Cardiff Economics Working Papers

What causes banking crises? An empirical investigation

Vo Phuong Mai Le, David Meenagh and Patrick Minford

What causes banking crises? An empirical investigation[*]

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

Patrick Minford (Cardiff University and CEPR)[§]

April 2013

Abstract

We add the Bernanke-Gertler-Gilchrist model to a modified version of the Smets-Wouters model of the US in order to explore the causes of the banking crisis. We then extract the model's implied residuals on US unfiltered data since 1984 to replicate how the model predicts the crisis. The main banking shock tracks the unfolding 'sub-prime' shock. This shock worsens the banking crisis but 'traditional' shocks explain the bulk of the crisis; the non-stationarity of the productivity shock plays a key role. Crises occur when there is a 'run' of bad shocks; based on this sample they occur on average once every 40 years and when they occur around half are accompanied by financial crisis. Financial shocks on their own, even when extreme, do not cause crises — provided the government acts swiftly to counteract such a shock as happened in this sample.

1   Introduction

Since the banking crisis macroeconomic models have come under severe criticism, not merely in the popular media but also among economists and policymakers, for failing to predict the crisis. While clearly the models deny that it is possible to predict crises, so that this criticism is ill-founded, nevertheless the majority did fail — much more seriously — to predict the possibility of crisis because they contain no mechanisms that could produce it. They had no banking sector, so that a fortiori no banking crisis could occur. Furthermore they embodied only stationary shocks so that permanent shocks to the level of trend output, such as appear to characterise crisis episodes, were not examined; true, in the background there was possibly a non-stationary trend in productivity (typically removed by filtering from the model data) but there was not much focus on this in practice; Figure 1 illustrates. A further issue concerns their ability to fit the facts; economists such as Heckman have attacked the lack of empirical content in macro models, implying that it is hardly surprising they could give little guidance to policy in the crisis.
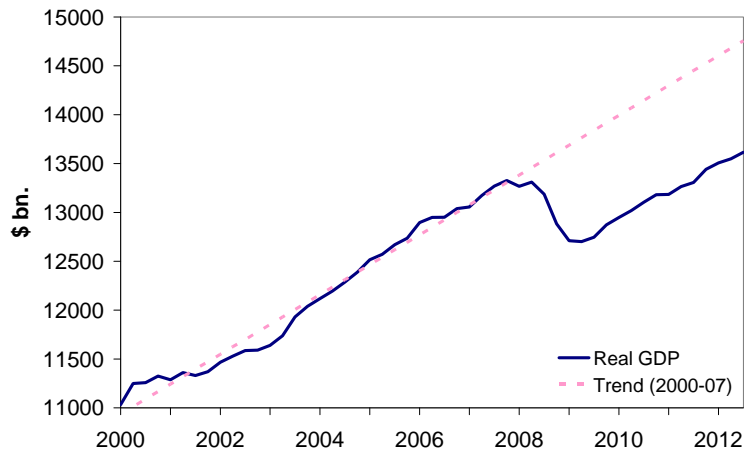

Figure 1: US real GDP and pre-crisis trend

In this paper we address these issues, building on recent work: we integrate a banking sector model into a widely-used DSGE model; we re-estimate the model and test the model against the data; and we apply it to unfiltered data so that nonstationary shocks are included. Finally we use it to give an account of what produced this banking crisis and relate this finding to the current policy debate.

Since the crisis, much work has been done to incorporate a banking sector into DSGE models. The banking models involve a friction in the intermediation process so that the interest margin required reflects the risk of loan default. This risk varies with the state of the economy. The exact transmission between the state and the risk-premium differs across the various banking models that have been proposed but that of Bernanke, Gertler and Gilchrist (1999) has been designed to capture many of the common features of these models; it is their set-up that we therefore use here as a representative one. In it, the IS curve now includes a variable risk-premium which is related to the economic state.

An important part of the explanation for 'crisis' episodes lies in the permanence of crisis shocks; thus after the Great Depression US output failed to catch up with its previous level for a decade, and similarly after the oil crisis of the mid-1970s the level of output was permanently reduced so that initial estimates of 'excess capacity' based on the previous output trend had to be revised downwards sharply. In the latest Great Recession it appears that much the same is happening; for example excess capacity in the UK is now officially estimated to be around 3%, whereas if the previous output trend level prevailed it would be around 13%. Thus the trend level seems to take a permanent hit in these crisis episodes. Furthermore the same appears to be true in reverse for periods of strong growth, such as the US in the late 90s and early 2000s; the output trend in these periods gets shifted upwards permanently. This suggests that the productivity and perhaps other shocks hitting the economy are non-stationary. Hence here our aim is to model the economy under potentially nonstationary shocks.

As the banking crisis originated and was at its most damaging in the US, we focus our efforts on models of the US economy. Our strategy is to take a well-known and empirically relatively successful model of the US, that of Smets and Wouters (2007, SW), and add to it the banking model due to Bernanke, Gertler and Gilchrist (1999, BGG). Variants of the combination of SW and BGG have been

used in recent papers by Christiano, Motto & Rostagno (2010) for the US and Eurozone separately; Gilchrist et al. (2009) for the US alone and Fahr et al. (2011) for the Eurozone alone. They find that shocks that come from the financial sector have an important role in explaining macroeconomic fluctuations. All these authors use the Bayesian approach to estimating the model parameters. Similar findings are reported by Jermann and Quadrini (2012).

What distinguishes our work from these papers are the two main departures noted above. First, we introduce non-stationary shocks, which we argue give us insight into the nature of crises. Second, we estimate and test the overall models we use against the unfiltered data by indirect inference. If, as we find, the original (usually Bayesian) parameters do not pass the test, we search for parameters that get closest to the data according to this test; and we only finally use a parameter set that passes the test. As part of this empirical search we find that a hybrid model works best, in which there are both imperfectly (New Keynesian, NK) and perfectly competitive (New Classical) sectors in the labour and product markets, with NK weights that turn out to be quite low — this builds on the work of Le et al. (2011). We think this empirical hurdle is necessary in this particular area because while it is possible to construct models that generate large financial accelerator effects by suitable choice of priors, the parameter sets that pass this test of overall fit to the data do not give anywhere near such large effects; it seems they cannot be found in the data.

One recent paper by Stock and Watson (2012) uses a dynamic factor model to show that the Great Recession is explained by the same shocks that explained previous postwar recessions. They also show that there has been a slowdown in underlying growth, and that the financial shocks are important, but no more important than before. So a new financial crisis shock is not needed. Stock and Watson's findings are similar to what we will report in what follows.

The paper is organised as follows. We begin in section 2 by giving a brief account of the method of indirect inference we use here. In section 3 we explain the modified SW model and the BGG banking sector model. In section 4, we apply the model as re-estimated to the original unfiltered data and consider what light it sheds on the causes of this banking crisis as well as of crises and banking crises in general. Section 5 concludes and draws out some implications for policy.

2 The method of indirect inference

We evaluate the models' capacity in fitting the data using the method of indirect inference originally proposed in Minford, Theodoridis and Meenagh (2009) and subsequently with a number of refinements by Le et al. (2011) who evaluate the method using Monte Carlo experiments. The approach employs an auxiliary model that is completely independent of the theoretical one to produce a description of the data against which the performance of the theory is evaluated indirectly. Such a description can be summarised either by the estimated parameters of the auxiliary model or by functions of these; we will call these the descriptors of the data. While these are treated as the 'reality', the theoretical model being evaluated is simulated to find its implied values for them.

Indirect inference has been widely used in the estimation of structural models (e.g., Smith, 1993, Gregory and Smith, 1991, 1993, Gourieroux et al., 1993, Gourieroux and Monfort, 1995 and Canova, 2005). Here we make a further use of indirect inference, to evaluate an already estimated or calibrated structural model. The common element is the use of an auxiliary time series model. In estimation the parameters of the structural model are chosen such that when this model is simulated it generates estimates of the auxiliary model similar to those obtained from the actual data. The optimal choices 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. Common choices of this function are the actual coefficients, the scores or the impulse response functions. In 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 a true model of the economy, the null hypothesis — with the performance of the auxiliary model when estimated from the actual data. If the structural model is correct then its predictions about the impulse responses, moments and time series properties of the data should statistically match those based on the 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.

The testing procedure thus involves first constructing the errors implied by the previously estimated/calibrated structural model and the data. These are called the structural errors and are backed

out directly from the equations and the data[1]. These errors are then bootstrapped and used to generate for each bootstrap new data based on the structural model. An auxiliary time series model is then fitted to each set of data and the sampling distribution of the coefficients of the auxiliary time series model is obtained from these estimates of the auxiliary model. A Wald statistic is computed to determine whether functions of the parameters of the time series model estimated on the actual data lie in some confidence interval implied by this sampling distribution.

Following Meenagh et al. (2012) we use as the auxiliary model a VECM which we re-express as a VAR(1) for the three macro variables (interest rate, output gap and inflation) with the productivity residual entered as an exogenous non-stationary process (this having the effect of achieving cointegration[2]). Thus our auxiliary model in practice is given by: $y_t = [I - K]y_{t-1} + \gamma \overline{x}_{t-1} + gt + v_t$ where $\overline{x}_{t-1}$ is the stochastic trend in productivity, gt are the deterministic trends, and v_t are the VECM innovations. We treat as the descriptors of the data the VAR coefficients (on the endogenous variables only, $I - K$) and the VAR error variances ($var[v]$). The Wald statistic is computed from these[3]. Thus effectively we are testing whether the observed dynamics and volatility of the chosen variables are explained by the simulated joint distribution of these at a given confidence level. The Wald statistic is given by:

$$(\Phi - \overline{\Phi})' \sum\nolimits_{(\Phi\Phi)}^{-1} (\Phi - \overline{\Phi}) \tag{1}$$

where Φ is the vector of VAR estimates of the chosen descriptors yielded in each simulation, with $\overline{\Phi}$ and $\sum_{(\Phi\Phi)}$ representing the corresponding sample means and variance-covariance matrix of these calculated across simulations, respectively.

The joint distribution of the Φ is obtained by bootstrapping the innovations implied by the data and the theoretical model; it is therefore an estimate of the small sample distribution[4]. Such a distribution is generally more accurate for small samples than the asymptotic distribution; it is also shown to be consistent by Le et al. (2011) given that the Wald statistic is 'asymptotically pivotal'; they also showed it had quite good accuracy in small sample Monte Carlo experiments[5].

This testing procedure is applied to a set of (structural) parameters put forward as the true ones (H_0, the null hypothesis); they can be derived from calibration, estimation, or both. However derived, the test then asks: could these coefficients within this model structure be the true (numerical) model generating the data? Of course only one true model with one set of coefficients is possible. Nevertheless we may have chosen coefficients that are not exactly right numerically, so that the same model with other coefficient values could be correct. Only when we have examined the model with all coefficient values that are feasible within the model theory will we have properly tested it. For this reason we later extend our procedure by a further search algorithm, in which we seek other coefficient sets that could do better in the test.

Thus we calculate the minimum-value full Wald statistic for each period using a powerful algorithm based on Simulated Annealing (SA) in which search takes place over a wide range around the initial values, with optimising search accompanied by random jumps around the space[6]. In effect this is indirect

---

[1] Some equations may involve calculation of expectations. The method we use here is the robust instrumental variables estimation suggested by McCallum (1976) and Wickens (1982): we set the lagged endogenous data as instruments and calculate the fitted values from a VAR(1) — this also being the auxiliary model chosen in what follows.

[2] See Appendix 2.

[3] We do not attempt to match the time trends and the coefficients on non-stationary trend productivity; we assume that the model coefficients yielding these balanced growth paths and effects of trend productivity on the steady state are chosen accurately. However, we are not interested for our exercise here in any effects on the balanced growth path, as this is fixed. As for the effects of productivity shocks on the steady state we assume that any inaccuracy in this will not importantly affect the business cycle analysis we are doing here — any inaccuracy would be important in assessing the effect on the steady state which is not our focus. Thus our assessment of the model is as if we were filtering the data into stationary form by regressing it on the time trends and trend productivity.

[4] The bootstraps in our tests are all drawn as time vectors so contemporaneous correlations between the innovations are preserved.

[5] Specifically, they found that the bias due to bootstrapping was just over 2% at the 95% confidence level and 0.6% at the 99% level. They suggested possible further refinements in the bootstrapping procedure which could increase the accuracy further; however, we do not feel it necessary to pursue these here.

[6] 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. Our method used our standard testing method: we take a set of model parameters (excluding error processes), extract the resulting residuals from the data using the LIML method, find their implied autoregressive coefficients ($AR(1)$ here) and then bootstrap the implied innovations with this full set of parameters to find the implied Wald value. This is then minimised by the SA algorithm.

inference estimation of the model; however here this estimation is being done to find whether the model can be rejected *in itself* and not for the sake of finding the most satisfactory estimates of the model parameters. Nevertheless of course the method does this latter task as a by-product so that we can use the resulting unrejected model as representing the best available estimated version. The merit of this extended procedure is that we are comparing the best possible versions of each model type when finally doing our comparison of model compatibility with the data.

Before we proceed to carry out our tests and estimation, we should explain why we do not use the much more familiar 'direct inference' estimation and testing procedures here. In direct inference one fits a structural model directly to the data, either by classical 'frequentist' FIML or by the now popular Bayesian ML. The likelihood that is maximised in FIML is derived from the size of the reduced form errors. In Bayesian ML it is derived from this plus the priors — effectively the resulting ML parameters are a weighted average of the FIML values and the priors, where the weights depend on the prior distributions and the extent to which the FIML values differ from the priors. The FIML values are essentially those that give the best current forecasting performance for the model (i.e. minimising the size of the reduced form errors). One can develop overall tests of the model specification under direct inference by creating, in the FIML case, a Likelihood Ratio against some benchmark model, a natural one being an unrestricted VAR; in the Bayesian case Del Negro and Schorfheide (2006) have proposed the DSGE-VAR weight as a measure of model closeness to the data (this is the weight on the prior model's implied VAR, as combined with the unrestricted VAR, that maximises the likelihood). This can also be treated as a specification test of the overall model, even though usually Bayesians are reluctant to talk about 'testing' the model as whole.

Such tests are compared with the indirect inference tests using Monte Carlo experiments with an SW model, in Le et al. (2012). They find that the tests compare quite different features of model performance. The direct ones check (in-sample) forecasting ability, while the indirect one checks the model's causal structure. For policy purposes we are most interested in using DSGE models for simulation of the effects of policy changes and hence in their causal structure. Typically forecasting is done by other means.

Both tests can still be used to test a model's specification and hence its causal structure, even if the direct method checks it via forecasting performance. But Le et al. also find that, viewed as a test of model specification, the power of direct inference tests in small samples is much lower than that of indirect inference. In other words they discriminate rather weakly against false models. This is presumably because forecasting is only weakly related to good specification; bad models with a lot of ad hoc lags and added exogenous variables forecast better than models based on good theory, which are restricted to having only structural shock processes as their exogenous variables. Furthermore false models will generate false structural shock processes which may well partly compensate for the specification error in the model's forecasting performance. Meanwhile the indirect inference test's power against false models allows one to discover rather accurately what features of the data behaviour a model can replicate and what not; this in turn can be helpful in thinking about respecification.

In estimation both FIML and indirect estimators are consistent and asymptotically normal. But as we have seen the latter's power is greater in small samples so that it should also give more reliable results from estimation in small samples.

For these reasons we use the indirect inference procedure here both to estimate the model on our available small samples and to test its specification.

3   The SW and BGG models

3.1   The SW 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. SW 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 made various tests of their model. Subsequently Del Negro, Schorfheide, Smets and Wouters (2007) further examined it by considering the extent to which its restrictions help to explain the data. It should be noted that none of these exercises in evaluating the SW model were a test of specification in the classical sense. Le et al. (2011) proposed such a test, a Wald test based on indirect inference which compares the model's VAR representation with the VAR coming from the data, and showed that over the full post-war sample the original SW New Keynesian (NK) model was rejected. In addition, they examined an alternative version in which prices and wages were fully flexible but there was a simple one-period information delay for labour suppliers. This 'New Classical (NC)' version was also rejected. They also proposed a hybrid model that merged 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. They assumed that the size of each sector depended on the facts of competition and did not vary in the sample but they allowed the degree of imperfect competition to differ between labour and product markets. The basic idea was that economies consist of product sectors where rigidity prevails and others where prices are flexible, reflecting 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. It turned out that this combined model got much closer to the data for the full sample, when the rigidity was quite limited.

Essentially, the NK model generated too little nominal variation while the NC model delivered too much. However the hybrid model was able to reproduce the variances of the data; and it is this key feature that enables it to match the data overall more closely. Nevertheless, it failed to match certain features of the data, notably the behaviour of interest rates in relation to other major macro variables. In view of this failure, it seemed 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 the 1950s to the 2000s. They therefore tested for structural change during this period and duly found parameter breaks in two places: 1965 and 1984. These were natural places to find such breaks because of 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 with the shift towards interest rate setting that followed from the adoption of (implicit) inflation targeting.

Le et al. (2011) found that for the third and last sub-period (1984Q3–2004Q2), a version of the model very close to SW's original NK model was not rejected by the Wald test on the main macro variables' behaviour. Accordingly it is this version of the model on a sample from 1984 that we use here. Being very close to the original version, with a high degree of nominal rigidity both in labour and product markets, it behaves very like a standard New Keynesian model. In it, 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 Tobin's 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.

3.2 The BGG model of the banking sector together with the SW model

The BGG financial sector produces certain changes in the model of Smets and Wouters (2007) in the form used here as modified by Le et al. (2011) but much remains unchanged.

The household sector is unchanged. Households maximise a utility function by choosing goods and labour over an infinite life. They exhibit some consumption habit behaviour. A part of labour is supplied to an imperfect labour market where households act as price-setters and the rest is supplied to a perfectly competitive labour market. This results in a hybrid wage equation, where the aggregate wage is the weighted average of wages obtained in the perfect and imperfect labour markets. Thus the aggregate wage equation and consumption Euler equation remain unchanged.

In the government sector both monetary and fiscal policy also remain the same.

The BGG model incorporation divides the production side into three distinct participants: as previously, retailers and intermediate goods producers (now called entrepreneurs for a reason described later) and in addition, capital producers. Retailers function in the same way as before, operating in perfect competition to produce final goods by aggregating differentiated intermediate products using the Dixit-Stiglitz technology. With the assumption that retail output is made up of a fixed proportion of intermediate goods in an imperfectly competitive market and intermediate goods sold competitively, the aggregate price is a weighted average of prices received in the two types of market. As a result, the aggregate price equation is unchanged. Capital producers operate in a competitive market and take prices as given. They buy final consumption goods and transform them into capital to be sold on to entrepreneurs.

The difference in BGG lies in the nature of entrepreneurs. They still produce intermediate goods, but now they do not rent capital from households (who do not buy capital but only buy bonds or deposits) but must buy it from capital producers and in order to buy this capital they have to borrow from a bank which converts household savings into lending. On their production side, entrepreneurs face the same situation as in Le et al. (2011). They hire labour from households for wages that are partly set in monopolistic, partly in competitive labour markets; and they buy capital from capital producers at prices of goods similarly set in a mixture of monopolistic and competitive goods markets. Thus the production function, the labour demand and real marginal cost equations are unchanged. It is on their financing side that there are major changes. Entrepreneurs buy capital using their own net worth, pledged against loans from the bank, which thus intermediates household savings deposited with it at the risk-free rate of return. The net worth of entrepreneurs (n_t) is kept below the demand for capital by a fixed death rate of these firms $(1-\theta)$; the stock of firms is kept constant by an equal birth rate of new firms. Entrepreneurial net worth therefore is given by the past net worth of surviving firms plus their total return on capital (cy_t) minus the expected return (which is paid out in borrowing costs to the bank) on the externally financed part of their capital stock — equivalent to

$$n_t = \theta n_{t-1} + \frac{K}{N}\left(cy_t - E_{t-1}cy_t\right) + E_{t-1}cy_t + enw_t \tag{2}$$

where $\frac{K}{N}$ is the steady state ratio of capital expenditures to entrepreneurial net worth and θ is the survival rate of entrepreneurs. Those who die will consume their net worth, so that entrepreneurial consumption (c_t^e) is equal to $(1-\theta)$ times net worth. In logs this implies that this consumption varies in proportion to net worth so that:

$$c_t^e = n_t \tag{3}$$

In order to borrow, entrepreneurs have to sign a debt contract prior to the realisation of idiosyncratic shocks on the return to capital: they choose their total capital and the associated borrowing before the shock realisation. The optimal debt contract takes a state-contingent form to ensure that the expected gross return on the bank's lending is equal to the bank opportunity cost of lending. When the idiosyncratic shock hits, there is a critical threshold for it such that for shock values above the threshold, the entrepreneur repays the loan and keeps the surplus, while for values below it, he would default, with the bank keeping whatever is available. From the first order conditions of the optimal contract, the external finance premium is equated with the expected marginal product of capital which under constant returns to scale is exogenous to the individual firm (and given by the exogenous technology parameter); hence the capital stock of each entrepreneur is proportional to his net worth, with this proportion increasing as the expected marginal product rises, driving up the external finance premium. Thus the external finance premium increases with the share of the firm's capital investment that is financed by borrowing:

$$E_t cy_{t+1} - (r_t - E_t \pi_{t+1}) = \chi\left(qq_t + k_t - n_t\right) + epr_t \tag{4}$$

where the coefficient $\chi > 0$ measures the elasticity of the premium with respect to leverage. Entrepreneurs leverage up to the point where the expected return on capital equals the cost of borrowing from

financial intermediaries. The external finance premium also depends on an exogenous premium shock, epr_t. This can be thought of as a shock to the supply of credit: that is, a change in the efficiency of the financial intermediation process, or a shock to the financial sector that alters the premium beyond what is dictated by the current economic and policy conditions.

Entrepreneurs buy capital at price qq_t in period t and uses it in $(t+1)$ production. At $(t+1)$ entrepreneurs receive the marginal product of capital rk_{t+1} and the ex-post aggregate return to capital is cy_{t+1}. The capital arbitrage equation (Tobin's Q equation) becomes:

$$qq_t = \frac{1-\delta}{1-\delta+R_*^K} E_t qq_{t+1} + \frac{R_*^K}{1-\delta+R_*^K} E_t rk_{t+1} - E_t cy_{t+1} \tag{5}$$

The resulting investment by entrepreneurs is therefore reacting to a Q-ratio that includes the effect of the risk-premium. There are as before investment adjustment costs. Thus, the investment Euler equation and capital accumulation equations are unchanged from Le et al. (2011). The output market-clearing condition becomes:

$$y_t = \frac{C}{Y} c_t + \frac{I}{Y} inn_t + R_*^K k_y \frac{1-\psi}{\psi} rk_t + c_y^e c_t^e + eg_t \tag{6}$$

4 What does the model with financial rigidity say about the origins of the banking crisis?

4.1 Estimation and model fit

The model that integrates the banking sector is estimated using the method of indirect inference as set out in Le et al. (2011) for the 1984–2009 period. The estimated model is tested against the data using the main macroeconomic variables, output, inflation and the interest rate. We use a test of whether the model can match the time series properties of the data jointly. The model is found to fit the data well with a Wald t-statistic of 1.56. The estimated parameters can be found in Table 1. Impulse response functions to key variables when the model is applied to non-stationary data are shown in Figure 2. Note that the second set of IRFs in Figure 2 are due to a non-stationary productivity shock. Figure 3 shows that the model generates 95% confidence intervals for the implied VAR responses that encompass or come close to encompassing the data-based VAR responses for a monetary shock; these VAR responses (the rest are shown in Appendix 3) are jointly encompassed by the model so that individual ones may lie slightly outside.

The estimated coefficients are substantially different from the starting set for which we used Smets and Wouters (2007) Bayesian estimated parameters of the non-banking model and the original BGG calibrated parameters for the two banking equations and low NK weights reflecting the initial findings of Le et al. (2011)[7]. One can think of the latest set of non-banking parameters along a spectrum between a fully New Keynesian and a fully New Classical model. Overall the model is fairly New Classical with NK weights of 0.56 in the labour market and 0.09 in the goods market. However, this is partly counterbalanced by a greater rigidity of wages and prices and more wage indexation in the NK sector than in the starting values, and by greater elasticities of labour supply and capacity utilisation, as well as lower capital adjustment cost, all of which raise output responses to shocks. Taylor Rule responses are also all smaller so again implying less dampening of output responses. In the banking equations the main change is a more than halving of the response of the spread to net worth and the return on capital. This underlines the point that the data in general do not support large financial transmission effects, a factor we revert to below that partly explains why our paper finds a much smaller role than others do for financial shocks.

---

[7] We use these instead of the post 1984 re-estimated weights because the crisis period injects turbulence into the Great Moderation period, more like the original full sample.

Model Coefficients: 1984Q3-2009Q2			
		Starting coef	Estimated coef
Steady-state elasticity of capital adjustment	φ	5.74	7.5744
Elasticity of consumption	σ_c	1.38	1.2716
External habit formation	λ	0.71	0.6512
Probability of not changing wages	ξ_w	0.70	0.7533
Elasticity of labour supply	σ_L	1.83	2.8327
Probability of not changing prices	ξ_p	0.66	0.8398
Wage indexation	ι_w	0.58	0.9404
Price indexation	ι_p	0.24	0.1213
Elasticity of capital utilisation	ψ	0.54	0.1988
Share of fixed costs in production (+1)	Φ	1.50	1.6841
Taylor Rule response to inflation	r_p	2.50	1.8886
Interest rate smoothing	ρ	0.60	0.7742
Taylor Rule response to output	r_y	0.08	0.0381
Taylor Rule response to change in output	$r_{\Delta y}$	0.22	0.1133
Share of capital in production	α	0.19	0.1435
Proportion of sticky wages	ω^w	0.40	0.5624
Proportion of sticky prices	ω^r	0.10	0.0874
Elasticity of the premium with respect to leverage	χ	0.04	0.0279
Quarterly steady-state inflation[†]	$\bar{\pi}$	0.78	0.7800
Discount factor[†]	β	0.9984	0.9984
Steady-state hours worked[†]	$\overline{L}$	0.53	0.5300
Quarterly steady-state output growth[†]	$\bar{\gamma}$	0.43	0.4300
WALD (Y, π, R)		104.0786	21.5148
T-stat		9.1080	1.5560

[†]Fixed parameter

Table 1: Coefficient Estimates (1984Q3-2009Q2)
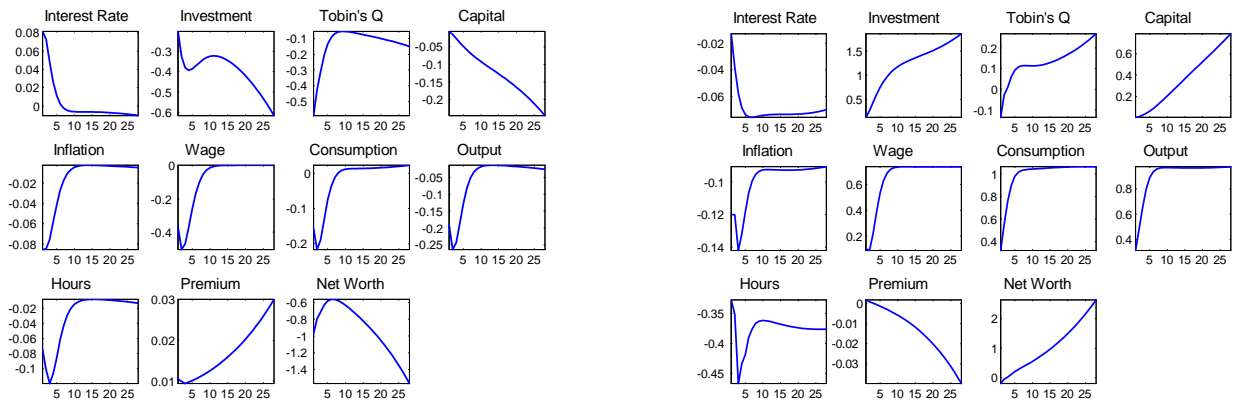
IRFs for a Monetary Shock

IRFs for a Non-Stationary Productivity Shock
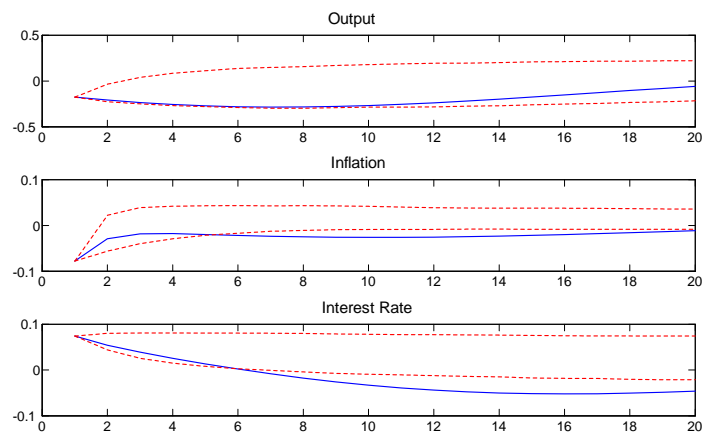
Figure 2: IRFs for key variables

Figure 3: VAR IRFs for a monetary shock

4.2 Error properties on unfiltered data

Having established that the model that integrates the banking sector fits the data, we now go on to apply it to the recent crisis episode in the US. To do this we extract the model errors from the unfiltered data (shown in Figure 4) and fit to each an AR time-series process over the period. Table 2 shows the status of each error and also the AR parameters estimated by our procedure. We find that productivity unambiguously has a unit root. The external premium, government and labour supply shocks are all ambivalent; their ADF statistics do not reject non-stationarity, while their KPSS statistics do not reject stationarity. All are heavily affected by recent observations, mainly the crisis period, and the theory is clear that they should be stationary; thus government spending would be limited by some limit on tax rates and both the labour supply and the premium by the ending of the crisis. We have accordingly treated them as either stationary or trend-stationary. For the AR coefficient we take the estimated value; where this is unity, we set it just below, so that while technically stationary it is very highly persistent over the crisis period. The other shocks are all clearly either stationary or trend-stationary.
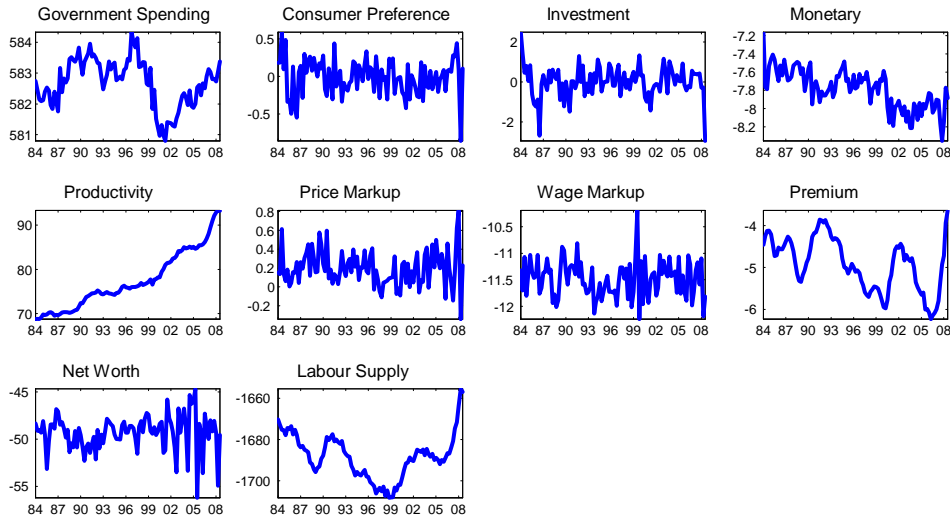

Figure 4: Equation Errors from 1984-2009.

	ADF p-value[+]	KPSS statistic	Conclusion	Coefficient	
				AR	Prod. shock
Government Spending	0.3946	0.2707	Stationary	0.8803	0.1931
Consumer Preference	0.0000	0.1290	Stationary	0.0346	
Investment	0.0000	0.0468	Stationary	0.3165	
Taylor Rule	0.0000	0.0919	Trend Stat.	0.3971	
Productivity	1.0000	1.2796***	Nonstat.	0.2547	
Price Mark-up	0.0000	0.2025	Stationary	0.0550	
Wage Mark-up	0.0000	0.1302	Stationary	0.0078	
Labour Supply	0.6096	0.2884	Trend Stat.	0.9990	
Premium	0.0631	0.0641	Trend Stat.	0.9990	
Net Worth	0.0000	0.1450	Stationary	0.0656	

[+] p-value of 0.05 is the 5% confidence limit for rejecting the unit root.
**(***) KPSS rejects stationarity at 5%(1%).

Table 2: Stationarity of Shocks and AR Parameters(1984Q3-2009Q2)

Plainly the crisis had international ramifications but we cannot identify the causality of these in a US-only model. The shocks that show up in the model are partly coming from these international effects; most obviously commodity price shocks that enter through the 'price mark-up' here are themselves

responding to the US crisis. A further, similar limitation of our account is our inability to analyse connections between the shocks to the model. No doubt the banking shocks we identify had simultaneous and lagged effects on the non-banking shocks; but also vice versa, the non-banking on the banking. The sample episode is too short to establish which way such effects might go or even if they exist, tempting as it might be to run some regressions to detect them. The model assumes that each shock is separate from the others and only related to its own past. The model then disentangles how each shock works through the economy to affect final outcomes. Anyone that wished to take matters further would have to model the interactions of the shocks themselves through a wider model, such as one of political economy.

4.3 The errors driving the episode

We begin by showing the behaviour of the main model errors (i.e. the total accumulated innovations) during the crisis episode, which we treat as 2006Q1 to 2009Q2. These are shown in Figure 5. We have not included the 'recovery period' at this stage, though this would be interesting; our focus is on the period when the economy first went into recession and then bottomed out.
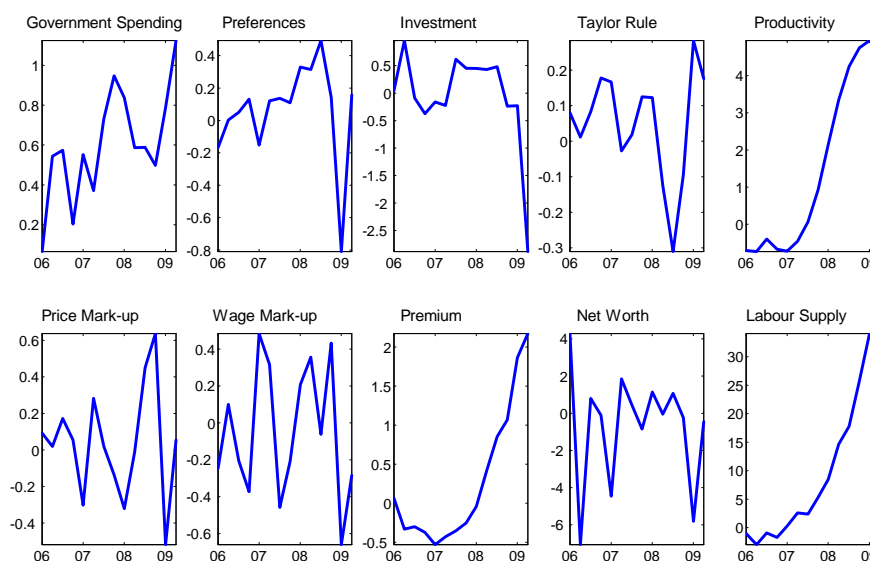

Figure 5: Accumulated Shocks from 2006Q1–2009Q2

We can immediately single out from these errors four key ones which behaved in a particularly persistent way during the crisis period: productivity (as we would expect), the Taylor Rule, labour supply and the External Finance Premium. These four turn out to have been the main drivers of the economy during this episode.

Productivity grew powerfully in the early stages of the period but stalled and fell in the heart of the crisis.

The Taylor Rule error forced up the interest rate steadily against the Rule's dictates: this was the operation of the zero bound.

Labour supply refers to the competitive sector wage equation which suffered from upward 'wage push' (equivalently a fall in labour supply) throughout the period until the heart of the crisis when it was partially reversed, presumably by the extreme labour market conditions. According to the model this error is the product of real wage 'push'. This may have resulted from the remarkable rise in oil and commodity prices over this period, which sharply reduced real wages; seen as abnormal it may have caused a reduction of labour supply (a real wage resistance) through intertemporal substitution. It could also be that the collapse of employment, particularly in construction, led to the redundancy of unskilled and (through FIFO) younger workers in the main; the fact that older and skilled workers were more likely to be kept on may thus have raised the average wage bill, even while productivity soared. This would suggest we should link the productivity and wage push errors. There is no obvious other cause: the policies of the Obama administration (on union power and Obamacare) only came in for the last

two quarters of our period when wage push fell back. As this was the competitive sector it should not have been wage 'stickiness'; indeed the fact that it accumulated and only went into reverse after 3 years suggests it cannot have been.

Finally, the error in the External Finance Premium equation is large, rising and persistent, only easing off slightly just before the end of the episode. What this reveals is that, even when a banking sector is included in the model, it cannot account for the behaviour of the Premium. There is a large additional and cumulative shock at work, specific to the episode. This error can be thought of as reflecting the fear of a 'run on the banks', emanating from the specific circumstances affecting banks in the crisis: the sub-prime write-off, the Lehman collapse and so on.

It is worth dwelling on the possible sources of this shock, which we will call 'the Sub-Prime' shock since plainly it is associated with the way large amounts of sub-prime mortgage debt were first bundled up as securities, then sliced up by 'risk status' and repackaged as non-transparent Collateralised Debt Obligations, then resold around the world to credulous banks mostly in Europe, to be held in unknown quantities on these banks' balance sheets, creating huge uncertainty about which banks were at risk — as in the card game, which turns on who will wind up holding the Queen of Spades.

The Sub-Prime shock reached its peak after the collapse of Lehman. After this governments all over the world, including the US, were forced to bail out their insolvent banks and the size of the shock eventually diminished (after our sample ends) as this government backing allayed concerns over the remaining banks' viability. Thus the shock we find here seems to be associated with government intervention both in its initial propagation and in its cleaning-up afterwards.

4.4 A stochastic variance decomposition of the episode

We next look at the variance decomposition of such episodes. Again, we are using unfiltered data when performing this analysis which treats the episode stochastically — that is, we take the shocks in the episode and replay them by redrawing them randomly and repeatedly with replacement to see what a typical crisis episode would be like. Our variance decomposition is therefore for such a typical episode.

	Int. rate	Inv.	Inf.	Real Wages	Cons.	Output	Emp.
Government Spending	2.8	0.0	0.6	1.0	0.3	1.4	1.1
Preferences	3.5	0.0	1.2	13.1	2.2	2.7	0.9
Investment	2.6	2.1	1.0	2.2	0.3	1.1	0.7
Taylor Rule	4.1	0.2	6.7	9.5	0.8	1.8	0.8
Productivity	6.5	0.5	17.1	19.3	27.6	29.5	11.3
Price Mark-up	5.2	0.1	22.5	7.8	0.4	1.0	0.4
Wage Mark-up	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Labour Supply	20.9	1.5	37.3	43.4	56.6	54.0	77.4
Premium	45.9	83.6	11.4	3.1	9.8	6.9	6.0
Net Worth	8.5	12.0	2.2	0.2	2.1	1.6	1.5
TOTAL	100	100	100	100	100	100	100
Banking Shocks	54.3	95.6	13.6	3.3	11.9	8.5	7.4
Non-Banking Shocks	45.7	4.4	86.4	96.7	88.1	91.5	92.6

Table 3: Variance Decomposition for Crisis Period

What we see from Table 3 is that only 9% of the output variance is due to financial shocks; and the rest is due to the usual non-banking shocks. For investment the share of financial shocks is very high (96%); but this gets dampened in its effect on GDP partly because interest rates react to them and partly because it is a small part of GDP. Accordingly we see that interest rates are also highly affected (54%) by the financial shocks. As for inflation only 14% comes from this financial side.

What we see here is that there was a distinct role for financial shocks (essentially the Sub-Prime shock) in such episodes. However, the bulk of the variation comes from the other shocks: labour supply

and productivity, plus for inflation the price mark-up. We can think of the crisis as being the result mainly of normal poor macro shocks (slowing of productivity growth, wage push, oil shock) with the Sub-Prime shock adding a nasty extra twist (about 1% extra off GDP at the bottom of the recession in real time as we will see in the next section). Thus turning it from a nasty (moderate crisis) episode into a 'Great Recession' episode.

The failure of fiscal policy to show up through the government spending effect is not really surprising: the fiscal response largely took the form of transfers, such as financial bail-outs (of AIG, Fannie and Freddie) and 'cash for clunkers'. Such transfers have no identifiable effect within this model; but one can think of them as already embodied in the Sub-Prime shock as a mitigating response, as discussed above; since clearly the risk premium and net worth shocks would have been much worse without this government cash infusion, which is acting like a supply of credit to both the cash-starved banks and the private sector. Thus these direct banking shocks are recorded net of this public response; unfortunately we have no way of disentangling the total banking shock from the mitigating effect of such government direct intervention in the financial system. We suggested above that the peaking of the Sub-Prime shock after Lehman was due to the government intervention. We could use this possibly as an identifying device for that part of the shock due to fiscal intervention; however we have not done so here as there are only three observations after Lehman.

What we have found here is in contrast to findings in the recent literature, where it has been found that the financial shocks are very important. Jermann and Quadrini (2012) examine the effects of financial shocks on linear detrended US data for the 1984–2010 period. They find that financial shocks contribute to almost half of the volatility of output and the productivity shock only contributes to 4.1% of output volatility. Christiano et al. (2010) augment a standard monetary DSGE model to include financial frictions for the US and EU using detrended data from $1985 - 2008$. They find that financial friction shocks account for a substantial portion of economic fluctuations. The main reason that we find that the productivity shock plays such an important role is that we have used non-stationary data for our analysis. This gives us a non-stationary productivity shock, so that any change in productivity is permanent. Another reason is that we are using a model that is fully estimated to conform to the data behaviour as much as possible.

We show in Table 4 how the variance decompositions in our model would differ if a) if we used the original SWBGG model with stationary data b) we used stationary data with our existing model. If we go back to SWBGG's original coefficients and stationary data, the share of banking shocks in output variance rises to 79.5%, much in line with the results of other authors. If we replace these coefficients by the ones we have estimated here, the share drops to 52.3%, illustrating the way in which parameters that can replicate the sample's data behaviour give smaller effects of financial transmission. Finally, once we then include nonstationary shocks, the share drops sharply to the one we report of 8.5%.

We also estimate and test our model against the data using indirect inference, which is a strict test of whether the model can match the time series properties of the data. The authors above generally use either calibrated parameters or estimate them using Bayesian maximum likelihood; we have found that such estimates can lie a fair way away from parameters flexibly reestimated as here and are generally rejected by our tests. Thus it may well be that a further contributing factor to the difference in our results lies in the closeness of our parameters to the data. In particular, as noted at the start, the effects of financial frictions appear to be much smaller in the data than in some calibrated models.

4.5   Accounting for this particular banking crisis episode

We can also decompose what actually happened in the precise episode that occurred according to the model as a result of these shocks. We do this in the charts that follow for the main macro variables.

If we focus first on output (Figure 6), we see that the Sub-Prime shock contributes about 1% to the downturn by 2009Q2, the bottom of the recession. The main other negative element is the labour supply shock. The Taylor Rule and investment shocks tip output down further. However, what stops the downturn from turning into a rout is a strong positive productivity shock — the overall swing from peak to trough is just under 6%.

We show three lines on the chart: the total predicted, the predicted total without financial shocks, and the total predicted without either financial shocks or any financial transmission. The last two hardly differ, showing that financial transmission of non-financial shocks is modest: the effect of the non-financial shocks is occurring through the usual non-financial channels. The financial effect is coming from the financial shocks themselves.

If we turn to interest rates (Figure 7), we see in the final 2 quarters, 2009Q1 and 2009Q2, how the

	Output		
	Stationary data Original coeffs	Stationary data II estimated coeffs	Unfiltered data II estimated coeffs
Government Spending	0.4	1.8	1.4
Preferences	2.3	18.2	2.7
Investment	1.8	6.4	1.1
Taylor Rule	1.0	5.1	1.8
Productivity	13.7	4.3	29.5
Price Mark-up	0.2	0.4	1.0
Wage Mark-up	1.0	0.0	0.0
Labour Supply	0.0	11.4	54.0
Premium	5.2	5.1	6.9
Net Worth	74.3	47.2	1.6
TOTAL	100	100	100
Banking Shocks	79.5	52.3	8.5
Non-Banking Shocks	20.5	47.7	91.5

Table 4: Comparison of Output Variance Decompositions for Crisis Period

productivity and Sub-Prime shocks were pulling rates down, but this was offset to some extent by the Taylor Rule policy shocks which moved from negative in the second half of 2008 to positive in the first half of 2009, reflecting the effect of the zero bound: policymakers would have liked to ease by another 0.6% per annum.

Finally on inflation (Figure 8), we see how the price mark-up shock has both negative and positive effects during the episode as oil and commodity prices first surged, and later fell back sharply as the recession took hold at the end of 2008 — see Figure 9. The labour supply and productivity shocks largely offset each other, implying that inflation remained surprisingly stable during the episode; by the end the Taylor Rule, investment and Sub-Prime shocks were driving inflation down, close to the deflationary range (below -0.5% means that if inflation had otherwise been at a 2% target, it would have gone negative for the quarter.)

The overall interpretation coming from this analysis is of a crisis triggered by severe exogenous shocks, and exacerbated by a large financial shock, itself offset by large fiscal intervention. The zero bound added some further pressure. This cannot be described as a crisis 'created by the financial system'.

4.6 What is and causes a (financial) crisis?

If we take a longer perspective than just this crisis, we can ask: what is a crisis and what causes it, according to our analysis of this US sample? Let us define a 'crisis' as a severe downturn in output, a large part of which is permanent; and a financial crisis as a crisis in which there is also a financial collapse of some sort. What does this model have to say in general about the causes of these? We examine this question by inspecting the bootstrap experience (potential scenarios over the period) from the model and its normal shocks; for this we use the shocks from the period $1984 - 2007$ so that we do not reuse the shocks from this crisis period itself. Again, this analysis is done on unfiltered data. Plainly we know that these shocks generate crisis; and we want to discover whether this experience is unique.

We find the following regularities:

a) Crisis is a normal part of capitalism: this economy will generate crises regularly from 'standard' shock sequences. We illustrate this from some of the bootstrap simulations/scenarios produced from the shocks of the $1984 - 2007$ period (i.e. sans crisis) in Figures 10 and 11. In around half of them there were quite serious interruptions of activity, which satisfy the definition of crisis. If we define a crisis as a fall in GDP of at least 1% with at least 5 years, before output returns to its previous peak, then we find that a crisis on average will occur about every 67 years.
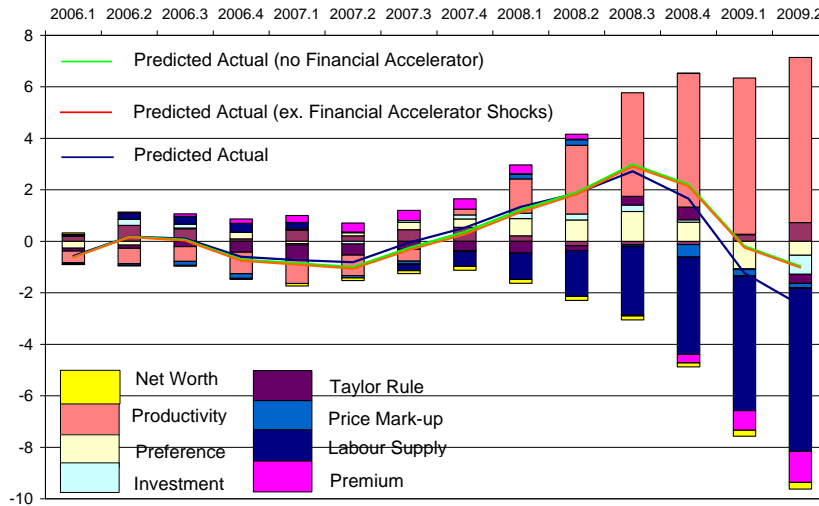
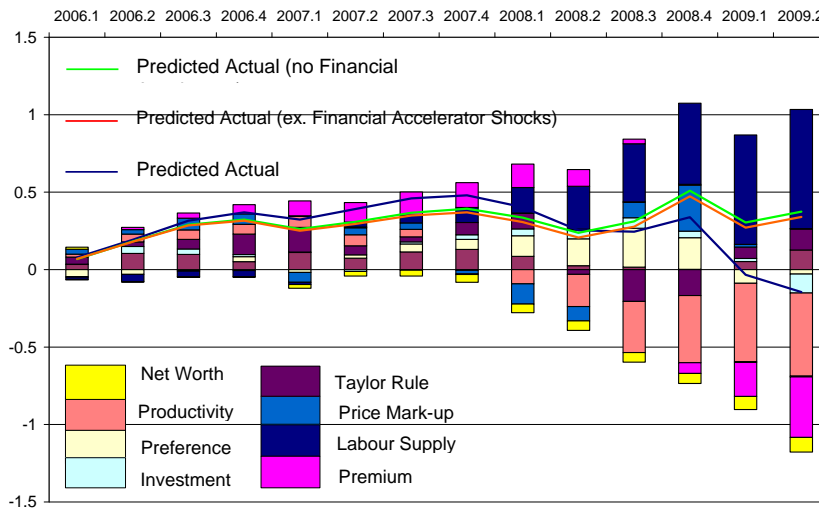Figure 6: Shock Decomposition for Output During the Banking Crisis Episode

Figure 7: Shock Decomposition for Interest Rate During the Banking Crisis Episode

We also ran the scenarios with the full set of shocks including the crisis period, $1984 - 2009$; this not surprisingly produced a somewhat greater average frequency, of one about every 61 years. Since the main shock specific to the crisis period of $2006 - 9$ is the Sub-Prime shock, one can think of this extra frequency as the result of this shock being included.

If one defines a Great Depression as a fall in GDP of 5% or more lasting for 5 years or more (before GDP returns to its previous level) then they occur on average once every 200 years if we use the sample without crisis shocks and the frequency rises to 121 years if the crisis shocks are included.

Plainly these figures are affected by the nature of the sample shocks; here we have used the experience of the last three decades, which apart from the crisis itself was the period of the Great Moderation. As we know that the variance of shocks in this period was markedly lower than in earlier post-war US history, extending our sample backwards in time would no doubt change our estimates in detail. However, the last three decades seems the most relevant experience for today's policymakers.

b) When there is crisis, there is also a financial crisis about a tenth of the time, when one uses the shocks for the non-crisis period. We measure this here by the appearance of an abnormal premium rise accompanying a crisis fall in output. This is shown for the same scenarios by showing the corresponding external premium behaviour.
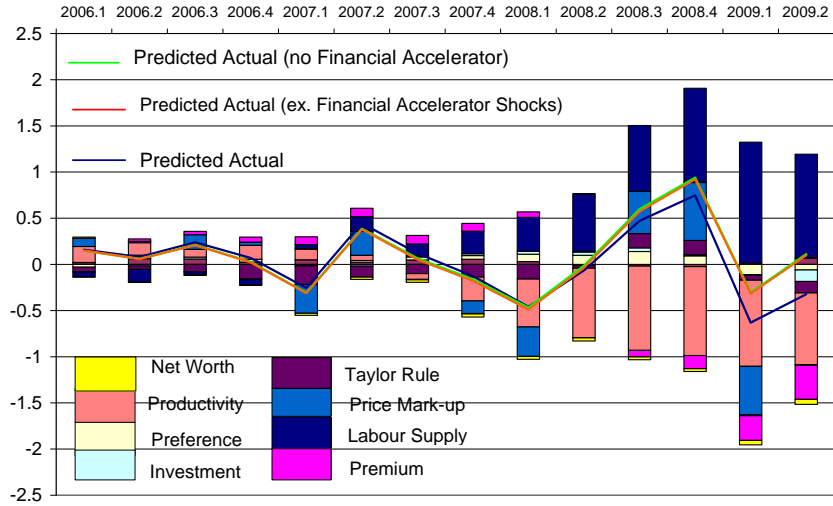
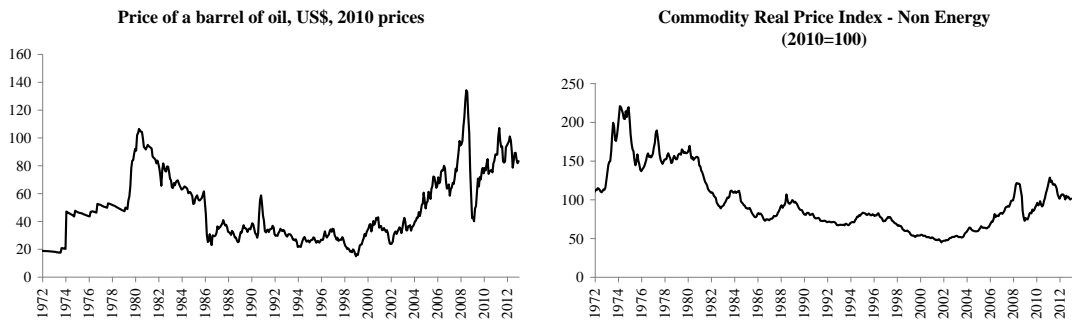Figure 8: Shock Decomposition for Inflation Rate During the Banking Crisis Episode

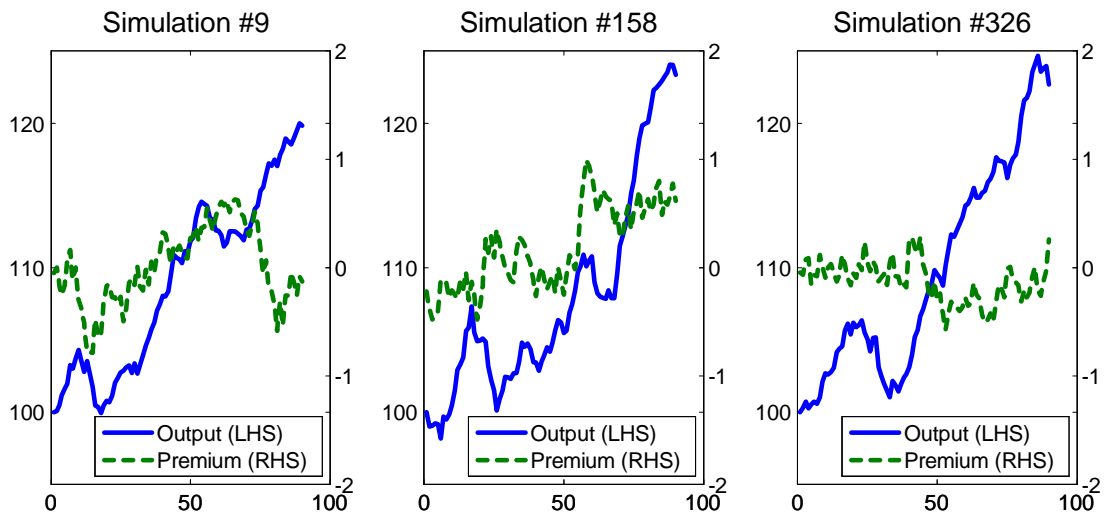Figure 9: History of Commodities Prices

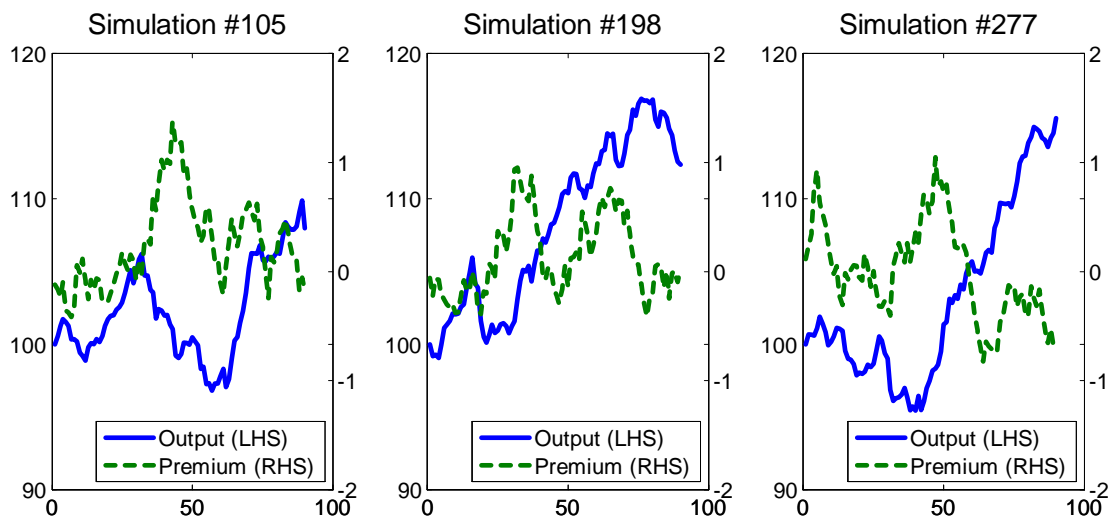Figure 10: Crises Not Accompanied by Financial Crisis

Figure 11: Crises Accompanied by Financial Crisis

c) An extreme financial shock is not required to produce a financial crisis. This is evident from the charts above since the financial shocks from $1984 - 2007$ used for these simulations were none of them extreme and yet we clearly got several financial crises. Figure 12 shows the premium shock during this sample period, the only financial shock that contributed to the crisis; as can be seen it varies on a small scale, compared with its severity over the crisis (to the right of the red line).
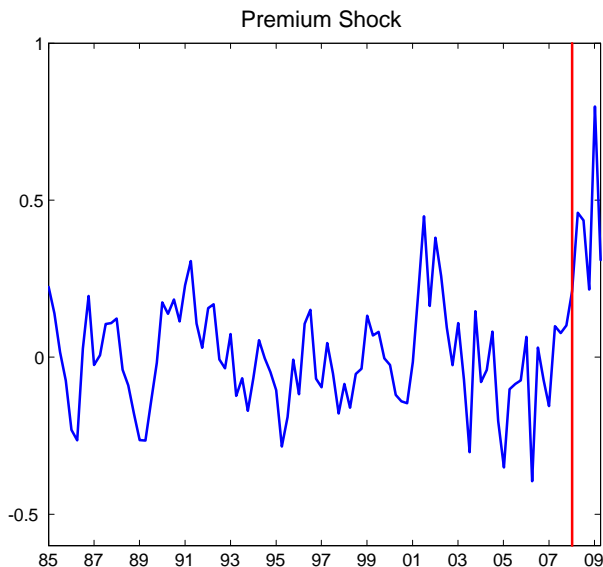
Figure 12: Premium Shock for the 84-09 Period

d) A financial shock is not sufficient to produce a crisis, even though it produces a rise in the premium. To check this point we redid these scenarios with just the two financial shocks including the crisis period values; thus this shock series includes both normal and extreme financial shocks. If financial crisis can be the result of extreme financial shocks, we should obtain a few at least. However we obtained absolutely no crises at all. What we see from the three simulations in Figure 13 is that even though our financial shock series is effectively non-stationary it does not cause a crisis; all it does is cause run-ups in the financial premium, but these do not count as financial crises if there is no accompanying crisis (i.e. there is a partially-permanent downturn in output). Here we should emphasise that the extreme financial shocks in the sample included the effects of massive government intervention, which occurred largely because of the experience of the Great Depression when there was no such intervention; thus this particular finding relies crucially on the assumption that financial shocks are accompanied by vigorous lender-of-last-resort activity by governments.

5   Conclusions

We have taken the Smets-Wouters model of the US, derived from Christiano et al. (2005), but here in the form as modified by Le et al.(2011) to allow for more heterogeneity in price/wage behaviour, and we have integrated into it the banking/financial accelerator model of Bernanke et al. (1999) in order to discover how far the banking crisis might have been caused by non-banking and by banking shocks. We began by estimating the model to get it as close as possible to the data on the indirect inference test we are using.

We then used the model with its re-estimated parameters to carry out an accounting exercise in the shocks causing the crisis episode. This was done on unfiltered data, allowing for non-stationary shocks. We did a variance decomposition to establish what a typical crisis generated by these shocks if redrawn randomly would be caused by. We then looked at the decomposition for this particular episode. Finally we ran a variety of simulations bootstrapped from different sets of the shocks in our sample (over the last three decades, on the grounds that this is of most relevance today) to shed light on the causes of crisis and banking crisis.
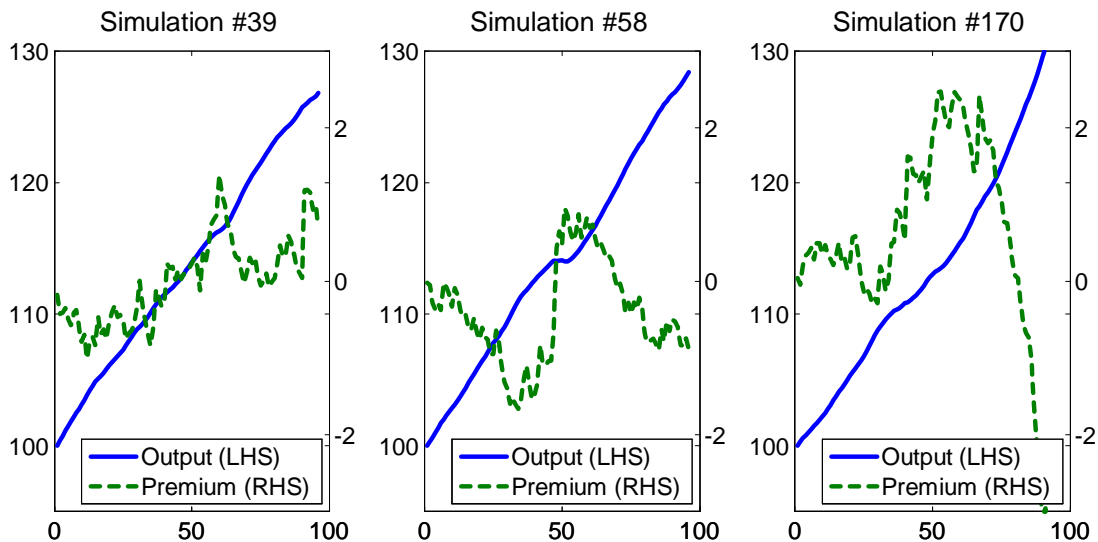
Figure 13: Simulation with only Financial Shocks

Our conclusion is perhaps rather startling: the banking crisis was mainly the result of non-banking shocks impacting through the usual non-financial channels on the US economy. The main non-bank shocks to output were from productivity (largely positive) and labour supply (negative from 'wage push'). Monetary policy shocks, apparently related to the zero bound problem, also contributed. The banking shock was a specific result of the Sub-Prime process; it contributed a further 1% drop in output; however, government direct fiscal action through transfers was designed to alleviate this shock.

We interpret these results as telling us that the banking system is integral to the functioning of the capitalist economy but that it is essentially responsive to the economy; nor does its transmission mechanism worsen economic instability. The sources of boom and slump remain those identified in non-banking models: shocks to productivity, including importantly those coming from the commodity sector, and to some extent shocks to the household sector mainly via labour supply. However we have also identified an independent shock to the banking sector, the Sub-Prime shock. This did not emerge from 'normal banking behaviour', as is made clear by its status as a pure shock. It was alleviated by government fiscal intervention; this was clearly an important element in this financial crisis.

The model also tells us that crises are regular occurrences in capitalist economies and that they frequently have as their by-product financial crisis in the sense that the premium rises sharply — a similar finding that the recent Great Recession was a product of the same shocks, including financial, that occurred earlier during the post-war period, but on this occasion unusually large, is made by Stock and Watson (2012) using a dynamic factor VAR. These crises/financial crises occur in spite of there being no extreme financial shocks such as occurred in the recent episode; so serious financial shocks are not required for crises to happen. Furthermore, extreme financial shocks on their own of the type identified in this sample do not cause crises; all they do is cause temporary recessions. Thus both crises and financial crises result from non-financial shocks; financial shocks if extreme enough will add an extra layer of recession. Again, we must stress the caveat that the financial shocks identified in this sample all occurred in a political environment where the government acted as lender of last resort; absent this, the scale of these shocks would have been no doubt very different.

Is there then a role for regulation of bank behaviour in such a system? Plainly regulation could not have stopped this bank crisis since it was not caused by bank behaviour. In some countries other than the US (e.g. Spain and Australia) banks were prevented from buying sub-prime CDOs by central banks that did not permit the 'special vehicles' through which these were usually held. In the US the crisis could have been prevented by limits on sub-prime loans in place of the encouragement that seems to have been given to them by Washington. The experience of other countries suggests that their regulators should have stopped the CDOs based on them being passed across the Atlantic so easily or else that other measures, such as creating far greater transparency in instruments such as these CDOs, should have been adopted. This points to a regulative system that puts backstop prudential limits in place.

None of this has much to do with the currently proposed regulation of the banking system in the US

20

and elsewhere. For example the Dodd-Frank legislation, which seems motivated by the aim of 'stopping future crises', represents a huge intervention in banking activity, that seems likely to badly distort and even stifle the bank transmission mechanism.

As far as fiscal and monetary policy go, we have found in common with many others that the zero bound is a problem, though its quantitative effect in this episode seems to have been small. We have also found that public spending shocks have had little effect. However fiscal policy mainly operated in this episode via large transfers to the banking and non-bank business sector; these transfers are wrapped up in the 'banking shocks' since they impacted directly on the credit risk-premium. Fiscal policy was effectively policy for credit supply from the taxpayer (as was Quantitative Easing which started in 2008 Q4); it appears that this fiscal policy was effective, if we compare the episode with the Great Depression where its absence seems to have led to a much larger shock.

References

[1] Bernanke, B. S., Gertler, M., Gilchrist, S. (1999). The financial accelerator in a business cycle framework, Handbook of Macroeconomics, vol. 1, edited by J.B. Taylor and M. Woodford, ch. 21, 1341-1393, Elsevier.

[2] Canova, F. (2005). Methods for Applied Macroeconomic Research, Princeton University Press, Princeton.

[3] Christiano, L.J., Eichenbaum, M., Evans, C.L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Journal of Political Economy, 113(1), 1-45.

[4] Christiano, L.J, Motto, R. & Rostagno, M. (2010). Financial factors in economic fluctuations. Working Paper Series 1192, European Central Bank.

[5] Del Negro, M & Schorfheide, S. (2006). How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models. Economic Review, Federal Reserve Bank of Atlanta, issue Q 2, pages 21-37.

[6] Del Negro, M, Schorfheide, F., Smets, F & Wouters, R. (2007). On the Fit of New Keynesian Models. Journal of Business & Economic Statistics, 25 April, 123-143.

[7] Fahr, S., Motto, R., Rostagno, M., Smets, F. & Tristani, O. (2011). A monetary policy strategy in good and bad times: Lessons from the recent past. Working Paper Series 1336, European Central Bank.

[8] Gilchrist, S., Ortiz, A. & Zakrajsek, E. (2009). Credit Risk and the Macroeconomy: Evidence from an Estimated DSGE Model. Unpublished manuscript, Boston University.

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

[10] Gourieroux, C., Monfort, A., Renault, E. (1993). Indirect inference. Journal of Applied Econometrics 8, S85–S118.

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

[12] Gregory, A., Smith, G. (1993). Calibration in macroeconomics, in: Maddala, G. (Ed.), Handbook of Statistics vol. 11, Elsevier, St. Louis, Mo., pp. 703–719.

[13] Ingber, L. (1996) Adaptive simulated annealing (ASA): Lessons learned. Lester Ingber Papers 96as, Lester Ingber.

[14] Jermann, U & Quadrini, V. (2012). Macroeconomic Effects of Financial Shocks. American Economic Review, 102(1), 238–71.

[15] Kimball, M.S., 1995. The Quantitative Analytics of the Basic Neomonetarist Model. Journal of Money, Credit and Banking, 27(4), 1241–77.

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

[17] Le, V.P.M., Meenagh, D., Minford, P., Wickens, M. (2011). How much nominal rigidity is there in the US economy — testing a New Keynesian model using indirect inference. Journal of Economic Dynamics and Control 35(12), 2078–2104.

[18] Le, V.P.M., Meenagh, D., Minford, P., Wickens, M., (2012). Testing DSGE models by indirect inference and other methods: some Monte Carlo experiments. Cardiff Economics Working Paper E2012/15.

[19] McCallum, B.T., (1976). Rational expectations and the natural rate hypothesis: some consistent estimates. Econometrica 44, 4–52.

[20] Meenagh, D., Minford, P., Wickens, M. (2012). Testing macroeconomic models by indirect inference on unfiltered data. Cardiff Economics Working Papers E2012/17.

[21] Minford, P, Theordoridis, K, Meenagh, D. (2009). Testing a model of the UK by the method of indirect inference. Open Economies Review, 20(2), 265–291.

[22] Smets, F., Wouters, R. (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. Journal of the European Economic Association, 1(5), p1123–1175.

[23] Smets, F., Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. American Economic Review 97, 586–606.

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

[25] Stock, J. H. and Watson, M. W. (2012). Disentangling the Channels of the 2007-2009 Recession, Brookings Papers on Economic Activity.

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

Appendix 1:   Model Listing

Consumption Euler equation

$$c_t = \frac{\frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}} c_{t-1} + \frac{1}{1 + \frac{\lambda}{\gamma}} E_t c_{t+1} + \frac{(\sigma_c - 1)\frac{W_* L_*}{C_*}}{\left(1 + \frac{\lambda}{\gamma}\right)\sigma_c}\left(l_t - E_t l_{t+1}\right) - \left(\frac{1 - \frac{\lambda}{\gamma}}{\left(1 + \frac{\lambda}{\gamma}\right)\sigma_c}\right)\left(r_t - E_t p_{t+1}\right) + eb_t \quad (7)$$

Investment Euler equation

$$inn_t = \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}} inn_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1 + \beta\gamma^{(1-\sigma_c)}} E_t inn_{t+1} + \frac{1}{\left(1 + \beta\gamma^{(1-\sigma_c)}\right)\gamma^2 \varphi} qq_t + einn_t \quad (8)$$

Tobin Q equation

$$qq_t = \frac{1 - \delta}{1 - \delta + R_*^K} E_t qq_{t+1} + \frac{R_*^K}{1 - \delta + R_*^K} E_t rk_{t+1} - E_t cy_{t+1} \quad (9)$$

Capital Accumulation equation

$$k_t = \left(\frac{1 - \delta}{\gamma}\right) k_{t-1} + \left(1 - \frac{1 - \delta}{\gamma}\right) inn_t + \left(1 - \frac{1 - \delta}{\gamma}\right)\left(1 + \beta\gamma^{(1-\sigma_c)}\right)\left(\gamma^2\right)(\varphi)(einn_t) \quad (10)$$

Price Setting equation

$$rk_t = \omega^r \left[\begin{array}{c} \overline{\alpha\left(\frac{1}{1+\beta\gamma^{(1-\sigma_c)}\iota_P}\right)\left(\frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_p\right)\left(1-\xi_p\right)}{\xi_p\left((\phi_p-1)\epsilon_p+1\right)}\right)} \\ \left\{ \begin{array}{c} p_t - \frac{\beta\gamma^{(1-\sigma_c)}\iota_P}{1+\beta\gamma^{(1-\sigma_c)}\iota_P} E_t p_{t+1} - \frac{\iota_P}{1+\beta\gamma^{(1-\sigma_c)}\iota_P} p_{t-1} + \left(\frac{1}{1+\beta\gamma^{(1-\sigma_c)}\iota_P}\right) \\ \left(\frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_p\right)\left(1-\xi_p\right)}{\xi_p\left((\phi_p-1)\epsilon_p+1\right)}\right)\left((1-\alpha)w_t - ea_t\right) - ep_t \end{array} \right\} \end{array} \right]$$
$$+(1-\omega^r)\left[\frac{ea_t}{\alpha} - \frac{1-\alpha}{\alpha}w_t\right] \quad (11)$$

Wage Setting equation

$$w_t = \omega^w \left[\begin{array}{c} \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} E_t w_{t+1} + \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} w_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} E_t p_{t+1} - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} p_t \\ + \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} p_{t-1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} \left(\frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_w\right)\left(1-\xi_w\right)}{(1+\epsilon_w(\phi_w-1))\xi_w}\right) \\ \left(w_t - \sigma_l l_t - \left(\frac{1}{1-\frac{\lambda}{\gamma}}\right)\left(c_t - \frac{\lambda}{\gamma}c_{t-1}\right)\right) + ew_t \end{array} \right] +$$
$$(1-\omega^w)\left[\sigma_l l_t + \left(\frac{1}{1-\frac{\lambda}{\gamma}}\right)\left(c_t - \frac{\lambda}{\gamma}c_{t-1}\right) - (\pi_t - E_{t-1}\pi_t) + ew_t^S \right] \quad (12)$$

Labour demand

$$l_t = -w_t + \left(1 + \frac{1 - \psi}{\psi}\right) rk_t + k_{t-1} \quad (13)$$

Market Clearing condition in goods market

$$y_t = \frac{C}{Y}c_t + \frac{I}{Y}inn_t + R_*^K k_y \frac{1 - \psi}{\psi} rk_t + c_y^e c_t^e + eg_t \quad (14)$$

Aggregate Production equation

$$y_t = \phi\left[\alpha\frac{1 - \psi}{\psi}rk_t + \alpha k_{t-1} + (1 - \alpha)l_t + ea_t\right] \quad (15)$$

Taylor Rule

$$r_t = \rho r_{t-1} + (1 - \rho)(r_p p_t + r_y y_t) + r_{\Delta y}(y_t - y_{t-1}) + er_t \quad (16)$$

Premium

$$E_t cy_{t+1} - (r_t - E_t p_{t+1}) = \chi (qq_t + k_t - n_t) + epr_t \tag{17}$$

Net worth

$$n_t = \frac{K}{N} (cy_t - E_{t-1}cy_t) + E_{t-1}cy_t + \theta n_{t-1} + enw_t \tag{18}$$

Entrepreneurial consumption

$$c_t^e = n_t \tag{19}$$

Appendix 2:   Using the VECM as the auxiliary model

After log-linearisation a DSGE model can usually be written in the form

$$A(L)y_t = BE_t y_{t+1} + C(L)x_t + D(L)e_t \tag{A1}$$

where y_t are p endogenous variables and x_t are q exogenous variables which we assume are driven by

$$\Delta x_t = a(L)\Delta x_{t-1} + d + c(L)\epsilon_t. \tag{A2}$$

The exogenous variables may contain both observable and unobservable variables such as a technology shock. The disturbances e_t and ϵ_t are both iid variables with zero means. It follows that both y_t and x_t are non-stationary. L denotes the lag operator $z_{t-s} = L^s z_t$ and $A(L)$, $B(L)$ etc. are polynomial functions with roots outside the unit circle.

The general solution of y_t is

$$y_t = G(L)y_{t-1} + H(L)x_t + f + M(L)e_t + N(L)\epsilon_t. \tag{A3}$$

where the polynomial functions have roots outside the unit circle. As y_t and x_t are non-stationary, the solution has the p cointegration relations

$$\begin{aligned} y_t &= [I - G(1)]^{-1}[H(1)x_t + f] \\ &= \Pi x_t + g. \end{aligned} \tag{A4}$$

The long-run solution to the model is

$$\begin{aligned} \overline{y}_t &= \Pi \overline{x}_t + g \\ \overline{x}_t &= [1 - a(1)]^{-1}[dt + c(1)\xi_t] \\ \xi_t &= \Sigma_{i=0}^{t-1}\epsilon_{t-s} \end{aligned}$$

Hence the long-run solution to x_t, namely, $\overline{x}_t = \overline{x}_t^D + \overline{x}_t^S$ has a deterministic trend $\overline{x}_t^D = [1 - a(1)]^{-1}dt$ and a stochastic trend $\overline{x}_t^S = [1 - a(1)]^{-1}c(1)\xi_t$.

The solution for y_t can therefore be re-written as the VECM

$$\begin{aligned} \Delta y_t &= -[I - G(1)](y_{t-1} - \Pi x_{t-1}) + P(L)\Delta y_{t-1} + Q(L)\Delta x_t + f + M(L)e_t + N(L)\epsilon_t \\ &= -[I - G(1)](y_{t-1} - \Pi x_{t-1}) + P(L)\Delta y_{t-1} + Q(L)\Delta x_t + f + \omega_t \\ \omega_t &= M(L)e_t + N(L)\epsilon_t \end{aligned} \tag{A5}$$

Hence, in general, the disturbance ω_t is a mixed moving average process. This suggests that the VECM can be approximated by the VARX

$$\Delta y_t = K(y_{t-1} - \Pi x_{t-1}) + R(L)\Delta y_{t-1} + S(L)\Delta x_t + g + \zeta_t \tag{A6}$$

where ζ_t is an iid zero-mean process.

As

$$\overline{x}_t = \overline{x}_{t-1} + [1 - a(1)]^{-1}[d + \epsilon_t]$$

the VECM can also be written as

$$\Delta y_t = K[(y_{t-1} - \overline{y}_{t-1}) - \Pi(x_{t-1} - \overline{x}_{t-1})] + R(L)\Delta y_{t-1} + S(L)\Delta x_t + h + \zeta_t. \tag{A7}$$

Either equations (A6) or (A7) can act as the auxiliary model. Here we focus on (A7); this distinguishes between the effect of the trend element in x and the temporary deviation from its trend. In our models these two elements have different effects and so should be distinguished in the data to allow the greatest test discrimination.

It is possible to estimate (A7) in one stage by OLS. Meenagh et al. (2012) do Monte Carlo experiments to check this procedure and find it to be extremely accurate.

Appendix 3:   VAR IRFs

In this appendix we show how the model, given that it jointly predicts (within 95% bounds) the VAR coefficients that determine the IRFs of shocks on the three key macro variables, thereby also broadly predicts these IRFs. Because the Wald test is of the joint behaviour of the VAR coefficients and on the variances of the three variable residuals, there is not a perfect correspondence with the individual IRFs. However, it can be seen, as expected, that most of the IRFs lie mostly within the bounds.

It is the IRFs that policymakers are interested in, as pointed out by Christiano et al. (2005). They need to be assured that empirically the IRFs the model implies should appear in the data actually do so within statistical bounds (of course the IRFs implied for data behaviour reflect both the model structural IRFs and sample shock variations). Then they feel able to use the model's (structural) IRFs to determine the effect of shocks and of policies to offset shocks.

The VAR innovations are identified throughout by the model; we have no independent way of identifying the VAR innovations (any such ways suggested are based on some 'non-controversial' model restrictions; however, the model here is non-controversial in its current innovation structure and so we use it.) The testing kicks in on the variances of the VAR innovations and on the lagged effects of each variable (the VAR coefficients).
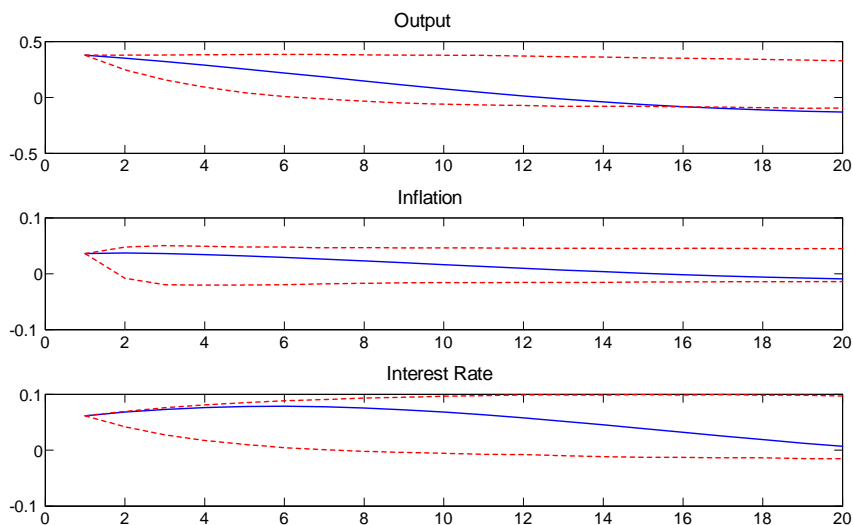

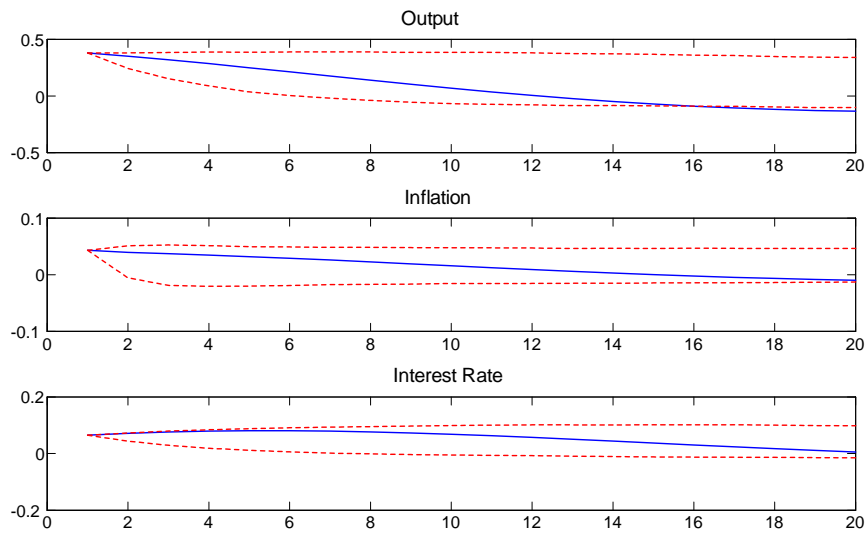
Figure 14: VAR IRFs for an exogenous demand shock

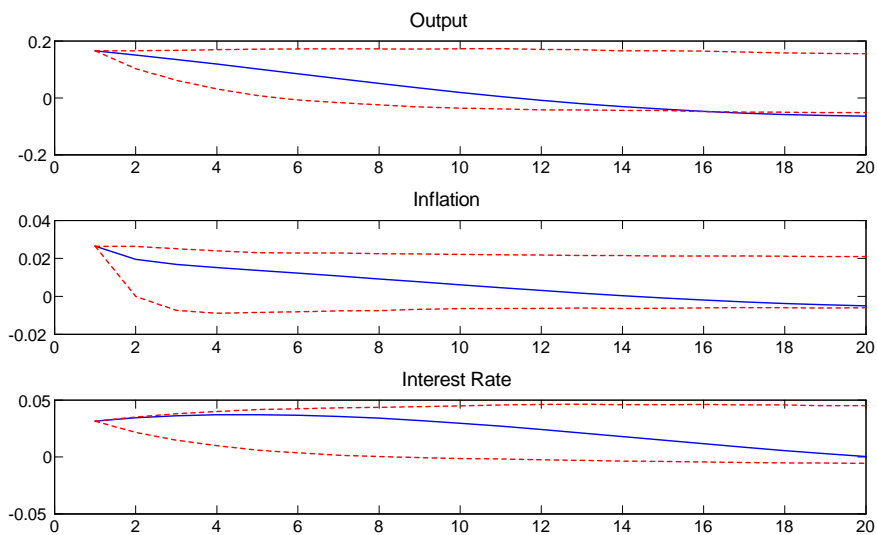Figure 15: VAR IRFs for a consumer preference shock

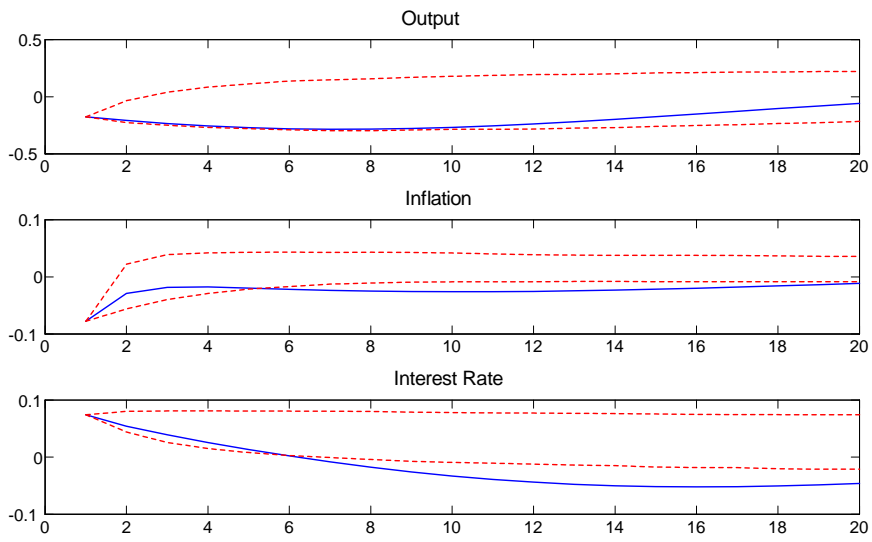Figure 16: VAR IRFs for an investment shock
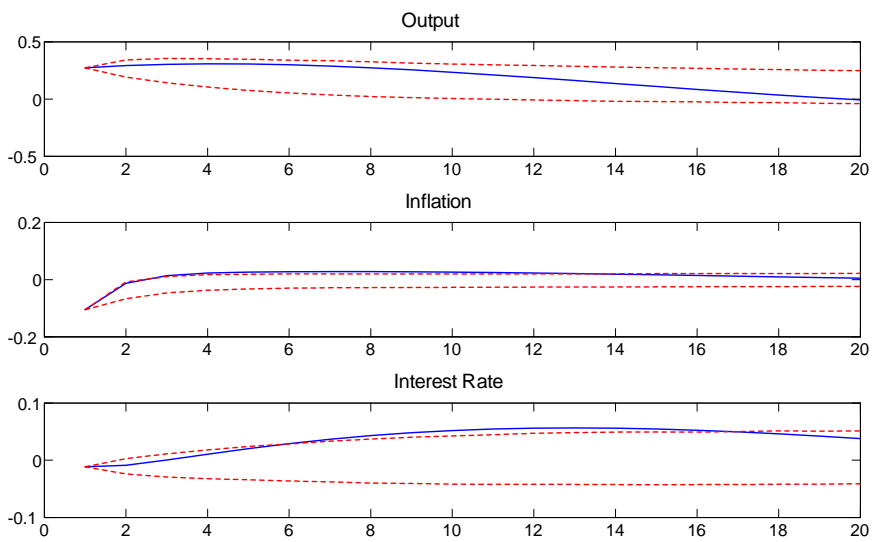
26

Figure 17: VAR IRFs for a monetary shock

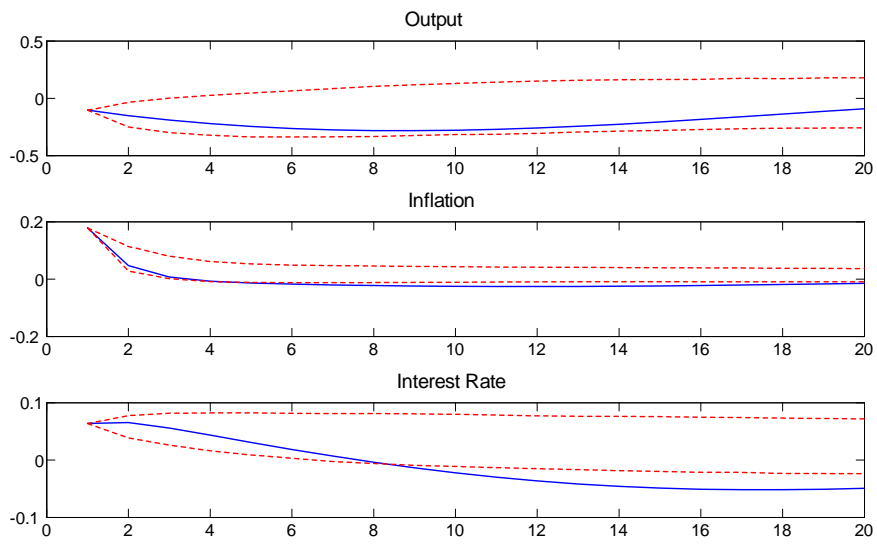Figure 18: VAR IRFs for a productivity shock
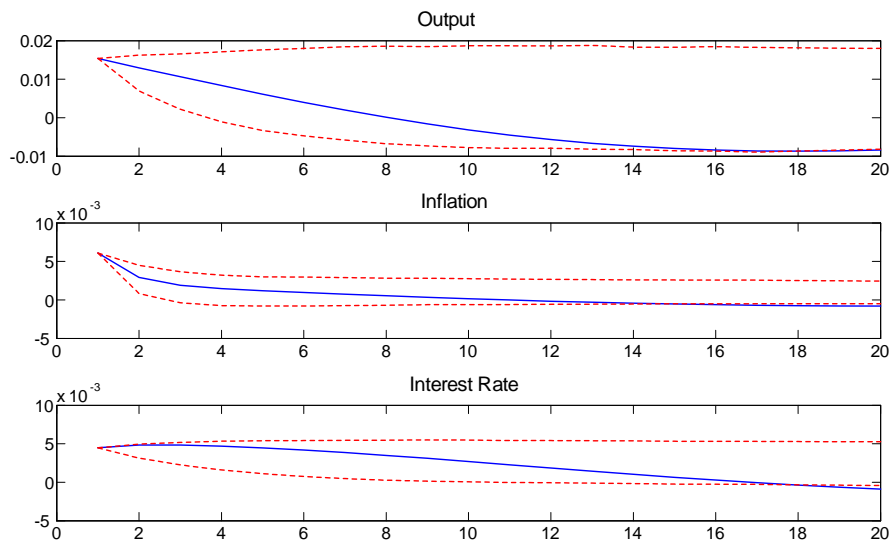
Figure 19: VAR IRFs for a price mark-up shock

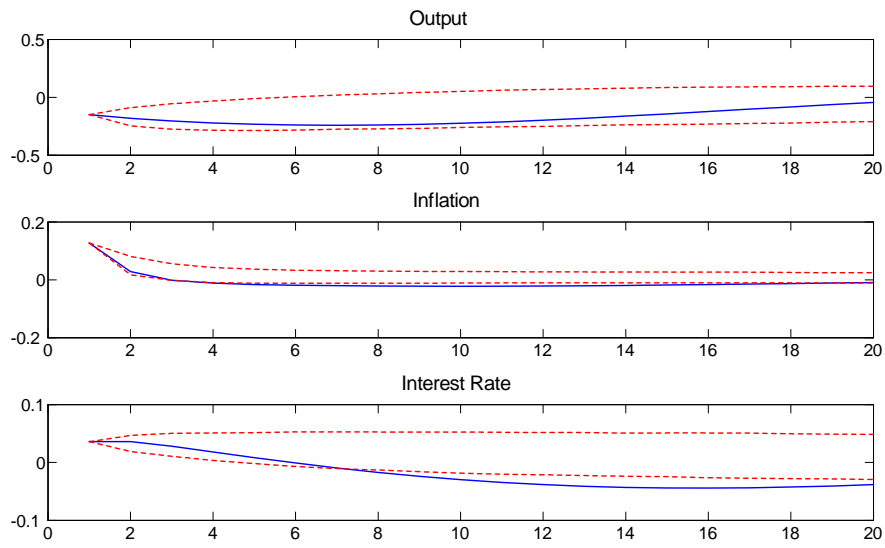Figure 20: VAR IRFs for a wage mark-up shock

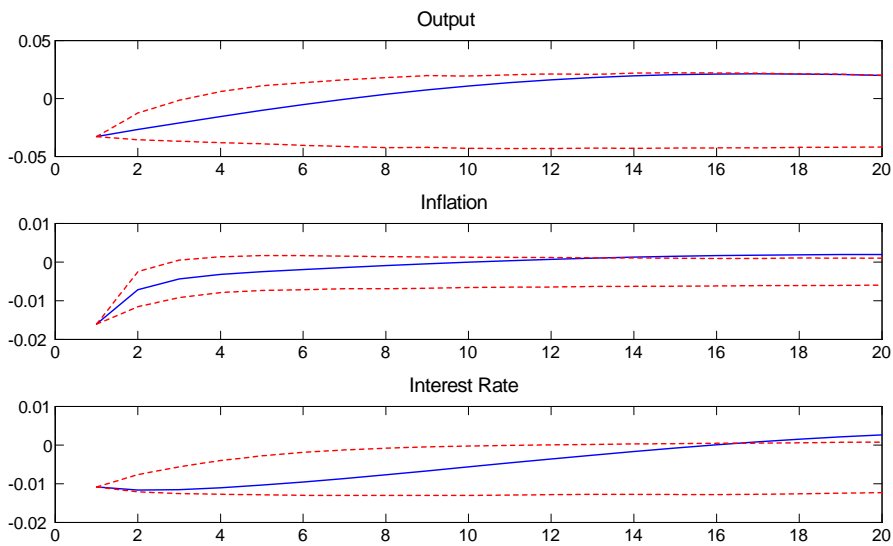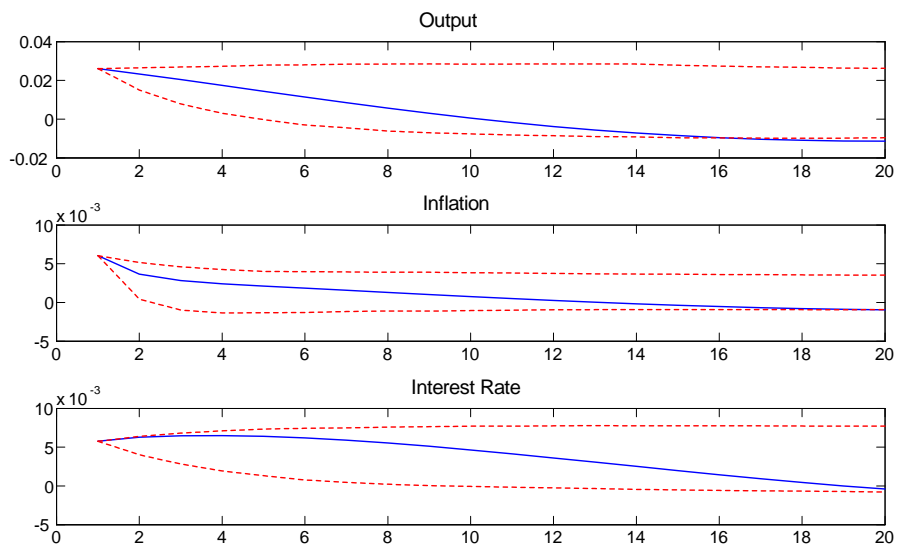Figure 21: VAR IRFs for a labour supply shock

Figure 22: VAR IRFs for a premium shock

29

Figure 23: VAR IRFs for a net worth shock