-05**	0.0000	0.0000	trend-stationary
R	-0.000186**	0.0027	0.0017	trend-stationary

Table 3: Stationarity Test

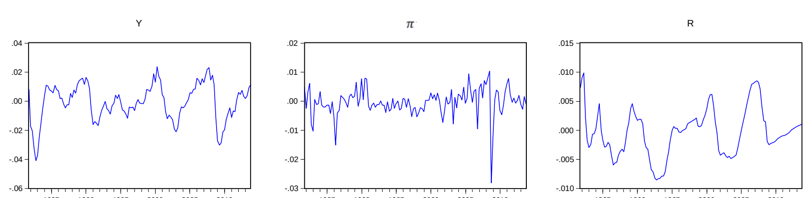

Figure 1: Time Paths of $\tilde{Y}$, π, R

4.2 Calibration

Table 4 shows the calibrated parameters used for the two models. The first part of the table shows the parameters that are common to both models, following Minford and Ou (2010); the second part shows the parameters individual to each model- the values of γ and ρ follow De Grauwe (2010). The structural errors backed out from model and data all are autoregressive; their AR(1) parameters are shown as $\rho_i(i = 1, \text{demand}; 2, \text{supply}; 3, \text{policy})$.

BF/RE	Parameters	Definitions	Values
BF/RE	a_1	real interest rate elasticity on output gap	0.50
	b_1	coefficient of output gap on inflation	2.36
	π^*	inflation target	0
	β	discount factor	0.99
	κ	coefficient of supply shock on inflation	0.42
	c_1	interest rate persistence parameter	0.8
	c_2	policy response to inflation	2.0
	c_3	policy response to output gap	0.1
BF	γ	intensity of choice parameter	1
	ρ	memory parameter	0.5
	ρ_1	autoregressive coefficient for demand error	0.69
	ρ_2	autoregressive coefficient for supply error	0.84
	ρ_3	autoregressive coefficient for policy error	0.18
RE	ρ_1	autoregressive coefficient for demand error	0.89
	ρ_2	autoregressive coefficient for supply error	0.86
	ρ_3	autoregressive coefficient for policy error	0.18

Table 4: Calibration of Behavioural and Rational Expectation Model

4.3 Test Results Based on Calibration

Our auxiliary model is the VAR(1), Equation 28,

$$\begin{bmatrix} \tilde{Y}_t \\ \pi_t \\ R_t \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \begin{bmatrix} \tilde{Y}_{t-1} \\ \pi_{t-1} \\ R_{t-1} \end{bmatrix} + \Omega_t \tag{28}$$

The VAR's nine coefficients represent the dynamic properties found in the data. We also look at the volatility properties as indicated by the variances. We consider these two properties both separately and together, calculating Wald statistics for each. We show these as the percentile where the data Wald lies in the Wald bootstrap distribution.

4.3.1 Behavioural Model

Table 5 shows the VAR estimates on the actual data and also the 95% bounds of the VAR estimates from the 1,000 bootstrap samples. It shows that five out of nine parameters lie outside the 95% bootstrapped bounds. They are the coefficients of the lagged interest rate on output, and of lagged inflation and lagged interest rates on inflation and interest rates. It is not surprising therefore that overall the model is strongly rejected by the dynamic properties of the data.

Categories	**Actual VAR Coefficients**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
β_{11}	0.9289	0.7635	0.9399	IN
β_{12}	0.0178	-0.1076	0.0320	IN
β_{13}	-0.2486	-0.1979	-0.0186	OUT
β_{21}	0.0372	-0.2033	0.3249	IN
β_{22}	0.1726	0.9653	1.1805	OUT
β_{23}	0.1419	-0.7550	-0.3694	OUT
β_{31}	0.0337	-0.0666	0.1483	IN
β_{32}	-0.0045	0.3747	0.4722	OUT
β_{33}	0.8776	0.4997	0.6630	OUT
Wald (Dynamics)	100%			

Table 5: Dynamic Properties of Behavioural Model Based on Calibration

Table 6 shows the volatility properties of the data and the behavioural model. The table shows that only the output variance can be captured by the model. The variances of inflation and interest rate in the data are far below the range of the 95% model bounds. Jointly the model -generated bounds on the variances are closer to the data, with the Wald percentile at 97.7%, still indicating rejection at 95%; this can be reconciled with the rejections of the two variances on their own by noting that the variance values generated by the model will be highly correlated; hence the lower 95% bound of the joint distribution will lie well below the individual 95% bounds of inflation and interest rates.

Nevertheless, when one combines the dynamic and volatility properties, the behavioural model is strongly rejected, with an overall Wald of 100%.

4.4 The Rational Expectations Model

Table 7 shows the test findings for the RE model. On its dynamic properities the model is not rejected, with a Wald of 92.6%. It is therefore fairly close to the data; individually, only one out of nine parameters lies outside the 95% bootstrapped bounds- the coefficient of the lagged interest rate on output.

Turning to the volatility properties, Table 8 shows that the model is not rejected by the data, with a Wald at 82.4%; individually, all the three variances lie well inside their 95% bounds.

Categories	**Actual Variances**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
$\mathrm{var}(\tilde{Y})$	0.1643	0.0787	0.2526	IN
$\mathrm{var}(\pi)$	0.0225	0.2296	0.8123	OUT
$\mathrm{var}(R)$	0.0166	0.1572	0.5729	OUT
Wald (Volatility)	97.7%			
Overall Wald	100%			

Table 6: Volatility and Full Properties of Behavioural Model Based on Calibration

Categories	**Actual VAR Coefficients**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
β_{11}	0.9289	0.7397	0.9296	IN
β_{12}	0.0178	-0.4235	0.1171	IN
β_{13}	-0.2486	-0.2237	0.3029	OUT
β_{21}	0.0372	-0.0777	0.0514	IN
β_{22}	0.1726	0.1372	0.4773	IN
β_{23}	0.1419	-0.0311	0.3035	IN
β_{31}	0.0337	-0.0268	0.0405	IN
β_{32}	-0.0045	-0.0339	0.1536	IN
β_{33}	0.8776	0.7960	0.9524	IN
Wald (Dynamics)	92.6%			

Table 7: Dynamic Properties of Rational Expectation Model Based on Calibration

When one combines the dynamics and volatility, Table 8 shows that the model is not rejected, with an overall Wald percentile of 95.0%.

Categories	**Actual Variances**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
var($\tilde{Y}$)	0.1643	0.0633	0.2471	IN
var(π)	0.0225	0.0122	0.0241	IN
var(R)	0.0166	0.0101	0.0411	IN
Wald (Volatility)	82.4%			
Overall Wald	95.0%			

Table 8: Volatility and Full Properties of Rational Expectation Model Based on Calibration

We bring all these results together in Table 9. It can be seen that, if we use our calibrated parameter values, only the rational expectations model fails to be rejected overall by the behaviour found in the data. However, it could be that this conclusion depends critically on the parameter values chosen and that the calibrated ones give a misleading impression. We accordingly now turn to the reestimation of these parameters.

Wald	**BF Model**	**RE Model**
Dynamics	100%	92.6%
Volatility	97.7%	82.4%
Overall	100%	95.0%

Table 9: Comparison of Behavioural and Rational Expectation Model Using Calibration

5 Indirect Inference Estimation

The main idea of indirect inference as an evaluation method is to see if the chosen parameter set θ_0 could have generated the actual data. However, if it cannot do so, another set of parameters could possibly have done so. If no set of parameters can be found under which the model fails to be rejected, then the model itself is rejected. Models that are already unrejected may also get closer to the data with alternative parameters. We now use indirect estimation on each model to obtain the set of parameters that maximises the chances of the model passing the test- in other words minimises the overall Wald statistic. For this purpose we use a powerful algorithm due to Ingber (1996) 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.

Table 10 and 11 show the estimation results for behavioural and rational expectation models respectively. For both models, apart from β (time preference) which is held fixed on a priori grounds, all the parameters are allowed to vary as required by each model. For the behavioural model, the estimated parameters are in Table 10. Most of the IS, Phillips Curve and Taylor Rule parameters need to vary generally by more than 40%, which implies that the original calibrated values were substantially at variance with the data's requirements. The parameters of expectation formation, γ, vary little however, suggesting that the problem lies with the expectations scheme itself and not with its parameter values. However, the memory coefficient, ρ,varies 41%, representing long memory of the representative agent. The autoregressive coefficients of the errors vary little, implying that the parameter changes largely offset each other in their effects on the left-hand-side variable in each equation; nevertheless the changes by affecting the model's transmission processes can change its implied behaviour substantially.

Table 11 shows the equivalent results for the rational expectations model. Similarly, most of the parameters change more than 30%.The effect of the output gap on inflation and the policy reaction to inflation and output gap, all of which increase sharply over the calibrated values. The autoregressive coefficients of the errors still vary little.

Parameters	**Estimates**	**Calibration**	**Variation**
a_1	0.7498	0.50	50%
b_1	2.8794	2.36	22%
k	0.6134	0.42	46%
c_1	0.4001	0.8	50%
c_2	2.6691	2.0	33%
c_3	0.0537	0.1	46%
γ	0.8525	1	15%
ρ	0.7056	0.5	41%
ρ_1	0.7003	0.69	1%
ρ_2	0.8674	0.84	3%
ρ_3	0.1903	0.18	6%

Table 10: Estimation of Behavioural Model

5.1 Testing Comparison Based on Estimated Parameters

Table 12 and Table 15 show how the test results on these estimated parameters. The behavioural model is still strongly rejected, with seven out of twelve parameters still outside the 95% bounds, and while it remains relatively close to the data's volatility it is rejected decisively on the dynamics as well as in total, with an overall Wald of 100%.

Though it is still strongly rejected overall, the behavioural model is now closer to the data. We can see this from the transformed Mahalanobis distance

Parameters	**Values**	**Calibration**	**Variation**
a_1	0.6486	0.50	30%
b_1	3.4941	2.36	48%
k	0.2437	0.42	42%
c_1	0.7363	0.8	8%
c_2	2.8025	2.0	40%
c_3	0.0593	0.1	41%
ρ_1	0.8846	0.89	1%
ρ_2	0.8810	0.86	2%
ρ_3	0.1834	0.18	2%

Table 11: Estimation of Rational Expectation Model

Categories	**Actual VAR Coefficients**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
β_{11}	0.9289	0.7423	0.9351	IN
β_{12}	0.0178	-0.4582	0.0264	IN
β_{13}	-0.2486	-0.1084	0.1932	OUT
β_{21}	0.0372	-0.0894	0.1143	IN
β_{22}	0.1726	0.4707	0.8462	OUT
β_{23}	0.1419	-0.2093	0.0605	OUT
β_{31}	0.0337	-0.0825	0.1760	IN
β_{32}	-0.0045	0.5385	1.0576	OUT
β_{33}	0.8776	0.2179	0.5726	OUT
$\text{var}(\tilde{Y})$	0.1643	0.0608	0.2483	IN
$\text{var}(\pi)$	0.0225	0.0229	0.0805	OUT
$\text{var}(R)$	0.0166	0.0554	0.1988	OUT

Table 12: Testing Details of Behavioural Model Using Estimated Parameters

Wald Percentiles	**Calibration**	**Estimation**
Dynamics	100%	100%
Volatility	97.7%	97.0%
Overall	100%	100%

Table 13: Comparison of Behavioural Expectation Model results under Calibration and Estimation

(TM) described above, which is a convenient transformation of the Wald statistic: it is a normalised t-statistic taking the value 1.645 at the 95% Wald percentile. Table 14 shows that the TM for the behavioural model improves materially after estimation.

Tsfmd Mahalanobis	**Calibration**	**Estimation**
Dynamics	34.81	6.10
Volatility	2.35	2.10
Overall	33.04	6.78

Table 14: Comparison TM of Behavioural and Rational Expectation Model Using Estimated Parameters

Table 15 shows that the rational expectations model improves to considerable closeness to the data behaviour. Only one individual parameters is now outside its 95% bounds and overall the model would not be rejected at 81.5% confidence (see Table 16).

Categories	**Actual VAR Coefficients**	**95% Lower Bound**	**95% Upper Bound**	**IN /OUT**
β_{11}	0.9289	0.7390	0.9358	IN
β_{12}	0.0178	-0.4102	0.1442	IN
β_{13}	-0.2486	-0.2241	0.2931	OUT
β_{21}	0.0372	-0.0692	0.0550	IN
β_{22}	0.1726	0.0494	0.3805	IN
β_{23}	0.1419	-0.0145	0.3244	IN
β_{31}	0.0337	-0.0305	0.0351	IN
β_{32}	-0.0045	-0.0569	0.1234	IN
β_{33}	0.8776	0.7967	0.9554	IN
$\mathrm{var}(\tilde{Y})$	0.1643	0.0647	0.2667	IN
$\mathrm{var}(\pi)$	0.0225	0.0092	0.0238	IN
$\mathrm{var}(R)$	0.0166	0.0092	0.0359	IN

Table 15: Testing Details of Rational Expectation Model Using Estimated Parameters

Wald Percentiles	**Calibration**	**Estimation**
Dynamics	92.6%	81.2%
Volatility	82.4%	76.9%
Overall	95.0%	81.5%

Table 16: Comparison of Rational Expectation Model under Calibrated and Estimated Parameters

In sum, we can see that while the behavioural model remains rejected overall, the rational expectations model has after estimation lowered the threshold at which it would not be rejected to 81.5%. It would seem that behavioural expectations are clearly rejected in favour of rational expectations in the context of a standard macroeconomic model.

6 Conclusion

This paper investigates whether behavioural expectations can improve on rational expectations in our understanding of recent macroeconomic behaviour. The banking crisis impelled many economists and commentators to question the standard New Keynesian model with rational expectations; one suggested improvement was that expectations could be formed in a behavioural manner. We have found in our work here that in fact this would be no improvement; indeed the standard model fits the behaviour found in the data, including the crisis period, rather well while the behavioural model is decisively rejected. Plainly our study is not a test of behavioural expectations in general. It is limited to macroeconomic data for the US economy, as measured by three key variables; it does not for example look at asset prices. It is also limited to a particular behavioural model, that of De Grauwe. Nevertheless macro behaviour is of importance to policymakers and the De Grauwe set-up is reasonably typical, so that our results have fairly wide relevance.

This is not to say that the standard model cannot be enriched in some way to improve our understanding of the events surrounding the crisis. In particular, our work makes no attempt to assess the shift in the economy's trend behaviour, as we abstract from trends in the usual way- others argue (eg Le, Meenagh and Minford (2012)), that shifts in trend were an important determinant of the US crisis. Nor does it attempt to model the behaviour of banks and how this was related to the economy in the crisis. Plainly these topics are important ones to investigate. However, our work here suggests that behavioural expectations are not a promising route to account for the banking crisis.

References

[1] Anderson, S., De Palma, A., and Thisse, J.F., 1992. Discrete Choice Theory of Product Differentiation. MIT Press, Cambridge, Mass.

[2] ap Gwilym, R., 2010. Can behavioral finance models account for historical asset prices?, Economics Letters 108(2), 187-189.

[3] Boswijk, H., Hommes, C., and Manzan, C., 2007. Behavioral Heterogeneity in Stock Prices. Journal of Economic Dynamics and Control, 31, 1938–1970.

[4] Branch, W., 2004. The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations. The Economic Journal, 114(497), 592–621.

[5] Brock, W., and Hommes, C., 1997. A Rational Route to Randomness, Econometrica 65, 1059-1095.

[6] Camerer, C., Loewenstein, G., and Prelec, D., 2005. Neuroeconomics: how neuroscience can inform economics, Journal of Economic Literature 63(1), 9-64.

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

[8] Carriero, A., 2008. A simple test of the New Keynesian Phillips Curve. Economics Letters, 100, 241–244.

[9] De Grauwe, P., 2010. Top-down versus bottom-up macroeconomics, CESifo Economic Studies 56(4), 465-497.

[10] De Jong, E., Verschoor, W., and Zwinkels, R., 2009. Behavioral Heterogeneity and Shift-Contagion: Evidence from the Asia Crisis. Journal of Economic Dynamics and Control, 33(11), 1929–1944.

[11] Della Vigna, S., 2009. Psychology and Economics: Evidence from the Field, Journal of Economic Literature 47(2), 315–372.

[12] Evans, G., and Honkapohja, S., 2001. Learning and expectations in macroeconomics. Princeton University Press.

[13] Frijns, B., Lehnert, T., and Zwinkels, R., 2010. Behavioral Heterogeneity in the Option Market. Journal of Economic Dynamics and Control, 34, 2273–2287.

[14] Gourieroux, C., Monfort, A., and Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 85-118.

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

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

[17] Gregory, A., and Smith, G., 1993. Calibration in macroeconomics, in: Maddala, G. (Ed.), Handbook of Statistics,11, Elsevier, St. Louis, Mo., 703-719.

[18] Hommes, C. H., 2011. The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab. Journal of Economic Dynamics and Control, 35, 1–24.

[19] 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.

[20] Juillard, M., 2001. DYNARE: a program for the simulation of rational expectations models. Computing in economics and finance 213. Society for Computational Economics.

[21] Kagel, J.H., Roth, A.E., 1995. Handbook of experimental economics. Princeton University Press, Princeton.

[22] Kirman, A., 2011., Learning in agent-based models. Eastern Economic Journal 37(1), 20-27.

[23] Le, V.P.M., Meenagh, D., and Minford, P., 2012. What causes banking crises: an empirical investigation, Cardiff University Economics working paper E2012_14.

[24] Le, V.P.M., Meenagh, D., Minford, P., and Wickens, M., 2011. How much nominal rigidity is there in the US economy? Testing a New Keynesian DSGE Model using indirect inference, Journal of Economic Dynamics and Control 35(12), 2078-2104.

[25] Le, V.P.M., Meenagh, D., Minford, P., and Wickens, M., 2012. Testing DSGE models by Indirect inference and other methods: some Monte Carlo experiments, Cardiff University Economics working paper E2012_15.

[26] McCabe, K., 2003. Neuroeconomics. in: L. Nadel, ed., Encyclopedia of Cognitive Science. New York: Macmillan Publishing. pp. 294-298.

[27] McCallum, B.T., 1976. Rational expectations and the natural rate hypothesis: some consistent estimates. Econometrica 44, 43-52.

[28] Minford, P., and Ou., Z., 2010. Testing the Monetary Policy Rule in US: a Reconsideration of Fed's Behavior, Cardiff University Working Paper Series E2009/12, October 2009, updated in 2010.

[29] Rudd, J., and Whelan, K., 2006. Can rational expectations sticky price models explain inflation dynamics. American Economic Review, 96, 303–320.

[30] Sargent, T., 1993. Bounded rationality in macroeconomics. Oxford University Press.

[31] Smith, A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8, 63-84.

[32] Taylor, J.B., 1993. Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195-214.

[33] Wickens, M.R., 1982. The efficient estimation of econometric models with rational expectations. Review of Economic Studies 49, 55-67.