The Power of Long-Run VARS

Christopher Gust & Robert Vigfusson
Federal Reserve Board

Second International Conference in memory of Carlo Giannini

Introduction/Summary

- **Question**: Can structural long-run VARs be used to help us choose amongst models? (statistical size and power)
- **Method**: Simulate many dataset from macro models and ask how frequently these VARS reject different null hypotheses. (some false)
- **Conclusions**: long-run VARs can frequently reject false models. More useful than the previous literature conjectured.

Statistical Properties

Size

- How often is a true null hypothesis mistakenly rejected?
- Researchers set a threshold for acceptable rejection rates.

Power

- How often is a false null hypothesis correctly rejected?
- A good test would have a high rejection rate for false models.
- Researchers less frequently report information about power.

Literature on long-run VARs

- Simulate data from RBC model and estimate a long-run VAR on simulated data.
- Chari Kehoe and McGrattan CKM claim that structural VARs with long-run restrictions do “not allow a researcher to distinguish between promising and unpromising classes of models"
- Christiano Eichenbaum and Vigfusson CEV
  - A response and critique of CKM.

CEV on size

- Simulate data from macro model.
- For each dataset, estimate impulse response and associated standard error. Construct confidence interval.
- Ask how frequently do we reject true model using these estimated confidence intervals.
- Results: Fairly good size, but we reject somewhat more often than the pre-declared threshold (nominal size).

Power

- Less is known about power.
- Tests can have good size but poor power.
- This paper uses methods in CEV to study power. A sequel but a necessary sequel.
- Issues still to resolve about power, especially given claims by
  - Faust and Leeper
  - Chari Kehoe McGrattan.

Faust and Leeper

- Faust and Leeper 1997 JBES

Proposition 1. Any test of $H_0\colon a_{ijk} = 0$ has significance level greater than or equal to maximum power.

- This true only if you don't restrict the DGP.
- Faust and Leeper

$F(1)$. The simplest solution they demonstrate is to assume that the model driving the data is a VAR with known maximum lag order, K. There is surely some K large enough to

Discussing Faust and Leeper

- For my simulated datasets, VARs have power greater than size.
- Why? Does this contradict Faust and Leeper?
- No contradiction.
- Faust and Leeper: very general DGPs. In particular, possible spikes at frequency zero.
- Here: DGPs are parameterized macro models, more restricted than DGPs discussed in Faust and Leeper.

Three DGPs Explored Here

1. Flexible price DSGE macro model with no real rigidities.
2. Flexible price DSGE models with varying levels of habit persistence and investment adjustment costs.
3. Sticky price and wage DSGE models with varying levels of real rigidities.

First Set of Experiments

- Simulate data sets from a standard RBC macro model. (Many times)
- For each simulated data set, estimate impulse responses and several other statistics.
- For each statistic, ask how often, do you reject
  - True DGP (a standard RBC macro model.) (size)
  - Other macro models that are not the RBC model. (power)

A Flexible Price DGE Model

- Similar to standard RBC model.
- Three estimated shocks
  - Technology shocks. (permanent)
  - Leisure shocks. (not permanent)
  - Investment tax shocks. (not permanent)
- Allow for habit persistence and investment adjustment costs that depend on the growth rate of investment.

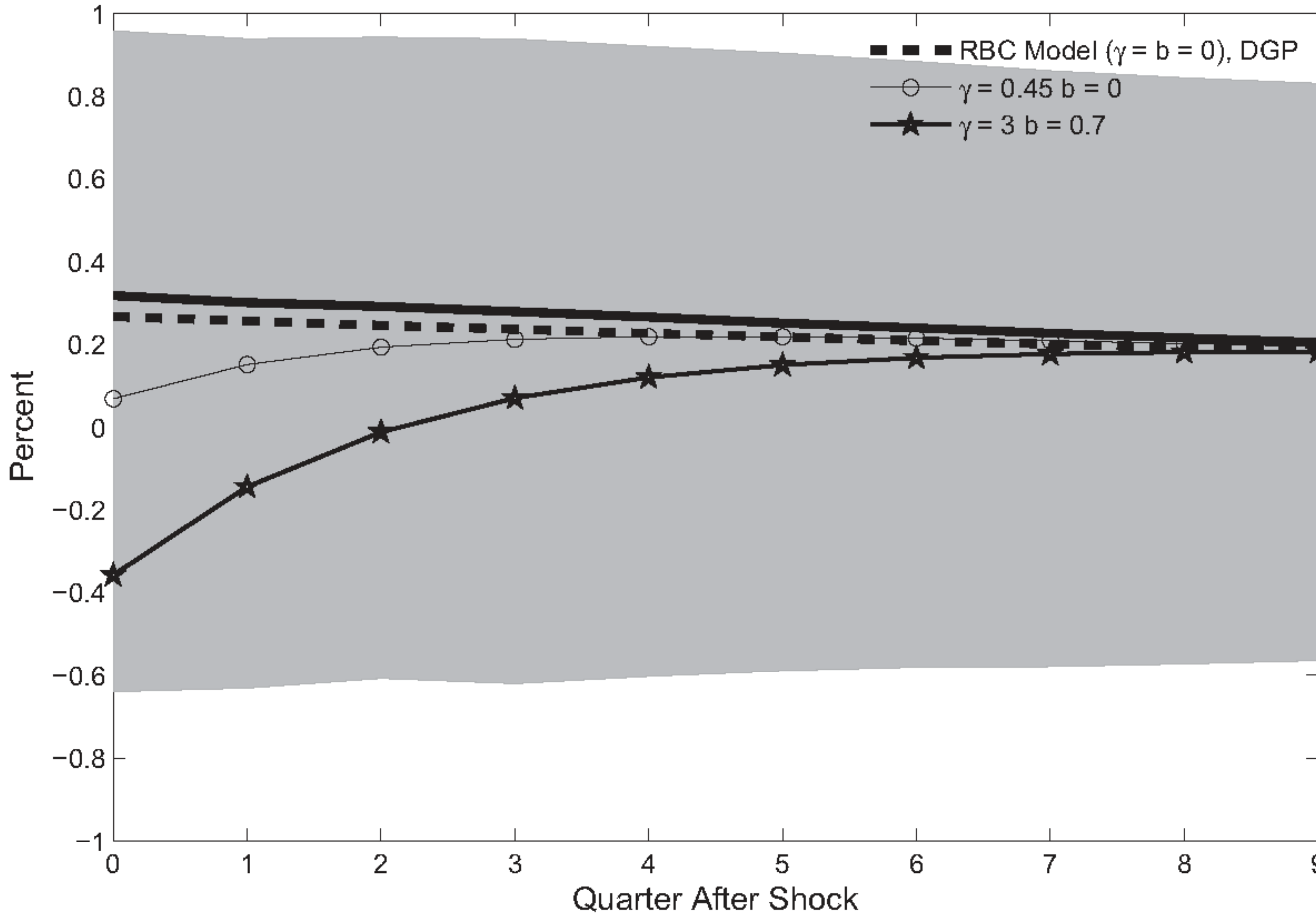
Estimate a Long-Run VAR

- Standard Estimation of a Long-run VAR.
- Estimate reduced-form VAR (four lags)
  - labor productivity growth
  - level of hours worked.
  - Investment to output ratio.
- Apply the standard long-run identification assumption that only technology shocks affect labor productivity in the long-run.
- Estimate standard error by bootstrap.

First Exercise

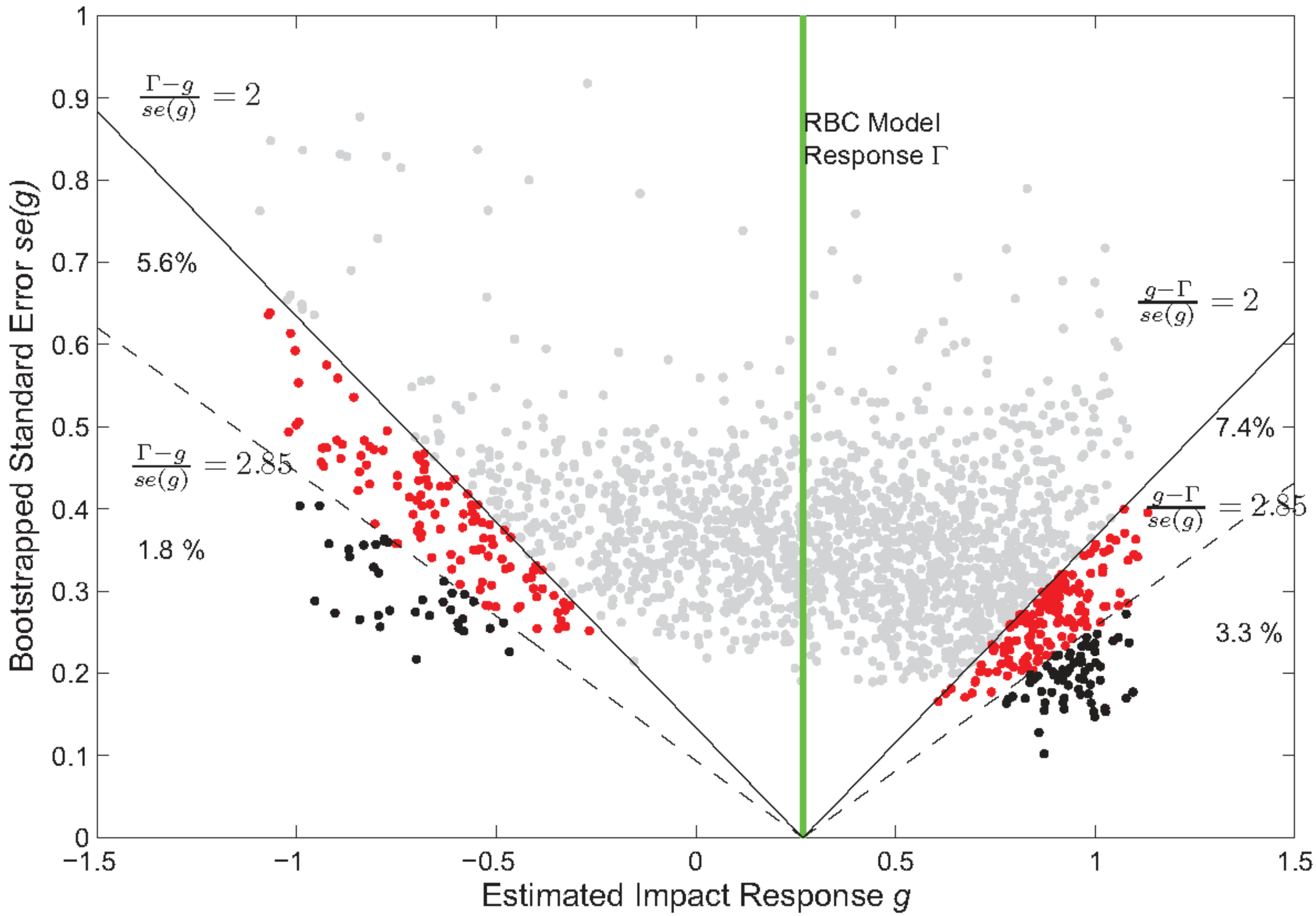
- Simulate data from a standard RBC model.
- For each simulation, Estimate VAR
  - Calculate impulse response of hours to technology shock.
  - Calculate Bootstrapped confidence interval.

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



Note Thick solid 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 one−standard−deviation technology shock

Figure 2: Testing The Impact Response of Hours.



Getting the Size Right.

- Each simulation of the model

G(i) response *se(i)* standard error

- Γ true model response
- Find the 95 percentile of $|G(i)-\Gamma|/se(i)$
- This is the size-adjusted critical value.
- Asymptotic critical value 1.96
- **Here** critical value is 2.96
- Rejection Rate Drops from 13 to 5 percent.

Figure 2: Testing The Impact Response of Hours.

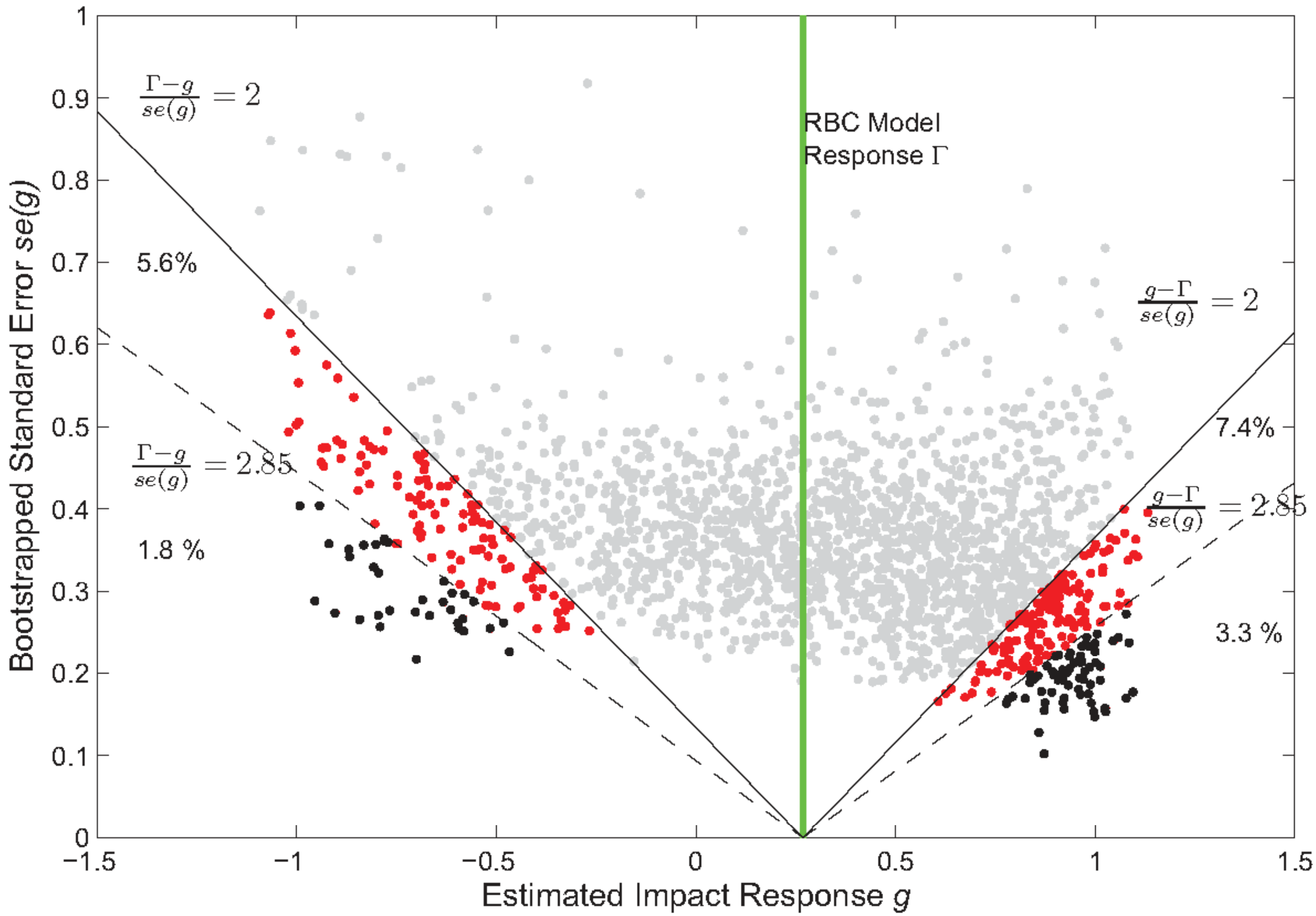
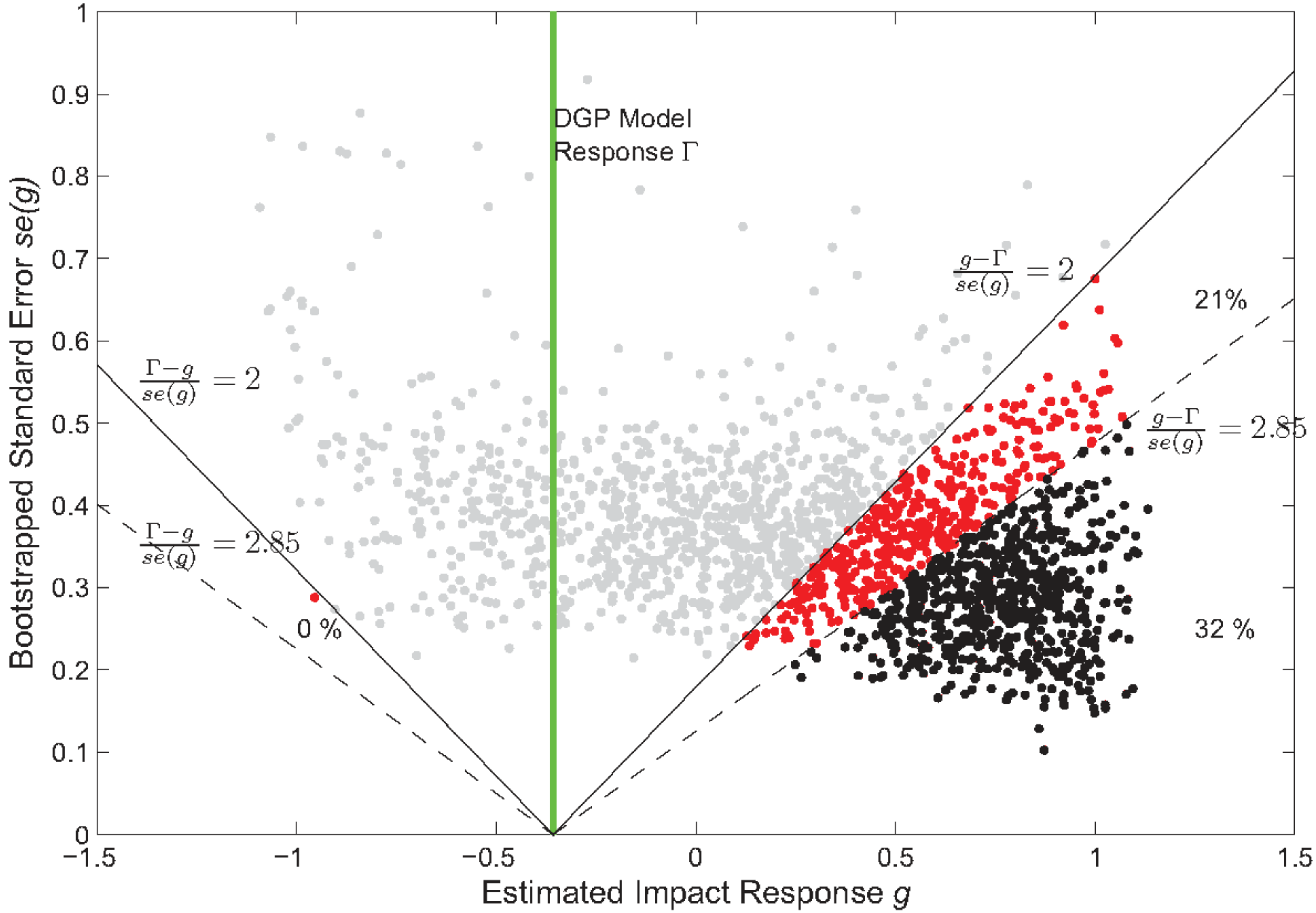


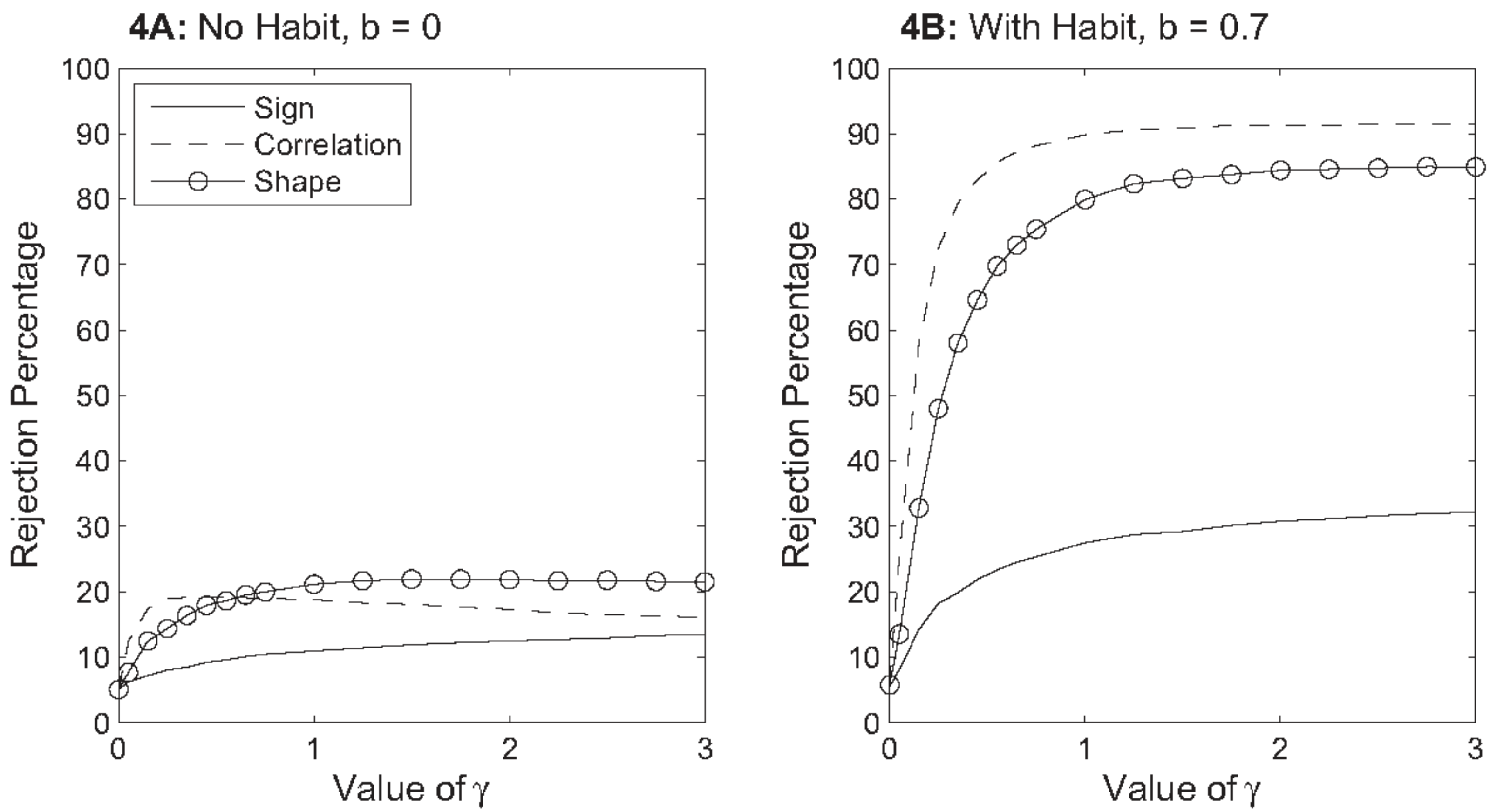
Figure 3: Testing The Impact Response of Hours using a false null hypothesis



How do we assess power.

- How to interpret rejection rates? Is 32 percent good or bad? (Clearly, not great)
- Easiest to do relative comparisons.
- Test the correlation of output growth.
- Advantages
  - Used before in the literature.
  - No identification necessary. Just a summary statistic from the model.

Figure 4 Rejection Rates For Different Tests when True DGP is RBC Model.



Is there anything else to be done with VARS?

- These results are not overly supportive of using VARs to discriminate amongst models.
- Can we use VARs in different ways to choose between models.
  - Look at shape.
  - Look at other variables

Shape

- Look at the response this period and the response six periods later.
- Construct a confidence ellipse based on Wald test.

$$\left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}\right]' V^{-1} \left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}\right] < \zeta$$

Figure 5: The Shape of The Hours Response

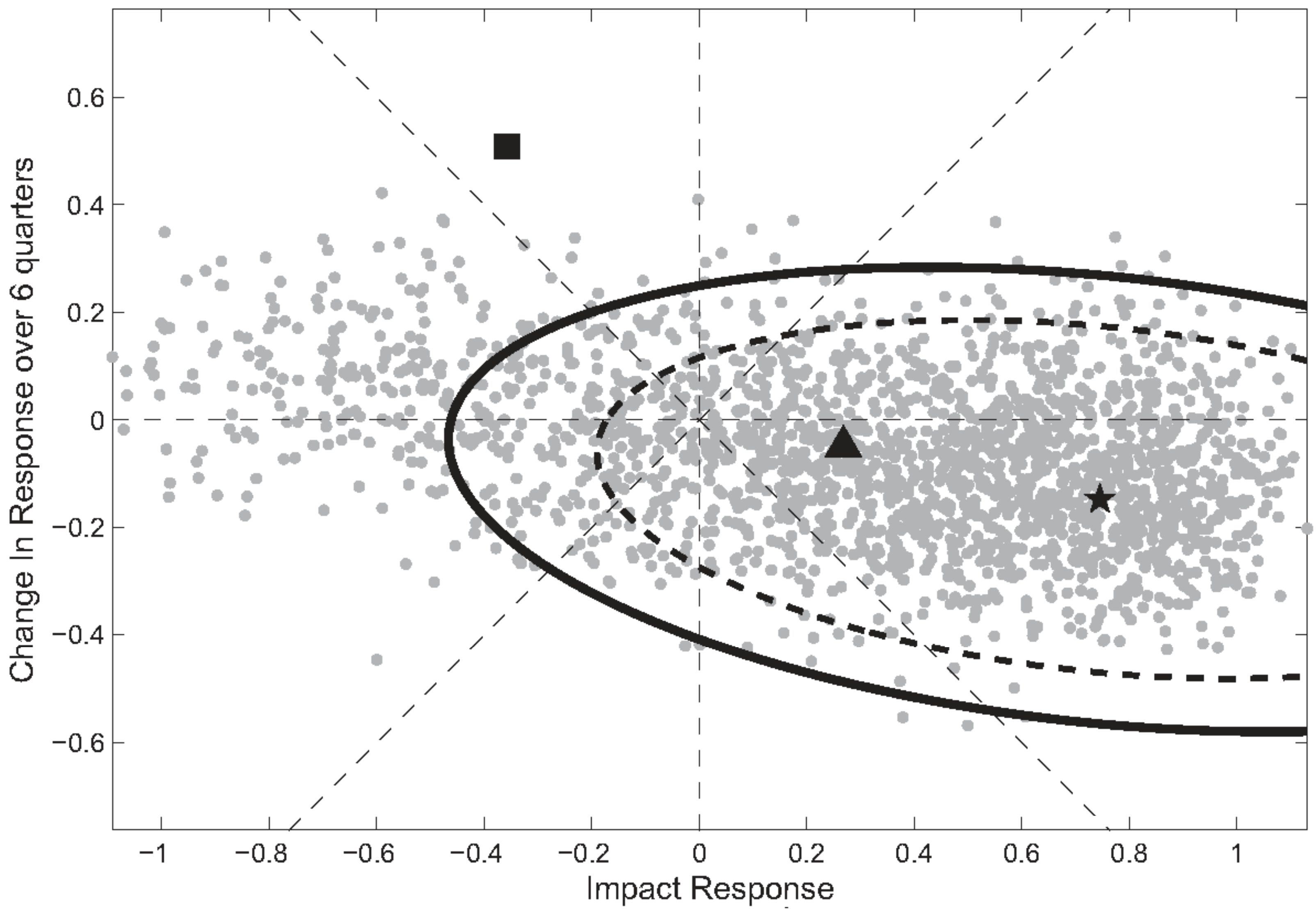
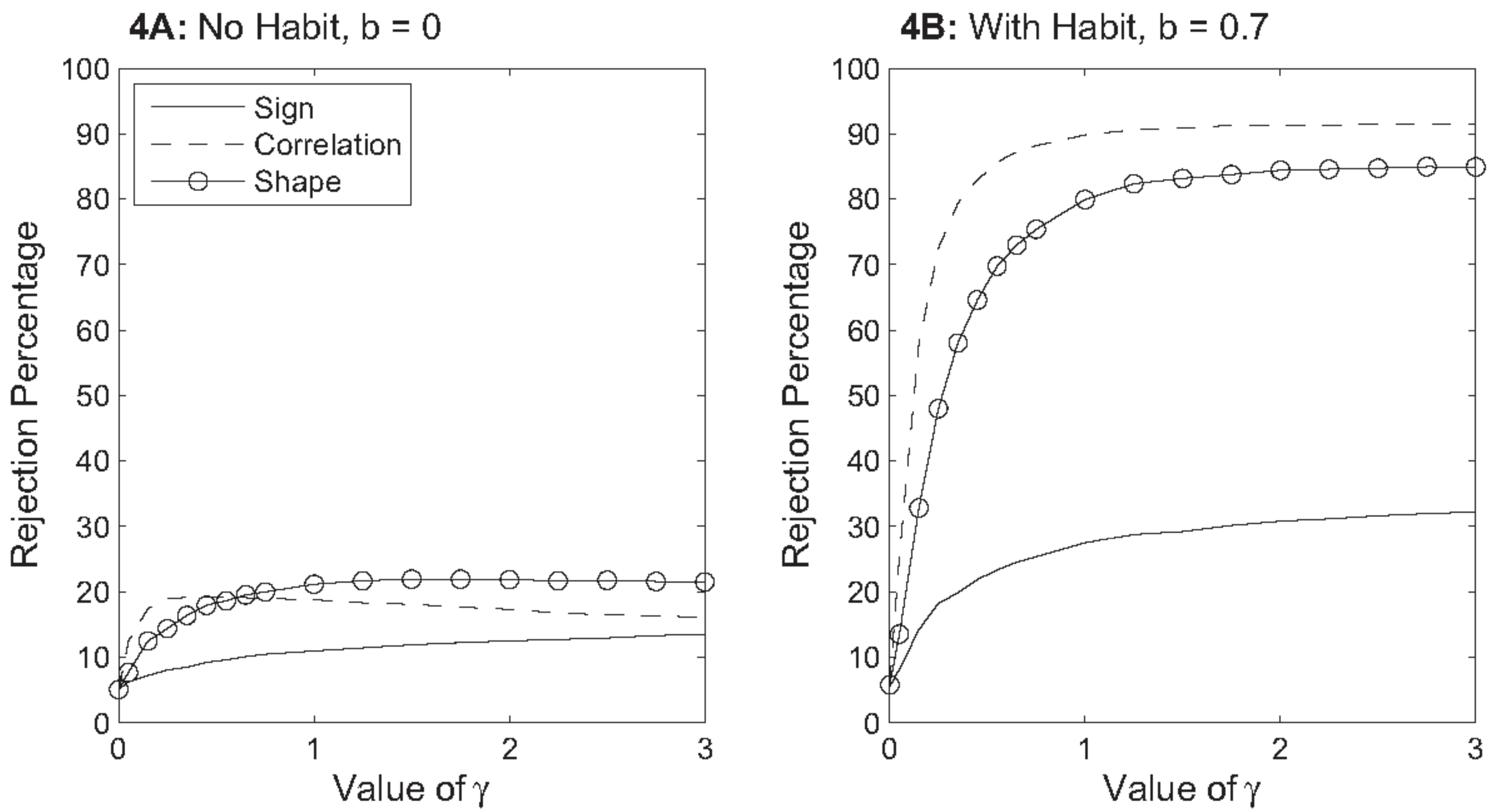


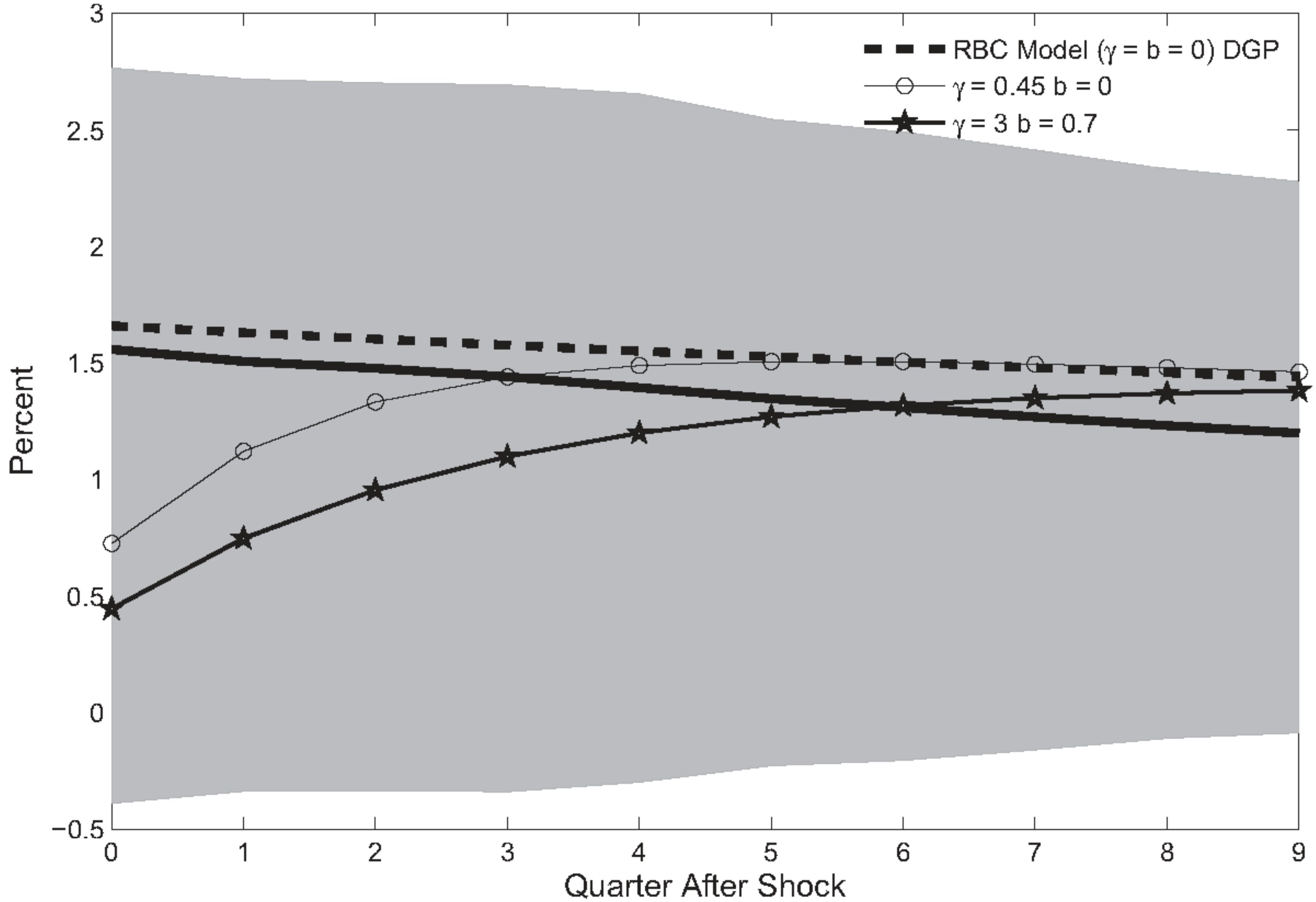
Figure 4 Rejection Rates For Different Tests when True DGP is RBC Model.



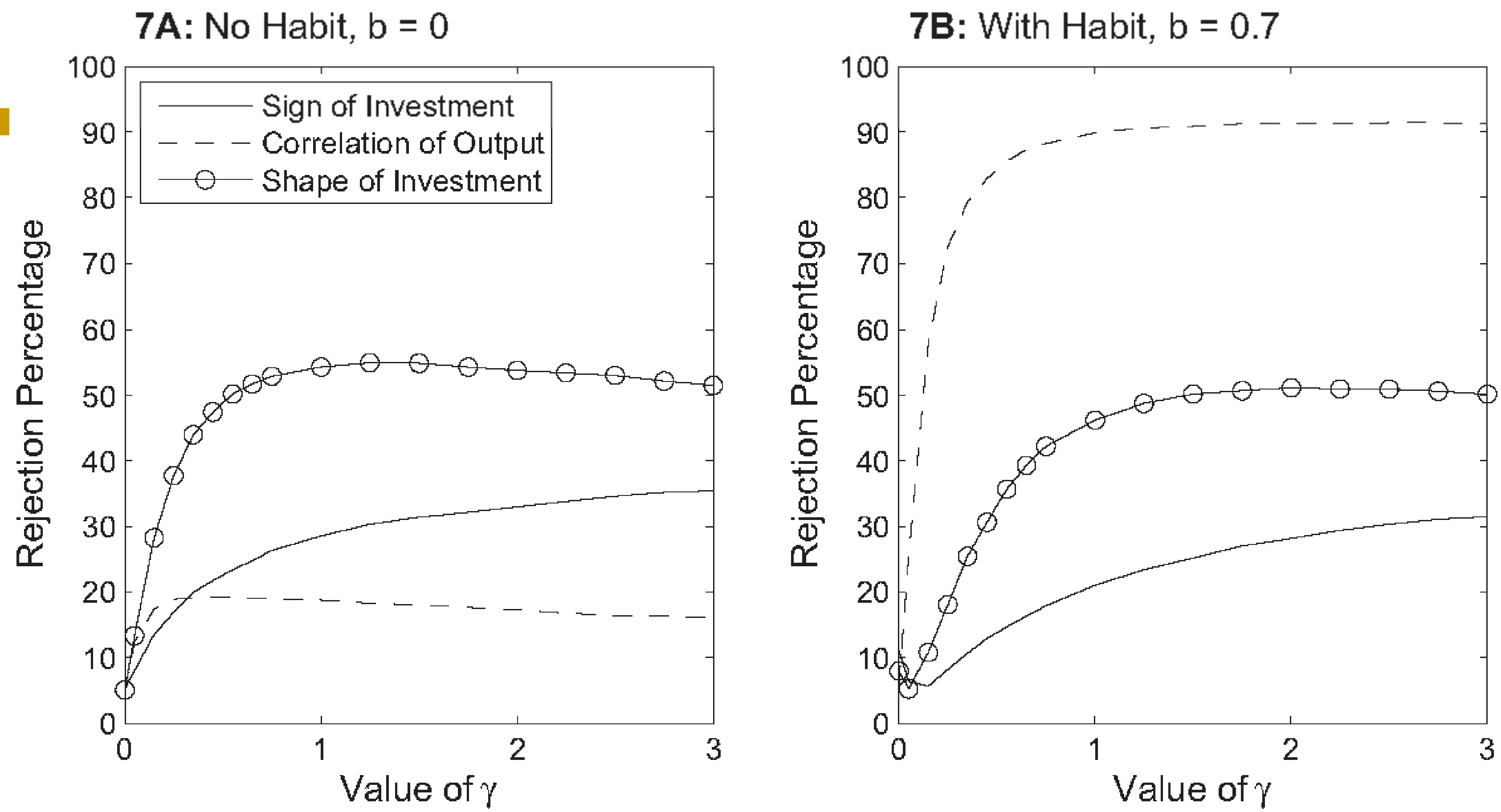
Test Investment Rather Than Hours.

- Estimate VAR of labor productivity, hours and investment over output.
- Standard confidence interval too tight. Rejects 13 percent of the time.
- Size-adjusted critical value for sign 2.53
- Size-adjusted critical value for shape 13

Figure 6: The response of investment to a technology shock estimated using data simulated from a RBC Model



Note Thick solid 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

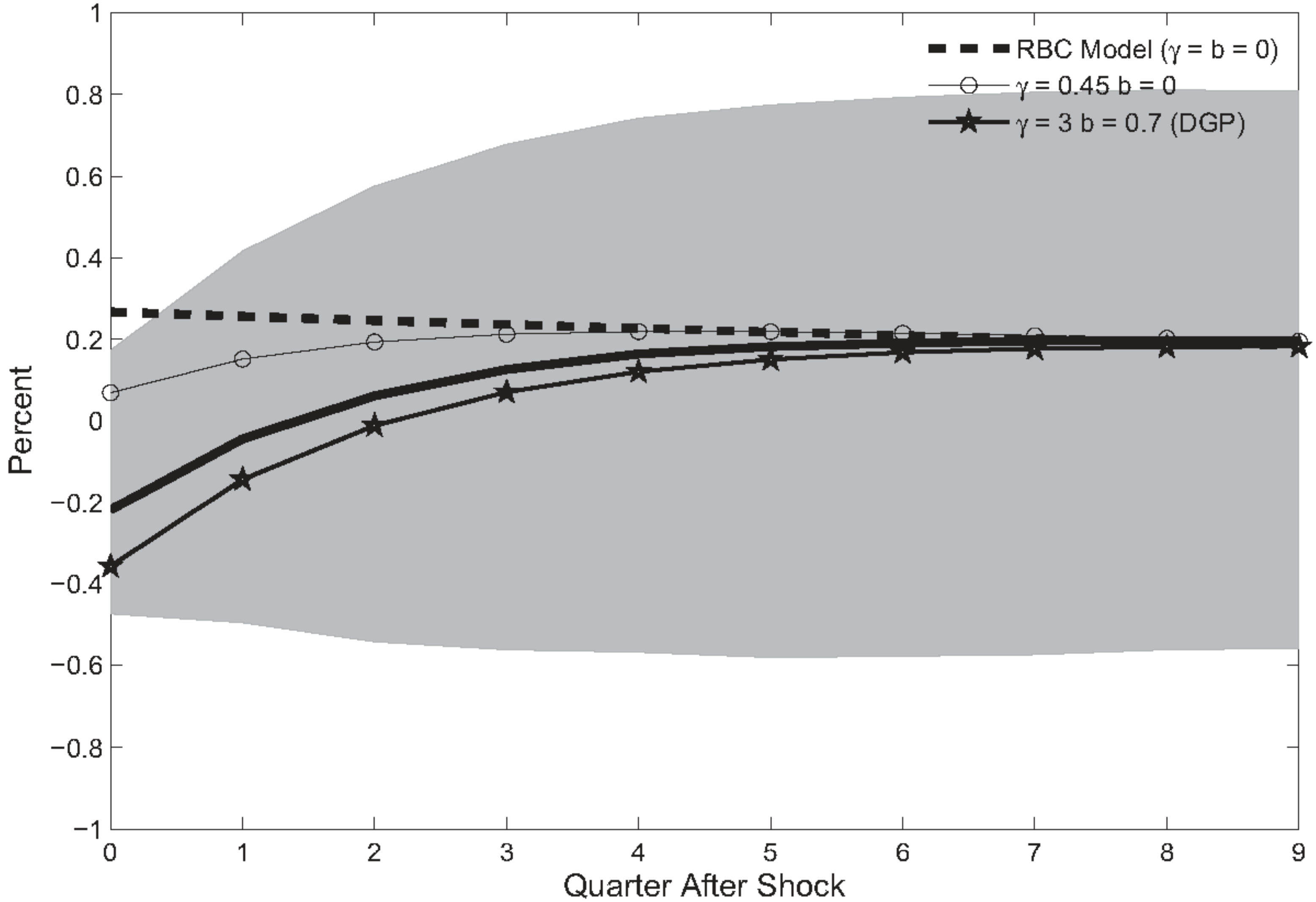


7A: No Habit, b = 0
7B: With Habit, b = 0.7
Sign of Investment
Correlation of Output
Shape of Investment
Rejection Percentage
Value of γ
0
10
20
30
40
50
60
70
80
90
100
0
1
2
3

Reversing the Role.

- Previous results simulated RBC model and studied rejection rates for macro models with real rigidities.
- Now, do the reverse.
  - Simulate Data from Macro Models with habit and investment adjustment costs.
  - Ask how often one rejects the RBC model.

Figure 8: 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 estimated response across 2000 simulations from a DSGE model with high real rigidities. Edges of grey area indicate 5th and 95th percentiles of all estimated responses to a one−standard−deviation technology shock

Rejection Rates of the RBC Model

Model Parameters		Hours		Investment		Output
b	gamma	Sign	Shape	Sign	Shape	Correlation
0	3	11	100	24	100	0
0.5	0.5	20	43	70	88	79
0.5	1.5	33	100	85	100	92
0.5	3	41	100	87	100	92
0.7	0.5	27	25	83	58	95
0.7	3	58	100	89	100	100

Sticky Price Models

- DGP (estimated by MLE)
- b=0.4 gamma = 2, theta = 0.75
- Test a variety of models

When b = 0.4 and gamma = 2

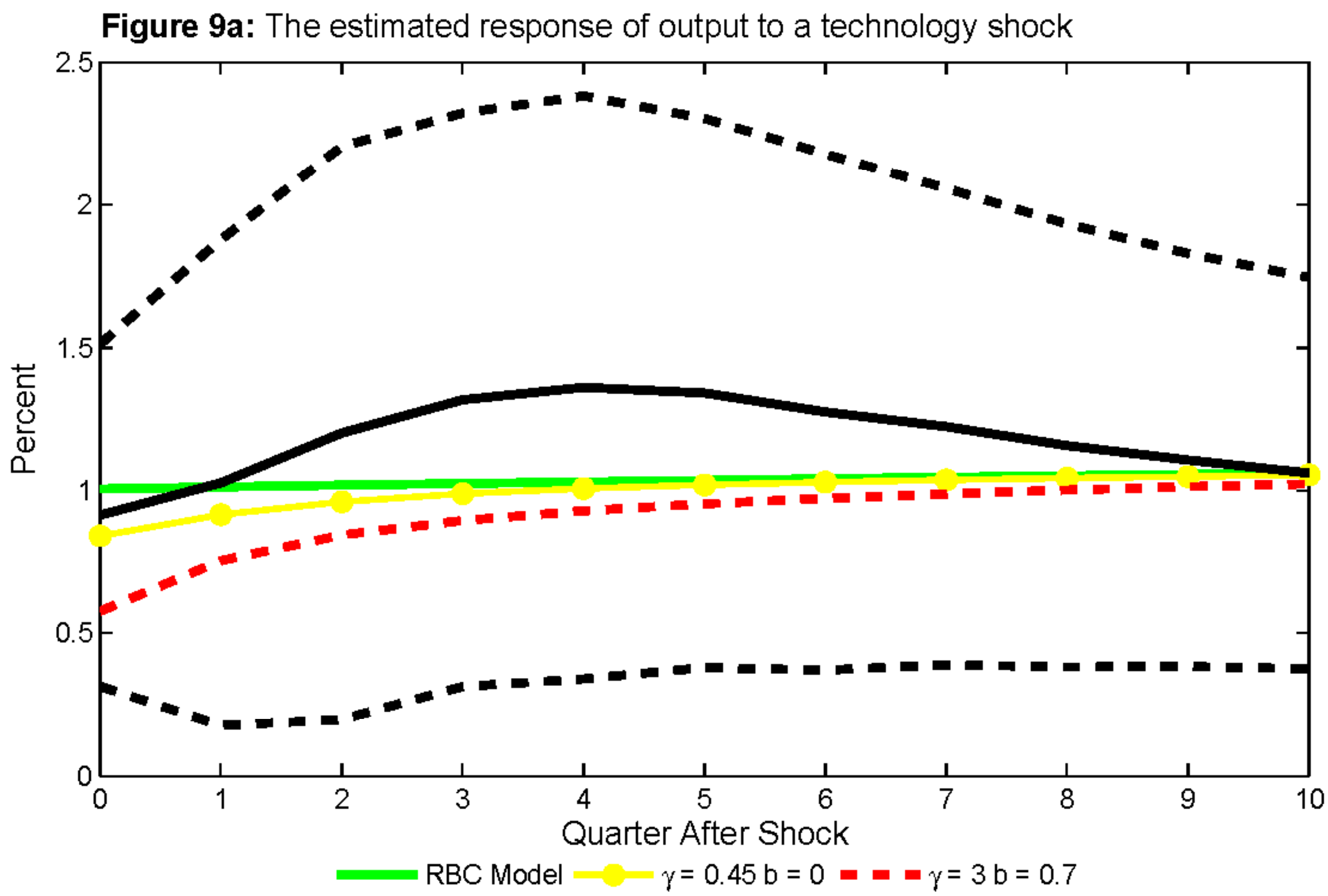
Theta	0.15	0.35	0.55	0.75*	0.9
Hours Impact	2	2	3	5	8
Hours Shape	2	2	3	5	8
Investment Impact	5	5	5	5	8
Investment Shape	**17**	**12**	**9**	5	10
Wage Impact	**32**	**20**	**11**	5	4
Wage Shape	**22**	**13**	**8**	5	4

When b = 0.4 and gamma = 0

Theta	0.15	0.35	0.55	0.75	0.9
Hours Impact	9	3	11	**100**	**100**
Hours Shape	13	3	13	**100**	**100**
Investment Impact	**73**	**27**	**19**	**100**	**100**
Investment Shape	**61**	**23**	**26**	**100**	**100**
Wage Impact	24	17	6	**29**	**94**
Wage Shape	17	11	6	**43**	**100**

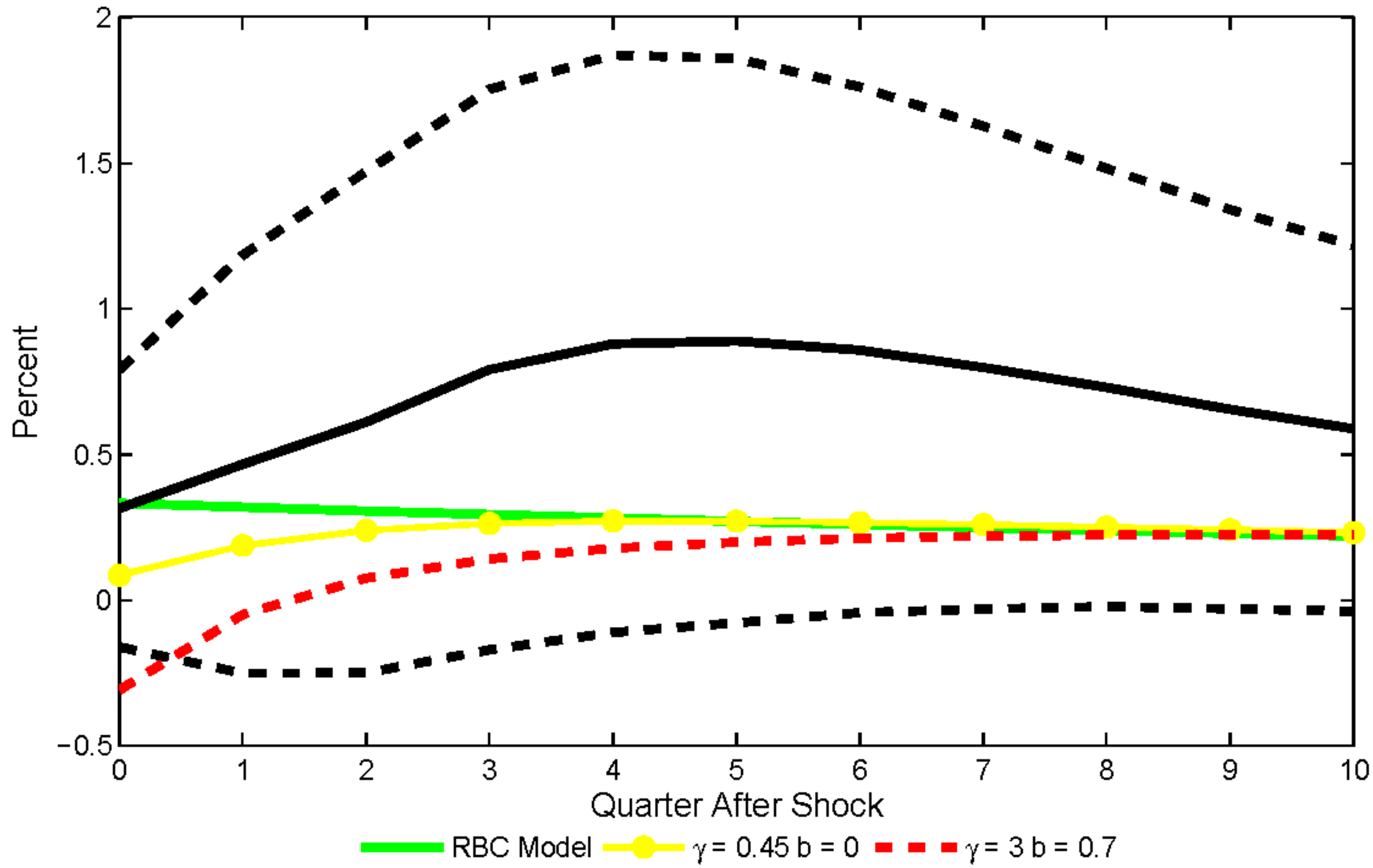
Empirical Application

- Estimate VAR Using U.S. Data
  - Labor Productivity
  - Hours Worked
  - Investment
  - 1959-2001
- Use Critical Values from Simulation
  - Sign 2.8 rather than 2
  - Shape 10 rather than 6



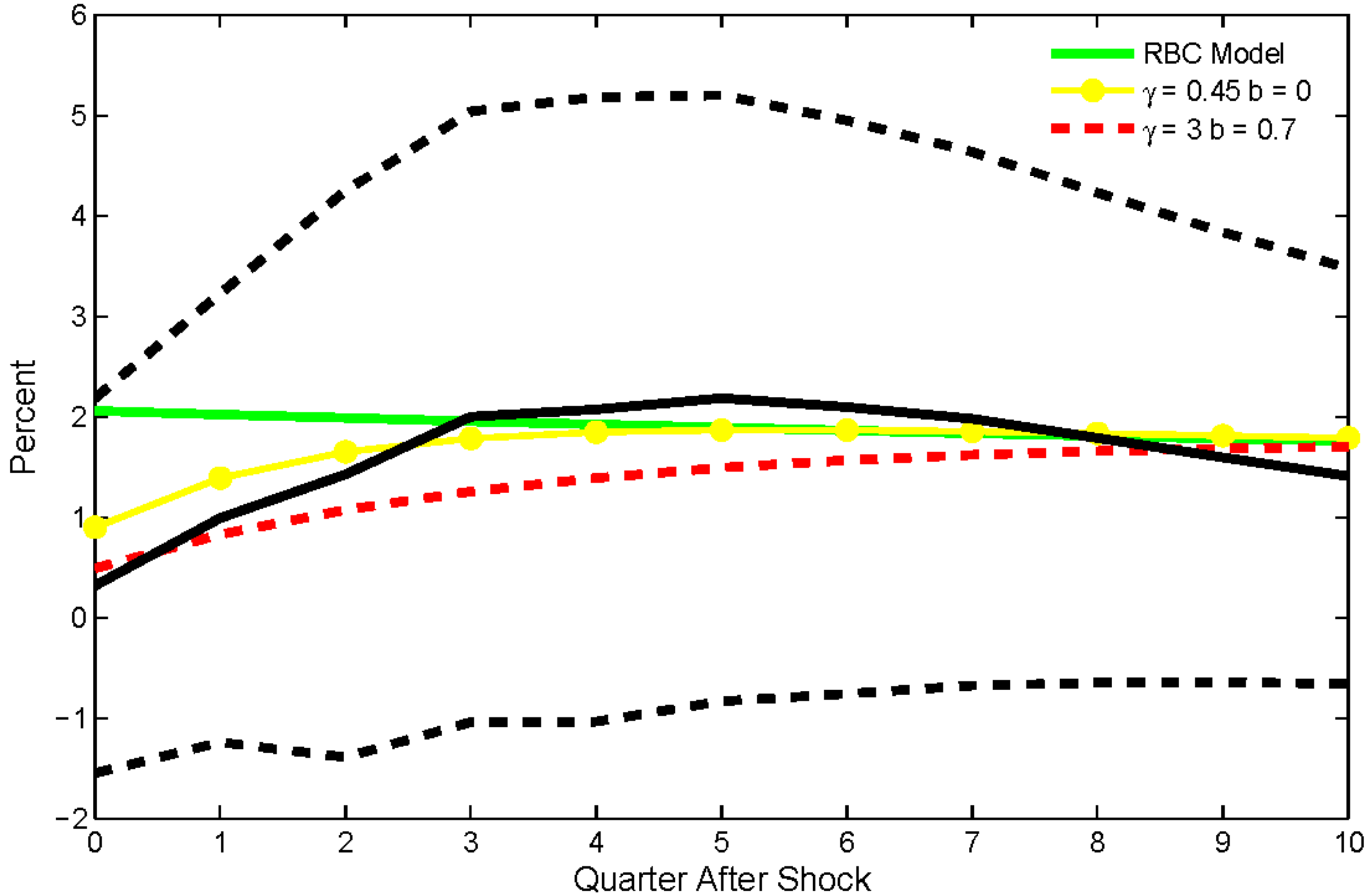
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 9b: The estimated response of hours worked to a technology shock



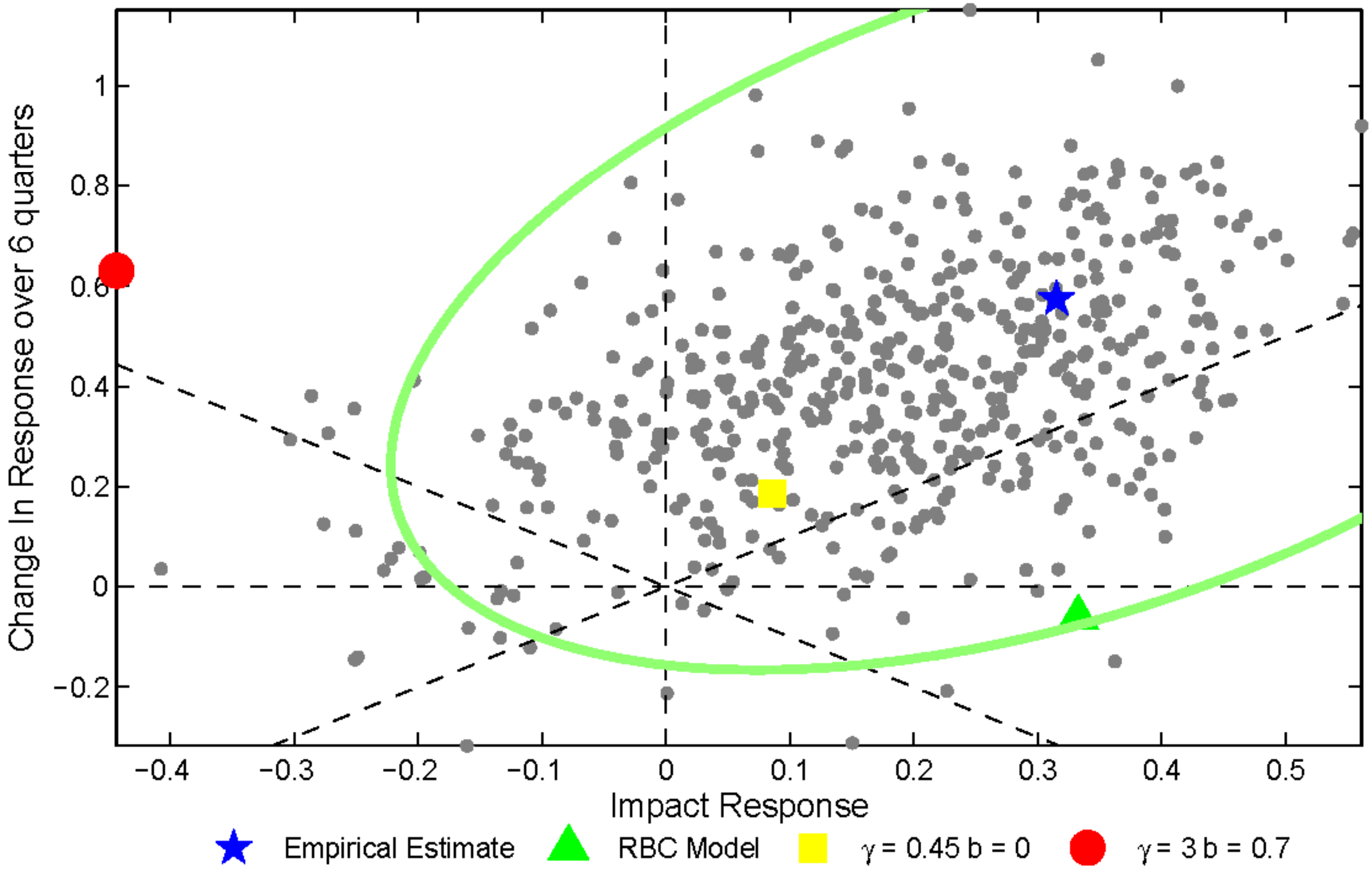
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 9c: The estimated response of Investment to a technology shock



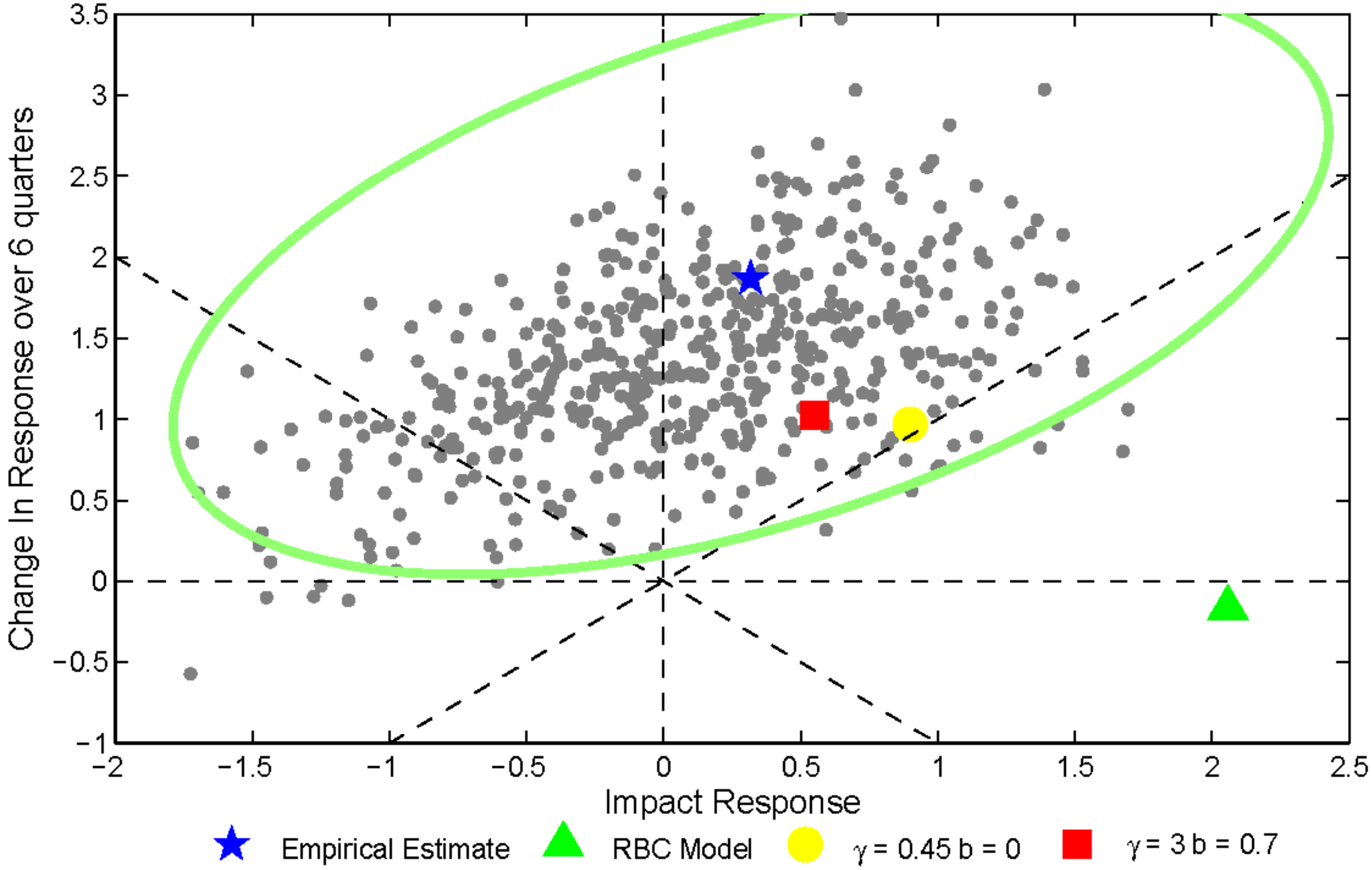
Note Thick black line is estimated response using a three variable VAR using U.S. data between 1954 to 2001. Edges of dashed areas indicate confidence interval of 2.8 standard deviations.

Figure 10a: 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 10b: 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.

Conclusions

- Impulse responses from long-run VARs can reject false models.
- Rejection rates increase the further away the false model is from the true data-generating model.
- Testing share is more powerful than testing just the sign on impact.

Conclusion (continued)

- Results should encourage us to find creative and new ways to test our models.

Possible improvements

- Feve and Guay (2009)
- Gospodinov (2008),
- Kascha and K. Mertens (2009)
- E. Mertens (2008).

Conclusions (continued)

- Overall, given these results on the power and size properties of long-run VARs, we conclude that
- VARs can be useful for discriminating between macro models and, therefore, should continue to be used in developing and testing business cycle theory.