The Power of Structural VARs*
(preliminary)

Robert Vigfusson[†]

November 1, 2007

Abstract

Are structural vector autoregressions (VARs) useful for discriminating between macro models? Recent assessments of VARs have shown that these statistical methods have good size properties. In other words, in simulation exercises, VARs will only infrequently reject the true data generating process. However, in assessing a statistical test, we often also care about power: the ability of the test to reject a false hypothesis. Much less is known about the power of structural VARs.

In this paper, I attempt to fill in this gap by exploring the power of long-run structural VARs against a set of DSGE models that vary in degree from the true data generating process. I report results for two tests: the standard test of checking the sign on impact and a test of the shape of the response. For the models studied here, testing the shape is a more powerful test than simply looking at the sign of the response. In addition, relative to an alternative statistical test based on sample correlations, I find that the shape-based tests have greater power. Given the results on the power and size properties of long-run VARs, I conclude that these VARs are useful for discriminating between macro models.

JEL Codes: C1

Keywords: Vector autoregression, dynamic stochastic general equilibrium model, confidence intervals, impulse response functions, identification, long run restrictions, specification error, sampling

---

*I thank colleagues at the Federal Reserve Board of Governors for useful comments. The views in this paper are solely those of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

[†]Federal Reserve Board of Governors. Email: robert.j.vigfusson@frb.gov

1 Introduction

Are structural vector autoregressions (VARs) useful for discriminating between macro models? Recent assessments of VARs have shown that these statistical methods have good size properties. In other words, in simulation exercises, VARs only infrequently reject the true data generating process. However, in assessing a statistical test, we often also care about power: the ability of the test to reject a false hypothesis. Much less is known about the power of structural VARs.

In this paper, I attempt to fill in this gap by exploring the power of long-run structural VARs against a set of DSGE models that vary in degree from the true data generating process. I report results for two tests: the standard test of checking the sign on impact and a test of the shape of the response. For the models studied here, testing the shape is a more powerful test than simply looking at the sign of the response. In addition, relative to a statistical test based on sample correlations, I find that the shape-based tests have greater power. Given the results on the power and size properties of long-run VARs, I conclude that these VARs are useful for discriminating between macro models.

Several recent papers have identified a technology shock in the U.S. macroeconomic data using a long-run restriction. These papers include Gali (1999); Francis and Ramey (2003); and Altig, Christiano, Eichenbaum, and Linde (2004). Identifying how the economy responds to a technology shock has the potential to be useful to help us determine how to best model the economy.

Although these restrictions are popular, some researchers have criticized using these identification restriction. In particular, discussions of power in long-run VARs may appear pointless given Faust and Leeper's proposition that any test of an impulse response has significance level greater than or equal to maximum power. (Faust and Leeper 1997 p. 347). However, Faust and Leeper's claim is that the model has such power properties against very general data generating processes (DGP). Here, the set of DGP is restricted to be only those DGPs that are consistent with being generated by a particular class of macroeconomic models. Faust and Leeper themselves note that fixing the DGP to be a finite-ordered VAR with maximum lag length K would be a sufficient restriction to make their proposition inapplicable. However, they view such a restriction as being implausible. The gain from the current paper is to show that, for data simulated from popular macroeconomic models, which implies an infinite order VAR, these long-run VARs do have power to reject false null hypothesis at a rate greater than the size of the test.

To place the current paper in the literature, it is useful to review three recent papers that have used simulation evidence from macro models to study these long-run VARs: Erceg, Guerrieri, and Gust (2004) [EGG]; Chari Kehoe and McGrattan (2007) [CKM]; and Christiano, Eichenbaum, and Vigfusson 2006 [CEV]. (For convenience, I refer to these papers by the bracketed initials.) The three papers differ in their choice of DGP and also their choice of what results to report. CKM only report results for parameterizations of RBC model without any adjustment costs and they only report on the distribution of the point estimate. The EGG paper reports results for two different models, a RBC model and also a sticky price model. They compare the models by looking at the relative frequency of the hour's response having positive point estimates. The CEV paper differs from the other two in that it also reports estimated bootstrapped standard errors.[1] Both EGG and CKM report on the fraction of point estimates that are the wrong sign when compared with the true response from the DGP. Reporting this fraction does not match up well with standard econometric practice which tests a null hypothesis using both the point estimate and the associated standard error from a single data set. In studying statistical size properties, the CEV paper and the current paper better match standard econometric practice. As such, these papers provide information that should be more useful to the applied researcher.

The current paper goes beyond the CEV paper in reporting power properties of long-run VARs. In addition, this paper also emphasizes the usefulness of looking at the shape of the impulse responses. As such, the current paper's main contributions are to suggest new tools for applied researchers (the emphasis on shape) and to give applied researchers renewed confidence in applying these methods.

The next two sections describe how to estimate a long-run VAR and the macro models that are used as data generating processes. These sections should be familiar to readers of the previous papers. Section 5 presents the simulation results. Section 6 reports on an empirical application of the methods to U.S. data. Section 7 concludes.

[1] The working paper version of the EGG paper reported size but not power properties of impulse responses. However, the size-related material was removed before journal publication.

2 Estimating A Vector Autoregression with A Long Run Identification Assumption

Here, as in Galí (1999), a technology shock is identified as a permanent shock to productivity. These shocks and resulting impulse responses are computed in the following manner. Consider the following structural vector autoregression (VAR) for a vector of variables Y_t :

$$A_0 Y_t = A(L) Y_{t-1} + \begin{pmatrix} \varepsilon_t^z \\ v_t \end{pmatrix} \tag{1}$$

The fundamental shocks ε_t^z and v_t (where v_t has $n-1$ elements) are assumed to be independent, have mean zero and have variances equal to one. The vector Y_t consists of n elements. The first element is the growth rate of labor productivity, denoted by Δa_t. The next $n-1$ elements are the other variables in the VAR x_t. As such Y_t can be written as

$$Y_t = \begin{pmatrix} \Delta a_t \\ x_t \end{pmatrix}. \tag{2}$$

Given this structural VAR, we can invert A_0 to construct the reduced form VAR

$$Y_t = A_0^{-1} A(L) Y_{t-1} + A_0^{-1} \begin{pmatrix} \varepsilon_t^z \\ v_t \end{pmatrix}, \tag{3}$$

where the reduced form VAR coefficients $A_0^{-1}A(L)$ are denoted by $B(L)$ and the reduced form errors are denoted by u_t.For notational simplicity let C denote A_0^{-1}. The mapping between structural shocks and reduced form errors is

$$u_t = C \begin{pmatrix} \varepsilon_t^z \\ v_t \end{pmatrix}. \tag{4}$$

Denote the variance covariance matrix of u_t, $Eu_t u_t'$ by V and note by assumption that V equals CC'. As such the reduced form is the following:

$$Y_t = B\left(L\right) Y_{t-1} + u_t. \tag{5}$$

Galí identifies the technology shock by assuming that only the technology shock ε_t^z can have a permanent effect on the level of productivity z_t. All other shocks are assumed to have no long-run effect. This restriction is referred to the exclusion restriction as it excludes the other shocks from having any long run effect on the level of productivity. This restriction imposes a restriction on the moving average representation of the data. Denote the moving average representation by:

$$Y_t = [I - B(L)]^{-1} C \begin{pmatrix} \varepsilon_t^z \\ v_t \end{pmatrix}. \tag{6}$$

The exclusion restriction implies that that each element in the top row of the sum of moving average coefficients equals zero except for the first element. In other words, we have the following restriction

$$[I - B(1)]^{-1} C = \begin{bmatrix} C_{11} & \underline{0} \\ \text{numbers} & \text{numbers} \end{bmatrix}, \tag{7}$$

where $\underline{0}$ is a row vector. To identify C_{11} requires the additional sign restriction that a positive technology shock increases labor productivity which implies that C_{11} is positive. No additional restrictions are required.

To compute the dynamic effects of ε_t^z, we require $B_1, ..., B_q$ and C_1, the first column of C. The symmetric matrix, V, and the B_i's can be computed using ordinary least squares regressions. However, the requirement that $CC' = V$ is not sufficient to determine a unique value of C_1. There are many matrices, C, that satisfy $CC' = V$ as well as the exclusion and sign restrictions. However, in all cases, the first column, C_1, of each of these matrices is the same. In particular we can compute a C that satisfies these restrictions as the following

$$C = [I - B(1)]\, D \tag{8}$$

where D is the lower triangular matrix such that

$$DD' = [I - B(1)]^{-1} V [I - B(1)']^{-1} = S_Y(0). \tag{9}$$

where $S_Y(\omega)$ denote the spectral density of Y_t at frequency ω that is implied by the q^{th} order VAR. The use of the spectral density at frequency zero to identify a technology shock is closely connected to the critique of Faust and Leeper.

3 Models

The model presented here is very similar to the standard quantitative dynamic flexible-price model presented in Christiano Eichenbaum and Vigfusson (2006). The model, however, has two additional features. The first is the addition of habit persistence in the utility function. Thus, the previous period's level of consumption affects current utility. Habit persistence results in a slower response by consumption. The second feature is adding investment adjustment costs to the model. Increasing investment is expensive and therefore an economic agent will have an incentive to smooth out investment. Christiano, Eichenbaum, and Evans (2005) use a similar specification to generate improved dynamics in a sticky price model.

3.1 The Utility Function

The model has a representative agent who chooses consumption C and the fraction of time spent working H to maximize utility, where utility is defined as

$$E_t \sum^{j} (\beta (1+g))^j (\log (C_{t+j} - bC_{t+j-1}) + \eta \log (1 - H_{t+j})) \tag{10}$$

The coefficient b describes the degree of habit persistence in the model. The parameter β is the discount rate, g is the growth rate of population, and η controls the trade-off between consumption and leisure. The agent maximizes utility subject to the budget constrain that consumption and investment I_t must equal the return r_t from capital K_t and income from working $(1 - \tau_{lt}) w_t H_t$.

$$C_t + (1 + \tau_{x,t}) I_t \leq (1 - \tau_{lt}) w_t H_t + r_t k_t \tag{11}$$

The capital accumulation equation is the following

$$(1+\gamma) K_{t+1} = (1-\delta) K_t + \left(1 - S\left(\frac{I_t}{I_{t-1}}\right)\right) I_t, \tag{12}$$

where S is the function that determines the cost of changing investment. The value of S and its first derivative are zero along a steady state growth path and the parameter γ denotes the second derivative of S evaluated in steady state.

The production function is standard

$$C_t + I_t = K_t^{\alpha} \left(Z_t H_t\right)^{1-\alpha} \tag{13}$$

where Z_t denotes the level of technology.

There are three shocks.

$$\begin{aligned} \log z_t &= \mu_Z + \sigma_z \varepsilon_t^z & (14) \\ \tau_{lt} &= (1-\rho_l)\,\bar{\tau}_l + \rho_l \tau_{lt-1} + \sigma_l \varepsilon_t^l & (15) \\ \tau_{xt} &= (1-\rho_x)\,\bar{\tau}_x + \rho_x \tau_{xt-1} + \sigma_x \varepsilon_t^x & (16) \end{aligned}$$

where z_t equals the growth in technology Z_t/Z_{t-1}. Each of the shocks $\varepsilon_t^z, \varepsilon_t^l$, and ε_t^x is independent and identically distributed with mean zero and variance equal to one. The values of $\mu_Z, \bar{\tau}_l$,and $\bar{\tau}_x$ are the average values of the shocks. One could describe the shocks τ_{lt} and τ_{xt} as labor and capital tax rates respectively. However, estimation that matches the model to observed non-tax variables implies that these variables τ_{lt+1} and τ_{xt+1} are much more variable than observed labor and capital tax rates. The values of the auto-regressive parameters ρ_l and ρ_x are both constrained to be less than one.

4 The Problem with Long-Run VARs.

There are two problems when you estimate a long-run VAR using data simulated from a macro model. The first problem is that the true data generating process is not a finite-ordered VAR rather it is a infinite-ordered VAR model. In particular, the model has the following log-linearized solution where ξ_t are the model's state variables (such as capital), ε_t are the fundamental shocks, and Y_t are the model's observed variables (such as investment and hours worked). The solution to the model can be given by the following system of equations

$$\begin{aligned} \xi_t &= F\xi_{t-1} + G\varepsilon_t & (17) \\ Y_t &= H\xi_t & (18) \end{aligned}$$

Note that the autoregressive nature of the first equation is without log of generality as we can stack variables in the state variable vector (such as k_{t+1}, k_t, and k_{t-1}).

Given this system of equations, one can derive the following infinite ordered VAR for the observed variables Y_t (CEV 2006). The data generating process for Y_t is

$$Y_t = HF\left(I - ML\right)^{-1} GC^{-1} Y_{t-1} + C\varepsilon_t \tag{19}$$

where L is the lag operator and the following matrices are defined

$$C = HG \tag{20}$$
$$M = \left(I - DC^{-1}H\right) F \tag{21}$$

Two additional assumptions are required. The first assumption is that the matrix C be square and invertible. For C to be square requires that there be as many economic fundamental shocks as there are observed variables. If there were fewer economic shocks than observed variables, then the variance covariance matrix of Y_t is singular.[2] The second assumption is that M^j converges to zero as j goes to infinity. This assumption rules out explosive solutions. Note if $B\left(L\right)$ denotes the infinite ordered polynomial for the autoregressive terms on Y_t, then we have that

$$B\left(L\right) = HF\left(I - ML\right)^{-1} GC^{-1}. \tag{22}$$

Given this definition, the jth term of $B\left(L\right)$, B_j, equals HFM^jGC^{-1}. For the system to be non-explosive, the value of M^j must converge to zero as j goes to infinity. By satisfying this require, we would have that B_j converges to zero.

This infinite ordered VAR is typically approximated by a finite order VAR $\hat{B}\left(L\right)$ of order p where $\hat{B}_q$ equals zero for all q greater than p. There has been some debate about the ability of a finite ordered VAR to approximate the dynamics of the infinite order VAR. The short-run identification results in CEV (2006), however, suggest that a finite-ordered VAR can do fairly well at capturing the short-run dynamics. The individual estimated VAR coefficients $\left\{\hat{B}_i\right\}_{i=1}^p$ are close estimates of the individual population coefficients $\{B_i\}_{i=1}^p$ and, with a relatively small value of p, do a good job of minimizing the variance of the one-step ahead forecast errors. However, as was described in Sims and further

[2] When there are more shocks than variables, one can still derive a VAR. However, identifying all the shocks becomes more difficult. See Sims and Zha (2006) for more details.

discussed in CEV, the sum of the estimated coefficients $\hat{B}(1)$ (or equivalently $\sum_{i=1}^{p}\hat{B}_i$) may not be close to the true sum $B(1)$ ($\sum_{i=1}^{\infty} B_i$).

An inability to match the long-run sum is a particular problem for the long-run identification assumption since the value of C the mapping between reduced form shocks u_t and fundamental shocks ε_t requires knowing D where D, as already defined in equation 9, is the following

$$DD' = [I - B(1)]^{-1} V [I - B(1)']^{-1} = S_Y(0). \tag{23}$$

Because the value of D is a function of $B(1)$, the inability of the sum of the finite-order VAR's coefficients $\hat{B}(1)$ to match $B(1)$ is a problem particular to the long-run identifying assumptions.

As was mentioned in the introduction, a discussion of power of long-run VARs may appear pointless given Faust and Leeper's proposition that any test of an impulse response identified with a long-run restriction has significance level greater than or equal to maximum power. (Faust and Leeper 1997 p. 347). The identification of the long-run VAR depends on knowing the matrix D. The matrix D is a function of the spectrum at frequency zero and knowing the spectrum at frequency zero is what underlies Faust and Leeper's claim about the problems with long-run VARs. As was discussed in Faust (199x), the confidence interval on a single point on the spectrum is unbounded because, under the assumption of a very general data generating processes, one can not rule out a spike at that single point. If one can restrict the DGP sufficiently to rule out these spikes, then the spectrum and hence the matrix D will be better behaved. For example, Faust and Leeper themselves note that fixing the DGP to be a finite-ordered VAR with maximum lag length K would be a sufficient restriction to make their proposition inapplicable. However, they view such a restriction as being implausible.

In the following sections, the set of data generating processes (dgp's) is restricted to the set of infinite-ordered VARs that arise from macro-economic models. For these DGPs, I will show that long-run VARs do have power to reject false null hypothesis at a rate greater than the size of the test. Although these macro models are perhaps not fully descriptive of the data, these models are more plausible dgp's than the fixed lag VARs.

5 Model calibrations and Simulation Experiments

To simulate data from the model requires values for the models parameters. To make my results reported here comparable to CKM (2007) and CEV (2006), most model parameters are set at values that they use. See Table 1 for the values of $\{\beta, \theta, \delta, \tau_x, \tau_l, g, \psi, \mu_z, \tau_l\}$. In the first set of simulations, the values of habit parameter b and the investment adjustment costs parameter γ are set equal to zero. In subsequent simulations, I relax this restriction and simulate data from a model with the coefficient of habit persistence b and the degree of investment adjustments costs γ fixed at the values ($b = 0.7$ and $\gamma = 3$) that are reported in Christiano, Eichenbaum and Evans (2005).

As in CEV(2006), the variance and auto-correlation of the model's shocks are estimated by standard maximum likelihood methods. Define the observed vector of variables to be the following

$$Y_t = \begin{pmatrix} \Delta y_t - \Delta h_t \\ h_t \\ i_t - y_t \end{pmatrix} \tag{24}$$

where $\Delta y_t - \Delta h_t$ is the growth rate of labor productivity, h_t is the level of per capita hours worked and $i_t - y_t$ is the ratio of investment to output expressed in logs. All data is from the United States for the period 1959 to 2001. Labor productivity and hours worked are measured for the business sector. The ratio of investment to output is measured using the nominal share of total investment in GDP. Given these observed variables and the model structure results in the model set-up described by equations 17 and 18. The model can then be estimated by applying the Kalman filter approach in Hamilton (1994, Section 13.4). Estimated model coefficients match those found in CEV (2006) and are reported in Table 1.

5.1 Simulation Evidence With Data Generated from a RBC model

All simulations are done 2000 times with a sample size of 200 observations. For each simulated data set, I estimate a three variable VAR where the three variables are the growth rate of labor productivity, the log level of per capita hours worked and the ratio of investment to output expressed in log. For each VAR, I fixed the lag length at four. Based on past experience, applying more sophisticated algorithms for choosing lag length does not provide substantially different results. By applying the long-run identifying assumption, for each data set, I identify the responses to a one-standard deviation increase in technology.

For each simulated data set, I estimate a bootstrapped standard error by simulating the estimated VAR 200 times where the vector of economic shocks at time s are drawn with replacement from the estimated set of residuals and the starting values come from that particular data set. The bootstrap standard deviation is estimated as the sample statistic coming from the distribution of the bootstrapped impulse responses.

Figure 1 reports, for the benchmark VAR estimated using data simulated from a RBC model, the response of hours worked to a permanent shock to labor productivity with size equal to one-standard deviation.

The gray area indicates the sampling distribution of the estimated impulse responses. The edges of the gray area indicate the 5th and 95th percentile of all the estimated impulse responses. These intervals are wide which is typical of structural VARs that are identified with a long-run restriction (see CEV 2006).

Figure 1 also reports the true impulse responses from several parameterizations of flexible price DGE models that have real rigidities in the form of investment adjustment costs and habit persistence. These other responses all lie within the gray area, which, as previously mentioned, indicates sampling uncertainty. One, therefore, might be tempted to conclude that these impulse responses are unable to discriminate between the different parameterizations and, as such, that the statistical test of the hours response has poor power.[3] The rest of this paper will show that a conclusion that these long-run VARs have poor power would be overly pessimistic.

Figure 2 reports a scatter plot of the estimated impact response of hours to a technology shock versus the corresponding estimated bootstrapped standard error for each of the 2000 simulations. For any given simulation, we can determine whether an econometrician observing only that simulation would reject the true null hypothesis that the hours response on impact matches the response from a RBC model. If the econometrician had assumed that the estimated impulse response has an asymptotic normal distribution centered around the true response, then she would use a standard critical value of 2 and falsely reject the true null hypothesis 18 percent of the time. Given that the rejection

[3] The argument on page 19 in CKM seems to be a claim of poor power. However, CKM never report the estimated statistical sampling uncertainty that a researcher would estimate when faced with only a single data set. Rather they just report the distribution of the point estimates. In terms of my Figure 2, they report the distribution of the values on the x-axis but do not calculate the values on the y-axis.

rate is greater than the nominal size of 5 percent, I calculate the critical value (2.8), where an econometrician observing only a single data set would correctly fail to reject the null hypothesis 95 percent of the time and reject the true model only 5 percent of the time.[4]

Figure 3 reports the same scatter plot of estimated impulse responses and standard errors. However, Figure 3 reports the results for testing whether the estimated impulse response matches the response from a DGE model with high levels of habit and investment adjustment cost. Even with the size-corrected critical values, the false model is correctly rejected 32 percent of the time. Using the standard critical value would lead to an even greater rejection rate of 53 percent.

Figure 4 reports the distribution of the absolute value of the test statistics for data simulated from the RBC model when the tested null responses is the RBC model. In addition for the same simulated data, the Figure reports the distribution of the test statistics when the tested null response is the DGE model with high levels of habit and investment adjustment cost. Regardless of the critical value chosen, the DGE model with real rigidities are rejected far more often than are the RBC model.

Reporting the equivalent of Figure 3 for all possible parameterizations is not feasible, as such, rejection rates are summarized by Figure 5 which report for low degrees of habit (Figure 5a) and high degrees of habit (Figure 5b), the rejection rates for different values of γ. When γ equals 0 and b equals 0, then the rejection rate is the likelihood of rejecting the true model and is by construction 5 percent. When γ equals 3 and b equals 0.7, the rejection rate is 32 percent.

As can be seen in Figure 5, the test is much more likely to reject a false null hypothesis than a true null hypothesis. As such, the evidence implies that the set of DGP's studied here are more restrictive than the general class of DGP's for which Faust and Leeper's proposition of low power for long-run VARs would hold true. The rejection rates do increase with γ and b. However, these rejection rates are not large. Even for parameterizations with a fairly large degree of habit persistence and investment adjustment costs, the test has a size-corrected rejection rate of under 40 percent.

To be able to discuss the usefulness of these long-run VARs, we need to determine whether the rejection rates reported in Figure 5 are high or low. Because statistical

[4] An alternative approach would be to experiment with the various proposed modifications of how to construct confidence intervals. However, as many different methods have been proposed, I leave explorations of the properties of these different methods for future work.

power is only infrequently reported, for comparison purposes, it would be useful to have a benchmark from the literature on statistical testing of DGE models. One such statistic is the correlation of output growth.[5] The inability of the RBC model to match the correlation of output growth has been an important statistic in casting doubt on the basic RBC model. Papers that discuss the correlation of output growth include Cogley and Nason (1995) and Christiano and Vigfusson (2003).

Results for this unconditional statistic help put the performance of the impulse response analysis in context. Figure 5 also reports results for the output correlation. The power of these correlation statistics are somewhat better than the impulse responses. The rejection rates increase as both γ and b increase. With no habit persistence in consumption, the rejection rates using the correlation are about 20 percent for models with moderate or high degrees of investment adjustment costs γ. For models with a high degree of habit, the correlations reject the false models much more frequently.

Overall, this evidence suggests that the power of testing using the correlation is better than testing using just the impact response of hours worked. The rest of this paper shows that other applications of VARs can be more informative. In particular, Figure 5 has an additional set of lines that, for certain parameterizations, have better rejection rates than the rejection rates from the correlation test. The next section describes these lines.

5.2 Shape of Hours

In the monetary structural VAR literature, the delayed responses to monetary policy shocks is what moved the literature from the New Classical models of immediate responses to models with rigidities (Woodford 2003, p. 173). For technology shocks, I will argue that, here too, the shape of the response can better differentiate between models than the impact response.

Figure 6 reports, for data generated from a standard RBC model, a scatter plot of the response of hours on impact and the change in the response six periods later. Around each of these responses, one could construct a confidence ellipse. Assuming that these responses are drawn from a multivariate normal distribution, then the formula for the confidence ellipse can be easily derived from a simple wald test. Given an estimated

[5]The confidence intervals were constructed using the standard method from the matlab statistics toolbox. Confidence intervals were constructed using the result that $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ is approximately normal with a variance equal to $\frac{1}{N-3}$. As with other statistics reported in this paper, the critical value was size-corrected.

vector $\hat{\mu}$ and a variance covariance matrix for $\hat{\mu}$ denoted by $\hat{V}$, then a point x lies in the 95 percent confidence ellipse if x satisfies the following inequality

$$[x - \hat{\mu}]' \hat{V}^{-1} [x - \hat{\mu}] < \xi_{95} \tag{25}$$

where ξ_{95} is the critical value. According to standard asymptotic theory, the statistic is distributed chi-squared with 2 degrees of freedom. As was done for the impact responses, the test needs to be size-corrected. In the simulations the value of ξ_{95} is 10 rather than the standard value of 6. To give some degree of the magnitude of the correction, Figure 6 reports, for one single simulation, the estimated confidence intervals using the two different critical values. In either case, one would fail to reject the true null hypothesis that the response matches the response from the RBC model (the green dot). Although, the size-corrected confidence set is much wider than the non-size corrected interval, for this particular simulation, one would reject the false null hypothesis that the response matches the response from a model with high degrees of habit and investment adjustment costs (the red dot).

Returning to Figure 5, we can compare the power properties of the shape test versus the power properties of the tests of the impact response and the correlation. The rejection rates for tests using the shape of hours are much better than the rejection rates for tests using the sign and are comparable to rejection rates testing the correlation of output growth.

5.3 Investment Response

As was mentioned in CEV, the variance of hours explained by technology shocks is very low. In the benchmark parameterizations studied here, technology shocks account for less than 1 percent of the variance of hours. However, technology shocks do account for 22 percent of the variance of the ratio of investment to output. As such, a natural question to ask is whether looking at the response of investment rather than hours is more informative. I report below that looking at investment does appear to be more informative.

Figure 7 plots the impulse responses estimated for investment from the same benchmark three variable VAR (with variables: labor productivity growth, hours worked and investment to output ratio). The various model responses again lie within the large gray area. Also of note is that, unlike with hours, the investment response is not monotonically increasing with respect to both habit and investment adjustment costs. For a given degree of habit persistence, the size of the investment response declines as the investment adjustment costs increase. However, for a high degree of habit persistence, the investment response is larger than would be the case with a low degree of habit. The economics behind this reversal is that, with a high degree of habit, the utility maximizing behavior is to invest more in order to avoid too quick of an increase in consumption.

Figure 8 reports the power properties of the two tests of investment and also reproduces from Figure 5 the rejection rates when one tests using the correlation of output. The shape of the investment response seems more useful than the impact response and the investment responses appear to be quite useful in discriminating between models.

5.4 Changing the True Data Generating Process

For the results reported above, the true data generating process is the RBC model. Given data simulated from the RBC model, I reported the rejection rate of the false null hypothesis that the data was generated from models with real rigidities. Of course, we are also interested in the opposite case where the data is simulated from a model with real rigidities and the false null hypothesis of the RBC model is tested. Reversing the role of the two models is particularly relevant as most empirical work favors models with various degrees of adjustment costs. (See ACEL and Smets and Wouters (2003) for examples).

For this exercise, I generate data from the model with high investment adjustment costs and habit persistence with values of γ and b set at the values estimated in Christiano, Eichenbaum and Evans (2002). Using data generated from this model, the statistical tests frequently reject the standard RBC model. The rejection rates are much greater than the rejection rates observed when data is simulated from the RBC model and the test is of the model with high adjustment costs.

Figure 9 reports the impulse response of hours worked. Absent the accommodating monetary reaction embodied in the ACEL model and given the high level of costs associated with adjusting either the level of consumption or investment, a technology shock actually drives down hours. The improvement in productivity causes the consumer to increase leisure rather than increase consumption or investment. The average estimated response is somewhat biased away from the true response. However, using the standard critical value, the rejection rate is 24 percent which is similar to the results presented above. As can be seen in Table 2, when using the size-corrected critical value, the power of the test to reject the now false RBC model is 58 percent. Comparing these results with the results from Figure 5, the test is much more powerful when the data is simulated from the model with real rigidities than when the data is simulated from the RBC model.

Figure 10 reports the shape of the hours response. Here the estimated responses are clustered around the true model response. Table 2 reports the power of testing the shape of the hours response using the size-adjusted critical values. The false RBC model is rejected 90 percent of the time.

Figure 11 and 12 report the same results for the investment response. Table 2 also reports the results for these models. As with the RBC model, the size and power properties are better for testing investment than testing hours. Using the investment response, one can almost always reject the RBC model. Finally Table 2 also reports a test using the correlation of output growth. Here too the statistic has a high degree of power.

Table 3 and Table 4 summarizes across several different parameterizations to reject the RBC model. For low levels of habit, testing the initial period of the hours response is not very informative but testing the shape of the hours response can be very informative. When there is no habit, the initial investment response is also not very informative. In models with a higher degree of habit, the initial investment response is much more informative.

Table 4 presents the values of the critical values that result in tests with correct size. For each test statistic, the critical values are fairly constant across parameterizations. As such, it seems reasonable to suppose that using the average critical value from this table is a good way to size correct when the true data generating process is unknown.

6 An Empirical Application

As an empirical application, I take a VAR similar to that estimated in Christiano Eichenbaum and Vigfusson (2003) and ask what parameterizations can be ruled out and which can be allowed.

The VAR is the same three variable system used above in terms of labor productivity, hours worked and the ratio of investment to output. The difference is that rather than the data being simulated from a RBC model, the data is the empirical data from the United States. In particular, productivity is measured as hourly labor productivity in the business sector, hours worked is per capita business hours worked, and investment is the ratio of private investment to GDP. In this section, hours enters the VAR in levels. The discussion of how to treat hours and other low frequency movements in estimating these VARs is beyond the scope of the current paper and is instead addressed in Christiano Eichenbaum and Vigfusson (2003). The sample period is 1954 to 2001.

Empirical responses are reported in Figure 13a,13b, and 13c. In addition, for comparison sake, model impulse responses are reported for a few model parameterizations. A 95 percent confidence interval is constructed for the empirical responses using the estimated bootstrap error and a size-adjusted critical value that is the average of the values reported in Table 5. Given the width of the confidence interval, I fail to reject any of the models. Figure 14a, however, shows that, even when using the large size-adjusted critical value, a test based on the shape of the hours response does lead to a rejection of both extremes of no real rigidities and large real rigidities. However, the investment response shown in Figure 14b is less informative.

Based on these results, the most promising way to model the response to a technology shock is to allow for delayed hump-shaped responses to the technology shock. Future work will be directed towards determining whether these delayed responses are best modeled as the result of nominal or real rigidities.

7 Conclusions

Impulse responses from long-run VARs can reject false models. As expected, these rejection rates increase the further away the false model is from the true data-generating model. In addition, these rejection rates vary depending on what variables are studied. Overall, however, this paper shows that these long-run VARs can be informative about which models are to be preferred. For the models studied here, testing the shape is a more powerful test than simply looking at the sign of the response. In addition, relative to an alternative statistical test based on sample correlations, I find that the shape-based tests have greater power.

These results should encourage us to find creative and new ways to test our models. The conclusion is not to abandon our tools but to find ways to improve their use. Overall, given these results on the power and size properties of long-run VARs, I conclude that these VARs can be useful for discriminating between macro models and, therefore, should continue to be used in developing and testing business cycle theory.

References

Altig, David, Lawrence Christiano, Martin Eichenbaum, and Jesper Linde (2005). "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle," NBER Working Paper Series 11034. Cambridge, Mass.: National Bureau of Economic Research, January.

Chari, V. V., Patrick J. Kehoe, and Ellen McGrattan (2007). "A Critique of Structural VARs Using Real Business Cycle Theory," Working Paper Series 631. Minneapolis: Federal Reserve Bank of Minneapolis, May.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, vol. 113 (February), pp. 1–45.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson (2003). "What Happens after a Technology Shock?" NBER Working Paper Series 9819. Cambridge, Mass.: National Bureau of Economic Research, July.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson (2006). "Assessing Structural VARS" NBER Macro Annual. Cambridge, Mass.: National Bureau of Economic Research.

Christiano Lawrence J. and Robert J. Vigfusson (2003) "Maximum likelihood in the frequency domain: the importance of time-to-plan", *Journal of Monetary Economics*, vol. 50, issue 4, pp. 789-815

Cogley, Timothy and Nason, James M, (1995). "Output Dynamics in Real-Business-Cycle Models," *American Economic Review*, vol. 85(3), June, pp. 492-511

Erceg, Christopher J., Luca Guerrieri, and Christopher Gust (2005). "Can Long-Run Restrictions Identify Technology Shocks?" *Journal of the European Economic Association*, vol. 3 (December), pp. 1237–78.

Faust, Jon, and Eric Leeper (1997). "When Do Long-Run Identifying Restrictions Give Reliable Results?" *Journal of Business and Economic Statistics*, vol. 15 (July), pp. 345–53.

Francis, Neville, and Valerie A. Ramey (2005). "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," *Journal of Monetary Economics*, vol. 52 (November), pp. 1379–99.

Gali, Jordi (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, vol. 89 (March), pp. 249–71.

Hamilton, James D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Kehoe Patrick J. (2006) “How to Advance Theory with Structural VARs: Use the Sims-Cogley-Nason Approach”, *NBER Macro Annual.* Cambridge, Mass.: National Bureau of Economic Research.

Sims, Christopher (1972). “The Role of Approximate Prior Restrictions in Distributed Lag Estimation,” *Journal of the American Statistical Association*, vol. 67 (March), pp. 169–75.

Sims, Christopher and Tao Zha (2006) “Does Monetary Policy Generate Recessions?” *Macroeconomic Dynamics*, 10(2), pp. 231-72

Smets, Frank, and Raf Wouters (2003). “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, vol. 1 (September), pp. 1123–75.

Vigfusson, Robert J. (2004). “The Delayed Response to a Technology Shock: A Flexible Price Explanation,” International Finance Discussion Paper Series 2004-810. Washington: Board of Governors of the Federal Reserve System, July.

Woodford, Michael M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.

Tables

Table 1: Model Parameter Values

Calibrated			
β	$0.98^{1/4}$	ψ	2.5
θ	0.33	g	$1.01^{1/4} - 1$
δ	$1 - (1 - .06)^{1/4}$	μ_z	$1.016^{1/4} - 1$
τ_x	0.3	τ_l	0.242
Estimated			
σ_z	0.00968		
σ_l	0.00631	ρ_l	0.9994
σ_x	0.00963	ρ_x	0.9923

Table 2: Size and Power when DGP is DGE Model with High Adjustment Costs

Test	Rejection Rate		Critical
	RBC Model Size Adjusted*	True Model Not-size Adjusted	Value
Impact Hours Response	58%	23%	3.2
Shape of Hours Response	99%	29%	16.5
Impact Investment Response	100%	10%	2.4
Shape of Investment Response	100%	13%	9.84
Correlation of Output Growth	100%	10%	2.55

*Rejection Rates are for the size-adjusted critical values given in column (iii)

Table 3: Rejection Rates of the RBC Model

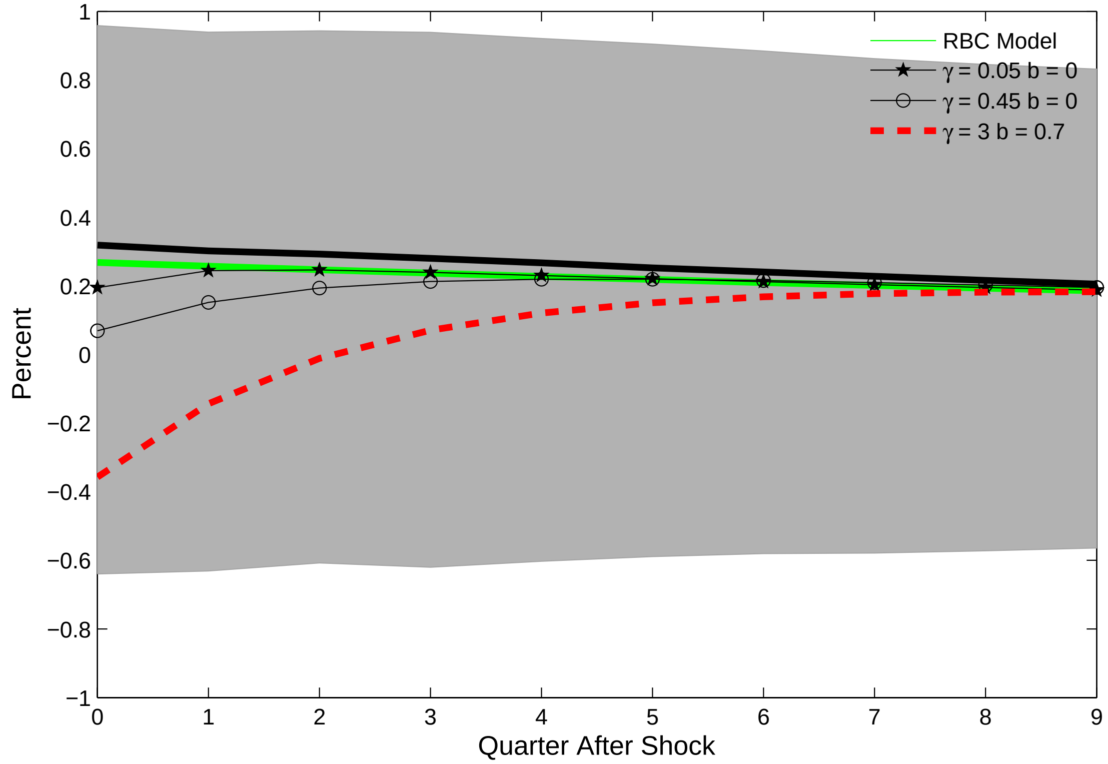
Model Parameters		Test Statistic Hours		Investment		Output
b	γ	Sign	Shape	Sign	Shape	Correlation
0	0.5	0.05	0.81	0.12	1.00	0.00
0	1.5	0.08	1.00	0.21	1.00	0.01
0	3	0.11	1.00	0.24	1.00	0.00
0.5	0.5	0.20	0.43	0.70	0.88	0.79
0.5	1.5	0.33	1.00	0.85	1.00	0.92
0.5	3	0.41	1.00	0.87	1.00	0.92
0.7	0.5	0.27	0.25	0.83	0.58	0.95
0.7	3	0.58	1.00	0.89	1.00	1.00

Results are reported for tests done on data simulated using the macro model described in the paper with parameters in the first two columns on the left. For each set of model parameters, 2000 simulations are done. For each parameterization, size-adjusted critical values are used to test the false null hypothesis null hypothesis that the data were generated by an RBC model.

Table 4: Size Adjusted Critical Values

Model Parameters		Test Statistic Hours		Investment		Output
b	γ	Sign	Shape	Sign	Shape	Correlation
0	0	3.20	16.5	2.4	9.84	2.55
0	0.5	2.96	10.80	2.36	8.38	2.96
0	1.5	2.81	10.60	2.27	8.36	3.11
0	3	2.75	9.93	2.31	8.78	3.06
0.5	0.5	2.77	10.56	2.36	8.12	2.64
0.5	1.5	2.83	12.55	2.32	9.06	2.64
0.5	3	2.87	13.42	2.40	9.34	2.63
0.7	0.5	2.87	11.65	2.52	9.13	2.56
0.7	3	3.20	16.47	2.39	9.85	2.55

Figure 1: The response of hours worked to a technology shock estimated using data simulated from a RBC Model



Note Thick black line is average response over 2000 estimated responses using data simulated from a RBC model. Edges of grey area indicate 5th and 95th percentiles of all estimated responses to a technology shock

Figure 2: Testing The Impact Response of Hours.

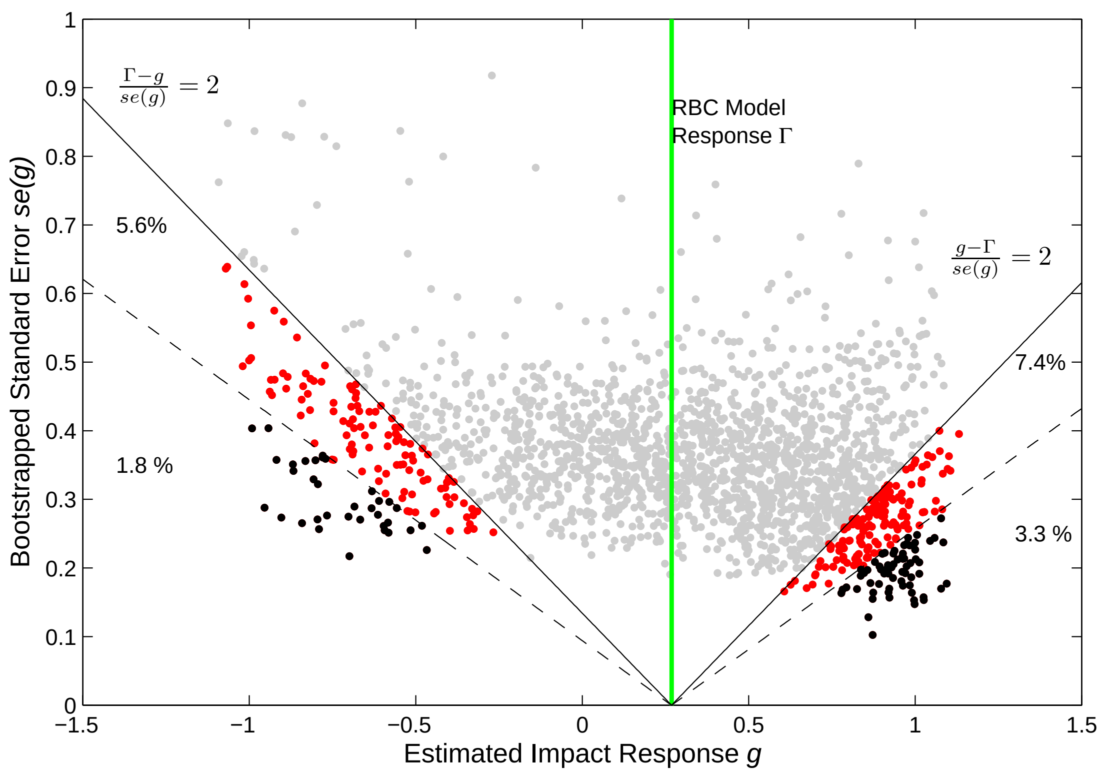


Figure 3: Testing The Impact Response of Hours using a false null hypothesis

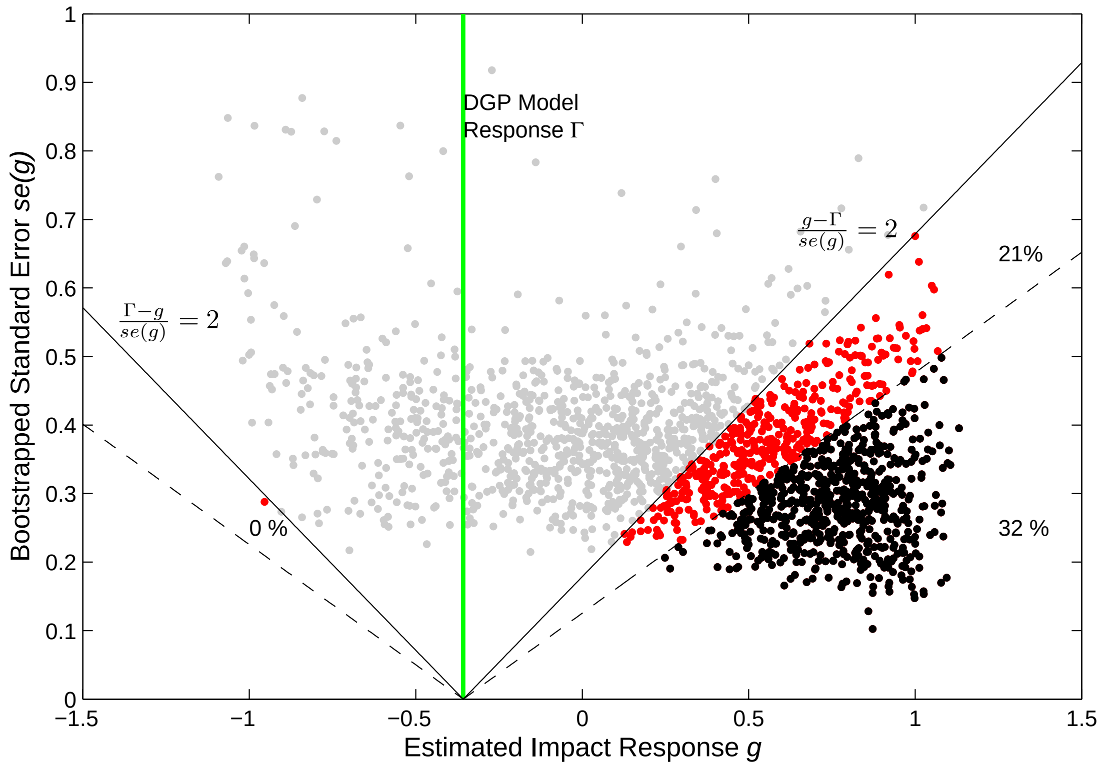


Figure 4: Cumulative Distribution Function for Testing Impact Effect.

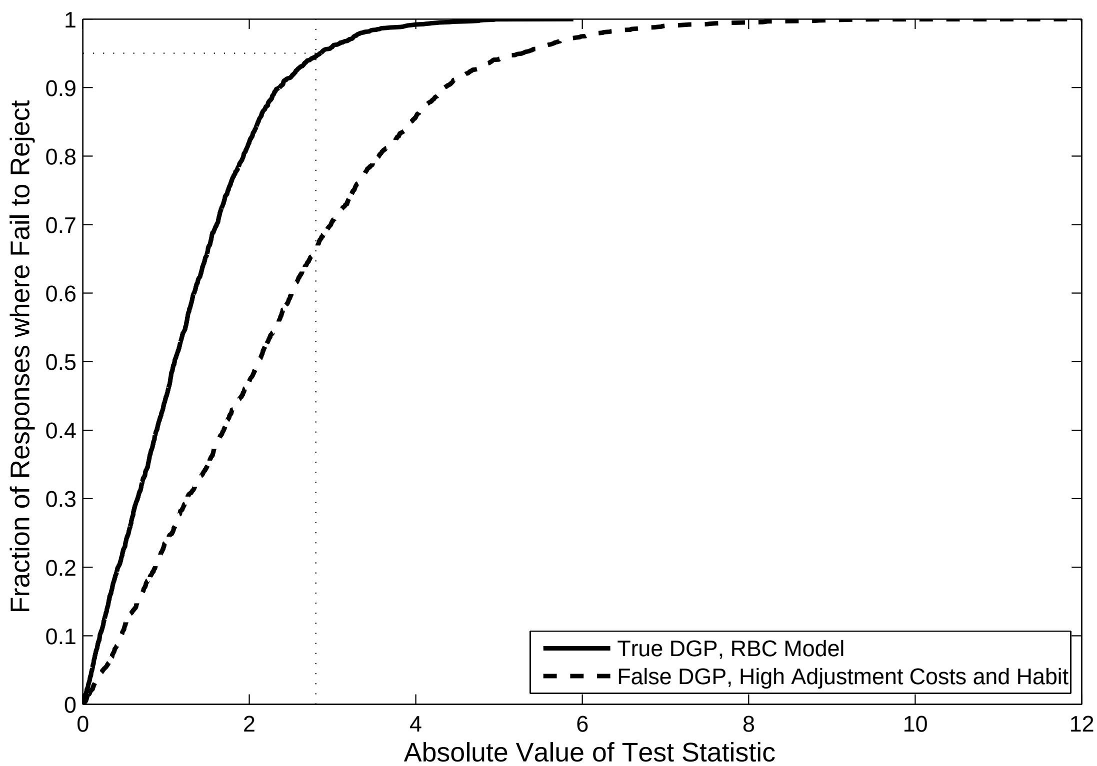


Figure 5 Rejection Rates For Different Tests when True DGP is RBC Model.

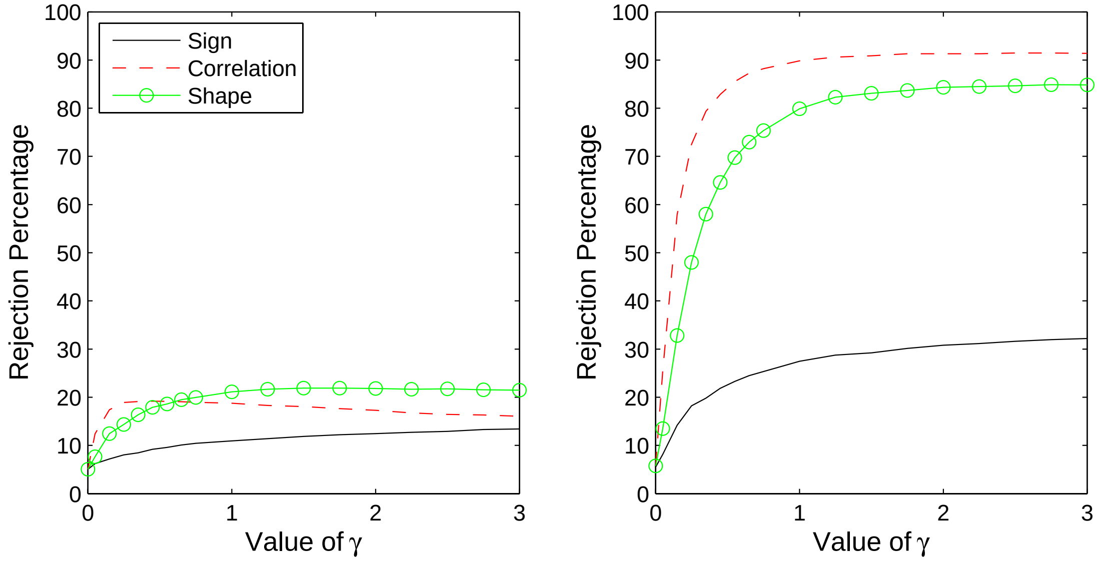


Figure 6: The Shape of The Hours Response

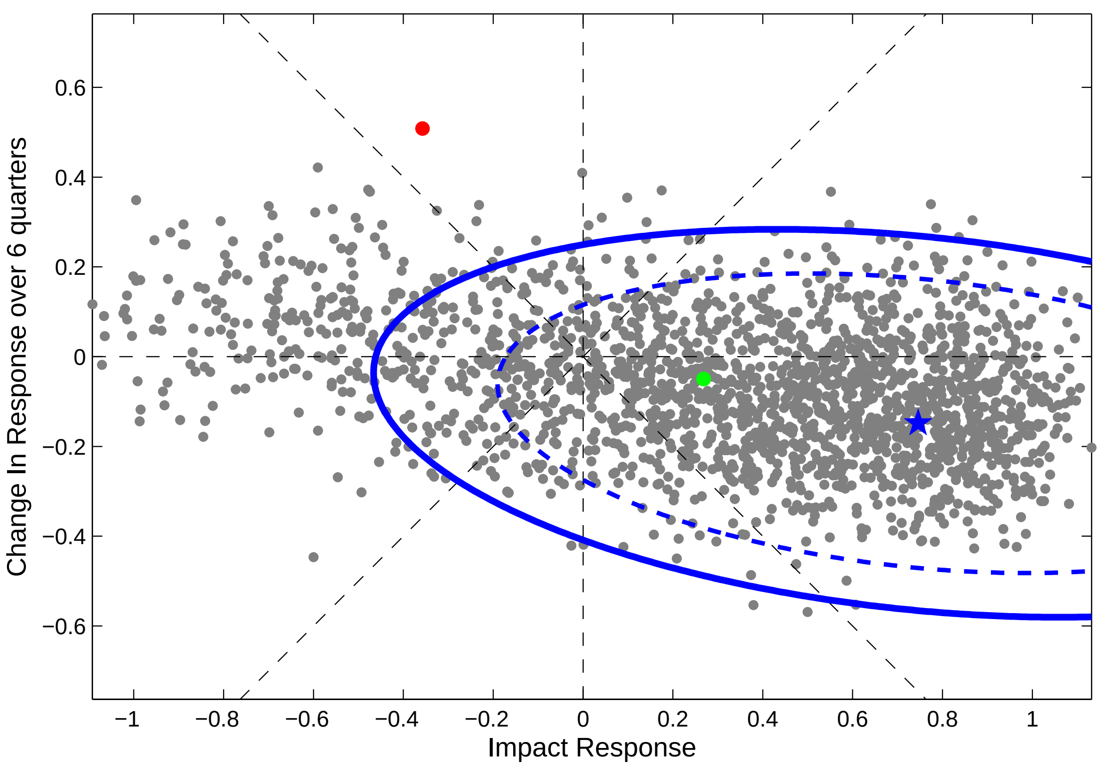
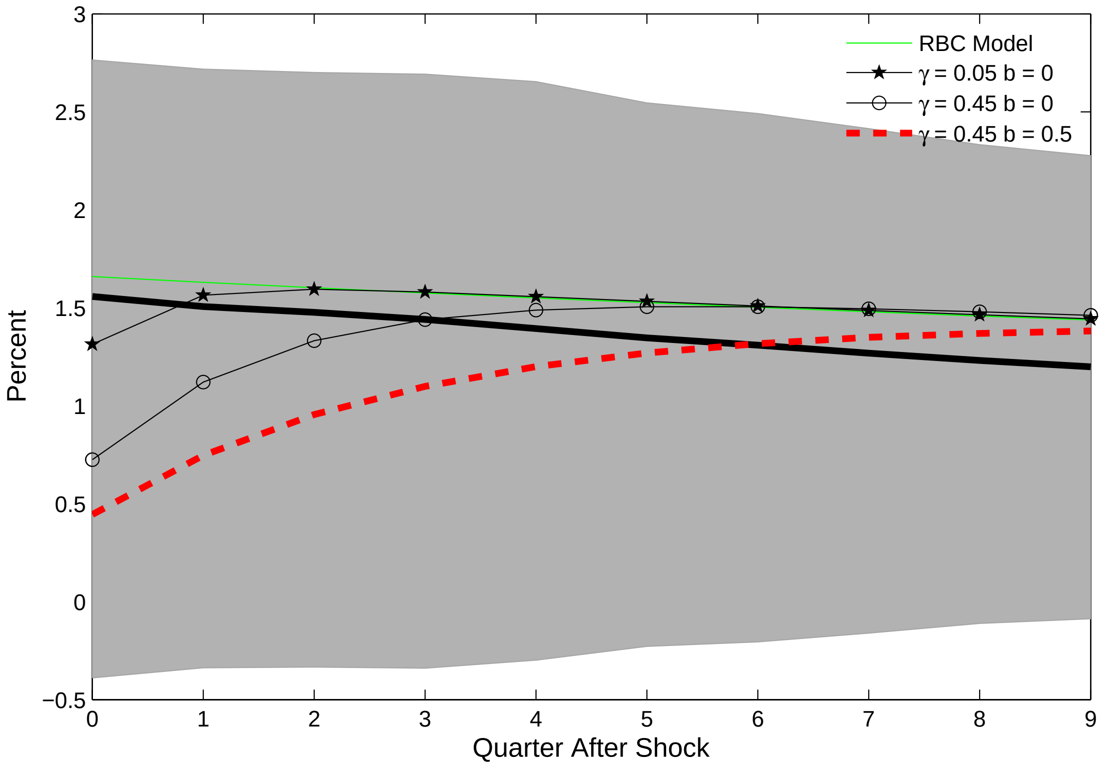


Figure 7: The response of investment to a technology shock estimated using data simulated from a RBC Model



Note Thick black line is average response over 2000 estimated responses using data simulated from a RBC model. Edges of grey area indicate 5th and 95th percentiles of all estimated responses to a technology shock

Figure 8 Rejection Rates For Testing Investment Response.

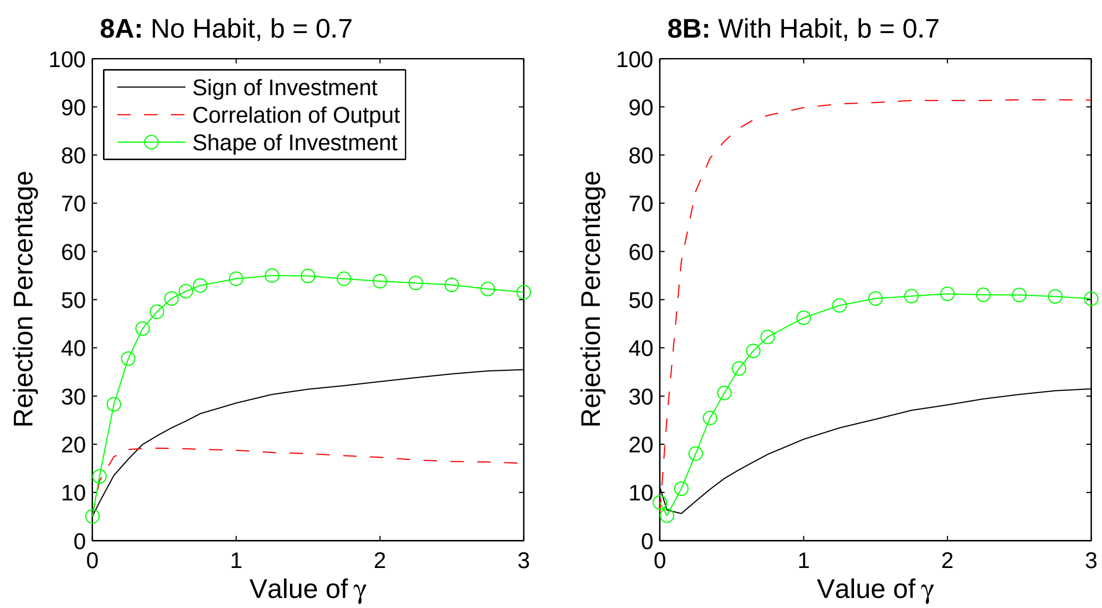
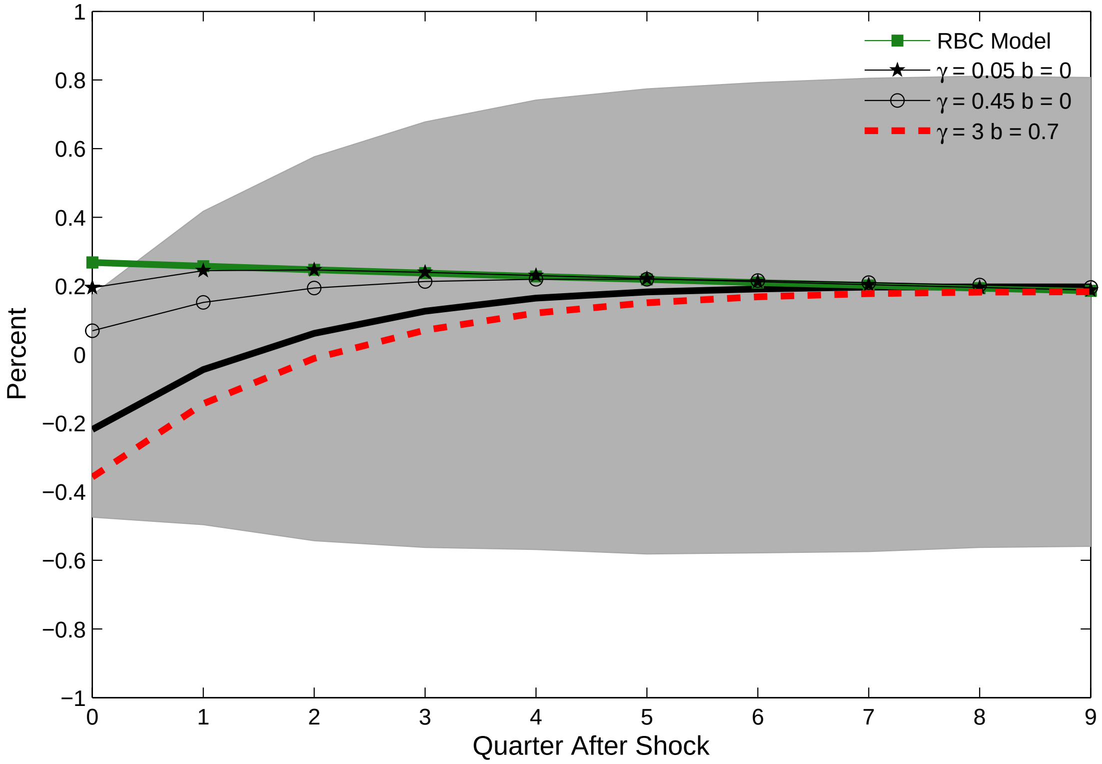


Figure 9: The response of hours worked to a technology shock estimated using data simulated from a Model with High Investment Adj. Costs and Habit



Note Thick black line is average response over 2000 estimated responses using data simulated from a DGE model. Edges of grey area indicate 5th and 95th percentiles of all estimated responses to a technology shock

Figure 10: The Shape of The Hours Response with Data Simulated From DGE Model with High Adjustment Costs

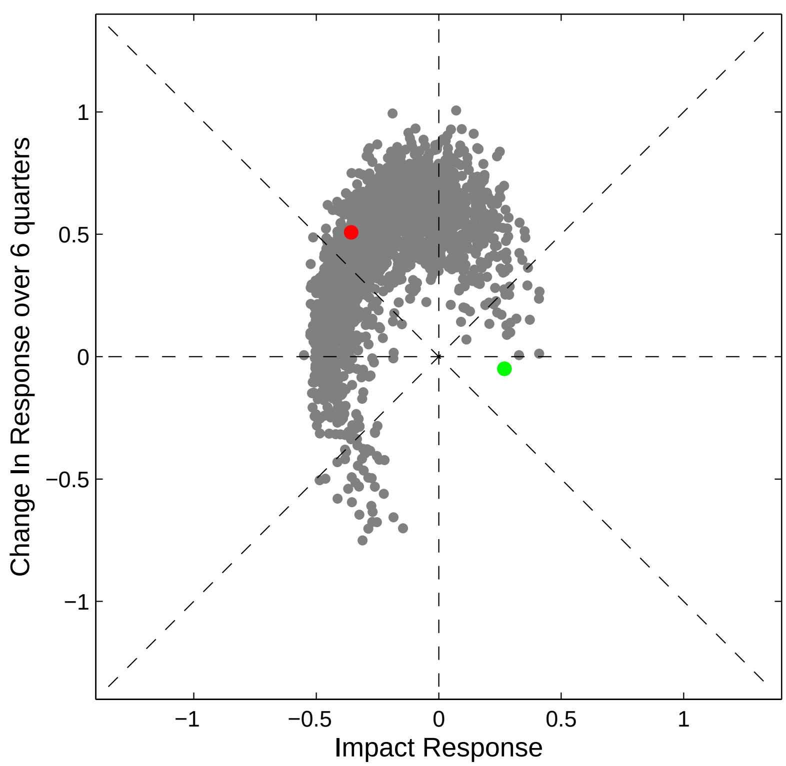
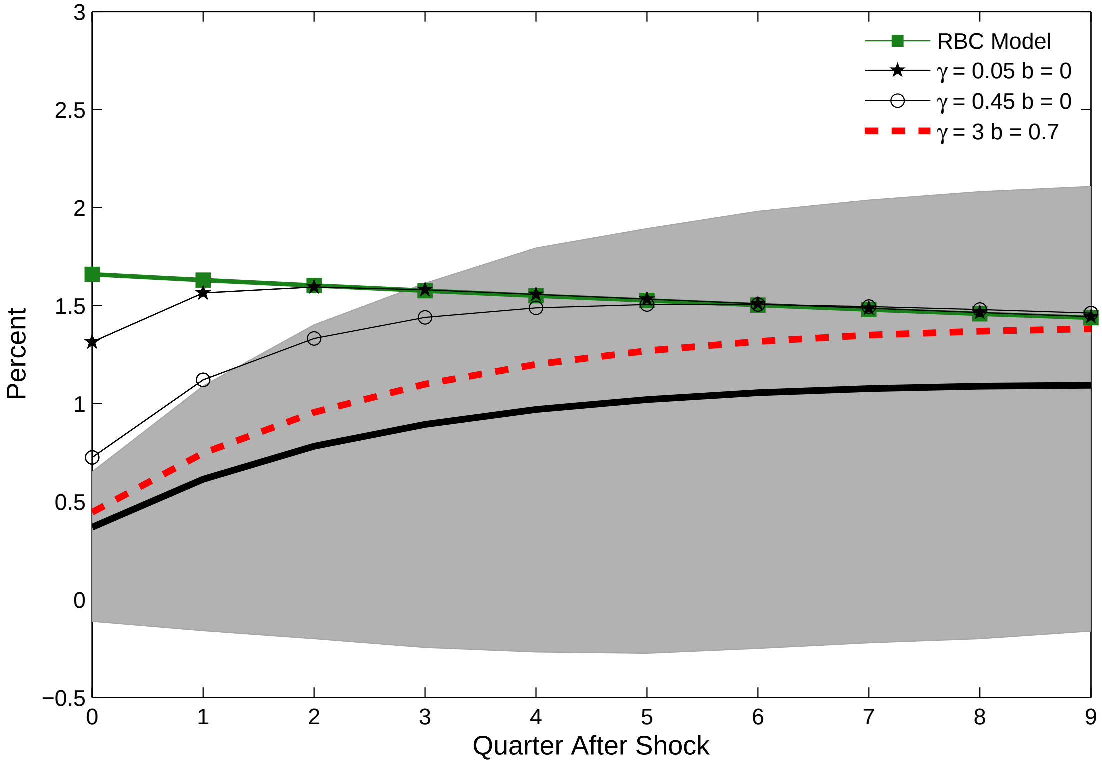


Figure 11: The response of investment to a technology shock estimated using data simulated from a Model with adjustment costs



Note Thick black line is average response over 2000 estimated responses using data simulated from a DGE model. Edges of grey area indicate 5th and 95th percentiles of all estimated responses to a technology shock

Figure 12: The Shape of The Investment Response with Data Simulated From DGE Model with High Adjustment Costs

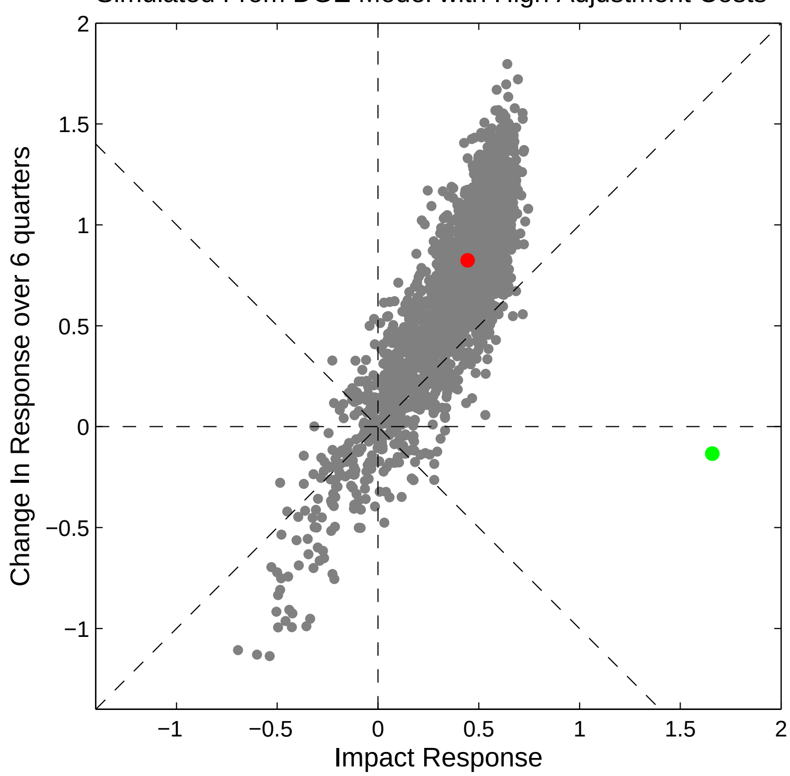
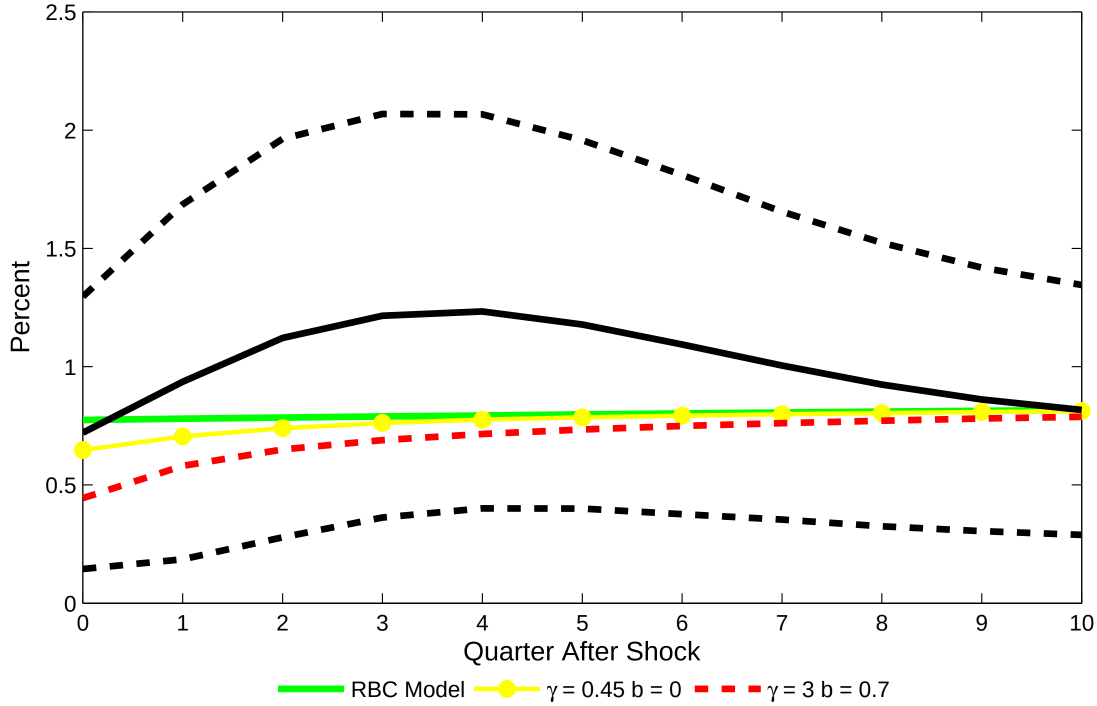
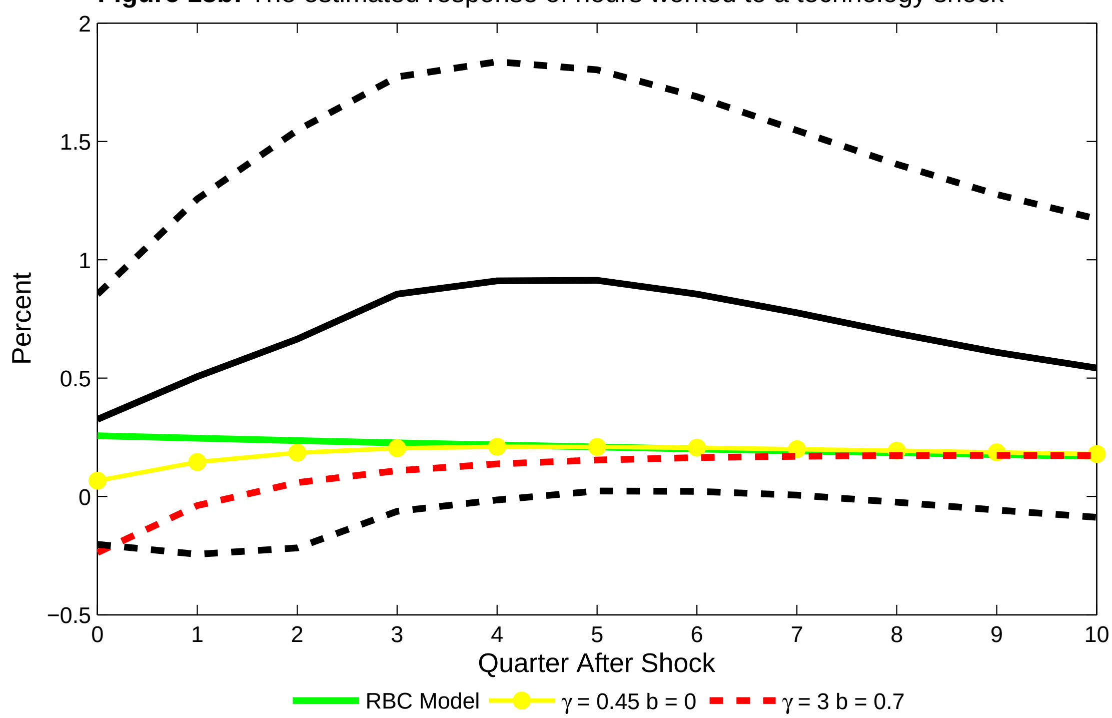


Figure 13a: The estimated response of output to a technology shock



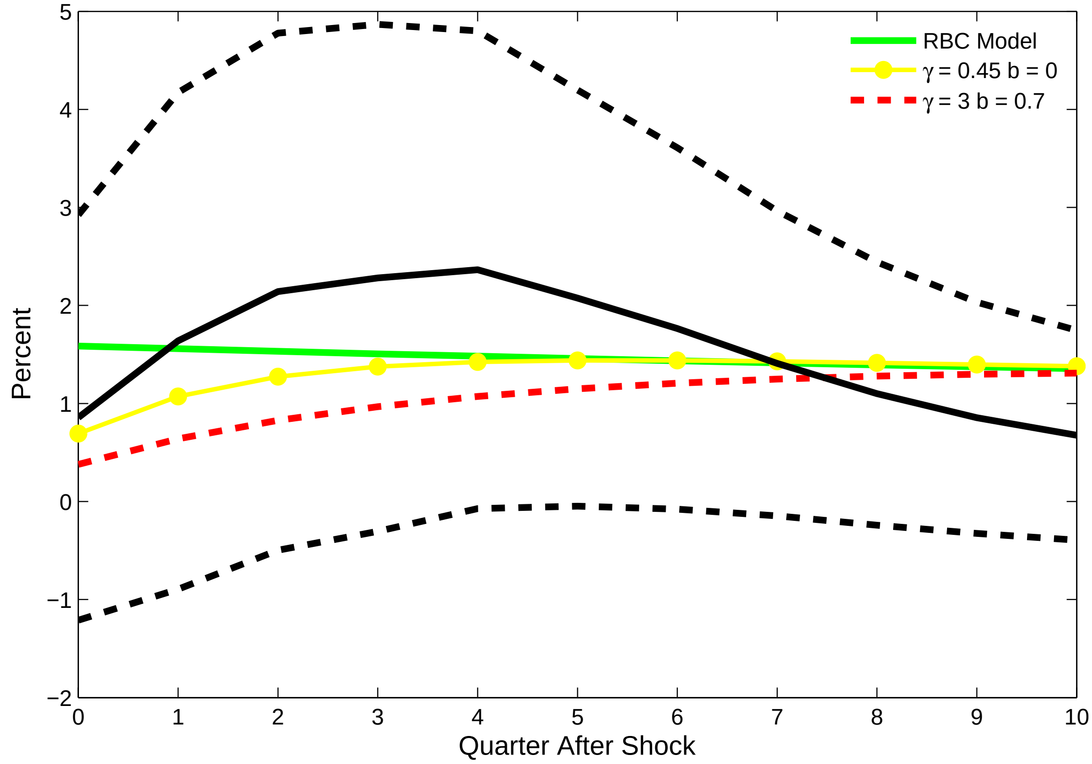
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 13b: The estimated response of hours worked to a technology shock



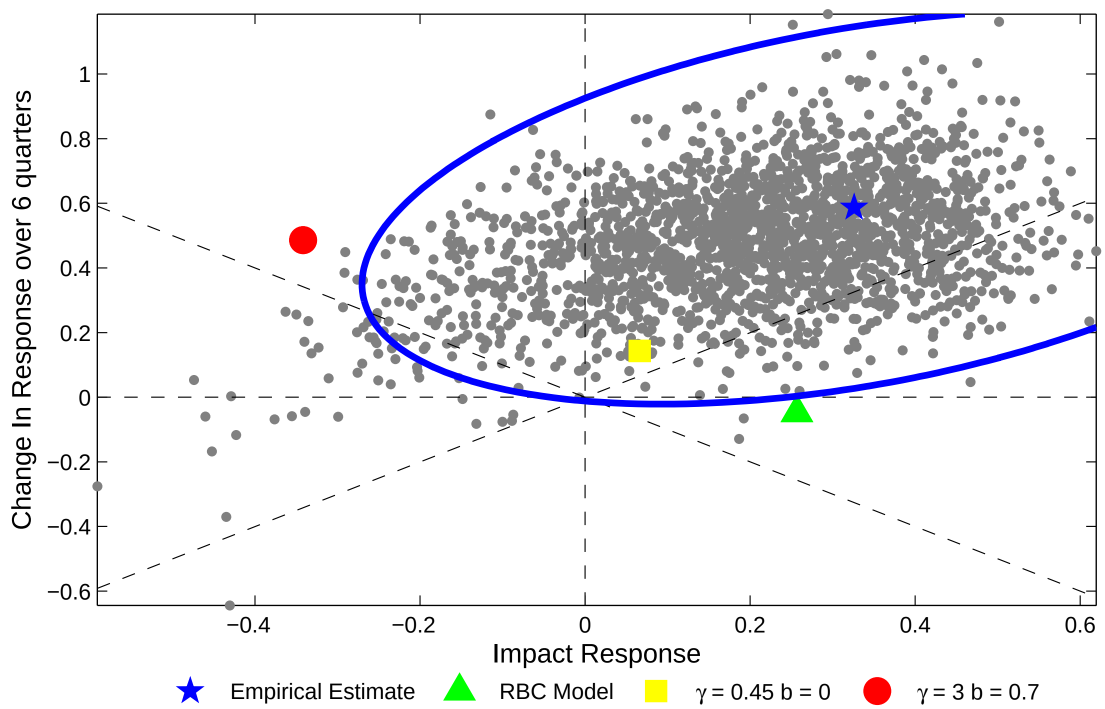
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 13c: The estimated response of Investment to a technology shock



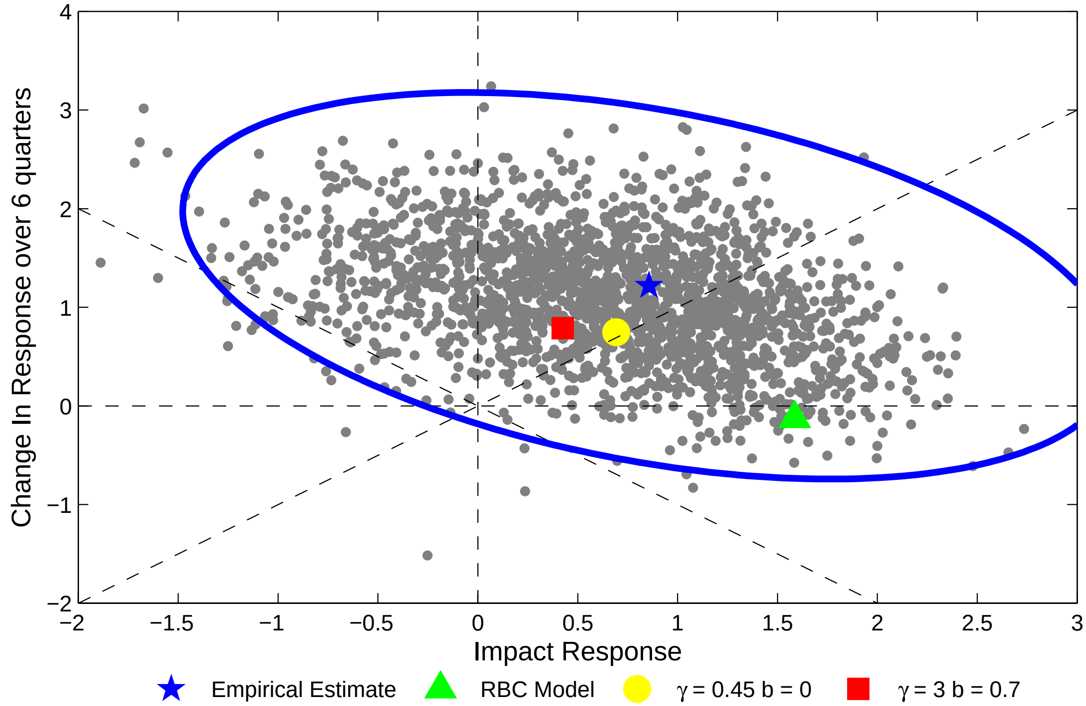
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 14a: The Shape of The Hours Response In An Estimated VAR



NoteGrey dots indicate responses from bootstrap simulations using empirical VAR.
Blue ellipse indicates confidence interval around point estimate.

Figure 14b: The Shape of The Investment Response In An Estimated VAR



Note: Grey dots indicate responses from bootstrap simulations using empirical VAR.
Blue ellipse indicates confidence interval around point estimate.