8
Empirical Issues

8.1 Introduction

A theoretical model can only convince if it passes some form of empirical testing. This is also the case with the behavioral model discussed in this book. The problem in macroeconomics is how to devise a credible empirical test of the model. The history of macroeconomics is littered with examples of models which passed econometric testing procedures with flying colors, to be found wanting later.

The empirical testing tradition in macroeconomics has consisted in estimating an econometric model that embodies the equations of the theoretical model and then to perform dynamic simulations given the exogenous variables. Measures of goodness of fit then allowed the researcher to decide about the empirical validity of the theoretical model.

This approach is now severely criticized. First, it does not pass the Lucas critique as the estimated parameters of the structural model are not invariant with respect to the policy regime (Lucas 1976). Second, an attempt to estimate a small-scale model, like the one presented in this book, is likely to encounter problems of missing variables and incomplete dynamics. This is likely to lead to a misspecified model. Some researchers have reacted to this by adding autoregressive processes in the error terms as a cheap way to deal with these specification errors. We have argued that some of the existing DSGE models suffer from this problem. This approach is not attractive. The main reason is that it does not allow us to find out whether the model is rejected by the data or not (see Juselius and Franchi 2007). We will therefore not try to do this.

We will follow the approach of indirect inference, i.e., we ask what the predictions of the theoretical model are and confront these predictions with the data. Of course, it should be stressed from the start that a lot of uncertainty will continue to prevail about the empirical validity of the behavioral model.

Let us list the main predictions of the behavioral model.

1. Output movements are correlated with measures of optimism and pessimism, i.e., when market sentiments are optimistic (pessimistic) output increases (decreases).

2. Output movements are not normally distributed and show fat tails.
3. An interest rate increase leads to a temporary decline in output and inflation (as in other models). These effects, however, are time dependent (depend on market sentiments). This leads to different impulse responses depending on the timing of the shock.

We now check whether these three predictions are corroborated by empirical evidence.

8.2 The Correlation of Output Movements and Animal Spirits

The concept of animal spirits, i.e., waves of optimism and pessimism, has played a central role in the behavioral macroeconomic model presented in the previous chapters. The first question that arises here is whether there is an empirical counterpart for this concept. The answer is that there is, and it is widely used in day-to-day macroeconomic analysis. Many countries use survey-based consumer and/or business sentiment indicators as a tool with which to analyze the business cycle and as a predictive instrument.

The best-known sentiment indicator in the United States is the Michigan Consumer Confidence indicator, which has been in use since the 1950s. The first measures of consumer confidence were developed by George Katona in the late 1940s (see Katona 1951). Since then similar indicators have been implemented in a large number of countries (see Ludvigson (2004) and Curtin (2007) for an evaluation). Typically, sentiment indicators are constructed on the basis of a number of questions of how the individual perceives the present and the future economic conditions. Thus, these surveys produce two indices, one concerning present conditions, and one about future economic conditions. I will concentrate on the latter here, because this comes closest to the concept of optimism and pessimism used in this book, which is forward looking. The structure of these questions usually presents the individual with a discrete choice between good, bad, and neutral. An example from the Michigan indicator is the question: "Do you think that during the next twelve months, we'll have good times financially or bad times or are you uncertain? [good times/uncertain/bad times]." The answers are then transformed into an index by computing the divergence between "good times" and "bad times" answers.

The question that is addressed in this section is to what extent these sentiment indicators behave in a way that is consistent with our behavioral macroeconomic model. In figure 8.1 the Michigan Consumer Confidence indicator is shown, together with the U.S. output gap (quarterly data) during 1970–2009. The correlated movements of the sentiment index and the output gap are striking. The correlation coefficient was found to be 0.6. Note that in the simulations reported in chapter 1 the correlation coefficient between the output gap and the fraction of optimists was typically around 0.85. The lower correlation observed in reality is

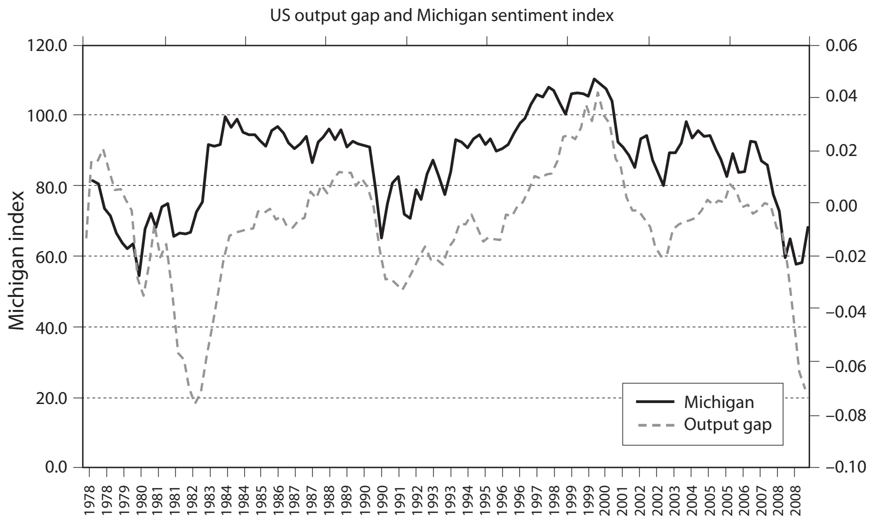

Figure 8.1. U.S. output gap and Michigan sentiment index (1970–2009).

Table 8.1. Pairwise Granger causality tests: behavioral model.

Null hypothesis	Obs	F-statistic	Probability
Output does not Granger cause optimism	1948	31.0990	5.1E-14
Optimism does not Granger cause output		32.8553	9.3E-15

Source: Calculated from the simulated output gap and animal spirits.

related to the fact that the survey-based sentiment indicators and the output gap have a lot of noise.

A typical feature of this correlation in the theoretical model is that the causality goes both ways, i.e., animal spirits affect output and output feeds back on animal spirits. We illustrate this by performing a Granger causality test on the simulated output gaps and the fractions of optimists obtained from the basic behavioral model of chapter 1 (see table 8.1). It shows that one cannot reject the hypotheses that animal spirits Granger cause the output gap and that the output gap Granger causes the animal spirits. Can one find the same structure in the relation between the observed GDP growth rates and the Michigan Consumer Confidence indicator? The answer is provided in table 8.2. It shows that one cannot reject the hypothesis that the Michigan Consumer Confidence indicator Granger causes the U.S. output gap and that the U.S. output gap Granger causes the Michigan Confidence indicator. Curtin (2007) has shown that in a majority of a sample of more than 50 countries one finds two-way causality between the confidence indicator and GDP growth (or another proxy for the business cycle).

Table 8.2. Pairwise Granger causality tests: U.S. data (1970–2009).

Null hypothesis	Obs	F-statistic	Probability
Michigan does not Granger cause GDP	123	15.83	0.00001
GDP does not Granger cause Michigan		4.83	0.0096

Source: Calculated from U.S. Department of Commerce, Bureau of Economic Analysis, and University of Michigan: Consumer Sentiment Index.

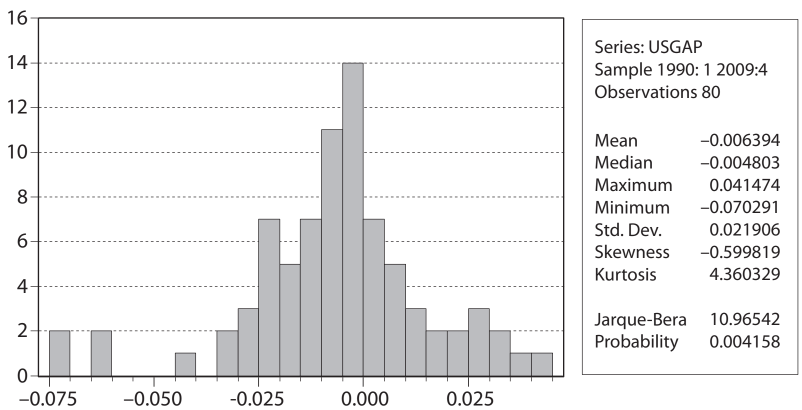

Figure 8.2. Frequency distribution of U.S. output gap.

8.3 Model Predictions: Higher Moments

In chapter 1 we showed that the behavioral model predicts that the output gap is not normally distributed and exhibits fat tails. This feature of the higher moments of the output gap is generated endogenously in the model. It is not the result of imposing such a feature on the stochastic shocks hitting the economy. We interpreted this result to mean that the model predicts that occasionally extreme movements in output can occur as a result of an endogenous dynamics. In chapter 1 we already confronted this prediction with data from the United States and concluded that indeed the distribution of the U.S. output gap during the postwar period was not normal. In this section we look at other countries, i.e., the United Kingdom and Germany. Unfortunately, the sample period is shorter and only starts in 1990. For the sake of comparability we also present the U.S. data for this shorter period. The results are shown in figures 8.2–8.4. They confirm what we found earlier, the output gap in these countries is not normally distributed (see the Jarque–Bera test, which rejects normality), and it exhibits fat tails. The latter means that the output gap is occasionally subjected to large changes that would not be observed if they were normally distributed.

One could object to this empirical analysis that the large shocks observed in the output gaps can also be the result of large exogenous shocks. In other words,

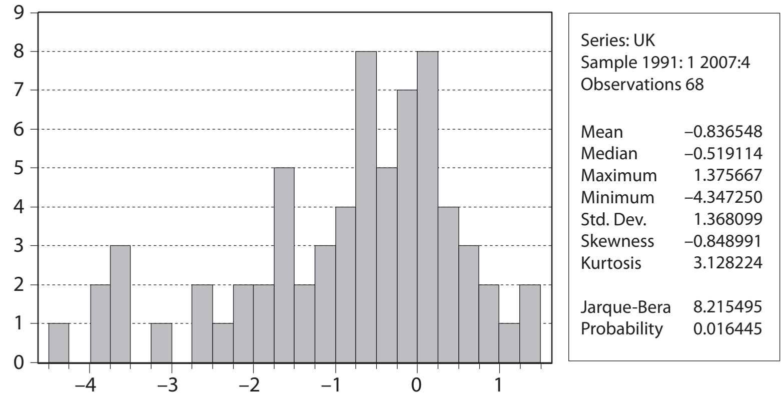

Figure 8.3. Frequency distribution of U.K. output gap.

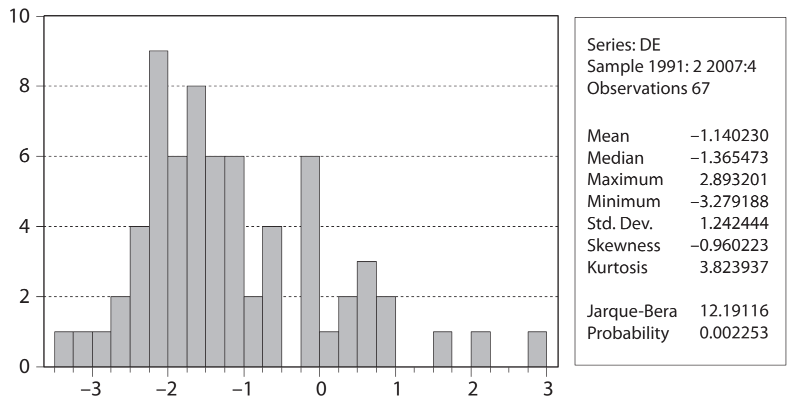

Figure 8.4. Frequency distribution of German output gap.

the evidence provided in figures 8.2–8.4 is also consistent with the view (implicit in DSGE models) that the macroeconomy does not produce large (nonnormally distributed) shocks, and thus that the observed large movements in output gaps must come from outside the model. Put differently, in the DSGE world large movements in output must be the result of large exogenous shocks.

The claim made here is not that the economy cannot sometimes be hit by large shocks, but that a theory that claims that large movements in output can *only* occur because of exogenous shocks is not a powerful theory. It necessitates finding a new exogenous explanation for every large boom and bust observed in output.[1] Put differently, for every boom or every bust a new story has to be told. Such a theory has very little predictive power. It amounts to a sophisticated story-telling

[1] For an illustration of this approach, see Favero (2001, pp. 73–75). Favero finds that there are many outliers in a VAR model of the U.S. economy, but interprets this to be the result of exogenous shocks (e.g., oil shocks). He then adds many dummies and, no surprise, the errors become normal.

exercise. Our theory allows for an explanation that is generated within the model. It is, therefore, more powerful.

We also observe from a comparison of figures 8.2–8.4 that not only are there fat tails, but also that the shape of the distribution is very different across countries. This implies that on the basis of the observations in these countries it is very unlikely that reliable statistical inferences can be drawn. This suggests that, as was noted in chapter 1, uncertainty in the sense of Knight is an important dynamic driving the macroeconomic movements.

8.4 Transmission of Monetary Policy Shocks

Empirical testing in macroeconomics has been very much influenced by Sims's (1980) seminal contribution (see Favero 2001). The basic idea is that theoretical models make predictions about the effects of policy shocks and that these predictions can be confronted with the data. This can be done by estimating a VAR of the macroeconomic variables and the policy variable. In the context of our model this consists in estimating a VAR of inflation, output gap, and the interest rate. This VAR then allows us to estimate an impulse response of inflation and output gap on interest rate shocks. This impulse response obtained from the data is then compared with the impulse response predicted by the theoretical model. It is important that, in doing so, the empirical impulse response is theory-free, i.e., does not use theory to impose identifying restrictions. After all, this approach is based on the idea of confronting theoretical predictions of the effect of a policy shock on inflation and output with the impulse response of inflation and output following a policy shock as detected from the data. In practice, this is not always easy to do, because restrictions on the parameters of the VAR must be imposed to be able to identify the impulse responses. The condition therefore has been to impose restrictions that use the least possible theory, or, put differently, that are used in the largest possible class of theoretical models. The Choleski decomposition (see, for example, Favero (2001) for explanation) is generally considered as the most theory-free set of restrictions.

Another popular set of identifying restrictions was proposed by Blanchard and Quah (1990). This consists in imposing restrictions on the long-term effects of demand and supply shocks based on theoretical insights. The latter are that demand shocks (e.g., a monetary shock) have only temporary effects. The problem with the Blanchard–Quah restrictions is that they exclude a priori a class of models that allow for multiple equilibria. In these models demand shocks can have permanent effects.

We now confront the theoretical impulses obtained from our behavioral model with the empirical ones. As a first step, we estimated a VAR model with three variables (output, inflation, and short-term interest rate) for the United States, using a Choleski decomposition (with ordering of inflation, output, interest rate). We then computed the impulse responses of output to an increase in the short-term interest

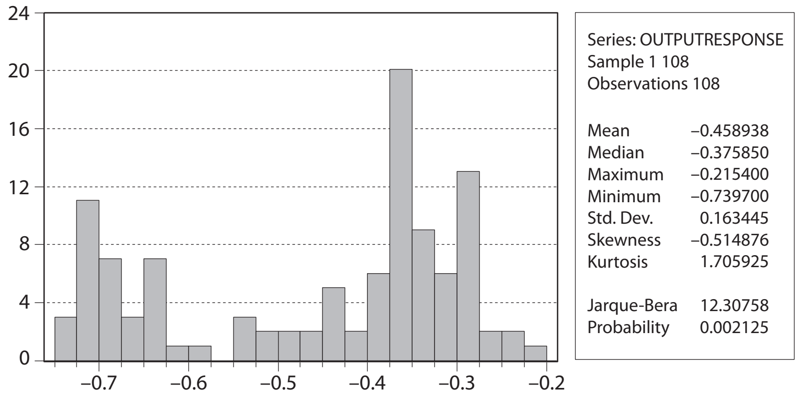

Figure 8.5. Distribution of short-term output response to shock Federal Funds rate.

rate (the Federal Funds rate). One of the main predictions of the behavioral model is that the impulse responses are very much influenced by the timing of the shock. We tested the empirical validity of this prediction by computing different impulse responses over different sample periods. We allowed for rolling sample periods of 30 years starting in 1972, and moving up each month. For each of these sample periods we computed the short-term output effect of an increase in the Federal Funds rate, where short-term refers to the effect after one year. We show the distribution of these short-term effects in figure 8.5. We find a wide range of short-term effects to the same policy shock (between −0.2% and −0.7% for a 1 standard deviation shock in the interest rate). In addition, we find that the distribution of these output responses is not normal. The Jarque–Bera test overwhelmingly rejects normality. We stressed in chapters 1 and 2 that this feature makes it very difficult for individual agents to make statistical inferences about the likely effects of a policy shock. On the whole the empirical results confirm the theoretical prediction of the behavioral model, i.e., the timing of the shock matters a great deal and affects how the same policy shock is transmitted into the economy. In addition, the nonnormality in the distribution of these shocks transforms risk into uncertainty (see chapter 2).

As in the case of the nonnormal distribution of the output gap, it must be admitted that the evidence of a nonnormal distribution of the short-term output effects of monetary policy shocks is not necessarily in contradiction with the DSGE model. In the framework of that model, the evidence provided here can be interpreted as arising from changes in policy regime. Ever since the famous Lucas critique (Lucas 1976), it has been well-known that changes in policy regime change the structural parameters of the standard demand and supply equations, and thus also change the transmission of policy shocks (the impulse responses). In this interpretation, the evidence of nonnormal distribution of the short-term output effects of a monetary policy shock is consistent with the view that there have been different changes in the policy regime during the sample period. These changes then produce nonnormal distributions of these effects.

Again we have two radically different interpretations of the same empirical evidence (which is not unusual in economics). The claim made in this book is that the interpretation given in the behavioral model is simpler than the one provided in the DSGE model. In the latter, the theoretical model predicts that, provided the policy regime does not change, a policy shock will always have the same effect. With noise in the data, the estimated effects of these shocks should be normally distributed. If we observe nonnormality, this must be produced outside the model, in this case by exogenous changes in the policy environment. Thus for every deviation from normality, the DSGE modelers must invoke a special event that has occurred outside the model. Such a model has no predictive power, because deviations from the predicted normality is always due to special circumstances. In contrast, in our behavioral model, nonnormality of the effects of policy shocks are not deviations from the rule, they are the rule.

We conclude that our behavioral model makes predictions that stand the test of the confrontation with the data. This does not mean, of course, that the empirical tests discussed here are in any way definitive. They should be considered as preliminary tests. Much more empirical testing will be necessary to elevate the behavioral model to the status of a serious alternative to the mainstream macroeconomic models.

8.5 Conclusion

Since its inception, booms and busts have characterized capitalism. The central issue in macroeconomics, therefore, is why these booms and busts in economic activity and in prices occur. Every macroeconomic theory must be capable of explaining these facts.

The explanation given by mainstream macroeconomics, in particular by new Keynesian rational expectations macroeconomics, fails to impress. In essence, the story told by mainstream macroeconomics is that these fluctuations arise because of large exogenous shocks. The latter have the effect of forcing rational agents to change their optimal consumption and production plans, but since they cannot adjust their plans instantaneously, prices and output adjust with lags. It is the combination of external shocks and slow adjustment that produces cyclical movements.

Thus, why did the world get into a recession in 2008–9? The answer given by the builders of the new Keynesian rational expectations model is that in 2007 a large external shock arose in the form of a sudden and unexpected increase in risk aversion. This change in risk perception then, like a hurricane, worked its way through the economy and produced a deep recession. In this sense mainstream macroeconomics has produced a "hurricane theory" of the business cycle.

The failure of mainstream macroeconomics to provide an endogenous explanation of booms and busts, in which a bust is the result of a previous boom, and a boom the result of a previous bust, has everything to do with the underlying paradigm of mainstream macroeconomics. This is the paradigm of the fully informed utility-maximizing agent who does not make systematic mistakes. Large booms or large busts can only be created by large external shocks to which these rational agents will then react.

I have argued that we need another macroeconomics that has the ambition of developing an endogenous explanation of the business cycle. I have tried to do so in this book. The behavioral model developed in this book allows us to better understand the recent macroeconomic developments in the world. The "Great Recession" of 2008–9 was not the result of an exogenous shock but resulted from excessive optimism that built up before 2008 and led to unsustainable consumption and investment. When the turnaround occurred, pessimism set in and led to a deep recession.

The behavioral model developed in this book was based on an enlarged concept of rationality. In mainstream macroeconomics, rationality is narrowly defined as utility maximization of agents who do not exhibit cognitive limitations, and as a result can solve incredibly complex problems of optimization and information processing. The starting point of the behavioral model presented in this book is that agents have limited cognitive abilities. These limitations force them to use simple rules (heuristics). Rationality was then introduced as a willingness of agents to learn by switching to alternative rules in order to improve their performance. Thus, moving away from the narrow rationality concept of mainstream macroeconomics does not imply that one is condemned to model irrationality where everything becomes possible.